On the Estimation of Treatment Effects with Endogenous Misreporting*

Pierre Nguimkeu†, Augustine Denteh† and Rusty Tchernis†,§

†Department of Economics, Georgia State University
§IZA & NBER

July 29, 2016

Abstract

Participation in social programs is often misreported in survey data, complicating the estimation of the effects of those programs. In this paper we propose a model to estimate treatment effect under endogenous participation and endogenous misreporting. We show that failure to account for endogenous misreporting can result in the estimates of the treatment effect having opposite sign from the true effect. We present an expression for the asymptotic bias of both OLS and IV estimators and discuss the conditions under which sign reversal may occur. We provide a method of eliminating this bias when researchers have access to information related to both participation and misreporting. We establish the consistency and asymptotic normality of our estimator and assess its small sample performance through Monte Carlo simulations. An empirical example is given to illustrate the method.

Keywords: Treatment effect, Misclassification, Endogeneity, Binary regressor, Partial observability, Bias.
JEL Classification: C35, C51,

---

*We thank Jason Abrevaya, Victor Chernozhukov, Steven Durlauf, Kei Hirano, Arthur Lewbel, Nikolas Mittag, Robert Moffitt, Tom Mroz, Alistair O′Malley, Seth Richards-Shubik, Gary Solon, and conference participants at the Tinbergen Institute (Amsterdam, The Netherlands), IZA (Bonn, Germany), CERGE-EI (Prague, Czech Republic), the Southern Economic Association (New-Orleans), and the American Society of Health Economists (Philadelphia) for helpful comments. Corresponding author: Pierre Nguimkeu, Department of Economics, Andrew Young School of Policy Studies, Georgia State University, P.O. Box 3992, Atlanta, GA 30302-3992, e-mail: nnguimkeu@gsu.edu.

1 Introduction

This paper proposes a solution to the problem of identification and estimation of treatment effects in parametric regressions when participation is endogenously misreported. In particular, we provide a two-step estimation procedure that consistently estimates the conditional average treatment effect. Participation in social programs is substantially misreported in survey data, sometimes with misreporting levels as high as 50% (Meyer et al. 2009). When a binary regressor is misreported (or misclassified), the measurement error is necessarily negatively correlated with the underlying true value of the regressor, thus making the classical measurement error assumptions implausible[1]. While earlier papers (Aigner 1973, Lewbel 2007) show that exogenous misreporting leads to attenuation bias, we demonstrate that the effects of endogenous misreporting are much more severe. To our knowledge, this paper is the first attempt to address endogenous misreporting.

Misreporting occurs when program participants report not receiving treatment when they actually did ("false negatives") or vice versa ("false positives"). False negatives are pervasive in practice and in many empirical studies. For example, Lynch et al. (2007) and Meyer & Goerge (2011) report that validation studies always find high rates of false negatives in the Food Stamps Program ranging from 20% to 40%. Marquis & Moore (2010) and Meyer & Goerge (2011) find up to 50% rate of false negatives in the CPS Annual Social and Economic Supplement.

False negatives are not confined to government programs. For example, according to Bound (1991), there are a number of reasons to be suspicious of any survey response to questions concerning self-evaluated health, not only because respondents are being asked for subjective judgments, but also because responses may not be independent of the outcomes we may wish to use them to explain.

[1] For empirical papers that discuss non-classical measurement errors with continuous explanatory variables, see, e.g., Stephens & Unayama (2015), Haider & Solon (2006) and the references therein.

Brachet (2008) argues that in health-related surveys, self-reported smoking status is significantly misreported, with false negatives ranging from 3.4% in some studies to 73% in others. Other instances of false negatives can be found in the development literature where a firm's formality status is often misreported, with informal firms more likely to falsely report their statuses (see Gandelman & Rasteletti 2013). In contrast, false positives are low; Meyer & Goerge (2011) find that only less than 1% of non-recipients report food stamp receipt.

The existing literature has focused on accounting for random (exogenous) misreporting when participation is exogenous. For instance, Aigner (1973) considers misclassification in exogenous binary regressors, shows that OLS estimates are biased downwards, and proposes a technique based on knowledge of the misclassification probabilities to consistently estimate the parameters of interest. More recently, Lewbel (2007) examines the identification and estimation of the treatment effect of a misclassified binary regressor in nonparametric and semiparametric regressions. Lewbel reaches the same attenuation-bias result that Aigner (1973) finds and introduces assumptions that identify the conditional average treatment effect of the misclassified binary regressor.

Some attempts have been made to address (exogenous) misreporting when treatment selection (participation) is endogenous. In estimating the effect of Supplemental Nutrition Assistance Program (SNAP) on health outcomes, Kreider et al. (2012) use auxiliary administrative data on the size of SNAP caseloads to address misreporting by bounding the average treatment effect under increasingly stronger assumptions. While this partial identification approach identifies favorable treatment effects with their tightest bounds, it does not yield point estimates, as such its relevance for policy making may not be widespread. In the education literature, Kane et al. (1999) address misreporting when estimating returns to schooling by proposing a generalized method of moments (GMM) estimator that relies on the existence of two categorical reports of educational attainment, and so may have limited applicability. In estimating the effects of maternal smoking on infant health, Brachet (2008) proposes a two-step GMM estimator, that essentially follows Hausman et al. (1998) and Kane et al. (1999). An admitted weakness of Brachet's approach is the assumption that misreporting probabilities are independent of covariates, conditional on treatment status.

This paper has three salient contributions. First, we propose a model of endogenous misreporting and endogenous participation. We only analyze the case of false negatives at this stage, which is the predominant case of misreporting described in Meyer et al. (2009). Second, we show that OLS and IV estimators are inconsistent when participation is endogenous and even when participation is exogenous. We provide theoretical expressions for these biases and simulation evidence showing that OLS estimates of treatment effects can be of opposite signs from the true effects (sign reversal). Third, we propose an estimator that is root-n consistent and asymptotically normal and show that it performs remarkably well in small samples.

The rest of the paper is organized as follows. Section 2 presents the model of endogenous misreporting and shows the inconsistency of OLS and IV estimators. Section 3 develops the proposed estimator. Section 4 provides Monte Carlo simulations, Section 5 contains an empirical application and Section 6 concludes. Proofs are collected in the appendix.

2 Framework

This section describes the proposed model and associated framework, and presents our estimation strategy.

2.1 Model with Endogenous Misreporting

Consider the following specification of the usual treatment effects model. The outcome variable, y_i, is related to the k–vector of exogenous covariates, x_i, and the (true) participation indicator, δ_i^*, by

$$y_i = x_i'\beta + \delta_i^*\alpha + \epsilon_i, \tag{1}$$

and we model participation as

$$\delta_i^* = \mathbf{1}\left(z_i'\theta + v_i \geq 0\right), \tag{2}$$

where α is a scalar capturing the treatment effect of interest, β and θ are parameter vectors of sizes $k \times 1$ and $q \times 1$ respectively, z_i is a q-vector of observed covariates, and ϵ_i and v_i are possibly correlated error terms.

The researcher does not observe the true participation indicator δ_i^* but only a possibly misclassified surrogate, δ_i, contaminated by a misreporting unobserved dummy variable, d_i, such that $\delta_i = \delta_i^* d_i$. In other words, an individual correctly reports her treatment status only if $d_i = 1$ and reports not receiving treatment otherwise. We assume that misreporting, d_i, is related to a p-vector of observable covariates w_i such that

$$d_i = \mathbf{1}\left(w_i'\gamma + u_i \geq 0\right) \tag{3}$$

where γ is a parameter vector of size $p \times 1$ and u_i is the error term. Hence, the observed participation, δ_i, can be modeled by

$$\delta_i = \delta_i^* d_i = \mathbf{1}\left(z_i'\theta + v_i \geq 0,\ w_i'\gamma + u_i \geq 0\right). \tag{4}$$

Our modeling of misreported participation is therefore similar to Poirier (1980)'s partial observability model. No restrictions are imposed on x_i. However, we require the covariates z_i and w_i to be different but possibly overlapping, to avoid the local identification problems discussed in Poirier (1980). The joint distribution of the error terms is given by

$$(\epsilon_i, u_i, v_i) \sim N(0, \Sigma), \quad \text{with} \quad \Sigma = \begin{pmatrix} \sigma^2 & \varphi_u\sigma & \varphi_v\sigma \\ \varphi_u\sigma & 1 & \rho \\ \varphi_v\sigma & \rho & 1 \end{pmatrix}, \tag{5}$$

where σ^2 is the variance of ϵ_i, and φ_u, φ_v, ρ are the correlations between ϵ_i and u_i, ϵ_i and v_i, u_i and v_i, respectively. Define the joint CDF of $(-u, -v)$ by

$$F(\underline{u}, \underline{v}, \rho) = \Pr[-u_i \le \underline{u}, \quad -v_i \le \underline{v}], \quad \text{for any } -\infty < \underline{u},\ \underline{v} < +\infty.$$

We make the following basic assumptions, which are standard in the literature.

Assumption 1. *The vectors of regressors x_i and z_i are orthogonal to the error terms ϵ_i, u_i and v_i, and the vector of regressors w_i is orthogonal to u_i and v_i.*

Assumption 2. *The $k \times k$ matrix $\mathbb{E}(x_i x_i')$ is nonsingular (and hence finite).*

It is important to notice that unlike x_i and z_i, the exogeneity requirement does not apply to w_i, the covariates associated with misreporting in equation (3). This could be of substantial interest in practice where exogenous covariates are often difficult to find. In this framework, participation and misreporting are allowed to be endogenous, with the latter only in one direction (i.e., only false negatives). While we assume jointly normal disturbance terms for simplicity, normality is not needed and the following discussion would hold for other absolutely continuous distributions.

Our estimation strategy relies on observing z and w. We recognize that exclusion restrictions for participation and misreporting may be difficult to obtain in practice and our suggestion is to rely on different data sources. For instance, exclusion restrictions for participation may come from qualification laws (eligibility requirements) for program participation. Covariates w could include peculiar features of the survey in question and its administration such as survey date, length of survey, etc., and the proportion of questions to which the individual refused to respond.

2.2 Bias due to Endogenous Misreporting

We first show that a naive OLS estimator of the treatment effect is biased and may assume a sign opposite to the true effect. Since the true participation status δ_i^* is unobserved but only δ_i is observed, the model with reported participation status estimated by the researcher is given by

$$y_i = x_i'\beta + \delta_i\alpha + \varepsilon_i. \tag{6}$$

Given the true outcome equation defined by equation (1), equation (6) implicitly implies that we have

$$\varepsilon_i = \epsilon_i + (\delta_i^* - \delta_i)\,\alpha. \tag{7}$$

For a random sample of size n, equation (6) can be re-written in the matrix form as

$$y = X\beta + \delta\alpha + \varepsilon, \tag{8}$$

where $y = [y_1, \ldots, y_n]'$, $X = [x_1, \ldots, x_n]'$, $\delta = [\delta_1, \ldots, \delta_n]'$, and $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_n]'$.

Denote by $\widehat{\alpha}_{LS}$ the OLS estimator obtained by naively estimating equation (6) using reported participation δ_i. Then, we have the following result.

Theorem 1. *Under Assumptions 1 and 2, the ordinary least squares estimator, $\widehat{\alpha}_{LS}$, is biased and inconsistent, and the asymptotic bias is given by*

$$\text{plim}(\widehat{\alpha}_{LS} - \alpha) = \frac{A - \alpha B}{C}, \tag{9}$$

with

$$A = \mathbb{E}\left[\sigma\varphi_v\phi\left(-z_i'\theta\right)\Phi\left(\frac{w_i'\gamma - \rho z_i'\theta}{\sqrt{1-\rho^2}}\right) + \sigma\varphi_u\phi\left(-w_i'\gamma\right)\Phi\left(\frac{z_i'\theta - \rho w_i'\gamma}{\sqrt{1-\rho^2}}\right)\right],$$

$$B = \mathbb{E}(\delta_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}[(\delta_i^* - \delta_i)x_i] \quad \textit{and} \quad C = \mathbb{E}(\delta_i) - \mathbb{E}(\delta_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}(\delta_i x_i),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the pdf and cdf of the standard normal.

Proof. See Appendix. □

Note that since the denominator in (9), C, is always positive (by the Cauchy-Schwarz Inequality, see, e.g. Tripathi (1999)), the sign of the asymptotic bias only depends on the numerator of the expression. For example, if $B > 0$, then $\text{plim}(\widehat{\alpha}_{LS}) < \alpha$ for all $\alpha > A/B$ (i.e. there is an attenuation bias). Also notice that if $B - C > 0$, then $\text{plim}(\widehat{\alpha}_{LS})$ and α have opposite signs whenever α lays between 0 and $A/(B - C)$. Figure 1 depicts the regions where bias and sign switching occur. Note that sign-switching can occur even when participation is exogenous, i.e., $\varphi_v = 0$.

This result shows that the bias related to endogenous misreporting is not merely an attenuation bias as found in many other studies (e.g., Lewbel 2007). Rather, it emphasizes that under endogenous misreporting the estimated treatment effect can possibly assume an opposite sign, yielding misleading policy prescriptions. This sign reversal would generally occur when misreporting is severe and the direction of its correlation with outcome is opposite to the direction of the treatment effect. For example, in the food stamp participation and obesity relationship, much em-

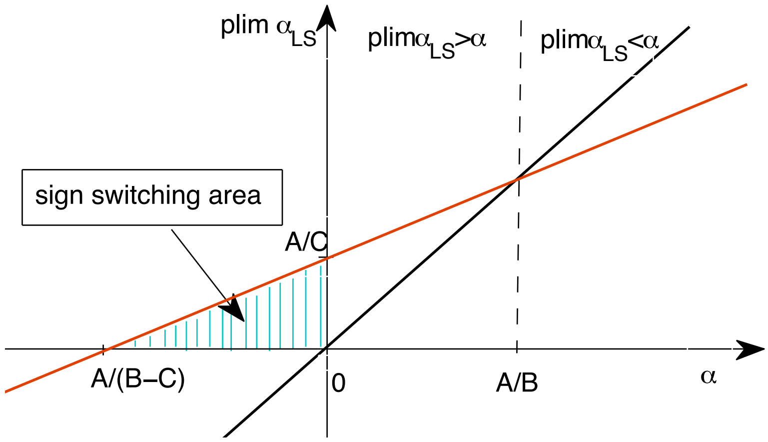

Figure 1: Illustration of the OLS bias

pirical work have relied on self-reported food stamp participation and have found a positive or no effect on obesity. But, if people who are overweight are also more likely to misreport food stamp participation (i.e. A positive) and since, as mentioned above, misreporting in food stamp is very severe in the data (i.e. B positive and large), then we could observe a positive relationship between food stamp participation and obesity (i.e. $\text{plim}\hat{\alpha}_{LS} > 0$) even if the true effect is negative (i.e. $\alpha < 0$).

In the next section, we provide an estimation strategy that allows consistent estimation of the treatment effect, α. But first, we examine how well an IV estimation strategy would perform in our framework.

2.3 IV Estimator under Endogenous Misreporting

The misreporting mechanism described above shows that in equation (6), the regressor δ_i is correlated with the error term ε_i as implied by equation (7). Thus, equation (1) can be seen as a regression with an endogenous binary regressor, even if true participation is exogenous and only misreporting is endogenous. So it may be tempting to suppose that if an instrument is present, then a standard IV estimator will address the issue raised in our framework. Here, we show that this is not the case.

Suppose we have access to a valid instrumental variable, z_i, such that $\mathbb{E}[z_i \varepsilon_i] = 0$ and $\text{Cov}(z_i, \delta_i) \neq 0$, and assume, for simplicity, that z_i is a scalar so that α is just identified. Then the (simple) instrumental variable estimator is given by

$$\widehat{\alpha}_{IV} = (z' M \delta)^{-1} z' M y,$$

where $M = I - X(X'X)^{-1}X'$ is the orthogonal projection matrix onto the null space of X.

We can show using the same reasoning as above that,

$$\text{plim}(\widehat{\alpha}_{IV}) = \frac{\mathbb{E}(z_i \delta_i^*) - \mathbb{E}(z_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}(x_i \delta_i^*)}{\mathbb{E}(z_i \delta_i) - \mathbb{E}(z_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}[x_i \delta_i]} \alpha. \tag{10}$$

Thus, the IV estimator of α is inconsistent, and we cannot sign the bias in general. However, in the special case where misreporting is uncorrelated with true participation and the other covariates, it can be shown that,

$$\text{plim}(\widehat{\alpha}_{IV}) = \frac{\alpha}{\mathbb{E}[d_i]} = \frac{\alpha}{\Pr[d_i = 1]} > \alpha.$$

Hence, in this specific scenario, the IV estimator is upwardly biased. This result is similar to those obtained by Loewenstein & Spletzer (1997), and Black et al. (n.d.). We now present an estimation procedure that delivers consistent and asymptotically normal estimates for the treatment effect, α.

3 The Proposed Estimator

Recall that our objective is to estimate α in the outcome equation (1), where true (and possibly endogenous) participation status, δ_i^*, is unobserved, but only a possibly misreported (and possibly endogenous) participation status, δ_i, is observed. The proposed estimation strategy proceeds in the following two steps.

Step 1: With the joint distribution of u_i and v_i given by $F(u, v, \rho)$, use the partial observability probit model given by equation (4) to estimate the parameter vectors θ and γ. Then, compute the predicted probability for person i's true participation status as $\hat{\delta}_i^* = \Phi(z_i'\hat{\theta})$.

Step 2: Estimate equation (1) by substituting $\hat{\delta}_i^*$ for δ_i^*. Assuming correct model specification and distribution of the error terms, the resulting two-step estimator of α is consistent. Moreover, with standard regularity assumptions, this estimator is asymptotically normal.

3.1 First-Step Estimation

Following Poirier (1980), the parameters γ, θ and ρ can be jointly estimated from the joint distribution of the error terms using the binary choice model defined by

$$\Pr[\delta_i = 1 | w_i, z_i] = \Pr\left[-u_i \leq w_i'\gamma, \ -v_i \leq z_i'\theta\right] = F\left(w_i'\gamma, z_i'\theta, \rho\right) = P_i(\gamma, \theta, \rho).$$

The log-likelihood function of this model is given by

$$L_n(\gamma, \theta, \rho) = \sum_{i=1}^{n} \delta_i \ln P_i(\gamma, \theta, \rho) + (1 - \delta_i) \ln\left(1 - P_i(\gamma, \theta, \rho)\right).$$

Assuming correct distributions, the maximum likelihood estimates of the parameters (γ, θ, ρ) are consistent and asymptotically normal, with the covariance matrix consistently estimated with the inverse of the information matrix. In particular, for the parameter θ, the MLE $\hat{\theta}$ is consistent and asymptotically normal, i.e.

$$\hat{\theta} \xrightarrow{p} \theta \quad \text{and} \quad \sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta),$$

where the asymptotic variance of $\hat{\theta}$ is obtained from the information matrix equality as

$$V_\theta = \left\{ \mathbb{E} \left[\frac{1}{P_i(1 - P_i)} \frac{\partial P_i}{\partial \theta} \frac{\partial P_i}{\partial \theta'} \right] \right\}^{-1}. \tag{11}$$

From this expression, a consistent estimator for the variance matrix can be obtained as

$$\widehat{V}_\theta = \left[\frac{1}{n} \sum_{i=1}^{n} \frac{1}{\widehat{P}_i(1 - \widehat{P}_i)} \frac{\partial \widehat{P}_i}{\partial \theta} \frac{\partial \widehat{P}_i}{\partial \theta'} \right]^{-1}, \tag{12}$$

where $\widehat{P}_i = P_i(\hat{\gamma}, \hat{\theta}, \hat{\rho}) = F\left(w_i'\hat{\gamma}, z_i'\hat{\theta}, \hat{\rho}\right)$. For the normal case, the gradient takes a fairly simple form

$$\frac{\partial \widehat{P}_i}{\partial \theta} = \phi(z_i'\hat{\theta})\Phi\left(\frac{w_i'\hat{\gamma} - \hat{\rho}z_i'\hat{\theta}}{\sqrt{1 - \hat{\rho}^2}}\right) z_i.$$

Notice that only the specification of the marginal distribution of v is necessary for the parametric identification and estimation of the model in the second step. If the distribution of u or the joint distribution of (u, v) are unknown, then one can still obtain a consistent estimator of θ in the first-step by using a semiparametric approach such as the series expansion of the joint PDF of (u, v) proposed by Gallant & Nychka (1987) or the single equation multiple index model described in Ichimura & Lee (1991).

3.2 Second-Step Estimation

In the second step, we compute the predicted values of true unobserved participation δ_i^*, given by $\hat{\delta}_i^* = \Phi(z_i'\hat{\theta})$ in the outcome equation in lieu of δ_i^* and estimate the new model given by

$$y_i = x_i'\beta + \hat{\delta}_i^* \alpha + \eta_i. \tag{13}$$

Using the same approach as above, the second step estimator is obtain as

$$\begin{aligned}\widehat{\alpha}_{2S} &= (\hat{\delta}^{*\prime} M \hat{\delta}^*)^{-1} \hat{\delta}^{*\prime} M y \\ &= \frac{\sum_{i=1}^n \Phi(z_i'\hat{\theta}) y_i - \sum_{i=1}^n \Phi(z_i'\hat{\theta}) x_i' [\sum_{i=1}^n x_i x_i']^{-1} \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n \Phi(z_i'\hat{\theta})^2 - \sum_{i=1}^n \Phi(z_i'\hat{\theta}) x_i' [\sum_{i=1}^n x_i x_i']^{-1} \sum_{i=1}^n x_i \Phi(z_i'\hat{\theta})}\end{aligned} \tag{14}$$

We have the following consistency result.

Theorem 2. *Under the model assumptions, the two-step estimator is consistent for α, that is, $\widehat{\alpha}_{2S} \xrightarrow{p} \alpha$.*

Proof. See Appendix. □

Notice that only the component $\hat{\theta}$ of the parameter vector is used at this second stage to predict the true unobserved participation status. The other components, $\hat{\gamma}$ and $\hat{\rho}$ are only used in the computation of the asymptotic variance estimator, as described below. We also have the following asymptotic normality result.

Theorem 3. *Under the model assumptions the two-step estimator is asymptotically normal, i.e.,*

$$\sqrt{n}(\widehat{\alpha}_{2S} - \alpha) \xrightarrow{d} N(0, \sigma_\alpha^2), \qquad \textit{with} \quad \sigma_\alpha^2 = \frac{\alpha^2 \mathbb{E}[\Lambda_i^2 \Phi(z_i'\theta)(1 - \Phi(z_i'\theta))]}{\mathbb{E}[\Lambda_i^2]^2} + \frac{\sigma^2}{\mathbb{E}[\Lambda_i^2]},$$

where

$$\Lambda_i = \Phi(z_i'\theta) - \mathbb{E}\left[\Phi(z_i'\theta) x_i'\right] \mathbb{E}[x_i x_i']^{-1} x_i$$

Proof. See Appendix. □

3.3 A consistent estimator for the asymptotic variance

Theorem 3 gives the asymptotic variance of the treatment effect estimator, $\hat{\alpha}_{2S}$. To perform inference based on $\hat{\alpha}_{2S}$ it is useful to find a consistent estimator of this variance. One could use

$$\hat{\sigma}_{\alpha}^2 = \frac{\hat{\alpha}_{2S}^2 \hat{\nu}^2}{\hat{q}^2} + \frac{\hat{\sigma}^2}{\hat{q}} \tag{15}$$

where $\hat{\nu}^2$, $\hat{q}$ and $\hat{\sigma}^2$ are obtained respectively by

$$\hat{\nu}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\Lambda}_i^2 \Phi(z_i'\hat{\theta}) \left(1 - \Phi(z_i'\hat{\theta})\right) \tag{16}$$

$$\hat{q} = \frac{1}{n}\sum_{i=1}^{n} \hat{\Lambda}_i^2 \tag{17}$$

$$\hat{\sigma}^2 \quad = \quad \frac{1}{n}\sum_{i} \left[\left(y_i - x_i'\hat{\beta} - \hat{\alpha}_{2S}\Phi(z_i'\hat{\theta}) \right)^2 + \hat{\alpha}_{2S}^2 \Phi(z_i'\hat{\theta}) \left(1 - \Phi(z_i'\hat{\theta})\right) \right], \tag{18}$$

with

$$\hat{\Lambda}_i = \hat{\delta}_i^* - \left(\frac{1}{n}\sum_{i=1}^{n} \hat{\delta}_i^* x_i' \right) \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i' \right)^{-1} x_i$$

It should be noted again here that the estimation of the variance uses the normal CDF $\Phi(\cdot)$ only because the normality of the error terms is assumed. Under other distributional assumptions, $\Phi(\cdot)$ and $\phi(\cdot)$ can be replaced by the corresponding CDF and PDF, respectively.

Summarizing, the outcome equation requires true participation status, δ^*, which is unobserved by the econometrician. Given the observed participation, δ, the first step in our estimation procedure amounts to a partial observability probit analysis on the indicator variable δ using both z and w, which are respectively the instrumental variables driving true participation and the covariates driving misreporting. The result of this analysis is an estimator, $\hat{\theta}$, of θ, the coefficient of z, which allows constructing a proxy $\hat{\delta}^*$ for truly being a participant. By construction, this proxy is purged from both endogeneity and misreporting, and is then used in lieu of δ^* in the outcome equation of interest to derive a reliable treatment effect estimator. The estimate $\hat{\theta}$ obtained from the first step can then be used along with the other model estimates to compute a consistent variance estimator for the treatment effect estimator.

4 Monte Carlo Simulations

This section presents the results of Monte Carlo simulations comparing the proposed two-step estimator (2S) with OLS and IV estimators. Our goal is to identify and consistently estimate α, the (conditional) average treatment effect of participation, δ^*, on an outcome, y, given by equation (1). However, since (true) participation is unobserved, our task reduces to consistently estimating α from equation (6) under the assumption that, observed (misclassified) participation, δ, arises according to the process described by equations (3) and (4).

4.1 Simulation setup

The data generating process is simulated as follows. The true treatment indicator, δ_i^*, is given by

$$\delta_i^* = \mathbf{1}\left(\theta_0 + \theta_1 z_i + v_i \geq 0\right), \quad \text{where } z_i \sim N(0,1), \quad \theta_1 = 10, \quad \theta_0 = 0.1.$$

The outcome equation y_i is given by

$$y_i = \beta_0 + x_i\beta_1 + \delta_i^*\alpha + \epsilon_i \quad \text{where } x_i \sim N(0,1) \quad \beta_0 = \beta_1 = 1, \quad \alpha = -0.2.$$

Note that $\alpha = -0.2$ is the true population treatment effect we seek to estimate.

As previously discussed, the econometrician only observes an error-ridden treatment indicator, δ_i, defined by

$$\delta_i = \delta_i^* \mathbf{1}\left(\gamma_0 + \gamma_1 w_i + u_i \geq c\right), \quad \text{where } w_i \sim N(0,2), \quad \gamma_1 = 100, \quad \gamma_0 = 0.2.$$

The parameter c is a threshold that determines the proportion of false negatives in the sample.[2] The disturbances ϵ_i, u_i and v_i are drawn from a trivariate distribution given by

$$(\epsilon_i, u_i, v_i) \sim N\left(0, \Sigma\right), \quad \text{where} \quad \Sigma = \begin{pmatrix} \sigma^2 & \varphi_u\sigma & \varphi_v\sigma \\ \varphi_u\sigma & 1 & \rho \\ \varphi_v\sigma & \rho & 1 \end{pmatrix}, \quad \sigma = 1, \quad \rho = 0.$$

The values of the correlation parameters φ_v and φ_u are varied in the simulations to examine how various degrees of the endogeneity of participation and misreporting affect the results. We estimate the treatment effect α and the associated bias using the naive OLS approach $\hat{\alpha}_{LS}$ and the proposed two-step approach $\hat{\alpha}_{2S}$. We also estimate the instrumental variable estimator $\hat{\alpha}_{IV}$ using both z and w as instruments.

[2]By appropriately choosing the value of c, one can simulate varying rates of misreporting.

4.2 Simulation Results

We report simulation results averaged over 1000 replications each with sample size 5000 for different levels of false negatives - $5\%, 10\%, 20\%, 40\%$ - for $\varphi_u \in \{0, 0.2, 0.8\}$ and $\varphi_v \in \{-0.5, 0, 0.5\}$, where φ_u and φ_v are the correlations of the outcome equation with misreporting and participation, respectively. The cases of exogenous participation and exogenous misreporting correspond to $\varphi_u = \varphi_v = 0$. Table (1) presents the results of the Monte Carlo simulations for OLS, IV, and the proposed two-step (2S) estimators. We report both the OLS estimates using the true treatment indicator, δ_i^* (True Participation) and the observed treatment effect δ_i (Observed Participation). Although δ_i^* is unobserved to the econometrician, these estimates provide a theoretical benchmark for the estimates obtained using the misclassified δ_i.

The naive OLS estimates, $\widehat{\alpha}_{LS}$, using δ_i (OLS Observed Participation) show that, not only is the OLS estimator inconsistent as asserted in Theorem 1, but also yields wrong (i.e. positive) signs, whether participation is exogenous or endogenous. Sign-switching is observed at all false negative rates i.e. 5%, $10\%, 20\%$ and 40% and is more pronounced at higher values of φ_u. These results persist even under the special case of exogenous misreporting ($\varphi_u = 0$). The IV estimates, $\widehat{\alpha}_{IV}$, are reported in the column (IV). In the estimation, the vector of instruments for δ_i is given by $[1\ x_i\ z_i\ w_i]$, since w_i is also exogenous in this setting.[3] When participation is endogenous, the results show, as we expect, that OLS is biased and inconsistent. However, perhaps surprisingly, the results show that the classic IV estimator is also inconsistent and sometimes worse, albeit keeping the correct

[3]This is actually a better set of simulations for the IV because the covariate w_i can be used as an additional instrument to improve the IV. Unreported simulations with w_i being endogenous, that is, the vector of instruments for δ_i is $[1\ x_i\ z_i]$, yielded worse results for the IV.

Table 1: Monte Carlo Simulations

False Negatives	φ_u	φ_v	OLS True Participation	OLS Observed Participation	IV	2S
5%	0	-0.5	-0.766	-0.724	-0.204	-0.201
		0	-0.200	-0.189	-0.202	-0.200
		0.5	0.365	0.345	-0.200	-0.198
	0.2	-0.5	-0.765	-0.712	-0.204	-0.200
		0	-0.201	-0.179	-0.204	-0.201
		0.5	0.366	0.357	-0.203	-0.200
	0.8	-0.5	-0.764	-0.679	-0.201	-0.198
		0	-0.200	-0.145	-0.202	-0.199
		0.5	0.365	0.388	-0.203	-0.201
10%	0	-0.5	-0.765	-0.689	-0.199	-0.201
		0	-0.199	-0.179	-0.197	-0.199
		0.5	0.365	0.329	-0.197	-0.199
	0.2	-0.5	-0.766	-0.672	-0.198	-0.201
		0	-0.199	-0.162	-0.199	-0.200
		0.5	0.363	0.345	-0.198	-0.201
	0.8	-0.5	-0.765	-0.617	-0.200	-0.202
		0	-0.200	-0.109	-0.194	-0.196
		0.5	0.365	0.400	-0.197	-0.200
20%	0	-0.5	-0.765	-0.624	-0.181	-0.201
		0	-0.200	-0.163	-0.173	-0.197
		0.5	0.364	0.297	-0.177	-0.201
	0.2	-0.5	-0.765	-0.592	-0.178	-0.201
		0	-0.200	-0.133	-0.178	-0.201
		0.5	0.365	0.328	-0.174	-0.199
	0.8	-0.5	-0.764	-0.503	-0.177	-0.199
		0	-0.202	-0.045	-0.180	-0.202
		0.5	0.363	0.415	-0.176	-0.201
40%	0	-0.5	-0.764	-0.526	-0.138	-0.199
		0	-0.200	-0.138	-0.141	-0.201
		0.5	0.366	0.251	-0.141	-0.200
	0.2	-0.5	-0.765	-0.481	-0.139	-0.200
		0	-0.200	-0.092	-0.137	-0.198
		0.5	0.365	0.297	-0.143	-0.202
	0.8	-0.5	-0.764	-0.342	-0.140	-0.199
		0	-0.201	0.046	-0.144	-0.203
		0.5	0.365	0.436	-0.139	-0.199

Notes. The true treatment effect is $\alpha = -0.2$. Each calibration in the Monte Carlo Design involved 1000 replications each of size 5000. We report results for four false negative rates ($5\%, 10\%, 20\%$, and 40%) i.e. the proportion of true participants who misreport their status. φ_v and φ_u are correlations that indicate the extents of endogeneity of participation and misreporting, respectively.

(negative) sign.[4]

In contrast, the proposed two-step estimator (2S), presented in the last column of Table 1, yields consistent estimates of the true treatment effect and by comparison, is superior to both the OLS and IV estimators under both endogenous and exogenous misreporting or participation. Interestingly, the proposed estimator remains accurate and performs remarkably well, even when the rate of false negatives is substantially high in the data.

There are few additional facts that are worth mentioning. Although only one set of parameter values are presented here, we also ran the model with different parameter values and distributions. While the magnitudes of the bias for OLS and IV were sensitive to the values of parameters the consistency of the proposed estimator (2S) was not affected by parameter choice. Finally, Lewbel (2007)'s estimator also worked well in our setting for the special cases where both participation and misreporting where exogenous.[5] However, Lewbel's estimator displayed large biases and sign reversals under some endogeneity cases, which is not surprising since this limitation is clearly emphasized in Lewbel (2007). These additional results are available from the authors upon request.

[4]The correct sign for the IV arises because misreporting and true participation are uncorrelated in this simulation setup. However, as shown in Section 2.3, we cannot generally sign the bias in the IV estimator.

[5]It is easy to slightly modify our set up to include the instrumental variable required by Lewbel's identification strategy. For that purpose, we added a binary indicator in the true participation equation, since, as explained by Lewbel (2007), only two points of support are needed for the instrument to identify the treatment effect if the rate of false positives is zero (as in our case).

5 The Impact of SNAP Participation on Obesity

We consider an empirical application examining the impact of the Supplemental Nutrition Assistance Program (SNAP) [6] on adult obesity when participation status is potentially misreported. We present estimates of SNAP's effect on Body Mass Index (BMI) based on OLS, IV, and our proposed two-step (2S) estimators using the 1998 wave of the restricted-use National Longitudinal Survey of Youth - 1979 (NLSY79).

The impact of SNAP on obesity is theoretically ambiguous. One the one hand, SNAP benefits loosen budget constraints of households, allowing them to make healthier food choices and engage in less-sedentary activities that could lead to reductions in weight status. On the other hand, the income effect from SNAP could lead to greater consumption of unhealthy food (more calories) and more sedentary activities that could lead to increases in weight status. A third possibility is that SNAP benefits could be too little to significantly affect weight outcomes. Depending on the proportion of recipients that are inframarginal and the types of food purchased, SNAP may positively or negatively impact obesity.

Previous studies on the relationship between SNAP and obesity have found positive effects (Meyerhoefer & Pylypchuk 2008, Baum 2011), negative effects (Hoyne et al. 2016, Burgstahler et al. 2012), and no effect (Fan 2010). These mixed results reflect two identification challenges in evaluating the causal effects of SNAP: self-selection and the misreporting of participation status. Although some previous studies have employed instrumental variable methods to address endogenous selection into SNAP, very few attempts have been made to also address misreporting of participation status. We demonstrate how our proposed estimator can be useful for empiricists in this context.

Misreporting of SNAP participation in national surveys has been well-documented

[6] SNAP was formerly known as the Food Stamp Program.

Table 2: Summary Statistics by SNAP Participation Status

	SNAP Participants		Non-SNAP Participants	
	Mean	Std.	Mean	Std.
Body Mass Index (BMI)	28.22	6.88	27.50	6.09
Age (years)	36.70	2.32	36.88	2.29
Hispanic	0.11	0.31	0.09	0.28
Black	0.35	0.48	0.20	0.40
Mother's Education ($>$ High School)	0.11	0.31	0.10	0.29
Female	0.72	0.45	0.54	0.50
Number of Observations	451		1767	

Notes. Summary statistics (weighted sample means and standard deviations) are based on the 1998 wave of the NLSY79, restricted to individuals or households with income below 250% of the federal poverty line.

with false negatives being more prevalent than false positives. For instance, false negatives for SNAP are estimated to be around $20 - 30\%$ in the 2001 and 2004 panels of the Survey of Income and Program Participation (SIPP) (Meyer et al. 2013), up to 50% in the 2002-2005 CPS Annual Social and Economic Supplement (Meyer et al. 2013), about $20 - 40\%$ in the 2001 American Community Survey (ACS) (Lynch et al. 2008), about 26% in the 2008-2010 ACS (Mittag 2013), and $23 - 45\%$ in the National Longitudinal Survey of Youth (NLSY) - 1979 cohort (Almada et al. 2016). However, false positives are low and typically less than 1% (Meyer & Goerge 2011).

Recall that the observed model to be estimated is given by

$$y_i = x_i'\beta + \delta_i\alpha + \varepsilon_i, \tag{19}$$

where y_i, the dependent variable is BMI [7], x_i is the set of independent variables, δ_i is the SNAP participation indicator that is endogenous and possibly systematically

[7] BMI is defined as weight in kilograms divided by height in meters squared.

misreported, and ε_i is the disturbance term. For the purpose of this illustration, x_i includes age, gender, race, and mother's education. Table 2 reports summary statistics (means and standard deviations) by SNAP participation status. The average BMI for SNAP participants is 28.22 while it is 27.50 for nonparticipants. The summary statistics indicate that SNAP participants are more likely to be Hispanic (0.11 vs. 0.09), black (0.35 vs. 0.20), and female (0.72 vs. 0.54).

To implement the 2S estimator, two sets of covariates need to be distinguished: covariates related to participation (z_i in equation (2)) and covariates related to misreporting (w_i in equation (3)). The covariates for participation are set equal to the independent variables in equation (19) in addition to two exclusion restrictions: whether the respondent's state uses a biometric identification technology –"Biometric"– and the percentage of SNAP benefits issued by the state via direct mail – "Direct Mail" (Meyerhoefer & Pylypchuk 2008). Intuitively, these state-level policies affect a person's SNAP participation decision by changing the transaction costs of participation but should not directly affect adult weight outcomes.

Our proposed estimator also exploits available information in the NLSY79 relating to why respondents might misreport their SNAP participation status. Interesting candidates in this regard are interview or survey characteristics. For the covariates related to misreporting (w_i), the set of independent variables in the outcome equation (19) are augmented with a binary indicator for whether the interview was conducted by telephone or in person.[8]

Table 3 reports OLS, conventional IV and our two-step (2S) estimates. The three approaches give completely different and sometimes opposite results. The OLS method estimates a positive and significant effect of SNAP participation on BMI. The average BMI difference between SNAP participants and non-participants

[8]About 24.6% of interviews conducted in the 1998 wave of the NLSY79 were by telephone.

Table 3: The Impact of SNAP on BMI

	Dependent Variable: BMI		
Variable	OLS	IV	2S
SNAP Dummy	0.2012	-5.6678	-1.4258
	(0.3569)	(5.6084)	(2.1954)
Age	0.0881	0.0472	0.0755
	(0.0581)	(0.0751)	(0.0615)
Hispanic	1.9062***	2.3877***	2.1534***
	(0.3431)	(0.5871)	(0.4990)
Black	1.7231***	2.4242***	2.0305***
	(0.3037)	(0.7500)	(0.5302)
Mother's Educ.	0.8391	0.9438	0.8715*
	(0.5894)	(0.6262)	(0.4906)
Female	0.4537	1.2398	0.6330*
	(0.2598)	(0.7921)	(0.3557)
Intercept	23.431***	25.2625***	24.2445***
	(2.1238)	(2.9384)	(2.4631)

Notes. Results are based on the 1996 wave of the NLSY79, restricted to individuals or households with income below 250% of the federal poverty level.
Standard errors in parenthesis. " *** "$p < 0.01$, " ** "$p < 0.05$, " * "$p < 0.1$

Table 4: First Stage Results for IV Estimator

Variable	Dependent Variable: SNAP Dummy
Biometric	0.0717***
	(.0260)
Direct Mail	0.0080
	(0.0254)
Global F-statistic	3.81**
Hansen J-statistic	0.002

Standard errors in parenthesis. " *** "$p < 0.01$, " ** "$p < 0.05$, " * "$p < 0.1$

is estimated at 0.2 BMI units, everything else equal. In contrast, our proposed estimator yields a negative (albeit insignificant) effect of SNAP participation on BMI. Food Stamp participants are found to be less heavier than non-participants, with a BMI difference of 1.43. Relative to the mean height of 65.75 inches, the 2S estimator suggests that SNAP participation is associated with a reduction of 8.75 pounds on average. Interestingly, however, all the remaining coefficients besides that on the SNAP participation indicator from both OLS and 2S are similar in sign and magnitude.

The conventional IV estimator, on the other hand, yields the same sign as the proposed estimator but is more than 3 times bigger in magnitude. This expansion bias or inflation of the IV is similar to results obtained in Brachet (2008)'s analysis of the relationship between smoking and infant health. The conventional IV estimator uses the same set of instruments for participation as our proposed estimator (i.e., "Biometric" and "Direct Mail"). Table 4 presents the first stage results for the IV estimator. The first stage regression yields significant global F-statistic and low and insignificant Hansen J-statistic, implying the instruments are sufficiently correlated with observed SNAP participation status.

These results from the empirical illustration corroborate the simulations results obtained in Section 4: (i) there is a possible sign reversal between the OLS and the 2S estimates [9]; (ii) the IV and the 2S estimators have the same sign but the IV has a larger magnitude. In addition, the results obtained from these methods lead to radically different and possibly contradictory policy advice.

This empirical illustration is undoubtedly limited and the results should be interpreted with caution since they are only suggestive. For instance, even if our proposed estimator estimates α consistently, it may not represent a causal

[9] Note that the sign reversal phenomenon obtained in this empirical illustration is neither a general result nor does its nonoccurrence invalidate the results herein. See Section 2.2 for further discussions.

effect of SNAP on obesity due to possible confounding by omitted variables in this application. Subject to these caveats, the discussions above suggest that the OLS and IV estimators are likely biased, with the OLS yielding a positive effect while the IV gives an implausibly large magnitude. These biases might be symptomatic of severe misreporting of SNAP participation.

6 Conclusion

This paper examines the identification and estimation of the conditional average treatment effect of a binary regressor in the presence of endogenous misreporting and possibly endogenous participation. We derive and prove the consistency and asymptotic normality of our proposed two-step estimator and show that OLS and IV estimators are inconsistent and may yield wrong (opposite) signs. We also provide Monte Carlo simulations to this effect, and apply our method to an empirical example examining the impact of food stamps participation on obesity using data from the National Longitudinal Survey of Youth (NLSY) - 1979 cohort. We find that the estimated effect of SNAP participation on BMI using our approach is negative, but turns out to be positive when misreporting is ignored (OLS approach).

Previous studies on misclassified binary regressors are mostly concerned with exogenous or random misreporting (Aigner 1973, Brachet 2008, Lewbel 2007, Mahajan 2006, Frazis & Loewenstein 2003), where it is commonly assumed that, misclassification probabilities depend only on the true treatment status and thus, independent of measurement errors and other regressors. Our two-step estimator relaxes this arguably strong assumption and shows that, when the researcher has access to information related to why individuals misreport, the treatment effect can be consistently estimated.

To our knowledge, this paper is the first attempt at addressing endogenous misreporting. This is important because of the prevalence of misreporting in public programs and survey data (Meyer et al. 2009, Bollinger 1996, Kane & Rouse 1995, Kane et al. 1999, Brachet 2008). While this paper focused on one-way endogenous misreporting when participation is possibly endogenous, future work should allow for bidirectional misreporting (i.e. false negatives and false positives). It would also be useful to show the level of dependence of our approach on distributional and functional form assumptions by considering parametric or semi-parametric estimation approaches.

Appendix A Proofs

A.1 Proof of Theorem 1

Proof.

Biasedness: By the Frisch-Waugh-Lovell Theorem, see, e.g. Davidson & MacKinnon (2004, page 68), the regression

$$My = M\delta\alpha + v$$

yields the same least squares estimate of α as the regression equation of interest (8). It follows that,

$$\widehat{\alpha}_{LS} = (\delta' M\delta)^{-1}\delta' My. \tag{20}$$

This implies that $\widehat{\alpha}_{LS} - \alpha = (\delta' M\delta)^{-1}\delta' M\varepsilon$.

Hence, $\mathbb{E}[\widehat{\alpha}_{LS} - \alpha|X, \delta] = (\delta' M\delta)^{-1}\delta' M\mathbb{E}[\varepsilon|X, \delta] \neq 0$, since $\mathbb{E}[\varepsilon|\delta, X] \neq 0$ by the correlation of ε and δ through u and v.

Inconsistency: We can write

$$\begin{aligned}\widehat{\alpha}_{LS} - \alpha &= (\delta' M \delta)^{-1} \delta' M \epsilon = \left(\frac{\delta' M \delta}{n}\right)^{-1} \frac{\delta' M \varepsilon}{n} \\ &= \left(\frac{\delta' M \delta}{n}\right)^{-1} \left(\frac{\delta' M \epsilon}{n} + \frac{\delta' M (\delta^* - \delta)\alpha}{n}\right) \quad \text{by Equation (7)} \end{aligned} \tag{21}$$

Notice that,

$$\frac{\delta' M \delta}{n} = \frac{\delta'[I - X(X'X)^{-1}X']\delta}{n} = \frac{\delta'\delta}{n} - \frac{\delta' X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\delta}{n}$$

Hence, by the Weak Law of Large Numbers and the Slutsky's lemma, we have

$$\frac{\delta' M \delta}{n} \xrightarrow{p} \mathbb{E}(\delta_i^2) - \mathbb{E}(\delta_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}(\delta_i x_i)$$

By a matrix extension of the Cauchy-Shwarz inequality (see Tripathi 1999), we know that $\mathbb{E}(\delta_i^2) - \mathbb{E}(\delta_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}(\delta_i x_i) > 0$. The Continuous Mapping Theorem then implies that

$$\left(\frac{\delta' M \delta}{n}\right)^{-1} \xrightarrow{p} \left[\mathbb{E}(\delta_i^2) - \mathbb{E}(\delta_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}(\delta_i x_i)\right]^{-1}. \tag{22}$$

Likewise, the term $\dfrac{\delta' M \epsilon}{n}$ can also be decomposed as

$$\frac{\delta' M \epsilon}{n} = \frac{\delta'[I - X(X'X)^{-1}X']\epsilon}{n} = \frac{\delta'\epsilon}{n} - \frac{\delta' X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\epsilon}{n}.$$

Then, using the same arguments as above we have

$$\frac{\delta' M \epsilon}{n} \xrightarrow{p} \mathbb{E}(\delta_i \epsilon_i) - \mathbb{E}(\delta_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}(x_i \epsilon_i) = \mathbb{E}(\delta_i \epsilon_i),$$

where the last equality follows from Assumption 1.

Using the expression of δ_i given by Equation (4) and the trivariate normality of (ϵ_i, u_i, v_i), it can be shown by integration that

$$\begin{aligned}
\mathbb{E}[\delta_i\epsilon_i] &= \mathbb{E}\left[\epsilon_i \mathbf{1}\left(z_i'\theta + v_i \geq 0, \quad w_i'\gamma + u_i \geq 0\right)\right] \\
&= \mathbb{E}\left[\Pr[u_i \geq -w_i'\gamma, \quad v_i \geq -z_i'\theta, \rho]\mathbb{E}\left[\epsilon_i | u_i \geq -w_i'\gamma, \quad v_i \geq -z_i'\theta\right]\right] \\
&= \mathbb{E}\left[\sigma\varphi_v\phi\left(-z_i'\theta\right)\Phi\left(\frac{w_i'\gamma - \rho z_i'\theta}{\sqrt{1-\rho^2}}\right) + \sigma\varphi_u\phi\left(-w_i'\gamma\right)\Phi\left(\frac{z_i'\theta - \rho w_i'\gamma}{\sqrt{1-\rho^2}}\right)\right],
\end{aligned}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDF of the standard normal. It follows that

$$\frac{\delta' M\epsilon}{n} \xrightarrow{p} \mathbb{E}\left[\sigma\varphi_v\phi\left(-z_i'\theta\right)\Phi\left(\frac{w_i'\gamma - \rho z_i'\theta}{\sqrt{1-\rho^2}}\right) + \sigma\varphi_u\phi\left(-w_i'\gamma\right)\Phi\left(\frac{z_i'\theta - \rho w_i'\gamma}{\sqrt{1-\rho^2}}\right)\right]. \quad (23)$$

Finally, using the same reasoning as above for the term $\frac{\delta' M(\delta^* - \delta)\alpha}{n}$, we have

$$\frac{\delta' M(\delta^* - \delta)\alpha}{n} \xrightarrow{p} -\alpha\mathbb{E}(\delta_i x_i')\mathbb{E}(x_i x_i')^{-1}\mathbb{E}[(\delta_i^* - \delta_i)x_i]. \quad (24)$$

The desired result follows by taking (24), (23) and (22) to Equation (21).

□

A.2 Proof of Theorem 2

Proof. We can write

$$\widehat{\alpha}_{2S} = (\hat{\delta}^{*\prime} M\hat{\delta}^*)^{-1}\hat{\delta}^{*\prime} M\delta^*\alpha + (\hat{\delta}^{*\prime} M\hat{\delta}^*)^{-1}\hat{\delta}^{*\prime} M\epsilon \quad (25)$$

By the exogeneity of X and Z given by Assumption 1, the consistency of $\hat{\theta}$, the continuity of $\Phi(\cdot)$ and the law of large numbers, we have

$$\frac{\hat{\delta}^{*\prime} M\epsilon}{n} \xrightarrow{p} \mathbb{E}[\Phi(z_i'\theta)\epsilon_i] = \mathbb{E}\left[\Phi(z_i'\theta)\mathbb{E}[\epsilon_i | z_i]\right] = 0,$$

so that the second term on the RHS of Equation (25) goes to zero. We also have, by Assumption 2, the consistency of $\hat{\theta}$, the continuity of $\Phi(\cdot)$ and the the law of large numbers,

$$\frac{\hat{\delta}^{*\prime} M \hat{\delta}^{*}}{n} \xrightarrow{p} \mathbb{E}\left[\Phi(z_i'\theta)^2\right] - \mathbb{E}\left[\Phi(z_i'\theta)x_i'\right]\mathbb{E}[x_i x_i']^{-1}\mathbb{E}\left[\Phi(z_i'\theta)x_i\right]$$

and

$$\begin{aligned}
\frac{\hat{\delta}^{*\prime} M \delta^{*}}{n} &\xrightarrow{p} \mathbb{E}\left[\Phi(z_i'\theta)\delta_i^*\right] - \mathbb{E}\left[\Phi(z_i'\theta)x_i'\right]\mathbb{E}[x_i x_i']^{-1}\mathbb{E}\left[x_i\delta_i^*\right] \\
&= \mathbb{E}\left[\Phi(z_i'\theta)\mathbb{E}[\delta_i^*|z_i]\right] - \mathbb{E}\left[\Phi(z_i'\theta)x_i'\right]\mathbb{E}[x_i x_i']^{-1}\mathbb{E}\left[x_i\mathbb{E}[\delta_i^*|z_i]\right] \\
&= \mathbb{E}\left[\Phi(z_i'\theta)^2\right] - \mathbb{E}\left[\Phi(z_i'\theta)x_i'\right]\mathbb{E}[x_i x_i']^{-1}\mathbb{E}\left[x_i\Phi(z_i'\theta)\right]
\end{aligned}$$

where the last display follows from the fact that $\mathbb{E}[\delta_i^*|z_i] = \Phi(z_i'\theta)$, as implied by Equation (2). Hence,

$$(\hat{\delta}^{*\prime} M \hat{\delta}^{*})^{-1}\hat{\delta}^{*\prime} M \delta^{*} = \left(\frac{\hat{\delta}^{*\prime} M \hat{\delta}^{*}}{n}\right)^{-1}\frac{\hat{\delta}^{*\prime} M \delta^{*}}{n} \xrightarrow{p} 1$$

so that

$$\widehat{\alpha}_{2S} \xrightarrow{p} \alpha$$

□

A.3 Proof of Theorem 3

Proof. We can write

$$\begin{aligned}\sqrt{n}(\widehat{\alpha}_{2S}-\alpha) &= \left(\frac{\hat{\delta}^{*\prime}M\hat{\delta}^{*}}{n}\right)^{-1}\left(\frac{\hat{\delta}^{*\prime}M(\delta^{*}-\hat{\delta}^{*})}{\sqrt{n}}\right)\alpha+\left(\frac{\hat{\delta}^{*\prime}M\hat{\delta}^{*}}{n}\right)^{-1}\frac{\hat{\delta}^{*\prime}M\epsilon}{\sqrt{n}} \\ &= q_n^{-1}\left[\sqrt{n}V_{1n}\alpha+\sqrt{n}V_{2n}\right] \qquad (26)\end{aligned}$$

where

$$q_n=\frac{\hat{\delta}^{*\prime}M\hat{\delta}^{*}}{n},\quad V_{1n}=\frac{\hat{\delta}^{*\prime}M(\delta^{*}-\hat{\delta}^{*})}{n},\quad \text{and}\quad V_{2n}=\frac{\hat{\delta}^{*\prime}M\epsilon}{n}$$

Denote $\hat{\Lambda}_i=\hat{\delta}_i^{*}-\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}_i^{*}x_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}x_ix_i'\right)^{-1}x_i$ and by $\Lambda_i=\Phi(z_i'\theta)-\mathbb{E}\left[\Phi(z_i'\theta)x_i'\right]\mathbb{E}[x_ix_i']^{-1}x_i$ its probability limit. We know, from the consistency results above that

$$q_n\xrightarrow{p}q=\mathbb{E}\left[\Phi(z_i'\theta)^2\right]-\mathbb{E}\left[\Phi(z_i'\theta)x_i'\right]\mathbb{E}[x_ix_i']^{-1}\mathbb{E}\left[\Phi(z_i'\theta)x_i\right]=\mathbb{E}[\Lambda_i^2]. \qquad (27)$$

Also, by a direct application of the central limit theorem,

$$\sqrt{n}V_{1n}\xrightarrow{d}N(0,\nu^2),\quad \text{where}$$

$$\nu^2 = \text{plim}\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i^2(\delta_i^{*}-\hat{\delta}_i^{*})^2=\mathbb{E}[\Lambda_i^2\Phi(z_i'\theta)(1-\Phi(z_i'\theta))] \qquad (28)$$

Likewise, by the central limit theorem,

$$\sqrt{n}V_{2n}\xrightarrow{d}N(0,\sigma_2^2),\quad \text{where}$$

$$\sigma_2^2 = \text{plim}\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i^2\epsilon_i^2 = \sigma^2\mathbb{E}[\Lambda_i^2] = \sigma^2 q \quad (29)$$

Finally, the asymptotic covariance term between the elements of $\sqrt{n}V_{1n}$ and $\sqrt{n}V_{2n}$ is

$$\text{plim}\frac{1}{n}\sum_{i=1}^{n}\hat{\Lambda}_i^2(\delta_i^* - \hat{\delta}_i^*)\epsilon_i = \mathbb{E}[\Lambda_i(\delta_i^* - \Phi(z_i'\theta))\epsilon_i] = 0$$

It then follows from Slutsky's Lemma, (26), (27), (28) and (29) that

$$\sqrt{n}(\widehat{\alpha}_{2S} - \alpha) \xrightarrow{d} N(0, \sigma_\alpha^2), \quad \text{where}$$

$$\sigma_\alpha^2 = \frac{\alpha^2\nu^2}{q^2} + \frac{\sigma^2}{q} = \frac{\alpha^2\mathbb{E}[\Lambda_i^2\Phi(z_i'\theta)(1 - \Phi(z_i'\theta))]}{\mathbb{E}[\Lambda_i^2]^2} + \frac{\sigma^2}{\mathbb{E}[\Lambda_i^2]}$$

□

References

Aigner, D. J. (1973), 'Regression with a binary independent variable subject to errors of observation', *Journal of Econometrics* **1**(1), 49–59.

Almada, L., McCarthy, I. & Tchernis, R. (2016), 'What can we learn about the effects of food stamps on obesity in the presence of misreporting?', *American Journal of Agricultural Economics* .

Baum, C. L. (2011), 'The effects of food stamps on obesity', *Southern Economic Journal* **77**(3), 623–651.

Black, D. A., Berger, M. C. & Scott, F. A. (n.d.), *Journal of the American Statistical Association* (451), 739–748.

Bollinger, C. R. (1996), 'Bounding mean regressions when a binary regressor is mismeasured', *Journal of Econometrics* **73**(2), 387–399.

Bound, J. (1991), 'Self-reported versus objective measures of health in retirement models', *Journal of Human Resources* **26**(1), 106–138.

Brachet, T. (2008), 'Maternal smoking, misclassification, and infant health', *Misclassification, and Infant Health (September 16, 2008)* .

Burgstahler, R., Gundersen, C., Garasky, S. et al. (2012), 'The supplemental nutrition assistance program, financial stress, and childhood obesity', *Agricultural and Resource Economics Review* **41**(1), 29.

Davidson, R. & MacKinnon, J. G. (2004), *Econometric theory and methods*, Vol. 5, Oxford University Press New York.

Fan, M. (2010), 'Do food stamps contribute to obesity in low-income women? evidence from the national longitudinal survey of youth 1979', *American Journal of Agricultural Economics* .

Frazis, H. & Loewenstein, M. A. (2003), 'Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables', *Journal of Econometrics* **117**(1), 151–178.

Gallant, A. & Nychka, D. (1987), 'Semi-nonparametric maximum likelihood estimation', *Econometrica* **55**, 363–390.

Gandelman, N. & Rasteletti, A. (2013), 'Credit constraints, sector informality and firm investments: Evidence from a panel of uruguayan firms'.

Haider, S. & Solon, G. (2006), 'Life-cycle variation in the association between current and lifetime earnings', *American Economic Review* **96**(4), 1308–1320.

Hausman, J. A., Abrevaya, J. & Scott-Morton, F. M. (1998), 'Misclassification of the dependent variable in a discrete-response setting', *Journal of Econometrics* **87**(2), 239–269.

Hoyne, H., Schanzenbach, D. W. & Almond, D. (2016), 'Long-run impacts of childhood access to the safety net', *The American Economic Review* **106**(4), 903–934.

Ichimura, H. & Lee, L.-F. (1991), 'Semiparametric least squares estimators of multiple index models', *Nonparametric and Semiparametric Methods in Econometrics* pp. 3–49.

Kane, T. J. & Rouse, C. E. (1995), 'Labor-market returns to two-and four-year college', *The American Economic Review* pp. 600–614.

Kane, T. J., Rouse, C. E. & Staiger, D. (1999), Estimating returns to schooling when schooling is misreported, Technical report, National bureau of economic research.

Kreider, B., Pepper, J. V., Gundersen, C. & Jolliffe, D. (2012), 'Identifying the effects of snap (food stamps) on child health outcomes when participation is endogenous and misreported', *Journal of the American Statistical Association* **107**(499), 958–975.

Lewbel, A. (2007), 'Estimation of average treatment effects with misclassification', *Econometrica* **75**(2), 537–551.

Loewenstein, M. A. & Spletzer, J. R. (1997), 'Delayed formal on-the-job training', *Industrial & labor relations review* **51**(1), 82–99.

Lynch, A. G., Marioni, J. C. & Tavaré, S. (2007), 'Numbers of copy-number variations and false-negative rates will be underestimated if we do not account for the dependence between repeated experiments', *American journal of human genetics* **81**(2), 4–18.

Lynch, V., Resnick, D. M., Staveley, J. & Taeuber, C. (2008), *Differences in esti-*

mates of public assistance recipiency between surveys and administrative records, Jacob France Institute, University of Baltimore.

Mahajan, A. (2006), 'Identification and estimation of regression models with misclassification', *Econometrica* **74**(3), 631–665.

Marquis, K. H. & Moore, J. C. (2010), 'Measurement errors in sipp program reports', *Survey Methodology* **1**.

Meyer, B. D. & Goerge, R. (2011), 'Errors in survey reporting and imputation and their effects on estimates of food stamp program participation', *US Census Bureau Center for Economic Studies Paper No. CES-WP-11-14* .

Meyer, B. D., Goerge, R. & Mittag, N. (2013), 'Errors in survey reporting and imputation and their effects on estimates of food stamp program participation', *Unpublished Manuscript* .

Meyer, B. D., Mok, W. K. & Sullivan, J. X. (2009), 'The under-reporting of transfers in household surveys: its nature and consequences', *National Bureau of Economic Research* .

Meyerhoefer, C. D. & Pylypchuk, Y. (2008), 'Does participation in the food stamp program increase the prevalence of obesity and health care spending?', *American Journal of Agricultural Economics* **90**(2), 287–305.

Mittag, N. (2013), 'A method of correcting for misreporting applied to the food stamp program', *US Census Bureau Center for Economic Studies Paper No. CES-WP-13-28* .

Poirier, D. J. (1980), 'Partial observability in bivariate probit models', *Journal of Econometrics* **12**, 209–217.

Stephens, M. & Unayama, T. (2015), 'Estimating the impacts of program benefits: Using instrumental variables with underreported and imputed data', *NBER Working Papers* (21248).

Tripathi, G. (1999), 'A matrix extension of the cauchy-schwarz inequality', *Economics Letters* **63**, 1–3.