econometrics

MDPI

Article

A Binary Choice Model with Sample Selection and Covariate-Related Misclassification

Jorge González Chapela

Academia General Militar, Centro Universitario de la Defensa de Zaragoza, 50090 Zaragoza, Spain; jorgegc@unizar.es

Abstract: Misclassification of a binary response variable and nonrandom sample selection are data issues frequently encountered by empirical researchers. For cases in which both issues feature simultaneously in a data set, we formulate a sample selection model for a misclassified binary outcome in which the conditional probabilities of misclassification are allowed to depend on covariates. Assuming the availability of validation data, the pseudo-maximum likelihood technique can be used to estimate the model. The performance of the estimator accounting for misclassification and sample selection is compared to that of estimators offering partial corrections. An empirical example illustrates the proposed framework.

Keywords: bivariate probit; misclassification; sample selection; lifetime migration

1. Introduction

Misclassification (or miscategorization) of a binary response variable and nonrandom sample selection are data issues frequently encountered by empirical researchers. It is by now well-known that errors in binary variables cannot be classical and thus lead to bias (see, for example, Hausman et al. 1998; Meyer and Mittag 2017). Additionally, regressions estimated on selected samples do not in general estimate the population parameters of interest (Heckman 1974, 1979). While there are methods for estimating models of misclassified binary responses (surveyed in Meyer and Mittag 2017), as a rule they assume the availability of a random sample. Likewise, the majority of the literature dealing with sample selection bias (surveyed in Vella 1998) assumes that the response variable of interest is measured accurately.

For cases in which misclassification of a binary response variable and nonrandom sample selection feature simultaneously in a data set, Arezzo and Guagnano (2019) formulated a model in which the probabilities of misclassification depend on the true values of the binary response but are otherwise independent of the covariates. In many cases, however, the assumption of conditionally random misclassification does not hold. Examples include classification errors varying across demographic groups or economic conditions in survey reports of unemployment status, participation in welfare programs, voter turnout, and debt repayment status (Levine 1993; Poterba and Summers 1995; Bollinger and David 1997, 2001; Bound et al. 2001; Davern et al. 2009; Katz and Katz 2010; Aller and Chapela 2013). Applying an estimator that assumes misclassification to be conditionally random when this assumption is false yields biased estimates, sometimes even more so than estimators that do not correct for misclassification (Meyer and Mittag 2017).

This paper generalizes Arezzo and Guagnano's (2019) formulation by allowing the misclassification probabilities to depend on covariates. When misclassification is not conditionally random, one can still obtain consistent estimates if validation data or a model of misclassification are available. Assuming the availability of validation data (or other out-of-sample information), a two-step approach can be used to estimate the parameters of interest (Bollinger and David 1997). In the first step, the misclassification probabilities are

check for updates

Citation: González Chapela, Jorge. 2022. A Binary Choice Model with Sample Selection and Covariate-Related Misclassification. *Econometrics* 10: 13. https://doi.org/10.3390/econometrics10020013

Academic Editor: Ryo Okui

Received: 18 April 2021
Accepted: 20 March 2022
Published: 23 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

predicted for specific subgroups from cross-tabulations or models of misclassification run on validation data. Second, the predicted probabilities of misclassification are incorporated into a model of a mismeasured binary response variable estimated on the observed data by the method of (pseudo-)maximum likelihood (ML) (e.g., Gourieroux et al. 1984). To this "predicted probabilities estimator" (PPE) (Meyer and Mittag 2017), we add the property that estimation in the second step can be conducted on a nonrandomly selected sample. If validation data are not available, then a model of misclassification dependent on covariates (or other observables) is needed (e.g., Aller and Chapela 2013). This model can be incorporated into the formulation developed by Arezzo and Guagnano (2019).

The sample selection mechanism considered in this paper is of the probit type. This setup applies to problems where units can be randomly drawn from the population, but due to some variables taking on particular values (incidental truncation) or unit/item nonresponse, we have missing data on the binary outcome of interest. For choice-based samples, Ramalho (2002) developed an estimator for a miscategorized discrete response variable under the assumption of conditionally random misclassification. Extending Ramalho's (2002) estimator to the case where misclassification is related to the covariates is left for future research.

Our model specification is based on the sample selection model introduced by Heckman (1974) under the assumption of bivariate normality between the outcome of interest and the selection propensity. Van de Ven and Praag (1981) and Dubin and Rivers (1989) extended Heckman's framework to deal with binary outcomes. For a misclassified binary outcome observed in a panel survey, Bollinger and David (2001) developed a model dealing with survey participation missed at random and misclassification dependent on covariates. Similarly, Katz and Katz's (2010) Bayesian procedure can account for data missed at random and covariate-related misclassification. On the other hand, Arezzo and Guagnano's (2019) model accounts for nonrandom selection but assumes conditionally random misclassification.

The rest of the paper is organized as follows. Section 2 formulates a bivariate probit model dealing with nonrandom sample selection and conditionally nonrandom misclassification and develops the proposed PPE. The finite-sample performance of the proposed and other related estimators is evaluated through simulation in Section 3. Section 4 contains an empirical application to a model of lifetime migration. Section 5 concludes the paper. Appendix A presents the technical material.

2. Model Specification

We start by specifying the probit model with sample selection of Van de Ven and Praag (1981). Then, we adapt it to manage covariate-related misclassification in the binary outcome of interest.

The probit model with sample selection can be written as

$$y_i^T = 1\left[X'_{1i}\beta_1 + \varepsilon_{1i} > 0\right] \tag{1}$$

$$s_i = 1\left[X'_{2i}\beta_2 + \varepsilon_{2i} > 0\right] \tag{2}$$

where $1[\cdot]$ is the indicator function, X_{1i} and X_{2i} are vectors of observed regressors, β_1 and β_2 are unknown vectors of parameters, and ε_{1i} and ε_{2i} are error terms from a bivariate normal distribution

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{21} \end{pmatrix} \sim N_2\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right\} \tag{3}$$

and independent of $X_i = (X'_{1i}, X'_{2i})'$. The binary outcome y_i^T is observed only when $s_i = 1$. A sample selection bias arises when $\rho \neq 0$ and probit is used to estimate β_1 in (1). Identification of $\beta = (\beta'_1, \beta'_2)'$ is assured whenever X_{2i} includes at least one variable that is not also in X_{1i}.

When y_i^T is observed, the density of $\left(y_i^T, s_i\right)$ is given by

$$P\left(y_i^T = y^T, s_i = 1\right) = \Phi_2\left(w_{1i}, X'_{2i}\beta_2, \rho_i^*\right) = P\left(y_i^T = y^T \middle| s_i = 1\right)\Phi\left(X'_{2i}\beta_2\right) \tag{4}$$

where $\Phi(\cdot)$ and $\Phi_2(\cdot)$ are the cdf of the standard and bivariate normal distributions, respectively. In expression (4), the first equality is written using notation adapted from Greene (2003): $w_{1i} = q_{1i}X'_{1i}\beta_1$, $q_{1i} = 2y_i^T - 1$, and $\rho_i^* = q_{1i}\rho$. The second equality (utilized below) is the multiplication rule of probabilities. When y_i^T is not observed, the density is simply

$$P(s_i = 0) = \Phi\left(-X'_{2i}\beta_2\right) \tag{5}$$

so the log-likelihood function for a sample of size N is

$$l(\beta, \rho) = \sum_{i=1}^{N}\left\{s_i \ln \Phi_2\left(w_{1i}, X'_{2i}\beta_2, \rho_i^*\right) + (1 - s_i) \ln \Phi\left(-X'_{2i}\beta_2\right)\right\}. \tag{6}$$

The estimator of (β, ρ) obtained by maximizing (6) is called Heckprobit (StataCorp 2019).

Now, let us suppose y_i^T may be mismeasured, that is, a true 1 may be misclassified as a 0 and a true 0 may be misclassified as a 1. We let y_i denote the mismeasured outcome indicator, which is observed only when $s_i = 1$.

Although we are interested in the probabilities of misclassification in the selected sample, if the selection mechanism is not operating in the validation data, only the probabilities of misclassification in the population can be calculated. To handle this case, we assume

$$P\left(y_i = 1 | s_i = 1, y_i^T = 0\right) = P\left(y_i = 1 | y_i^T = 0\right) = \alpha_{0i} \tag{7}$$

$$P\left(y_i = 0 | s_i = 1, y_i^T = 1\right) = P\left(y_i = 0 | y_i^T = 1\right) = \alpha_{1i}. \tag{8}$$

In these expressions, the first equality means that misclassification is conditionally independent from sample selection, or in other words that the misclassification mechanism operating in the population operates in the selected sample as well. The second equality defines what are usually called the conditional probabilities of misclassification, where conditional here refers to the true response. Overall, expressions (7) and (8) allow the conditional probabilities of misclassification in the selected sample to be calculated from validation data even if these data are not affected by selection.

The total probability theorem is used to derive the conditional (on selection) probability of an observed mismeasured outcome:

$$P(y_i = y | s_i = 1) = y_i\alpha_{0i} + (1 - y_i)\alpha_{1i} + (1 - \alpha_{0i} - \alpha_{1i})P\left(y_i^T = y^T \middle| s_i = 1\right). \tag{9}$$

The log-likelihood function based on observations (y_i, s_i) is

$$\begin{aligned} l(\alpha, \beta, \rho) &= \sum_{i=1}^{N}\{s_i \ln[P(y_i = y | s_i = 1)\Phi(X'_{2i}\beta_2)] + (1 - s_i) \ln \Phi(-X'_{2i}\beta_2)\} \\ &= \sum_{i=1}^{N}\left\{\begin{array}{c} s_i \ln\left[(y_i\alpha_{0i} + (1 - y_i)\alpha_{1i})\Phi(X'_{2i}\beta_2) + (1 - \alpha_{0i} - \alpha_{1i})\Phi_2\left(w_{1i}, X'_{2i}\beta_2, \rho_i^*\right)\right] \\ +(1 - s_i) \ln \Phi(-X'_{2i}\beta_2) \end{array}\right\} \end{aligned} \tag{10}$$

where α is the vector with $(\alpha_{0i}, \alpha_{1i})$, $i = 1, \ldots, N$, and the second equality follows from (9) and the second equality in (4).

The estimation framework provided by expression (10) encompasses several models that correct for misclassification and/or sample selection. When the binary outcome of interest is correctly classified, that is, $\alpha_{0i} = \alpha_{1i} = 0$, then (10) collapses to (6). When α_{0i} and α_{1i} are constants, that is, $\alpha_{0i} = \alpha_0$ and $\alpha_{1i} = \alpha_1$, then (10) becomes the log-likelihood function of Arezzo and Guagnano (2019). When $P(s_i = 1) = 1$, then (10) conforms to the log-likelihood function of a probit model accounting for misclassification (e.g., Meyer and Mittag 2017):

$$l(\alpha, \beta_1) = \sum_{i=1}^{N} \{\ln[(y_i\alpha_{0i} + (1 - y_i)\alpha_{1i}) + (1 - \alpha_{0i} - \alpha_{1i})\Phi(w_{1i})]\}. \tag{11}$$

In expression (10), the unknown parameters (α, β, ρ) are unidentified as there are $2N + \dim(\beta) + 1$ parameters. Assuming the availability of validation data (or other out-of-sample information), Bollinger and David's (1997) two-step procedure can be used to estimate (β, ρ). First, the conditional probabilities of misclassification are predicted from cross-tabulations or models of classification errors run on validation data. The prediction step does not require access to the validation data themselves and may be conducted by other researchers with direct access to the data. Second, after replacing α_{0i} and α_{1i} in (10) with their estimates, the resulting expression is maximized with respect to β and ρ. If α_{0i} and α_{1i} are consistently estimated, this PPE of (β, ρ) is consistent and asymptotically efficient. Standard errors need to be adjusted for the estimation of the conditional probabilities of misclassification unless these probabilities are precisely estimated or assumed to be known. Bollinger and David (1997) provided the formula for the asymptotic variance matrix. The bootstrap can also be used.

Function (10) is not globally concave in (β, ρ), and since estimators corresponding to local maxima may have no useful properties, a number of steps are taken to increase the chance that the maximum obtained is global.[1] Maximizations are conducted using the Newton–Raphson algorithm combined with the steepest ascent, providing the maximization routine analytical first and second derivatives of (10).[2] Probit estimates of the selection Equation (2), $\hat{\beta}_2^{\text{Probit}}$, provide the initial values for β_2 and are also utilized to calculate the inverse Mills ratio for the n observations with data available on y_i:

$$\hat{\lambda}_i = \frac{\phi\left(X'_{2i}\hat{\beta}_2^{\text{Probit}}\right)}{\Phi\left(X'_{2i}\hat{\beta}_2^{\text{Probit}}\right)}, \quad i = 1, \ldots, n \tag{12}$$

where ϕ is the standard normal pdf. Initial values for β_1 and ρ were obtained from Heckman's (1979) two-step method applied to the linear probability model (LPM)

$$P\left(y_i^T = 1 \middle| s_i = 1\right) = X'_{1i}\beta_1^{LPM} + \beta_\lambda\hat{\lambda}_i \tag{13}$$

augmented with misclassification:

$$P(y_i = 1|s_i = 1) = \alpha_{0i} + (1 - \alpha_{0i} - \alpha_{1i})\left(X'_{1i}\beta_1^{LPM} + \beta_\lambda\hat{\lambda}_i\right). \tag{14}$$

After replacing α_{0i} and α_{1i} with their estimates, Equation (14) is estimated by ordinary least squares without an intercept and constraining the coefficient of α_{0i} to unity. The initial value of β_1 is $\hat{\beta}_1^{LPM} \times 2.5$,[3] while that of ρ is

$$\hat{\rho}^{LPM} = \frac{\hat{\beta}_\lambda}{\hat{\sigma}} \tag{15}$$

where

$$\hat{\sigma}^2 = \frac{e'e + \beta_\lambda^2 \sum_{i=1}^{n} \omega_i}{n} \tag{16}$$

$$\omega_i = \hat{\lambda}_i\left(\hat{\lambda}_i + X'_{2i}\hat{\beta}_2^{Probit}\right) \tag{17}$$

and $e'e$ is the sum of squared residuals of regression (14) (Cameron and Trivedi 2005, p. 550).

3. Monte Carlo Study

This section reports the results of a Monte Carlo study designed to investigate the properties of the estimator proposed above and of other related estimators in situations characterized by different models of misclassification, probabilities of selection, and correlation between ε_{1i} and ε_{2i}. We used the data generating process considered in Arezzo and Guagnano (2019):

$$y_i^T = 1[\beta_{10} + \beta_{11}X_{11i} + \beta_{12}X_{12i} + \beta_{13}X_{13i} + \varepsilon_{1i} > 0] \tag{18}$$

$$s_i = 1[\beta_{20} + 0.8X_{21i} - 0.5X_{22i} + \varepsilon_{2i} > 0] \tag{19}$$

where X_{11}~.., X_{12} is a dummy variable equal to one with probability of $1/3$, X_{13}~$U(0,1)$, X_{21} and X_{22} are standard normal variates, and the parameters $(\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13})$ are set equal to $(-1,\ 0.2,\ 1.5,\ -0.6)$, respectively. The parameter β_{20} is set alternatively at $\{0.5,\ 2.18\}$, producing a moderate or a low amount of incidental truncation in the data (about 35% and 5%, respectively). The value of ρ is set alternatively at $\{0.2,\ 0.8\}$.

Three models of misclassification were considered. Misclassification Model 1 (MM1) is conditionally random misclassification with $\alpha_{0i} = 0.05$ and $\alpha_{1i} = 0.20$. Misclassification Models 2 and 3 (MM2 and MM3) allow α_{0i} and α_{1i} to depend on covariates. In MM2, the values taken by α_{0i} and α_{1i} are listed in Table 1, where the first (second) entry in a cell with numerical entries gives α_{0i} (α_{1i}). MM2 represents a case in which the conditional probabilities of misclassification are calculated from tabulations of validation data or other out of sample information, as in, for example, Levine (1993) and Poterba and Summers (1995). The values of α_{0i} and α_{1i} given in Table 1 ensure that the average probability of false positives (false negatives) is 0.05 (0.20).

Table 1. Misclassification probabilities in Misclassification Model 2.

X_{11} Values	X_{12} Values	
	$X_{12} = 0$	$X_{12} = 1$
$X_{11} < 1$	0.06, 0.20	0.08, 0.16
$X_{11} \geq 1$	0.03, 0.18	0.04, 0.28

MM3 has α_{0i} and α_{1i} predicted from probit models of classification errors run on validation data:

$$\alpha_{0i} = \Phi(-1.5 - 0.1X_{11i} - 0.1X_{12i}) \tag{20}$$

$$\alpha_{1i} = \Phi(-0.5 - 0.2X_{11i} - 0.3X_{12i}) \tag{21}$$

as in, for example, Bollinger and David (1997). The values of the parameters in expressions (20) and (21) ensure that the average probability of false positives (false negatives) is close to 0.05 (0.20). Note that in any of these three models of misclassification, $\alpha_{0i} + \alpha_{1i} < 1\ \forall i$, implying that the outcome of interest is not so mismeasured that its analysis should probably be abandoned (Hausman et al. 1998).

Seven estimators were compared in the Monte Carlo study.[4] Probit of y_i on X_{1i} run on the selected sample corrects neither for misclassification nor for sample selection. The estimators called HAS-Probit in Meyer and Mittag (2017) and the one we denote PP-Probit correct for misclassification only. Following Hausman et al. (1998), HAS-Probit assumes that $\alpha_{0i} = \alpha_0$ and $\alpha_{1i} = \alpha_1$, so α_0, α_1, and β_1 can be estimated by maximizing (11) on the selected sample. PP-Probit maximizes (11) with respect to β_1 on the selected sample after replacing α_{0i} and α_{1i} with $\hat{\alpha}_{0i}$ and $\hat{\alpha}_{1i}$ estimated in the first step. The fourth estimator, Heckprobit, corrects for sample selection only.

The other three estimators can correct for misclassification and sample selection. The estimators developed in Arezzo and Guagnano (2019) and this paper are denoted HAS-Heckprobit1 and PP-Heckprobit, respectively. HAS-Heckprobit2 is a modification of the former, allowing the misclassification probabilities to depend on X_{11} and X_{12} as specified in Table 1. If $r = \{1,2\}$ and $c = \{1,2\}$ are sets of indices for the rows and columns with numerical entries of Table 1, the conditional probability of an observed mismeasured outcome for an individual with values of X_{11} and X_{12} corresponding to the cell r,c is

$$P(y_{irc} = y|s_i = 1) = y_{irc}\alpha_0^{rc} + (1 - y_{irc})\alpha_1^{rc} + (1 - \alpha_0^{rc} - \alpha_1^{rc})P\left(y_{irc}^T = y^T \middle| s_i = 1\right). \quad (22)$$

Thus, the log-likelihood based on observations (y_{irc}, s_i) can be written as

$$l(\alpha, \beta, \rho) = \sum_{i=1}^{N} \left\{ \begin{array}{c} s_i \ln \left[\begin{array}{c} (y_{irc}\alpha_0^{rc} + (1 - y_{irc})\alpha_1^{rc})\Phi(X'_{2i}\beta_2) \\ +(1 - \alpha_0^{rc} - \alpha_1^{rc})\Phi_2(w_{1irc}, X'_{2i}\beta_2, \rho_i^*) \end{array} \right] \\ +(1 - s_i)\ln\Phi(-X'_{2i}\beta_2) \end{array} \right\}. \quad (23)$$

HAS-Heckprobit2 maximizes (23) over $\left((\alpha_0^{rc}, \alpha_1^{rc}), \beta, \rho\right)$ (a total of $8 + 7 + 1$ parameters) under the condition $\alpha_0^{rc} + \alpha_1^{rc} < 1$ for each cell r,c.

The analysis focused on the mean bias and the standard deviation across simulations. Bias is the relative difference between the estimate and the parameter value, averaged across simulations. For each estimator, the results of 12 different experiments are presented, combining three models of misclassification, two different β_{20}, and two different ρ. Each experiment consisted of 500 simulations with $N = 5000$.

The results of the Monte Carlo experiments for $\hat{\beta}$ and $\hat{\rho}$ are presented in Tables 2–4. These tables also list the number of convergences achieved in each experiment.[5] As expected (e.g., Hausman et al. 1998; Meyer and Mittag 2017), Probit presented significantly downward biased means and standard deviations. In terms of both bias and variance, Heckprobit offered little improvement. Under MM1, correcting for misclassification greatly reduced the bias, especially when ρ was small. PP-Probit tended to outperform HAS-Probit in terms of both bias and variance, but we stacked the deck in favor of PP estimators by using the true misclassification probabilities and by treating them as known parameters. Under MM2 and MM3, PP-Probit performed quite well, especially when ρ was small. In contrast, HAS-Probit was biased, sometimes significantly so. The bias of HAS-Probit tended to be upward under MM2 and was not clear-cut under MM3.

Both HAS-Heckprobit1 and PP-Heckprobit performed well when misclassification was random. Under this scenario, HAS-Heckprobit1 outperformed HAS-Heckprobit2, especially when the degree of incidental truncation in the data was moderate. The use of additional information in the form of $\hat{\alpha}_{0i}$ and $\hat{\alpha}_{1i}$ positively affected the accuracy of PP-Heckprobit vis à vis HAS-Heckprobit estimators. Under MM2 and MM3, HAS-Heckprobit1 generally obtained worse results in terms of bias for $\hat{\beta}_1$ and sometimes produced substantial convergence failures. However, it still yielded good results in terms of bias for $\hat{\beta}_2$. HAS-Heckprobit2 produced many convergence failures. This estimator outperformed HAS-Heckprobit1 under MM2 (which is not surprising) and MM3, in the latter case especially when the degree of incidental truncation in the data was moderate. In nearly all settings, PP-Heckprobit performed markedly better than the other estimators and showed few convergence failures. When estimation was conducted with PP-Heckprobit, the difference between the simulations mean and the true value of the parameter was usually below 1%.

Table 2. Monte Carlo simulation results for Misclassification Model 1.

	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
$\boldsymbol{\beta}_{20} = 0.5,\ \boldsymbol{\rho} = 0.2$														
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.146	0.058	−0.034	0.216	−0.081	0.080	−0.092	0.066	0.036	0.229	0.011	0.260	−0.004	0.092
$\hat{\beta}_{11}$	−0.416	0.018	0.071	0.055	0.005	0.024	−0.418	0.017	0.042	0.053	0.087	0.070	−0.002	0.024
$\hat{\beta}_{12}$	−0.273	0.052	0.053	0.302	0.005	0.081	−0.275	0.051	0.035	0.291	−0.146	0.494	−0.002	0.079
$\hat{\beta}_{13}$	−0.314	0.084	0.093	0.210	0.010	0.123	−0.316	0.083	0.067	0.212	0.128	0.257	0.003	0.122
$\hat{\beta}_{20}$							0.001	0.049	0.002	0.048	−0.001	0.048	0.001	0.049
$\hat{\beta}_{21}$							0.004	0.026	0.003	0.026	0.003	0.024	0.004	0.026
$\hat{\beta}_{22}$							0.002	0.023	0.001	0.023	0.003	0.025	0.002	0.023
$\hat{\rho}$							−0.328	0.076	−0.023	0.124	0.152	0.152	−0.008	0.109
Convergence	493		491		493		493		409		186		493	
$\boldsymbol{\beta}_{20} = 0.5,\ \boldsymbol{\rho} = 0.8$														
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.286	0.059	−0.190	0.212	−0.258	0.080	−0.096	0.060	0.004	0.126	−0.024	0.175	0.000	0.075
$\hat{\beta}_{11}$	−0.402	0.018	0.187	0.059	0.109	0.026	−0.432	0.017	0.004	0.031	0.017	0.044	0.005	0.024
$\hat{\beta}_{12}$	−0.230	0.052	0.168	0.330	0.113	0.090	−0.262	0.051	0.001	0.154	−0.060	0.301	0.003	0.082
$\hat{\beta}_{13}$	−0.273	0.085	0.216	0.238	0.106	0.126	−0.305	0.082	0.011	0.138	0.028	0.140	0.005	0.114
$\hat{\beta}_{20}$							−0.006	0.044	0.002	0.044	0.009	0.044	0.001	0.044
$\hat{\beta}_{21}$							0.003	0.025	0.002	0.025	0.002	0.024	0.002	0.025
$\hat{\beta}_{22}$							0.002	0.022	0.000	0.023	0.002	0.024	0.000	0.023
$\hat{\rho}$							−0.345	0.061	−0.007	0.092	−0.000	0.103	−0.005	0.072
Convergence	485		483		485		485		476		169		485	
$\boldsymbol{\beta}_{20} = 2.18,\ \boldsymbol{\rho} = 0.2$														
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.098	0.048	0.012	0.184	−0.017	0.067	−0.089	0.049	0.021	0.185	−0.059	0.203	−0.004	0.069
$\hat{\beta}_{11}$	−0.411	0.014	0.043	0.045	0.005	0.019	−0.412	0.014	0.038	0.045	0.025	0.056	0.001	0.019
$\hat{\beta}_{12}$	−0.270	0.042	0.031	0.249	0.004	0.064	−0.272	0.042	0.028	0.246	−0.135	0.375	−0.000	0.065
$\hat{\beta}_{13}$	−0.310	0.070	0.065	0.171	0.017	0.102	−0.311	0.069	0.065	0.169	0.025	0.205	0.013	0.102
$\hat{\beta}_{20}$							0.001	0.091	0.001	0.091	0.001	0.088	0.002	0.091
$\hat{\beta}_{21}$							0.003	0.045	0.004	0.045	0.002	0.044	0.003	0.045
$\hat{\beta}_{22}$							0.005	0.038	0.005	0.038	−0.002	0.039	0.004	0.038
$\hat{\rho}$							−0.323	0.156	0.112	0.258	0.108	0.246	0.038	0.239
Convergence	495		496		496		496		466		218		496	
$\boldsymbol{\beta}_{20} = 2.18,\ \boldsymbol{\rho} = 0.8$														
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.116	0.048	0.008	0.185	−0.039	0.068	−0.087	0.048	0.009	0.163	−0.041	0.162	−0.003	0.066
$\hat{\beta}_{11}$	−0.402	0.014	0.079	0.046	0.033	0.020	−0.413	0.014	0.020	0.039	−0.005	0.038	0.005	0.020
$\hat{\beta}_{12}$	−0.255	0.043	0.064	0.255	0.032	0.066	−0.265	0.043	0.010	0.207	−0.083	0.352	0.003	0.066
$\hat{\beta}_{13}$	−0.297	0.071	0.103	0.176	0.044	0.104	−0.306	0.070	0.030	0.153	−0.014	0.145	0.012	0.101
$\hat{\beta}_{20}$							0.002	0.090	0.003	0.090	0.005	0.091	0.003	0.089
$\hat{\beta}_{21}$							0.007	0.044	0.006	0.044	0.007	0.045	0.006	0.045
$\hat{\beta}_{22}$							−0.002	0.040	−0.004	0.040	−0.004	0.039	−0.003	0.040
$\hat{\rho}$							−0.356	0.151	−0.060	0.198	−0.116	0.210	−0.039	0.186
Convergence	493		494		494		494		424		162		462	

Table 3. Monte Carlo simulation results for Misclassification Model 2.

	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
$\boldsymbol{\beta}_{20} = 0.5,\ \boldsymbol{\rho} = 0.2$														
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.177	0.057	−0.067	0.251	−0.079	0.081	−0.123	0.066	0.016	0.262	−0.014	0.244	−0.002	0.094
$\hat{\beta}_{11}$	−0.513	0.016	0.266	0.086	0.005	0.023	−0.514	0.016	0.210	0.080	0.098	0.067	−0.001	0.023
$\hat{\beta}_{12}$	−0.300	0.052	0.412	0.792	0.007	0.083	−0.302	0.052	0.345	0.640	−0.227	0.558	−0.001	0.081
$\hat{\beta}_{13}$	−0.321	0.080	0.595	0.420	0.005	0.118	−0.323	0.080	0.518	0.387	0.187	0.299	−0.001	0.117
$\hat{\beta}_{20}$							0.001	0.049	0.004	0.048	−0.007	0.049	0.002	0.049
$\hat{\beta}_{21}$							0.004	0.026	0.004	0.026	0.003	0.026	0.004	0.026
$\hat{\beta}_{22}$							0.002	0.023	0.002	0.023	0.004	0.023	0.002	0.023
$\hat{\rho}$							−0.329	0.075	0.282	0.169	0.206	0.186	−0.013	0.106
Convergence	500		499		500		500		443		197		500	

Table 3. *Cont.*

	β_{20} = 0.5, ρ = 0.8													
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.323	0.058	−0.179	0.257	−0.258	0.080	−0.133	0.060	−0.056	0.127	−0.005	0.145	−0.000	0.076
$\hat{\beta}_{11}$	−0.503	0.017	0.493	0.088	0.109	0.025	−0.528	0.016	−0.008	0.032	0.021	0.040	0.005	0.022
$\hat{\beta}_{12}$	−0.267	0.051	0.709	0.979	0.114	0.091	−0.297	0.050	0.094	0.185	−0.100	0.313	0.003	0.083
$\hat{\beta}_{13}$	−0.279	0.083	0.837	0.407	0.105	0.123	−0.309	0.080	0.128	0.151	0.037	0.140	0.006	0.111
$\hat{\beta}_{20}$							−0.006	0.044	0.002	0.044	0.005	0.045	0.002	0.044
$\hat{\beta}_{21}$							0.004	0.025	0.002	0.025	0.002	0.025	0.002	0.025
$\hat{\beta}_{22}$							0.002	0.023	−0.000	0.023	0.003	0.023	0.000	0.023
$\hat{\rho}$							−0.351	0.060	0.086	0.079	0.023	0.092	−0.006	0.071
Convergence	500		498		500		500		480		254		500	
	β_{20} = 2.18, ρ = 0.2													
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.129	0.047	−0.020	0.213	−0.015	0.067	−0.119	0.048	−0.017	0.211	−0.066	0.173	−0.003	0.069
$\hat{\beta}_{11}$	−0.508	0.013	0.212	0.070	0.005	0.019	−0.509	0.013	0.189	0.068	0.015	0.046	0.001	0.019
$\hat{\beta}_{12}$	−0.296	0.041	0.335	0.484	0.005	0.065	−0.298	0.041	0.311	0.470	−0.185	0.395	0.000	0.065
$\hat{\beta}_{13}$	−0.320	0.068	0.521	0.338	0.011	0.101	−0.321	0.068	0.490	0.335	0.109	0.209	0.006	0.101
$\hat{\beta}_{20}$							0.001	0.091	0.001	0.091	0.003	0.093	0.001	0.091
$\hat{\beta}_{21}$							0.003	0.045	0.003	0.045	0.006	0.044	0.003	0.045
$\hat{\beta}_{22}$							0.005	0.038	0.005	0.038	0.009	0.038	0.005	0.038
$\hat{\rho}$							−0.330	0.155	0.309	0.349	0.106	0.287	0.006	0.234
Convergence	500		500		500		500		493		203		499	
	β_{20} = 2.18, ρ = 0.8													
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.148	0.047	−0.018	0.215	−0.037	0.068	−0.118	0.047	−0.063	0.184	−0.019	0.176	−0.001	0.066
$\hat{\beta}_{11}$	−0.501	0.013	0.267	0.072	0.033	0.019	−0.509	0.013	0.085	0.055	0.032	0.044	0.007	0.019
$\hat{\beta}_{12}$	−0.282	0.041	0.388	0.489	0.032	0.066	−0.292	0.041	0.203	0.375	−0.104	0.326	0.004	0.067
$\hat{\beta}_{13}$	−0.307	0.068	0.589	0.339	0.037	0.101	−0.315	0.068	0.304	0.265	0.075	0.171	0.010	0.098
$\hat{\beta}_{20}$							0.002	0.091	0.003	0.093	0.007	0.089	0.003	0.090
$\hat{\beta}_{21}$							0.006	0.045	0.010	0.045	0.012	0.042	0.005	0.044
$\hat{\beta}_{22}$							−0.002	0.040	0.002	0.039	−0.003	0.039	−0.002	0.039
$\hat{\rho}$							−0.365	0.152	0.027	0.176	−0.023	0.206	−0.047	0.186
Convergence	500		499		500		500		399		197		466	

Table 4. Monte Carlo simulation results for Misclassification Model 3.

	β_{20} = 0.5, ρ = 0.2													
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.041	0.061	0.155	0.218	−0.075	0.083	0.016	0.067	0.244	0.221	0.146	0.201	0.003	0.090
$\hat{\beta}_{11}$	−0.037	0.014	0.090	0.032	0.016	0.018	−0.041	0.014	0.100	0.031	0.094	0.029	0.008	0.018
$\hat{\beta}_{12}$	−0.184	0.053	−0.081	0.172	0.009	0.075	−0.187	0.053	−0.075	0.166	−0.137	0.192	0.002	0.076
$\hat{\beta}_{13}$	−0.273	0.087	−0.151	0.125	0.013	0.123	−0.277	0.087	−0.143	0.125	−0.022	0.149	0.005	0.122
$\hat{\beta}_{20}$							−0.002	0.046	−0.003	0.045	−0.004	0.041	−0.002	0.046
$\hat{\beta}_{21}$							0.003	0.026	0.003	0.026	0.001	0.027	0.003	0.026
$\hat{\beta}_{22}$							0.001	0.023	0.001	0.023	−0.002	0.025	0.001	0.023
$\hat{\rho}$							−0.276	0.076	−0.138	0.092	−0.097	0.094	0.005	0.103
Convergence	500		399		500		500		407		92		500	
	β_{20} = 0.5, ρ = 0.8													
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.155	0.059	−0.135	0.104	−0.254	0.082	0.049	0.060	0.172	0.125	0.127	0.138	0.004	0.076
$\hat{\beta}_{11}$	0.074	0.016	0.085	0.018	0.117	0.020	0.021	0.015	0.107	0.021	0.064	0.022	0.007	0.018
$\hat{\beta}_{12}$	−0.121	0.052	−0.110	0.073	0.114	0.081	−0.167	0.052	−0.108	0.096	−0.101	0.172	0.002	0.077
$\hat{\beta}_{13}$	−0.230	0.089	−0.216	0.091	0.106	0.130	−0.271	0.085	−0.191	0.105	−0.131	0.105	0.003	0.116
$\hat{\beta}_{20}$							0.004	0.048	0.004	0.049	0.003	0.050	0.002	0.049
$\hat{\beta}_{21}$							0.002	0.024	0.002	0.024	0.001	0.021	0.002	0.024
$\hat{\beta}_{22}$							−0.000	0.022	0.001	0.022	0.004	0.022	0.000	0.022
$\hat{\rho}$							−0.265	0.062	−0.201	0.081	−0.141	0.085	−0.000	0.068
Convergence	500		350		500		500		477		75		499	

Table 4. *Cont.*

	β_{20} = 2.18, ρ = 0.2													
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	0.001	0.051	0.054	0.137	−0.011	0.070	0.012	0.052	0.112	0.175	0.098	0.149	0.003	0.071
$\hat{\beta}_{11}$	−0.054	0.012	−0.019	0.019	0.008	0.015	−0.057	0.012	0.009	0.026	0.043	0.027	0.003	0.015
$\hat{\beta}_{12}$	−0.189	0.041	−0.163	0.100	0.004	0.059	−0.191	0.041	−0.137	0.140	−0.152	0.135	−0.001	0.059
$\hat{\beta}_{13}$	−0.282	0.072	−0.252	0.088	0.003	0.101	−0.284	0.072	−0.215	0.100	−0.108	0.129	−0.002	0.101
$\hat{\beta}_{20}$							0.003	0.093	0.003	0.092	0.006	0.096	0.004	0.093
$\hat{\beta}_{21}$							0.001	0.044	0.001	0.044	−0.010	0.039	0.001	0.044
$\hat{\beta}_{22}$							0.001	0.039	0.002	0.039	0.002	0.035	0.001	0.040
$\hat{\rho}$							−0.200	0.173	−0.286	0.181	−0.067	0.198	0.077	0.230
Convergence	500		356		500		500		387		86		499	
	β_{20} = 2.18, ρ = 0.8													
	Probit		HAS-Probit		PP-Probit		Heckprobit		HAS-Heckprobit1		HAS-Heckprobit2		PP-Heckprobit	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	−0.011	0.050	0.084	0.182	−0.031	0.069	0.022	0.049	0.095	0.148	0.139	0.148	0.003	0.066
$\hat{\beta}_{11}$	−0.025	0.012	0.032	0.025	0.039	0.015	−0.046	0.012	0.010	0.022	0.082	0.027	0.007	0.015
$\hat{\beta}_{12}$	−0.168	0.042	−0.119	0.134	0.033	0.060	−0.185	0.042	−0.143	0.117	−0.134	0.185	0.002	0.061
$\hat{\beta}_{13}$	−0.266	0.071	−0.213	0.094	0.033	0.101	−0.281	0.070	−0.224	0.096	−0.085	0.117	0.000	0.098
$\hat{\beta}_{20}$							0.003	0.090	0.006	0.095	0.008	0.087	0.003	0.090
$\hat{\beta}_{21}$							−0.003	0.044	0.003	0.046	0.003	0.041	0.002	0.044
$\hat{\beta}_{22}$							−0.003	0.040	0.009	0.040	0.009	0.040	0.004	0.040
$\hat{\rho}$							−0.186	0.168	−0.170	0.200	−0.139	0.201	−0.017	0.165
Convergence	500		377		500		499		294		58		460	

The results for the alpha parameters are presented in Table 5. Results are limited to cases where the estimator and the misclassification model match. HAS-Probit and HAS-Heckprobit1 performed well in terms of both bias and variance. HAS-Heckprobit2 performed well only in a few cases. Estimates for $\hat{\alpha}_0^{12}$ and $\hat{\alpha}_0^{22}$ were heavily biased and imprecise. A possible explanation for this behavior is the existence of local maxima, as only one set of starting values was tried in each simulation. (Using an alternative set of starting values for α_0^{12} and α_0^{22} did not offer an improvement.) The degree of bias and imprecision seemed to diminish as the number of convergences achieved by HAS-Heckprobit2 increased. However, increasing the number of simulations to 1000 hardly reduced the bias of $\hat{\alpha}_0^{12}$ and $\hat{\alpha}_0^{22}$ (results not shown).

Table 6 shows additional results for PP-Heckprobit when the value of ρ was set alternatively at $\{-0.9, -0.8, 0.9\}$. PP-Heckprobit's good properties were preserved. However, when $\beta_{20} = 2.18$ (low incidental truncation), the number of convergence failures increased somewhat at higher values of ρ.

Table 5. Monte Carlo simulation results for alpha parameters.

	HAS-Probit under Misclassification Model 1							
	$\beta_{20} = 0.5$, $\rho = 0.2$		$\beta_{20} = 0.5$, $\rho = 0.8$		$\beta_{20} = 2.18$, $\rho = 0.2$		$\beta_{20} = 2.18$, $\rho = 0.8$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\alpha}_0$	0.040	0.043	0.170	0.051	−0.011	0.037	0.054	0.037
$\hat{\alpha}_1$	−0.025	0.070	−0.041	0.054	−0.032	0.064	−0.042	0.060
Convergence	491		483		496		494	
	HAS-Heckprobit1 under Misclassification Model 1							
	$\beta_{20} = 0.5$, $\rho = 0.2$		$\beta_{20} = 0.5$, $\rho = 0.8$		$\beta_{20} = 2.18$, $\rho = 0.2$		$\beta_{20} = 2.18$, $\rho = 0.8$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\alpha}_0$	0.047	0.044	−0.058	0.033	−0.011	0.037	−0.075	0.034
$\hat{\alpha}_1$	−0.033	0.070	−0.036	0.041	−0.014	0.063	−0.045	0.057
Convergence	409		476		466		424	

Table 5. *Cont.*

	HAS-Heckprobit2 under Misclassification Model 2							
	$\beta_{20} = 0.5$ $\rho = 0.2$		$\beta_{20} = 0.5$ $\rho = 0.8$		$\beta_{20} = 2.18$ $\rho = 0.2$		$\beta_{20} = 2.18$ $\rho = 0.8$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\alpha}_0^{11}$	0.025	0.048	−0.007	0.032	−0.111	0.038	−0.042	0.036
$\hat{\alpha}_1^{11}$	0.179	0.220	−0.016	0.152	0.128	0.206	0.063	0.192
$\hat{\alpha}_0^{12}$	1.771	0.176	0.953	0.156	1.237	0.162	0.885	0.135
$\hat{\alpha}_1^{12}$	−0.285	0.124	−0.154	0.068	−0.206	0.114	−0.172	0.103
$\hat{\alpha}_0^{21}$	0.061	0.048	0.004	0.035	−0.280	0.039	−0.087	0.037
$\hat{\alpha}_1^{21}$	0.112	0.136	−0.010	0.100	0.046	0.111	0.010	0.106
$\hat{\alpha}_0^{22}$	3.738	0.172	2.249	0.148	2.451	0.147	2.137	0.130
$\hat{\alpha}_1^{22}$	−0.088	0.081	−0.041	0.045	−0.086	0.063	−0.052	0.059
Convergence	197		254		203		197	

Table 6. Monte Carlo simulation results for PP-Heckprobit.

	Misclassification Model 1											
	$\beta_{20} = 0.5$ $\rho = -0.9$		$\beta_{20} = 0.5$ $\rho = -0.8$		$\beta_{20} = 0.5$ $\rho = 0.9$		$\beta_{20} = 2.18$ $\rho = -0.9$		$\beta_{20} = 2.18$ $\rho = -0.8$		$\beta_{20} = 2.18$ $\rho = 0.9$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	0.005	0.100	0.004	0.105	−0.001	0.072	0.002	0.071	0.001	0.072	−0.001	0.066
$\hat{\beta}_{11}$	0.008	0.022	0.006	0.023	0.010	0.023	0.000	0.019	0.000	0.019	0.008	0.019
$\hat{\beta}_{12}$	0.005	0.086	0.003	0.086	0.004	0.079	0.003	0.066	0.001	0.067	0.005	0.065
$\hat{\beta}_{13}$	0.012	0.125	0.006	0.126	0.013	0.110	0.007	0.104	0.005	0.104	0.015	0.102
$\hat{\beta}_{20}$	0.001	0.047	−0.003	0.045	0.003	0.044	0.001	0.089	0.002	0.094	0.005	0.089
$\hat{\beta}_{21}$	0.002	0.024	0.002	0.023	0.002	0.026	0.007	0.045	0.008	0.044	0.009	0.044
$\hat{\beta}_{22}$	0.003	0.022	0.003	0.022	0.002	0.022	0.011	0.040	0.012	0.040	0.001	0.040
$\hat{\rho}$	−0.001	0.044	−0.000	0.060	−0.009	0.057	−0.012	0.094	−0.007	0.117	−0.055	0.152
Convergence	498		500		490		432		482		441	
	Misclassification Model 2											
	$\beta_{20} = 0.5$ $\rho = -0.9$		$\beta_{20} = 0.5$ $\rho = -0.8$		$\beta_{20} = 0.5$ $\rho = 0.9$		$\beta_{20} = 2.18$ $\rho = -0.9$		$\beta_{20} = 2.18$ $\rho = -0.8$		$\beta_{20} = 2.18$ $\rho = 0.9$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	0.004	0.102	0.004	0.106	−0.001	0.073	0.003	0.072	0.001	0.071	−0.001	0.066
$\hat{\beta}_{11}$	0.006	0.022	0.004	0.023	0.008	0.022	0.001	0.019	0.002	0.019	0.007	0.019
$\hat{\beta}_{12}$	0.005	0.086	0.004	0.087	0.005	0.083	0.004	0.065	0.002	0.066	0.006	0.064
$\hat{\beta}_{13}$	0.012	0.123	0.003	0.124	0.010	0.108	0.005	0.100	0.005	0.102	0.014	0.099
$\hat{\beta}_{20}$	0.001	0.047	−0.003	0.045	0.003	0.045	0.000	0.091	0.001	0.094	0.005	0.091
$\hat{\beta}_{21}$	0.002	0.024	0.002	0.023	0.002	0.026	0.007	0.045	0.008	0.044	0.011	0.043
$\hat{\beta}_{22}$	0.003	0.022	0.003	0.022	0.002	0.022	0.010	0.040	0.010	0.040	0.002	0.039
$\hat{\rho}$	−0.001	0.043	−0.001	0.057	−0.008	0.056	−0.016	0.088	−0.010	0.110	−0.064	0.157
Convergence	499		500		488		425		479		430	
	Misclassification Model 3											
	$\beta_{20} = 0.5$ $\rho = -0.9$		$\beta_{20} = 0.5$ $\rho = -0.8$		$\beta_{20} = 0.5$ $\rho = 0.9$		$\beta_{20} = 2.18$ $\rho = -0.9$		$\beta_{20} = 2.18$ $\rho = -0.8$		$\beta_{20} = 2.18$ $\rho = 0.9$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
$\hat{\beta}_{10}$	0.009	0.101	0.010	0.104	0.004	0.072	0.007	0.076	0.008	0.075	0.004	0.067
$\hat{\beta}_{11}$	0.006	0.017	0.008	0.018	0.006	0.018	0.005	0.015	0.005	0.015	0.006	0.015
$\hat{\beta}_{12}$	0.005	0.080	0.004	0.080	0.002	0.076	0.004	0.059	0.003	0.061	0.003	0.060
$\hat{\beta}_{13}$	0.005	0.127	0.002	0.127	0.001	0.111	0.003	0.102	−0.003	0.101	0.002	0.097
$\hat{\beta}_{20}$	0.001	0.045	−0.002	0.046	0.004	0.047	0.001	0.098	0.000	0.095	0.002	0.092
$\hat{\beta}_{21}$	0.003	0.026	0.004	0.026	0.003	0.024	0.007	0.045	0.005	0.044	0.005	0.043
$\hat{\beta}_{22}$	0.003	0.022	0.002	0.022	0.002	0.021	0.010	0.041	0.006	0.042	0.003	0.040
$\hat{\rho}$	−0.001	0.041	−0.003	0.056	−0.000	0.052	−0.016	0.091	−0.017	0.120	−0.022	0.127
Convergence	500		500		489		439		477		429	

4. Application to a Lifetime Migration Model

Data from the Survey of Financial Competences (referred to here by its Spanish abbreviation ECF) were used by González Chapela (2022) to investigate whether cross-region migrants in Spain are less impatient than individuals who choose to remain in their birth region. Previous studies on the empirical link between time preference and migration involve small samples or do not control for cognitive skills (Gibson and McKenzie 2011; Arcand and Mbaye 2013; Nowotny 2014; Goldbach and Schlüter 2018). The ECF (Banco de España and National Securities Market Commission 2018), conducted in Spain in 2016, enables the empirical link between time preference and migration to be investigated on the basis of a large sample and purged of the influence of cognitive skills.

The residential history available in the ECF is limited to the region of birth and the region of residence at the time of the survey, thus raising misclassification of lifetime cross-region migrant status. As argued by Molloy et al. (2011), some true migrants will have returned to their birth region after having spent time elsewhere, whereas individuals who moved when they were still a member of their parents' household are indistinguishable in the data from individuals who moved during their adult lives. The 2011 edition of the Spanish Population Census, conducted by the National Statistics Institute (INE, www.ine.es, accessed on 8 July 2020), provides a validation sample for migrant status. In addition to information on residence at birth and at the Census date, the Census indicates the year of arrival in the region of residence, which reveals interim moves between birth and the census date, and, in conjunction with information on minimum working age, provides a basis for inferring the autonomy of migration decisions.

The conditional probabilities of misclassifying migrant status, estimated using Census data, are incorporated into (11), which is intended to be maximized on a sample of 7129 Spanish natives drawn from the ECF. However, 433 of them have their birth region undisclosed by the ECF to preserve confidentiality, so maximization was conducted on the subsample of 6696 individuals with observable migrant status ($s_i = 1$). Individuals in the subsample are significantly different from individuals with undisclosed birth region in some observables. If individuals in the subsample were also selected in terms of unobservables affecting their true migrant status, the results would be contaminated by sample selection bias. The estimator developed in this paper makes it possible to conduct a sensitivity analysis of the results reported in González Chapela (2022) to sample selection bias.

We let y_i take on value 1 if individual i resides in a region other than that where he/she was born and value 0 if i resides in the same region where he/she was born. The population proportion of out-of-birth region residents is 14.5% in the ECF and 19.3% in the Census.[6] The proportion of true migrants, calculated in the Census, is 17.4%. The lower proportion of true migrants is the result of 27% of true migrants returning to their birth region (false nonmigrants) and 8% of true nonmigrants migrating nonautonomously as a child (false migrants). The probabilities of misclassification are not constant but vary with observed personal attributes. These probabilities are estimated on very large samples and hence are considered as known parameters.

The ECF included a Money Earlier or Later (MEL) task to measure time preferences (e.g., Cohen et al. 2020). Respondents were presented sequentially with two hypothetical binary choices between immediate and delayed monetary rewards. According to their choices, they were sorted into four groups, which are described in terms of required rates of return (*RRR*s):[7] below 4.9% (22.7% of the sample), between 4.9% and 9.8% (10.7%), between 9.8% and 44.9% (29.6%), and above 44.9% (37.0%). Higher levels of *RRR* reflect greater impatience.

Results were developed for six specifications of X_{1i}, corresponding to three functions of *RRR* and two sets of controls. Time preference was measured alternatively with indicators for *RRR* group, an indicator for an *RRR* > 9.8%, and a quadratic function of *RRR*.[8] The first set of controls comprised sex, single-year age group, and birth region. These characteristics are conceivably exogenous to an individual's mobility decisions and appear related in the data (in the case of age and sex) to time preference. The second set of controls added attained education, the number of books at home at the age of 10, cognitive skills, willingness to take risks in financial matters, and the marginal propensity to consume (MPC) from windfall income (in percentage terms). Education and the number of books read might simultaneously increase a person's ability to appreciate the future (Becker and Mulligan 1997) and to live in other places, confounding the relation of interest. Likewise, impatience appears systematically related to risk aversion and cognitive ability (e.g., Dohmen et al. 2010), which are strong predictors of geographic mobility (e.g., Jaeger et al. 2010; Bütikofer and Peri 2021). Krupka and Stephens (2013) found that measured rates of time preference are responsive to individuals' immediate economic conditions. In this respect, the MPC may control for individuals' economic resources at the interview date. X_{2i} comprises these same regressors except birth region (which is unknown for individuals with unobservable migrant status) and the individual's age (which was entered as a linear trend), plus the population of the region of residence on 1 January 2017 (taken from INE *Cifras Oficiales de Población*).

Before presenting the results, an issue requires some discussion. For the six specifications given above, PP-Heckprobit produced $\hat{\rho}$ near +1.0 with no or an implausible standard error. As demonstrated in Butler (1996), this problem is caused by a sample with no observations for which $s_i = 1$ and $X'_{1i}\hat{\beta}_1 \geq X'_{2i}\hat{\beta}_2$. A value of ρ on the boundary of the parameter space invalidates ML standard errors. Following Butler (1996), the standard errors for $\hat{\beta}$ can be calculated by re-estimating the model under the assumption that $\rho = +1.0$. The components of the likelihood function for $\rho = +1.0$ are presented in Table 7. These components are simple adaptations of those derived by Butler (1996) to a mismeasured response variable.

Table 7. Likelihood function for $\rho = +1.0$ in the probit model with sample selection and misclassification.

Condition	Likelihood Function
$s_i = 0:$	$\Phi(-X'_{2i}\beta_2)$
$s_i = 1$ and:	
$y_i = 0,\ X'_{1i}\beta_1 \geq X'_{2i}\beta_2:$	$\alpha_{1i}\Phi(X'_{2i}\beta_2)$
$y_i = 0,\ X'_{1i}\beta_1 < X'_{2i}\beta_2:$	$\alpha_{1i}\Phi(X'_{2i}\beta_2) + (1-\alpha_{0i}-\alpha_{1i})(\Phi(X'_{2i}\beta_2) - \Phi(X'_{1i}\beta_1))$
$y_i = 1,\ X'_{1i}\beta_1 \geq X'_{2i}\beta_2:$	$(1-\alpha_{1i})\Phi(X'_{2i}\beta_2)$
$y_i = 1,\ X'_{1i}\beta_1 < X'_{2i}\beta_2:$	$\alpha_{0i}\Phi(X'_{2i}\beta_2) + (1-\alpha_{0i}-\alpha_{1i})\Phi(X'_{1i}\beta_1)$

Notes: Entries for $s_i = 1$ show $P(y_i = y|s_i = 1)P(s_i = 1)$.

The estimation output yielded by PP-Heckprobit is presented in Table 8 for the six specifications pointed out above. For comparison purposes, Table 9 shows Probit estimates of the reduced-form Equation (2) and PP-Probit estimates of the outcome Equation (1). If ρ was zero, the sum of the log-likelihood values from these two equations would equal the log likelihood of the probit model with sample selection and misclassification.

Table 8. PP-Heckprobit estimates.

	(1)		(2)		(3)		(4)		(5)		(6)	
Explanatory Variables	**Selection**	**Outcome**	**Selection**	**Outcome**	**Selection**	**Outcome**	**Selection**	**Outcome**	**Selection**	**Outcome**	**Selection**	**Outcome**
$1(4.9\% < RRR \leq 9.8\%)$	−0.040	0.009	−0.062	0.001								
	(0.089)	(0.184)	(0.091)	(0.191)								
$1(9.8\% < RRR \leq 44.9\%)$	0.057	−0.250 *	0.018	−0.250								
	(0.070)	(0.150)	(0.071)	(0.158)								
$1(44.9\% < RRR)$	0.059	−0.166	−0.039	−0.141								
	(0.067)	(0.152)	(0.071)	(0.165)								
$1(9.8\% < RRR)$					0.071	−0.207 *	0.008	−0.194				
					(0.053)	(0.117)	(0.055)	(0.120)				
RRR									0.340	−1.334 *	0.155	−1.328 *
									(0.337)	(0.720)	(0.345)	(0.744)
RRR^2									−0.068	0.273 *	−0.034	0.273 *
									(0.070)	(0.148)	(0.071)	(0.153)
Male	0.015	−0.037	0.013	0.023	0.016	−0.040	0.015	0.020	0.016	−0.038	0.014	0.023
	(0.050)	(0.108)	(0.053)	(0.111)	(0.050)	(0.112)	(0.053)	(0.112)	(0.050)	(0.108)	(0.053)	(0.111)
Lower secondary education			−0.364 ***	−0.108			−0.364 ***	−0.114			−0.363 ***	−0.108
			(0.099)	(0.186)			(0.099)	(0.186)			(0.099)	(0.186)
Upper secondary			−0.239 **	−0.006			−0.235 **	−0.011			−0.239 **	−0.005
			(0.108)	(0.199)			(0.108)	(0.198)			(0.108)	(0.196)
Higher education			−0.434 ***	0.279			−0.424 ***	0.264			−0.431 ***	0.278
			(0.112)	(0.219)			(0.111)	(0.217)			(0.112)	(0.216)
11–25 books at home			−0.016	−0.149			−0.014	−0.147			−0.015	−0.146
			(0.081)	(0.183)			(0.081)	(0.183)			(0.081)	(0.180)
26–100			0.008	0.045			0.009	0.046			0.009	0.050
			(0.083)	(0.179)			(0.083)	(0.183)			(0.083)	(0.174)
101–200			−0.165	−0.130			−0.164	−0.125			−0.166	−0.128
			(0.102)	(0.262)			(0.101)	(0.267)			(0.102)	(0.259)
>200			−0.235 **	−0.200			−0.236 **	−0.202			−0.235 **	−0.201
			(0.096)	(0.270)			(0.096)	(0.268)			(0.096)	(0.265)
Numeracy skills			−0.129 **	0.076			−0.127 **	0.081			−0.129 **	0.076
			(0.055)	(0.146)			(0.055)	(0.147)			(0.055)	(0.145)
Reading comprehension			−0.041	0.047			−0.040	0.045			−0.040	0.047
			(0.036)	(0.068)			(0.036)	(0.068)			(0.036)	(0.068)
Cognitive reflection			0.046	−0.059			0.048	−0.068			0.047	−0.060
			(0.062)	(0.133)			(0.062)	(0.131)			(0.062)	(0.132)

Table 8. *Cont.*

Explanatory Variables	(1) Selection	(1) Outcome	(2) Selection	(2) Outcome	(3) Selection	(3) Outcome	(4) Selection	(4) Outcome	(5) Selection	(5) Outcome	(6) Selection	(6) Outcome
Risk score			−0.032	−0.092 *			−0.031	−0.093 *			−0.033	−0.093 *
			(0.020)	(0.049)			(0.020)	(0.049)			(0.020)	(0.048)
MPC (÷10)			0.001	−0.044 *			0.001	−0.043 *			0.001	−0.044 *
			(0.008)	(0.024)			(0.008)	(0.024)			(0.008)	(0.024)
Age	−0.001		−0.006 ***		−0.001		−0.006 ***		−0.001		−0.006 ***	
	(0.001)		(0.002)		(0.001)		(0.002)		(0.001)		(0.002)	
Region population (10^6)	0.213 ***		0.215 ***		0.213 ***		0.215***		0.213 ***		0.215 ***	
	(0.019)		(0.019)		(0.019)		(0.019)		(0.019)		(0.019)	
Intercept	1.014 ***	−0.897 ***	1.877 ***	−0.671	1.002 ***	−0.886 ***	1.850 ***	−0.638	0.993 ***	−0.847 ***	1.857 ***	−0.625
	(0.093)	(0.320)	(0.202)	(0.425)	(0.090)	(0.309)	(0.199)	(0.407)	(0.093)	(0.320)	(0.201)	(0.426)
Log-likelihood	−3653.97		−3609.54		−3654.31		−3610.53		−3654.28		−3609.91	

Notes: The number of observations is 7129. Regressors in the outcome equation include indicators for single-year age group and birth region. 1(·) is the indicator function. Robust standard errors are in parentheses. *: Significant at 10%. **: Significant at 5%. ***: Significant at 1%.

Table 9. Probit (selection equation) and PP-Probit (outcome equation) estimates.

Explanatory Variables	(1) Selection	(1) Outcome	(2) Selection	(2) Outcome	(3) Selection	(3) Outcome	(4) Selection	(4) Outcome	(5) Selection	(5) Outcome	(6) Selection	(6) Outcome
$1(4.9\% < RRR \leq 9.8\%)$	−0.038	0.015	−0.061	0.013								
	(0.089)	(0.181)	(0.090)	(0.192)								
$1(9.8\% < RRR \leq 44.9\%)$	0.053	−0.255 *	0.011	−0.248								
	(0.069)	(0.150)	(0.071)	(0.156)								
$1(44.9\% < RRR)$	0.055	−0.173	−0.045	−0.131								
	(0.067)	(0.155)	(0.071)	(0.166)								
$1(9.8\% < RRR)$					0.066	−0.216 *	0.002	−0.192				
					(0.053)	(0.126)	(0.055)	(0.120)				
RRR									0.318	−1.371 *	0.118	−1.325 *
									(0.336)	(0.738)	(0.344)	(0.744)
RRR^2									−0.064	0.281 *	−0.026	0.273 *
									(0.070)	(0.152)	(0.071)	(0.153)
Male	0.014	−0.040	0.008	0.023	0.015	−0.045	0.011	0.018	0.015	−0.044	0.009	0.023
	(0.050)	(0.109)	(0.053)	(0.113)	(0.050)	(0.117)	(0.053)	(0.113)	(0.050)	(0.111)	(0.053)	(0.114)

Table 9. *Cont.*

Explanatory Variables	(1) Selection	(1) Outcome	(2) Selection	(2) Outcome	(3) Selection	(3) Outcome	(4) Selection	(4) Outcome	(5) Selection	(5) Outcome	(6) Selection	(6) Outcome
Lower secondary education			−0.364 ***	−0.074			−0.364 ***	−0.081			−0.363***	−0.075
			(0.098)	(0.189)			(0.098)	(0.189)			(0.098)	(0.190)
Upper secondary			−0.237 **	0.010			−0.234 **	0.004			−0.236 **	0.009
			(0.108)	(0.192)			(0.107)	(0.190)			(0.108)	(0.192)
Higher education			−0.432 ***	0.328			−0.423 ***	0.310			−0.430 ***	0.325
			(0.111)	(0.216)			(0.111)	(0.210)			(0.111)	(0.214)
11–25 books at home			−0.016	−0.146			−0.013	−0.144			−0.014	−0.142
			(0.081)	(0.174)			(0.081)	(0.173)			(0.081)	(0.173)
26–100			0.015	0.052			0.016	0.054			0.015	0.059
			(0.082)	(0.158)			(0.083)	(0.158)			(0.082)	(0.156)
101–200			−0.155	−0.119			−0.155	−0.113			−0.156	−0.116
			(0.101)	(0.252)			(0.101)	(0.255)			(0.101)	(0.253)
>200			−0.231 **	−0.178			−0.232 **	−0.180			−0.231 **	−0.178
			(0.096)	(0.268)			(0.096)	(0.262)			(0.096)	(0.265)
Numeracy skills			−0.127 **	0.096			−0.125 **	0.101			−0.127 **	0.097
			(0.055)	(0.148)			(0.055)	(0.148)			(0.055)	(0.148)
Reading comprehension			−0.041	0.049			−0.040	0.047			−0.040	0.049
			(0.036)	(0.070)			(0.036)	(0.070)			(0.036)	(0.070)
Cognitive reflection			0.047	−0.070			0.049	−0.080			0.048	−0.071
			(0.062)	(0.136)			(0.062)	(0.135)			(0.062)	(0.136)
Risk score			−0.031	−0.090 *			−0.030	−0.091 *			−0.031	−0.091 *
			(0.020)	(0.050)			(0.020)	(0.050)			(0.020)	(0.050)
MPC (÷10)			0.001	−0.042 *			0.001	−0.042 *			0.001	−0.043 *
			(0.008)	(0.025)			(0.008)	(0.025)			(0.008)	(0.025)
Age	−0.001		−0.006 ***		−0.001		−0.006 ***		−0.001		−0.006 ***	
	(0.001)		(0.002)		(0.001)		(0.002)		(0.001)		(0.002)	
Region population (10^6)	0.208 ***		0.210 ***		0.208 ***		0.210 ***		0.208 ***		0.210 ***	
	(0.019)		(0.019)		(0.019)		(0.019)		(0.019)		(0.019)	
Intercept	1.025 ***	−0.927 ***	1.875 ***	−0.767 *	1.014 ***	−0.912 ***	1.848 ***	−0.724 *	1.006 ***	−0.872 ***	1.856 ***	−0.717
	(0.094)	(0.332)	(0.202)	(0.448)	(0.090)	(0.330)	(0.198)	(0.428)	(0.093)	(0.338)	(0.201)	(0.451)
Log-likelihood	−1470.68	−2189.85	−1438.48	−2178.17	−1470.77	−2190.05	−1439.06	−2178.58	−1470.86	−2189.96	−1438.75	−2178.27
Observations	7129	6696	7129	6696	7129	6696	7129	6696	7129	6696	7129	6696

Notes: Regressors in the outcome equation include indicators for single-year age group and birth region. 1(·) is the indicator function. Robust standard errors are in parentheses. *: Significant at 10%. **: Significant at 5%. ***: Significant at 1%.

As the selection equation was affected neither by sample selection nor by misclassification of the dependent variable, it is not surprising that $\hat{\beta}_2$ is similar in the two tables. The population of the region of residence was a strong predictor of having the birth region disclosed by the ECF, with the probability of disclosing the birth region increasing with population size. Attained education had a negative effect on this probability, with having a higher education showing the strongest effect. Having more than 200 books at home at the age of 10 was negatively associated with the probability of disclosing the birth region. Individuals' numeracy skills and (in the full specification) age exerted a negative effect as well.

Correcting for possible sample selection bias using PP-Heckprobit had little effect on the estimated coefficients and associated standard errors yielded by PP-Probit, thus leaving the main results of González Chapela (2022) almost unchanged. The *RRR* for financial flows and the probability of ever migrating appeared to be inversely related even after accounting for individuals' cognitive skills. Of course, the sample selection correction might be important in other contexts with different probabilities of selection and different correlation between error terms.

5. Conclusions

In many cases, the assumption that misclassification of a binary outcome is conditionally random does not hold. Applying an estimator that assumes conditionally random misclassification can make estimates worse when this assumption is false. This paper has extended Arezzo and Guagnano's (2019) formulation for a misclassified binary response variable affected by sample selection to incorporate misclassification related to the covariates. Assuming the availability of validation data, the proposed pseudo-maximum likelihood estimator is consistent and asymptotically efficient. A Monte Carlo study documents the good performance of the proposed estimator under different models of misclassification, probabilities of selection, and correlation between error terms. We have also illustrated our method with an empirical example, examining the impact of time preference on the propensity to migrate.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/econometrics10020013/s1. Program files (Stata 16) that generated the data used in the Monte Carlo study and that analyzed the dataset used in the illustrative study.

Funding: This research was funded by the Government of Aragón, grant number S32−20R.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The program files that generated the data used in the Monte Carlo study and that analyzed the dataset used in the illustrative study are included in the electronic Supplementary Materials. The dataset analyzed in the illustrative study was constructed from publicly available data published by the Banco de España, Spain's National Securities Market Commission, and Spain's National Statistics Institute. Instructions for how other researchers can obtain these data can be found in the Supplementary Materials.

Acknowledgments: Two anonymous reviewers provided helpful comments and suggestions.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Maximum Likelihood Evaluator

In this appendix, we derive the analytic first and second derivatives of the log-likelihood function (10) and write out our method-lf2 likelihood evaluator (see Gould et al. 2010) implemented in Stata (version 16).

The log-likelihood function (10) can be written more compactly as

$$l(\beta,\rho) = \sum_{i=1}^{N} s_i \ln[\mathrm{cc}_i \Phi(w_{2i}) + \mathrm{om}_i \Phi_2(w_{1i}, w_{2i}, \rho_i^*)] + (1-s_i)\ln\Phi(w_{2i}) \tag{A1}$$

where $\mathrm{cc}_i = y_i\alpha_{0i} + (1-y_i)\alpha_{1i}$, $w_{1i} = q_{1i}z_{1i}$, $w_{2i} = q_{2i}z_{2i}$, $z_{1i} = X'_{1i}\beta_1$, $z_{2i} = X'_{2i}\beta_2$, $q_{1i} = 2y_i^T - 1$, $q_{2i} = 2s_i - 1$, $\rho_i^* = q_{1i}q_{2i}\rho$, and $\mathrm{om}_i = 1 - \alpha_{0i} - \alpha_{1i}$. To accommodate the restriction that in general $-1 < \rho < 1$, maximization is not conducted directly with respect to ρ, but with respect to its inverse hyperbolic tangent (atanh ρ):

$$a \equiv \operatorname{atanh}\rho = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right). \tag{A2}$$

We first obtain the first derivatives of (A1) for $s_i = 1$, using Greene (2003, p. 711) as a reference:

$$\frac{\partial l}{\partial z_{1i}} = \sum_{i=1}^{N} q_{1i}\frac{\mathrm{om}_i k_{1i}}{d_i} \tag{A3}$$

$$\frac{\partial l}{\partial z_{2i}} = \sum_{i=1}^{N} q_{2i}\frac{\mathrm{cc}_i\phi(w_{2i}) + \mathrm{om}_i k_{2i}}{d_i} \tag{A4}$$

$$\frac{\partial l}{\partial a} = \sum_{i=1}^{N}\left[q_{1i}q_{2i}\frac{\mathrm{om}_i\phi_{2i}}{d_i}\right]\left(1-(\tanh a)^2\right) \tag{A5}$$

where

$$k_{1i} = \phi(w_{1i})\Phi\left[\frac{w_{2i} - \rho_i^* w_{1i}}{\sqrt{1-\rho_i^{*2}}}\right] \tag{A6}$$

the subscripts 1 and 2 in k_{1i} are reversed to obtain k_{2i}, $d_i = \mathrm{cc}_i\Phi(w_{2i}) + \mathrm{om}_i\Phi_2(w_{1i}, w_{2i}, \rho_i^*)$, $\phi_{2i} = \phi_2(w_{1i}, w_{2i}, \rho_i^*)$, ϕ_2 being the bivariate normal density, and

$$\tanh a = \frac{e^a - e^{-a}}{e^a + e^{-a}}. \tag{A7}$$

The first derivatives for $s_i = 0$ are all zero except

$$\frac{\partial l}{\partial z_{2i}} = \sum_{i=1}^{N} q_{2i}\frac{\phi(w_{2i})}{\Phi(w_{2i})}. \tag{A8}$$

To obtain the second derivatives for $s_i = 1$, we use the result

$$\delta_i\phi(w_{1i})\phi(v_{1i}) = \delta_i\phi(w_{2i})\phi(v_{2i}) = \phi_{2i} \tag{A9}$$

where $\delta_i = 1/\sqrt{1-\rho_i^{*2}}$, $v_{1i} = \delta_i(w_{2i} - \rho_i^* w_{1i})$, and $v_{2i} = \delta_i(w_{1i} - \rho_i^* w_{2i})$; see Greene (2003). Thus:

$$\frac{\partial^2 l}{\partial z_{1i}^2} = \sum_{i=1}^{N}\mathrm{om}_i\left[-\frac{w_{1i}k_{1i}}{d_i} - \frac{\rho_i^*\phi_{2i}}{d_i} - \frac{\mathrm{om}_i k_{1i}^2}{d_i^2}\right] \tag{A10}$$

$$\frac{\partial^2 l}{\partial z_{1i}\partial z_{2i}} = \sum_{i=1}^{N}\mathrm{om}_i q_{1i}q_{2i}\left[\frac{\phi_{2i}}{d_i} - \frac{\mathrm{om}_i k_{1i}k_{2i}}{d_i^2} - \frac{\mathrm{cc}_i\phi(w_{2i})k_{1i}}{d_i^2}\right] \tag{A11}$$

$$\frac{\partial^2 l}{\partial z_{1i}\partial a} = \sum_{i=1}^{N}\mathrm{om}_i q_{2i}\frac{\phi_{2i}}{d_i}\left[\rho_i^*\delta_i v_{1i} - w_{1i} - \frac{\mathrm{om}_i k_{1i}}{d_i}\right]\left(1-(\tanh a)^2\right) \tag{A12}$$

$$\frac{\partial^2 l}{\partial z_{2i}^2} = \sum_{i=1}^{N}\left[-\frac{\mathrm{om}_i w_{2i}k_{2i}}{d_i} - \frac{\mathrm{om}_i\rho_i^*\phi_{2i}}{d_i} - \frac{\mathrm{cc}_i w_{2i}\phi(w_{2i})}{d_i} - \frac{(\mathrm{om}_i k_{2i} + \mathrm{cc}_i\phi(w_{2i}))^2}{d_i^2}\right] \tag{A13}$$

$$\frac{\partial^2 l}{\partial z_{2i}\partial a} = \sum_{i=1}^{N} \mathrm{om}_i q_{1i} \frac{\phi_{2i}}{d_i}\left[\rho_i^* \delta_i v_{2i} - w_{2i} - \frac{\mathrm{om}_i k_{2i} + \mathrm{cc}_i \phi(w_{2i})}{d_i}\right]\left(1 - (\tanh a)^2\right) \quad \text{(A14)}$$

$$\begin{aligned}\frac{\partial^2 l}{\partial a^2} = \sum_{i=1}^{N} \Big\{ \mathrm{om}_i \tfrac{\phi_{2i}}{d_i} \big[\delta_i^2 \rho_i^* \big(1 - \delta_i^2 \big(w_{1i}^2 + w_{2i}^2 - 2\rho_i^* w_{1i} w_{2i}\big)\big) + \delta_i^2 w_{1i} w_{2i} \\ - \tfrac{\mathrm{om}_i \phi_{2i}}{d_i}\big]\left(1 - (\tanh a)^2\right)^2 - 2\tanh a \tfrac{\mathrm{om}_i q_{1i} q_{2i} \phi_{2i}}{d_i}\left(1 - (\tanh a)^2\right)\Big\}\end{aligned} \quad \text{(A15)}$$

The second derivatives for $s_i = 0$ are all zero except

$$\frac{\partial^2 l}{\partial z_{2i}^2} = -\sum_{i=1}^{N} \frac{\phi(w_{2i})}{\Phi(w_{2i})}\left[w_{2i} + \frac{\phi(w_{2i})}{\Phi(w_{2i})}\right]. \quad \text{(A16)}$$

Referring to α_{0i}, α_{1i}, and om_i as alpha0_i, alpha1_i, and om_i, respectively, our method-lf2 likelihood evaluator can be written as:

```
program PP_Heckprobit_lf2
  version 16.1
  args todo b lnfj g1 g2 g3 H
  tempvar z1 z2 q1 q2 w1 w2 rhos delta v1 v2 k1 k2 bden qf cc d
  tempname a
  mleval ‘z1’=‘b’, eq(1)
  mleval ‘z2’=‘b’, eq(2)
  mleval ‘a’=‘b’, eq(3) scalar
  quietly {
  gen double ‘q1’=(2*$ML_y1)-1
  gen double ‘q2’=(2*$ML_y2)-1
  gen double ‘w1’=‘q1’*‘z1’
  gen double ‘w2’=‘q2’*‘z2’
  gen double ‘rhos’=‘q1’*‘q2’*tanh(‘a’)
  gen double ‘delta’=1/sqrt(1-(‘rhos’^2))
  gen double ‘v1’=‘delta’*(‘w2’-(‘rhos’*‘w1’))
  gen double ‘v2’=‘delta’*(‘w1’-(‘rhos’*‘w2’))
  gen double ‘k1’=normalden(‘w1’)*normal(‘v1’)
  gen double ‘k2’=normalden(‘w2’)*normal(‘v2’)
  gen double ‘bden’=normalden(‘w2’)*normalden(‘v2’)*‘delta’
  gen double ‘qf’=(‘delta’^2)*‘rhos’*(1-((‘delta’^2)*((‘w1’^2) +(‘w2’^2)-(2*‘rhos’*‘w1’*‘w2’))))
  gen double ‘cc’=(alpha0_i*$ML_y1) + (alpha1_i*(1-$ML_y1))
  gen double ‘d’=(om_i*binormal(‘w1’,‘w2’,‘rhos’)) + (‘cc’*normal(‘w2’))
  replace ‘lnfj’=ln(‘d’) if $ML_y2==1
  replace ‘lnfj’=ln(normal(‘w2’)) if $ML_y2==0
  if (‘todo’==0) exit
  replace ‘g1’=om_i*‘q1’*‘k1’/‘d’ if $ML_y2==1
  replace ‘g1’=0 if $ML_y2==0
  replace ‘g2’=‘q2’*((om_i*‘k2’)+(‘cc’*normalden(‘w2’)))/‘d’ if $ML_y2==1
  replace ‘g2’=‘q2’*normalden(‘w2’)/normal(‘w2’) if $ML_y2==0
  replace ‘g3’=(om_i*‘q1’*‘q2’*‘bden’/‘d’)*(1-(tanh(‘a’)^2)) if $ML_y2==1
  replace ‘g3’=0 if $ML_y2==0
  if (‘todo’==1) exit
  tempvar h11 h12 h13 h22 h23 h33
  gen double ‘h11’=-om_i*((‘w1’*‘k1’/‘d’)+(‘rhos’*‘bden’/‘d’) + (om_i*((‘k1’/‘d’)^2))) if ///
$ML_y2==1
  replace ‘h11’=0 if $ML_y2==0
  gen double ‘h12’=om_i*‘q1’*‘q2’*((‘bden’/‘d’)-(om_i*‘k1’*‘k2’/(‘d’^2))
///-(‘cc’*normalden(‘w2’)*‘k1’/(‘d’^2))) if $ML_y2==1
  replace ‘h12’=0 if $ML_y2==0
  gen double ‘h13’=(om_i*‘q2’*‘bden’*((‘rhos’*‘delta’*‘v1’)-‘w1’
///-(om_i*‘k1’/‘d’))/‘d’)*(1-(tanh(‘a’)^2)) if $ML_y2==1
  replace ‘h13’=0 if $ML_y2==0
```

```
    gen double 'h22'=-(om_i*'w2'*'k2'/'d')-(om_i*'rhos'*'bden'/'d')
///-('cc'*'w2'*normalden('w2')/'d')-((((om_i*'k2')+('cc'*normalden('w2')))/'d')^2) if ///
$ML_y2==1
    replace 'h22'=-normalden('w2')*('w2'+(normalden('w2')/normal('w2')))/normal('w2') /// if
$ML_y2==0
    gen double 'h23'=(om_i*'q1'*'bden'*(('rhos'*'delta'*'v2')-'w2'-(((om_i*'k2')
///+('cc'*normalden ('w2')))/'d'))/'d')*(1-(tanh('a')^2)) if $ML_y2==1
    replace 'h23'=0 if $ML_y2==0
    gen double 'h33'=(((om_i*'bden'*('qf'+(('delta'^2)*'w1'*'w2')
///-(om_i*'bden'/'d'))/'d')*(1-(tanh('a')^2)))-(2*tanh('a')*(om_i*'q1'*'q2'*'bden'/'d')))
///*(1-(tanh('a')^2)) if $ML_y2==1
    replace 'h33'=0 if $ML_y2==0
    tempname d11 d12 d13 d22 d23 d33
    mlmatsum 'lnfj' 'd11'='h11', eq(1)
    mlmatsum 'lnfj' 'd12'='h12', eq(1,2)
    mlmatsum 'lnfj' 'd13'='h13', eq(1,3)
    mlmatsum 'lnfj' 'd22'='h22', eq(2)
    mlmatsum 'lnfj' 'd23'='h23', eq(2,3)
    mlmatsum 'lnfj' 'd33'='h33', eq(3)
    matrix 'H'=('d11','d12','d13'\'d12'','d22','d23'\'d13'','d23'','d33')
    }
end
```

Notes

1 See Train (2009) for a good treatment of numerical maximization.

2 These derivatives are shown in Appendix A along with the maximum likelihood evaluator.

3 Except the coefficient on the intercept included in X_{1i}, which is the LPM estimate of the coefficient on $(1 - \alpha_{0i} - \alpha_{1i})$ in (14) minus 0.5 multiplied by 2.5 (Amemiya 1981).

4 Probit and Heckprobit were implemented using the homonymous Stata commands. The other five estimators were implemented in Stata using programs written by the author and available in the supplementary material.

5 Convergence is accepted if the Hessian is negative definite and the scaled gradient is lower than 1^{-8}.

6 If all individuals with undisclosed birth region were out-of-birth region residents, the proportion of migrants (in this sense) in the ECF would rise to 17.6%. Given that the ECF is an individual survey, its lower migration rates might be the consequence of a greater probability of survey noncontact among movers, reducing the proportion of migrants in the sample. However, results in Imbens (1992) suggest that small amounts of endogenous sampling are unlikely to substantially alter estimated parameters. In addition, to guard against possible misspecification, our inference is based on robust estimators of variance.

7 When a respondent is indifferent between €m1 today and €m_2 in a year's time, the *RRR* necessary to induce her/him to forgo €m1 immediately is $2((m_2/m_1)^{1/2} - 1)$. This definition assumes semiannual compounding of the annual interest rate as a natural compromise between the types of compounding that Spaniards are most familiar with.

8 *RRR* is treated as a continuous variable by predicting the conditional mean for each *RRR* group from a lognormal curve fitted to the distribution of *RRR* data.

References

Aller, Carlos, and Jorge González Chapela. 2013. Misclassification of the dependent variable in a debt–repayment behavior context. *Journal of Empirical Finance* 23: 162–72. [CrossRef]

Amemiya, Takeshi. 1981. Qualitative response models: A survey. *Journal of Economic Literature* 19: 1483–536.

Arcand, Jean-Louis, and Linguere Mbaye. 2013. *Braving the Waves: The Role of Time and Risk Preferences in Illegal Migration from Senegal*. IZA Discussion Paper No. 7517. Bonn: Institute for the Study of Labor.

Arezzo, Maria Felice, and Giuseppina Guagnano. 2019. Misclassification in binary choice models with sample selection. *Econometrics* 7: 32. [CrossRef]

Banco de España and National Securities Market Commission. 2018. Survey of Financial Competences (ECF) 2016. Available online: https://app.bde.es/pmk/en/ecf/2016 (accessed on 17 January 2020).

Becker, Gary S., and Casey B. Mulligan. 1997. The endogenous determination of time preference. *Quarterly Journal of Economics* 112: 729–58. [CrossRef]

Bollinger, Christopher R., and Martin H. David. 1997. Modeling discrete choice with response error: Food Stamp participation. *Journal of the American Statistical Association* 92: 827–35. [CrossRef]

Bollinger, Christopher R., and Martin H. David. 2001. Estimation with response error and nonresponse: Food-Stamp participation in the SIPP. *Journal of Business & Economic Statistics* 19: 129–41.
Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. Measurement error in survey data. In *Handbook of Econometrics*. Edited by James J. Heckman and Edward Leamer. Amsterdam: Elsevier, vol. 5, pp. 3705–843.
Bütikofer, Aline, and Giovanni Peri. 2021. How cognitive ability and personality traits affect geographic mobility. *Journal of Labor Economics* 39: 559–95. [CrossRef]
Butler, John S. 1996. Estimating the correlation in censored probit models. *Review of Economics and Statistics* 78: 356–58. [CrossRef]
Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics. Methods and Applications*. Cambridge: CUP.
Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White. 2020. Measuring time preferences. *Journal of Economic Literature* 58: 299–347. [CrossRef]
Davern, Michael, Jacob A. Klerman, Jeanettec Ziegenfuss, Victoriad Lynch, and Georgee Greenberg. 2009. A partially corrected estimate of medicaid enrollment and uninsurance: Results from an imputational model developed off linked survey and administrative data. *Journal of Economic and Social Measurement* 34: 219–40. [CrossRef]
Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. Are risk aversion and impatience related to cognitive ability? *American Economic Review* 100: 1238–60. [CrossRef]
Dubin, Jeffrey A., and Douglas Rivers. 1989. Selection bias in linear regression, logit and probit models. *Sociological Methods & Research* 18: 360–90.
Gibson, John, and David McKenzie. 2011. The microeconomic determinants of emigration and return migration of the best and brightest: Evidence from the Pacific. *Journal of Development Economics* 95: 18–29. [CrossRef]
Goldbach, Carina, and Achim Schlüter. 2018. Risk aversion, time preferences, and out-migration. Experimental evidence from Ghana and Indonesia. *Journal of Economic Behavior and Organization* 150: 132–48. [CrossRef]
González Chapela, Jorge. 2022. Is There a Patience Premium on Migration? *Empirical Economics.* Available online: https://doi.org/10.1007/s00181-021-02196-z (accessed on 10 January 2022).
Gould, William, Jeffrey Pitblado, and William Sribney. 2010. *Maximum Likelihood Estimation with Stata*, 4th ed. College Station: Stata Press.
Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. Pseudo maximum likelihood methods: Theory. *Econometrica* 52: 681–700. [CrossRef]
Greene, William H. 2003. *Econometric Analysis*, 4th ed. Upper Saddle River: Prentice Hall.
Hausman, Jerry A., Jason Abrevaya, and Fiona M. Scott-Morton. 1998. Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics* 87: 239–69. [CrossRef]
Heckman, James. 1974. Shadow prices, market wages, and labor supply. *Econometrica* 42: 679–94. [CrossRef]
Heckman, James. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–62. [CrossRef]
Imbens, Guido W. 1992. An efficient method of moments estimator for discrete choice models with choice-based sampling. *Econometrica* 60: 1187–214. [CrossRef]
Jaeger, David, Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, and Holger Bonin. 2010. Direct evidence on risk attitudes and migration. *Review of Economics and Statistics* 92: 684–89. [CrossRef]
Katz, Jonathan N., and Gabriel Katz. 2010. Correcting for survey misreports using auxiliary information with an application to estimating turnout. *American Journal of Political Science* 54: 815–35. [CrossRef]
Krupka, Erin L., and Melvin Stephens Jr. 2013. The stability of measured time preferences. *Journal of Economic Behavior and Organization* 85: 11–19. [CrossRef]
Levine, Phillip B. 1993. CPS contemporaneous and retrospective unemployment compared. *Monthly Labor Review* 116: 33–39.
Meyer, Bruce D., and Nikolas Mittag. 2017. Misclassification in binary choice models. *Journal of Econometrics* 200: 295–311. [CrossRef]
Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. Internal migration in the United States. *Journal of Economic Perspectives* 25: 173–96. [CrossRef]
Nowotny, Klaus. 2014. Cross-border commuting and migration intentions: The roles of risk aversion and time preference. *Contemporary Economics* 8: 137–56. [CrossRef]
Poterba, James M., and Lawrence H. Summers. 1995. Unemployment benefits and labor market transitions: A multinomial logit model with errors in classification. *Review of Economics and Statistics* 77: 207–16. [CrossRef]
Ramalho, Esmeralda A. 2002. Regression models for choice-based samples with misclassification in the response variable. *Journal of Econometrics* 106: 171–201. [CrossRef]
StataCorp. 2019. *Stata: Release 16. Statistical Software*. College Station: StataCorp LLC.
Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*, 2nd ed. New York: CUP.
Van de Ven, Wynand, and Bernard van Praag. 1981. The demand for deductibles in private health insurance. *Journal of Econometrics* 17: 229–52. [CrossRef]
Vella, Francis. 1998. Estimating models with sample selection bias: A survey. *Journal of Human Resources* 33: 127–69. [CrossRef]