Semiparametric Identification of Binary Choice Models with Misreported Outcomes*

Orville Mondal† Rui Wang‡

September 20, 2022

Abstract

Our paper characterizes partial identification of a binary choice model when the binary dependent variable is potentially misreported. We propose two different approaches by exploiting different instrumental variables respectively. In the first approach, the instrument is assumed to only affect the true dependent variable but not misreporting probabilities. The second approach uses an instrument that only affects misreporting probabilities monotonically but does not influence the true dependent variable. These approaches neither impose distributional assumptions on unobserved disturbances nor assume parametric models for the misreporting process. We characterize conditional moment inequalities based on the identification results and demonstrate using simulations that the two approaches perform more robustly than the alternative parametric method. We apply the proposed approaches to study educational attainment using data from National Longitudinal Surveys in 1976.

Keywords: Binary choice model, Misreporting, Instrumental variable, Partial identification, Nonclassical measurement error

---

*We are grateful to Marc Henry, Keisuke Hirano, Sung Jae Jun, and Joris Pinkse for their valuable comments.

†Department of Economics, The Pennsylvania State University. Email: odm4@psu.edu

‡Department of Economics, The Ohio State University. Email: wang.16498@osu.edu

1 Introduction

This paper studies semiparametric identification of a binary choice model when the binary dependent variable is potentially misreported. Binary choice models have been widely used in empirical applications such as analyzing participation in social programs, employment status, and educational attainment. However, many applications rely on survey data such as the Survey of Income and Program Participation (SIPP) and the Current Population Survey (CPS), where the binary outcome variable may be misreported or misclassified in survey data due to interviewer or respondent errors. The problem of misreporting is well documented and many studies show that the misreporting probabilities can be significant. For example, Meyer, Mittag, and George (2020) show that the probability of misreporting participation in a food stamp program can range from 23% in the SIPP to 50% in the CPS.

Several papers have analyzed the bias induced by misreporting in different econometric models (Aigner, 1973; Bollinger and David, 1997; Kane, Rouse, and Staiger, 1999; Davern, Klerman, Ziegenfuss, Lynch, and Greenberg, 2009; Nguimkeu, Denteh, and Tchernis, 2019). Regarding a binary choice model, Meyer and Mittag (2017) show that misreporting in the binary dependent variable can lead to serious biases in parametric estimators. While misreporting might be pervasive in some widely used datasets, these datasets may remain valuable sources of information, often with no appropriate substitute. It is therefore vital to investigate what can still be learned from the contaminated data.

Tackling misreporting issues can be challenging. First of all, misreporting in a binary variable involves non-classical measurement errors since the measurement error is always negatively correlated with the true outcome. Moreover, misreporting may be due to unobserved respondents' incentives; for example, people who benefit from a food stamp program may conceal their participation out of a sense of shame. As such, misreporting probabilities can depend on observed characteristics in an unknown way.

This paper proposes two different approaches to identify a binary choice model with a potentially misreported dependent variable. We adopt the binary choice model in Manski (1985) but consider that the binary outcome variable may be subject to misreporting. Manski (1985) mainly exploits the relationship between the sign of the covariate index and the conditional expectation of the true outcome for identification. However, the conditional expectation of the true outcome is no longer identified if there is potential misreporting in the true outcome. Our identification strategy comprises two steps. In the first step, we introduce two different approaches to bound the conditional expectation of the true outcome. The second step derives partial identification for the binary choice model based on the relationship between the sign of the covariate index and the bounds characterized in the first step.

The first step of our identification analysis establishes bounds on the conditional expectation of the true outcome. We introduce two different approaches for the identification by exploiting the presence of different instruments respectively. In our first approach, the instrument is assumed to only affect the true outcome while not influencing misreporting probabilities. In the example of program participation, this instrument can be people's eligibility for the program. Variation in this instrument can be used to bound misreporting probabilities, then leading to bounds on the conditional expectation of the true outcome. The second approach uses an instrument that only affects misreporting probabilities monotonically, but does not influence the true outcome. One example of this variable could be interview styles in survey data such as in-person, phone, or email interviews. People are more likely to report the truth when they have an in-person interview than an email interview. We characterize bounds on the conditional expectation of the true outcome by exploiting the exclusion restriction and monotonicity of this instrument.

Our approaches are all semiparametric since they do not impose distributional assumptions on the unobserved disturbance in the binary choice model. Moreover, our methods can allow for a wide range of misreporting processes. First, we do not assume a parametric model for the misreporting process (such as a linear index model) and allow for arbitrary dependence between the true outcome and the misreporting process. Furthermore, we allow misreporting probabilities to depend on observed characteristics arbitrarily. This flexibility has practical value; for example, Bollinger and David (1997, 2001) demonstrate that misreporting probabilities are correlated with participants' characteristics such as demographic characteristics and family income.

We characterize conditional moment inequalities for the model parameter based on the identification results. We examine the finite sample performance of our method via simulations. For comparison, we also implement two alternative methods: the parametric method studied in Hausman, Abrevaya, and Scott-Morton (1998) and the maximum score estimator assuming no misreporting. The results show that our approaches can perform more robustly to distributional assumptions and the misreporting processes. As an empirical illustration, we apply our method to study a binary choice model of educational attainment using the data from National Longitudinal Surveys in 1976.

In the extension, we study the joint identifying power of using the two instrumental variables together. The two instruments jointly provide a new channel for identification, which can lead to a tighter bound on the conditional expectation of the true outcome compared to simply taking intersections over the bounds by using each instrument separately. Moreover, we characterize identification results with one-sided misreporting, which can be relevant in some applications. For example, Nguimkeu, Denteh, and Tchernis (2019) study one-sided misreporting in the application of SNAP participation and provide more discussions on potential applications of one-sided misreporting.

1.1 Related Literature

This paper contributes to the literature studying a binary choice model with a potentially misreported dependent variable. Hausman, Abrevaya, and Scott-Morton (1998) propose a modified maximum likelihood estimator to correct misclassification, while their method specifies a parametric distribution over unobserved disturbances and assumes misclassification rates to be independent of covariates. Bollinger and David (1997) allow misreporting rates to depend on covariates in a probit model, imposing parametric assumptions for both the true binary model and misreporting rates. Lewbel (2000) studies semiparametric identification by using a continuous instrumental variable that only affects the true outcome but is independent of misreporting rates. In a more recent paper, Meyer and Mittag (2017) propose different parametric estimators that either need a parametric model for misreporting rates or the availability of validation data. Our paper does not require parametric assumptions on either the true binary choice model or the misreporting rates. Moreover, we do not restrict the support of the two instruments used in this paper so our identification results hold even if the instruments are binary.

Our work also relates to other models with potential misreporting in a discrete variable. There is a large literature studying a regression model with misreporting in the binary or discrete regressor, dating back to Aigner (1973), Bollinger (1996), Frazis and Loewenstein (2003) which assume constant misreporting probabilities. Mahajan (2006), Lewbel (2007), DiTraglia and Garcia-Jimeno (2019) allow misreporting probabilities to be covariates dependent and focus on the identifying power of a binary instrument that only affects the true regressor but not misreporting rates. Hu (2008) and Molinari (2008) extend to misclassification in a discrete regressor. Hu (2008) uses an instrument for identification and Molinari (2008) explores different restrictions on misreporting probabilities such as upper bounds on misreporting rates. A more related paper in Nguimkeu, Denteh, and Tchernis (2019) employs two instruments jointly, one of which only affects the true regressor and the other only influences misreporting. Their paper achieves point identification under parametric distributions and one-sided misreporting.

The above literature considers homogeneous effects of the binary regressor given covariates. Several papers attempt to allow for heterogeneous treatment effects with a misreported treatment such as Kreider and Pepper (2007), Kreider, Pepper, Gundersen, and Jolliffe (2012), Battistin, De Nadai, and Sianesi (2014), Calvi, Lewbel, and Tommasi (2017), and Ura (2018). Their approaches either exploit a repeated measurement for the true treatment, auxiliary administrative data to restrict misreporting errors, or an instrument related to the true treatment. Our work focuses on misreporting in the binary dependent variable instead of the regressor. Moreover, we explore the identifying power of two different instruments respectively in a nonparametric way.

The rest of this paper is organized as follows. Section 2 introduces a binary choice model with a misreported dependent variable. Section 3 presents the identification results. Section 4 characterizes conditional moment inequalities and Section 5 examines the finite sample performance via simulations. Section 6 studies the application of educational attainment. Section 7 discusses two extensions and Section 8 concludes.

2 Baseline

Our analysis focuses on a binary choice model with possibly misreporting (or misclassification) in the binary outcome variable. Let $Y_i^* \in \{0,1\}$ denote the true binary outcome variable, and $Y_i \in \{0,1\}$ denote the observed outcome variable which may be subject to misreporting. The true outcome variable $Y_i^* \in \{0,1\}$ depends on a vector of observed covariate denoted as $X_i \in \mathcal{X}$, and a latent variable $\epsilon_i \in \mathbb{R}$. We study the following binary choice model:

$$Y_i^* = \mathbb{1}\{X_i'\beta_0 + \epsilon_i \geq 0\}. \tag{1}$$

The objective is to learn the unknown parameter β_0, which captures the effect of covariate X_i on the true outcome variable Y_i^*. Without misreporting (i.e. $Y_i^* = Y_i$), the identification and estimation of β_0 have been widely studied in the literature under various parametric and nonparametric assumptions on the latent variable ϵ_i. For example, the probit/logit model assumes ϵ_i to follow a standard normal/logistic distribution and to be independent of covariate X_i. Manski (1985) studies nonparametric identification of β_0 by assuming that the median of ϵ_i is zero conditional on covariate X_i.

Distinct from the above literature, our paper considers the situation where the observed outcome Y_i is a potentially misreported version of the true outcome Y_i^*. We allow for two-sided misreporting: underreporting occurs when $Y_i^* = 1$ but $Y_i = 0$, while overreporting occurs when $Y_i^* = 0$ but $Y_i = 1$. The measurement error $Y_i - Y_i^*$ is -1 when there is underreporting, and it is 1 when there is overreporting. This sort of measurement error is referred to as being "non-classical." Classical measurement error refers to when the measurement error is assumed to be independent of the true outcome Y_i^*. We study misreporting in a binary variable, then the measurement error is negatively correlated with the true outcome Y_i^* so it is "non-classical."

Due to potential misreporting, earlier identification results for model (1) such as Manski (1985) do not apply. Our paper proposes two different approaches to identify β_0, which are robust to potential misreporting in the binary outcome variable. Before describing the identification strategy, we first introduce a basic independence assumption on the unobserved term ϵ_i.

Assumption 1 (Median Independence)**.**

$$\mathrm{Med}(\epsilon_i \mid X_i) = 0.$$

Assumption 1 is used in Manski (1985) and it is a special case of the quantile regression model ($\tau = 0.5$). This assumption only requires a median independence condition instead of independence for the whole distribution. The probit/logit specification of ϵ_i (without misreporting) can be nested in Assumption 1.

Let $p^*(x) = \Pr(Y_i^* = 1 \mid x)$ denote the probability that the true outcome equals to one given $X_i = x$. As shown in Manski (1985), Assumption 1 implies the following identification conditions for β_0 when ϵ_i is continuously distributed:

$$\mathrm{sgn}\{x'\beta_0\} = \mathrm{sgn}\left\{p^*(x) - 0.5\right\}. \tag{2}$$

where $\mathrm{sgn}\{x\} = \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$ denotes the sign function. However, the identification result in (2) for β_0 does not apply to the present context since $p^*(x)$ is not identified when the true outcome Y_i^* is not observed. Our paper introduces different methods to (partially) identify the true conditional probability $p^*(x)$ using observed variables (X_i, Y_i). Then given bounds on $p^*(x)$, we can characterize identifying restrictions on the model parameter β_0.

3 Identification

Our identification strategy proceeds in two steps. The first step is to derive bounds for $p^*(x)$ as $p^*(x) \in [L(x), U(x)]$, where the bounds only depend on observed variables (X_i, Y_i) and thus can be identified from data. The analysis to bound $p^*(x)$ will be described in detail in a later section. Given bounds for $p^*(x)$, the second step is to derive identifying conditions on β_0.

We first assume that the true probability $p^*(x)$ is identified as $p^*(x) \in [L(x), U(x)]$ and show the main idea for the second step. As shown in condition (2), the sign of the covariate index $x'\beta_0$ is the same with the sign of $p^*(x) - 0.5$. From the sign of $p^*(x) - 0.5$, we can infer that its upper bound $U(x) - 0.5$ should be positive if its sign is positive and its lower bound $L(x) - 0.5$ should be negative if its sign is negative.

The following conditions display the relationship between the sign of the covariate index and the sign of the bounds:

$$
\begin{aligned}
x'\beta_0 \geq 0 &\Longrightarrow p^*(x) - 0.5 \geq 0 \Longrightarrow U(x) - 0.5 \geq 0, \\
x'\beta_0 \leq 0 &\Longrightarrow p^*(x) - 0.5 \leq 0 \Longrightarrow L(x) - 0.5 \leq 0.
\end{aligned} \tag{3}
$$

The above conditions form identifying restrictions for the parameter β_0 by using bounds on $p^*(x)$ which can be identified from the data. Compared to condition (2), condition (3) can only derive the sign of either the upper bound or the lower bound of $p^*(x) - 0.5$ but not both of them. Therefore, the identified set for β_0 characterized by condition (3) is generally larger than the one in condition (2), but this result is robust to misreporting in the outcome variable. Ignoring the misreporting issue may lead to misleading results for β_0 as shown in Meyer and Mittag (2017).

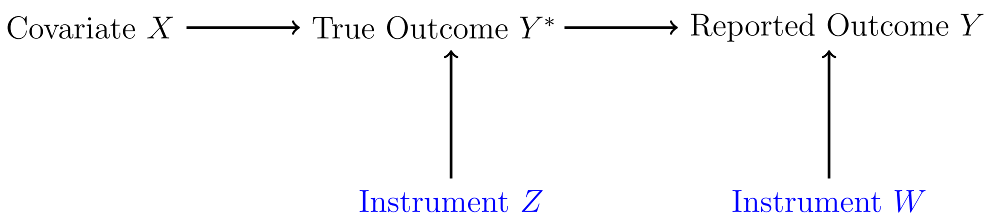
We now focus on the first step of the identification analysis, which is about establishing bounds on the true conditional probability $p^*(x) = \Pr(Y_i^* = 1 \mid x)$. Let $p(x) = \Pr(Y_i = 1 \mid x)$ denote the reported probability of $Y_i = 1$ given $X_i = x$ which is identified from the data. The reported probability $p(x)$ depends on two components: the true probability and misreporting probabilities. We need to distinguish the two components in order to learn the true probability from the reported probability.

We propose two different approaches to identify the true probability $p^*(x)$ by exploiting different exclusion restrictions. In the first approach, we use an instrument $Z_i \in \mathcal{Z}$ that only affects the true probability $p^*(x)$ but does not affect misreporting probabilities. The other approach uses an instrument $W_i \in \mathcal{W}$ that only affects misreporting probabilities but not the true probability $p^*(x)$. The two instruments can be either discrete or continuous, and our identification results hold even when the two instruments are binary. We first study the identifying power of the two instruments separately and then discuss how they can jointly identify the true probability $p^*(x)$ in the extension.

For simplicity of notation, we suppress subscript i for random variables in the following analysis. The following graph describes the relationship among all variables.

Figure 1 summarizes our model and main identification strategies. The objective is to learn the effects of covariate X on the true binary outcome Y^*, but we only observe the reported outcome Y which can be possibly misreported. We study the

Figure 1: Relationship among Variables



identifying power of two different instruments respectively: the first one is instrument Z that only affects the true outcome Y^* and the other is instrument W that only affects the reported outcome Y by affecting misreporting probabilities.

Next, we study the identification by exploiting each of the two instruments respectively for the true conditional probability $p^*(x)$.

3.1 Instrument Z

This section studies the identifying power of instrument Z that only affects the true choice probability but does not influence misreporting probabilities. Instrument Z enters the binary choice model (1) directly so it is a component of the covariate vector X. Therefore, we divide covariate X into two parts: instrument Z and the remaining covariates denoted as $\tilde{X} = X \setminus Z \in \tilde{\mathcal{X}}$.

Next we state some assumptions on instrument Z.

Assumption 2 (Exclusion)**.** *For any $x \in \mathcal{X}, y \in \{0, 1\}$,*

$$\Pr(Y = 1 - y \mid Y^* = y, x) = \Pr(Y = 1 - y \mid Y^* = y, \tilde{x}).$$

The exclusion restriction requires instrument Z to be independent of misreporting probabilities, so it only affects the reported probability by shifting the true probability. When the true outcome is participation in social programs, one example of instrument Z can be the eligibility for the program. Several papers such as Mahajan (2006), Ura (2018), and DiTraglia and Garcia-Jimeno (2019) exploit similar exclusion restrictions in different models with misreporting and provide more examples for this instrument in various applications.

The next assumption is about the extent of misreporting probabilities.

Assumption 3 (Degree of Misreporting)**.** *For any $\tilde{x} \in \tilde{\mathcal{X}}$,*

$$\Pr(Y = 0 \mid Y^* = 1, \tilde{x}) + \Pr(Y = 1 \mid Y^* = 0, \tilde{x}) \leq 1.$$

Assumption 3 is about the degree of misreporting so that the reported data is informative for the true probability. It requires that the misreporting errors are not too large so that the sum of two-sided misreporting probabilities is smaller than one. This assumption is consistent with the empirical evidence in multiple studies such as Meyer, Mok, and Sullivan (2009) and Meyer, Mittag, and George (2020), which document that one-sided misreporting probability is usually less than 50% in survey data. Actually, our identification analysis only needs the degree of misreporting to be known, and the analysis applies to the case where the sum of two-sided misreporting rates is larger than one.

Assumption 4 (Boundary Condition)**.** *The reported choice probability $p(x)$ satisfies that: $\sup_{z \in \mathcal{Z}} p(\tilde{x}, z) > 0$ and $\inf_{z \in \mathcal{Z}} p(\tilde{x}, z) < 1$ for any $\tilde{x} \in \tilde{\mathcal{X}}$.*

Assumption 4 is a boundary condition for the reported probability, which only requires that the supremum of the reported probability is bounded away from zero and the infimum of the reported probability is bounded away from one. This assumption can be satisfied as long as instrument Z strictly affects the reported probability, and takes on at least two values.

Under the above assumptions, we are ready to establish identification results for the true probability $p^*(x)$.

Proposition 1. *Under Assumptions 2-4, the sharp bounds for $p^*(x)$ are characterized as $p^*(x) \in [L_1(x), U_1(x)]$ for any $x \in \mathcal{X}$, where*

$$L_1(x) = \frac{p(x) - \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)}{1 - \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)}, \qquad U_1(x) = \frac{p(x)}{\sup_{z \in \mathcal{Z}} p(\tilde{x}, z)}.$$

Proposition 1 characterizes sharp bounds for the true probability $p^*(x)$ by using variation in instrument Z. The identifying power of instrument Z depends on the range of the reported probability $p(\tilde{x}, z)$ as one varies z while fixing $\tilde{x}$: a larger range of the reported probability leads to tighter bounds for the true probability. The idea is that a larger variation in the reported probability can provide smaller bounds for misreporting probabilities and lead to tighter bounds on the true probability. If the conditional reported probability can vary from zero to one when instrument Z changes, we can infer that there is no misreporting and achieve point identification as $p^*(x) = p(x)$.

To illustrate the intuition, we focus on the case where the reported probability can change from zero to one. When the reported probability is one (or zero), there are two possibilities: one is that the true conditional probability is one, and people with $Y^* = 1$ all report the truth (or misreport); the other is that the true probability is zero while people with $Y^* = 0$ all misreport (or report the truth). The possibility that everyone misreports can be rejected by Assumption 3 which requires the sum of two-sided misreporting probabilities to be smaller than one. Therefore, we can conclude that there is no misreporting, and point identification for the true probability is obtained.

3.2 Instrument W

This section provides identification for the true probability from a different channel by using instrument W. This instrument is assumed to only affect misreporting probabilities but not the true probability. In addition to covariate X in the true binary choice model, instrument W is a variable that only affects the misreporting process but is independent of the true choice model. Our approach does not impose any parametric models for the misreporting process so instrument W is allowed to affect misreporting probabilities nonparametrically.

The complete set of observed variables are (X, W, Y) in this section. Let $p_W(x, w) = \Pr(Y = 1 \mid x, w)$ denote the reported probability conditional on $(X, W) = (x, w)$. The subscript W in the function p_W is used to distinguish it from the function $p(x) = \Pr(Y = 1 \mid x)$.

The following are some assumptions on instrument $W \in \mathcal{W}$.

Assumption 5 (Exclusion)**.**

$$\Pr(Y^* = 1 \mid x, w) = \Pr(Y^* = 1 \mid x) = p^*(x),$$

for any $x \in \mathcal{X}$ and $w \in \mathcal{W}$.

Similar to Assumption 2 for instrument Z, Assumption 5 states a different exclusion restriction requiring instrument W to be independent of the true outcome. One example for instrument W could be the interview style in survey data such as email interview, phone interview, and in-person interview. Different interview styles affect respondents' probabilities of reporting the truth, but are unlikely to affect the true outcome.

Assumption 6 (Monotonicity)**.** *For any $x \in \mathcal{X}$, the following conditions hold for any $y \in \{0, 1\}$ and $w_1 > w_2 \in \mathcal{W}$,*

$$\Pr(Y = 1 - y \mid Y^* = y, x, w_1) \leq \Pr(Y = 1 - y \mid Y^* = y, x, w_2).$$

Assumption 6 is the monotonicity condition for instrument W: the misreporting probabilities are weakly decreasing with respect to the instrument. In the example of interview styles, people would be more likely to report the truth during an in-person interview than a phone interview. Assumption 6 is relatively weak since it only requires monotonicity for the average probability but can allow for violations of some populations.

Assumption 7. *(1) Degree of Misreporting: for any $x \in \mathcal{X}, w \in \mathcal{W}$,*

$$\Pr(Y = 1 \mid Y^* = 0, x, w) + \Pr(Y = 0 \mid Y^* = 1, x, w) \leq 1.$$

(2) Boundary Condition: the reported probability $p_W(x, w)$ is bounded away from zero and one: $0 < p_W(x, w) < 1$ for any $x \in \mathcal{X}, w \in \mathcal{W}$.

Assumption 7 is similar to Assumptions 3-4 for instrument Z. The difference is that all probabilities are also conditional on the additional variable W since the misreporting probabilities and the reported probability depend on instrument W in this scenario.

Under the above assumptions, the next proposition establishes identification results for the true probability $p^*(x)$.

Proposition 2. *Under Assumptions 5-7, the true probability $p^*(x)$ can be bounded as $p^*(x) \in [L_2(x), U_2(x)]$ for any $x \in \mathcal{X}$, where*

$$L_2(x) = \sup_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w) - \inf_{\tilde{w} \leq w} p_W(x, \tilde{w})}{1 - \inf_{\tilde{w} \leq w} p_W(x, \tilde{w})} \right\}, \quad U_2(x) = \inf_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w)}{\sup_{\tilde{w} \leq w} p_W(x, \tilde{w})} \right\}.$$

And the above bounds are sharp when instrument W is binary.

Proposition 2 characterizes the identifying power of instrument W that only affects misreporting probabilities monotonically but not the true probability. It also shows that the bounds $L_2(x)$ and $U_2(x)$ are the best possible, given the assumptions and data, when instrument W is binary. The identification results in Proposition 2 imply that $U_2(x) \geq L_2(x)$ for all x, which can serve as testable implications for our assumptions. It can be tested by using methods developed for conditional moment inequalities such as Andrews and Shi (2013).

Proposition 2 mainly exploits the exclusion and monotonicity of instrument W. Under the monotonicity condition, the misreporting probabilities at w can be bounded above by all upper bounds of misreporting probabilities evaluated at smaller values $\tilde{w} \leq w$. The identifying power of monotonicity is shown within the bracket of the bounds $(L_2(x), U_2(x))$. The exclusion restriction can help tighten the bounds for the true probability $p^*(x)$ by taking intersections of all bounds derived from any value of instrument W, whose identifying power is shown outside the bracket of bounds.

3.3 Identified Set for β_0

We have established bounds for the true probability $p^*(x)$ under two different scenarios in Sections 3.1 and 3.2. Given bounds on $p^*(x)$, we are ready to characterize identification results for the parameter β_0. For model (1), the parameter β_0 can be only identified up to a constant. Let β^1 denote the first element of the parameter β, and we normalize it as one: $\mathcal{B} = \{\beta : |\beta^1| = 1\}$. The following theorem characterizes identification results for the parameter β_0.

Theorem 1. *Under Assumption 1, an identified set $B_{I,t}$ for β_0 is characterized as $\beta_0 \in B_{I,t}$=$\{\beta \in \mathcal{B}$: β satisfies condition (4) for any $x\}$:*

$$
\begin{aligned}
x'\beta \geq 0 &\Longrightarrow U_t(x) - 0.5 \geq 0, \\
x'\beta \leq 0 &\Longrightarrow L_t(x) - 0.5 \leq 0.
\end{aligned}
\tag{4}
$$

where $t = 1$ under Assumptions 2-4 and $t = 2$ under Assumptions 5-7. The identified set $B_{I,t}$ is sharp when the bounds $[L_t(x), U_t(x)]$ for $p^*(x)$ are sharp.

Theorem 1 characterizes identification results for the true parameter β_0 based on the relationship between the sign of the covariate index and the sign of the bounds on the true probability $p^*(x)$. Therefore the size of the identified set $B_{I,t}$ depends on the informativeness of the bounds, and tighter bounds lead to a smaller identified set for β_0.

In Theorem 1, we provide semiparametric identification results for the model parameter β_0 without imposing parametric distributions over the latent variable ϵ. Moreover, our identification analysis mainly relies on the exclusion restriction or the monotonicity condition of the two instruments, while it is flexible about the misreporting process. We do not impose any parametric models for the misreporting process. Also, we allow for heterogeneous misreporting probabilities that can depend on covariates arbitrarily. This approach is shown to perform more robustly than the parametric method via simulations in Section 5.

4 Conditional Moment Inequalities

This section characterizes the identified set using conditional moment inequalities based on the identification results in Theorem 1. Then we can conduct estimation and inference for β_0 using established methods in the literature developed for general conditional moment inequalities.

To characterize conditional moment inequalities, we first conduct monotone transformations of the bounds $(L_t(x), U_t(x))$ by multiplying them by their respective (positive) denominators. This monotone transformation preserves the sign of those bounds so the identified set characterized by condition (4) stays unchanged. We focus on the case $t = 1$, and a similar analysis applies to the case $t = 2$. When $t = 1$, we define $(l_1(x), u_1(x))$ as follows:

$$
\begin{aligned}
l_1(x) &= (L_1(x) - 0.5)(1 - \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)) = p(x) - 0.5 \inf_{z \in \mathcal{Z}} p(\tilde{x}, z) - 0.5, \\
u_1(x) &= (U_1(x) - 0.5) \sup_{z \in \mathcal{Z}} p(\tilde{x}, z) = p(x) - 0.5 \sup_{z \in \mathcal{Z}} p(\tilde{x}, z).
\end{aligned}
$$

We look at the identifying restrictions with nonnegative covariate index $x'\beta \geq 0$ in condition (4), which has the following implication:

$$
x'\beta_0 \geq 0 \Longrightarrow U_1(x) - 0.5 \geq 0 \Longleftrightarrow u_1(x) \geq 0.
$$

Based on the above identifying restriction, we can construct a conditional moment inequality as follows:

$$
\begin{aligned}
E[m_1^1(X, Y, \beta_0) \mid X = x] &\equiv E[X'\beta_0 \mathbb{1}\{X'\beta_0 \geq 0\}(Y - 0.5 \sup_{z \in \mathcal{Z}} p(\tilde{X}, z)) \mid X = x] \\
&= x'\beta_0 \mathbb{1}\{x'\beta_0 \geq 0\} u_1(x) \geq 0.
\end{aligned}
$$

The conditional moment inequality for the nonpositive covariate index denoted as $E[m_1^2(X, Y, \beta_0) \mid X = x]$ can be constructed similarly:

$$
E[m_1^2(X, Y, \beta_0) \mid X = x] \equiv E[X'\beta_0 \mathbb{1}\{X'\beta_0 \leq 0\}(Y - 0.5 \inf_{z \in \mathcal{Z}} p(\tilde{X}, z) - 0.5) \mid X = x].
$$

Let the function g_1 collects the two conditional moment inequalities $g_1(x, \beta_0) = (E[m_1^1(X, Y, \beta_0) \mid X = x]; E[m_1^2(X, Y, \beta_0) \mid X = x])$. The identified set $B_{I,1}$ is characterized by the following conditional moment inequalities:

$$
B_{I,1} = \{\beta \in \mathcal{B} : g_1(x, \beta) \geq 0 \;\; \forall x\}.
$$

Now we look at the characterization of conditional moment inequalities for the case $t = 2$. Similarly we conduct monotone transformation of the bounds $(L_2(x), U_2(x))$, and define $(l_2(x), u_2(x))$ as follows:

$$l_2(x) = \sup_{w\in\mathcal{W}} \Big\{p_W(x,w) - 0.5 \inf_{\tilde{w}\leq w\in\mathcal{W}} p_W(x,\tilde{w}) - 0.5\Big\},$$
$$u_2(x) = \inf_{w\in\mathcal{W}} \Big\{p_W(x,w) - 0.5 \sup_{\tilde{w}\leq w\in\mathcal{W}} p_W(x,\tilde{w})\Big\}.$$

The identifying restriction with nonnegative covariate index in condition (4) has the following implication: for any x,

$$x'\beta_0 \geq 0 \Longrightarrow u_2(x) \geq 0 \Longleftrightarrow p_W(x,w) - 0.5 \sup_{\tilde{w}\leq w\in\mathcal{W}} p_W(x,\tilde{w}) \geq 0 \quad \forall w.$$

Based on the above condition, we can construct the following conditional moment inequality:

$$\begin{aligned} &E[m_2^1(X,W,Y,\beta_0) \mid X=x, W=w] \\ =&E[X'\beta_0 \mathbb{1}\{X_i'\beta_0 \geq 0\}(Y - 0.5 \sup_{\tilde{w}\leq w\in\mathcal{W}} p_W(X,\tilde{w}) \mid X=x, W=w] \geq 0 \quad \forall x, w. \end{aligned}$$

The conditional moment inequality based on nonpositive covariate index can be developed similarly. Let the function g_2 collect the two conditional moment inequalities, defined as

$$g_2(x,w,\beta_0) = \begin{cases} E[X'\beta_0 \mathbb{1}\{X'\beta_0 \geq 0\}(Y - 0.5 \sup_{\tilde{w}\leq w\in\mathcal{W}} p_W(X,\tilde{w})) \mid X=x, W=w], \\ E[X'\beta_0 \mathbb{1}\{X'\beta_0 \leq 0\}(Y - 0.5 \inf_{\tilde{w}\leq w\in\mathcal{W}} p_W(X,\tilde{w}) - 0.5) \mid X=x, W=w]. \end{cases}$$

The identified set $B_{I,2}$ is characterized as

$$B_{I,2} = \{\beta \in \mathcal{B} : g_2(x,w,\beta) \geq 0 \quad \forall (x,w)\}.$$

Theorem 2. *The identified set $B_{I,t}$ is characterized as follows: for $t = \{1,2\}$,*

$$B_{I,1} = \{\beta \in \mathcal{B} : g_1(x,\beta) \geq 0 \quad \forall x\},$$
$$B_{I,2} = \{\beta \in \mathcal{B} : g_2(x,w,\beta) \geq 0 \quad \forall (x,w)\}.$$

Theorem 2 characterizes the identified set $B_{I,t}$ using conditional moment inequalities. The literature has developed various methods to conduct inference for conditional moment inequalities such as Kim (2008), Andrews and Shi (2013), and Chernozhukov, Lee, and Rosen (2013). Our conditional moment inequalities in Theorem 2 also depend on the function $p(x)$ or $p_W(x, w)$ that needs to be consistently estimated. This situation is accommodated in Andrews and Shi (2013) (Section 8), which allows for conditional moment inequalities with a preliminary consistent estimator.

5 Simulation Study

This section examines the finite sample performance of the identified set $B_{I,t}$ for $t \in \{1, 2\}$ via Monte Carlo simulations. To assess the robustness of our method, we also implement two other established methods for comparison. One method proposed in Hausman, Abrevaya, and Scott-Morton (1998) accounts for potential misreporting in a binary choice model but assumes constant misreporting probabilities and parametric distributions. The second method does not allow for misreporting and estimates a standard binary choice model under Assumption 1. The simulation results show that the method in this paper can perform more robustly concerning misreporting process and parametric assumptions.

We adopt the approach in Chernozhukov, Hong, and Tamer (2007) and Andrews and Shi (2013) to estimate the identified set $B_{I,t}$. We focus on the identified set $B_{I,1}$ and a similar analysis can apply to $B_{I,2}$. Given the conditional moment inequalities in Theorem 2, we follow Andrews and Shi (2013) to transform conditional moment inequalities into unconditional moment inequalities using instrumental functions. We use indicator functions of hypercubes in the space of covariates as instrumental functions. The space of covariates is divided into k subintervals I and let $\mathcal{I}$ denote the collection of all hypercubes. Then the objective function can be established as follows:

$$Q(\beta) = \sum_{I \in \mathcal{I}} \sum_{j \in \{1,2\}} \left[E[m_1^j(X, Y, \beta) \mathbb{1}\{X \in I\}] \right]_-^2 ,$$

where $[x]_- = \min\{x, 0\}$.

The set of minimizers of the objective function Q equals to the identified set $B_{I,1}$

if $\mathcal{I}$ is a sufficiently rich collection of hypercubes. See Andrews, Kim, and Shi (2017) for more details and choices of instrumental functions.

Given a finite sample, let $\hat{p}$ denote an estimator for the conditional probability function $p(\cdot)$, which is estimated using the kernel method. We use a fourth-order Gaussian kernel function $k(u) = \frac{1}{2}(3 - u^2)\phi(u)$, where ϕ is the PDF of the standard normal distribution. The bandwidth is in the form $h_n = c \cdot \hat{\sigma}(X) \cdot n^{-1/5}$, where $c = 2.5$. Let function $\hat{m}_1^j$ denote the estimated moment function using the estimated function $\hat{p}$. The sample criterion function is defined as follows:

$$\hat{Q}_n(\beta) = \sum_{I \in \mathcal{I}} \sum_{j \in \{1,2\}} \left[\frac{1}{n} \sum_i \hat{m}_1^j(X_i, Y_i, \beta) \mathbb{1}\{X_i \in I\} \right]_-^2 .$$

Following Chernozhukov, Hong, and Tamer (2007), the estimated identified set $\hat{B}_{I,1}$ is given as

$$\hat{B}_{I,1} = \left\{ \beta \in \mathcal{B} : n\hat{Q}_n(\beta) \leq n \inf_{b \in \mathcal{B}} \hat{Q}_n(b) + \kappa_n \right\},$$

where κ_n is a positive sequence which should satisfy $\kappa_n \to +\infty$ and $\kappa_n/n \to 0$. It is chosen as $\kappa_n = 10^{-6} \cdot \log(n)$ in our simulation. Let $\hat{\beta}^l_{semi}$ and $\hat{\beta}^u_{semi}$ denote the estimated lower and upper bound of the model coefficient in our method.

To better evaluate our method, we implement two alternative methods for comparison. One is the parametric method allowing for potential misreporting as proposed in Hausman, Abrevaya, and Scott-Morton (1998). This method assumes the error term to follow a standard normal distribution and be independent of covariates. Moreover, it assumes that the misreporting probabilities are constant, which means that the misreporting process can be summarized by two constants (α_1, α_0) as follows:

$$\Pr(Y = 0 \mid Y^* = 1 \mid x) = \alpha_1, \quad \Pr(Y = 1 \mid Y^* = 0 \mid x) = \alpha_0.$$

This rules out the possibility that misreporting behavior may depend on covariates. Given these assumptions, the parameters $\beta_0, \alpha_1, \alpha_0$ can be estimated using the maximum likelihood method, and we use $\hat{\beta}_{par}$ to denote the estimator of the model coefficient in the method.

The second method we compare is the standard maximum score estimator which does not allow for misreporting. Given the assumptions in Manski (1985) and no misreporting assumption, the maximum score parameter defined as β_{mse} should satisfy the following conditional moment inequalities:

$$X'\beta_{mse}(E[Y \mid X] - 0.5) \geq 0.$$

We can transform the above conditional moment inequalities as unconditional moment inequalities using instrumental functions and build an objective function in the same way as described before. The minimizer of the sample criterion function is the estimator in this method, denoted as $\hat{\beta}_{mse}$.

Next, we present the performance of our two approaches and the two alternative methods.

5.1 Instrument Z

This section studies the performance of the identified set $\hat{B}_{I,1}$ by using instrument Z. The data generating process is described as follows. Instrument Z is uniformly distributed over the set $\{-1, -0.5, 0, 0.5, 1\}$, $\tilde{X}$ follow a uniform distribution over the interval $[0, 1]$, and the full covariate X is given as $X = [1; \tilde{X}; Z]$. The true parameter $\beta_0 = [1; 1.5; -1.5]$ and the true outcome is generated by $Y^* = \mathbb{1}\{X'\beta_0 \geq \epsilon\}$. We study two specifications of the error term ϵ: a standard normal distribution $\mathcal{N}(0, 1)$ and a Cauchy distribution $Cauchy(0, 0.5)$.

To describe the misreporting process, let $M_y \in \{0, 1\}$ denote the reporting variable given the true outcome $Y^* = y$, where $M_y = 1$ means reporting the truth. The reporting variable M_y follows a Bernoulli(α_y) distribution, where $\alpha_1 = 0.9$ and $\alpha_0 = 0.7$. The observed outcome Y is therefore given by $Y = M_1 \cdot Y^* + (1 - M_0) \cdot (1 - Y^*)$. The number of hypercubes is $k = \{20, 30, 40\}$ for the sample size $N = \{500, 1000, 2000\}$, respectively. The repetition number is $B = 300$.

To compare the performance of different methods, we report the root mean-squared error (rMSE) and median of absolute deviation (MAD) for the lower bound $\hat{\beta}^l_{semi}$ and upper bound $\hat{\beta}^u_{semi}$ in this paper, the parametric estimator $\hat{\beta}_{par}$, and the maximum score estimator $\hat{\beta}_{mse}$ without misreporting. Let β^k denote the kth element of the parameter β.

Tables 1 and 2 display the performance of the three different methods under different specifications of error term ϵ and different sample sizes. The results show that the lower and upper bound in our approach can perform uniformly better than the maximum score estimator assuming no misreporting regardless of the sample size and the specification of the error term. The parametric method does not necessarily perform better than our method even if the distribution of ϵ is correctly specified. It is because the parametric method also needs to estimate misreporting probabilities, while our method can directly estimate the coefficient β_0 without estimating misreporting probabilities by characterizing moment inequalities on β_0. Moreover, the parametric method performs much worse when the distribution of ϵ is misspecified (Cauchy), but the method in this paper has similar performances under different specifications. The results show that the method in this paper performs more robustly concerning distributional assumptions.

Table 1: Performance Comparisons for $\hat{\beta}^2$

Design	$\hat{\beta}^l_{semi}$		$\hat{\beta}^u_{semi}$		$\hat{\beta}_{par}$		$\hat{\beta}_{mse}$	
	rMSE	MAD	rMSE	MAD	rMSE	MAD	rMSE	MAD
	$N = 500$							
Normal	0.742	0.540	0.717	0.540	0.927	0.439	0.721	0.653
Cauchy	0.738	0.540	0.713	0.540	1.878	0.999	0.707	0.617
	$N = 1000$							
Normal	0.620	0.439	0.566	0.439	0.661	0.324	0.644	0.575
Cauchy	0.678	0.540	0.623	0.439	1.516	0.905	0.690	0.613
	$N = 2000$							
Normal	0.571	0.439	0.446	0.338	0.349	0.209	0.595	0.543
Cauchy	0.591	0.439	0.460	0.338	1.129	0.718	0.612	0.564

Table 2: Performance Comparisons for $\hat{\beta}^3$

Design	$\hat{\beta}^l_{semi}$		$\hat{\beta}^u_{semi}$		$\hat{\beta}_{par}$		$\hat{\beta}_{mse}$	
	rMSE	MAD	rMSE	MAD	rMSE	MAD	rMSE	MAD
	$N = 500$							
Normal	0.477	0.338	0.483	0.338	1.004	0.474	0.605	0.548
Cauchy	0.532	0.439	0.552	0.439	1.802	0.913	0.646	0.573
	$N = 1000$							
Normal	0.425	0.268	0.443	0.338	0.650	0.329	0.577	0.533
Cauchy	0.428	0.338	0.452	0.338	1.560	0.856	0.572	0.539
	$N = 2000$							
Normal	0.329	0.237	0.355	0.268	0.325	0.197	0.531	0.510
Cauchy	0.331	0.237	0.365	0.303	1.118	0.722	0.524	0.501

5.2 Instrument W

This section examines the performance of the identified set $\hat{B}_{I,2}$ by using instrument W. The data generating process is described as follows. Covariate X is given by $X = (1; \tilde{X})$, where $\tilde{X}$ follows a uniform distribution over $[-1, 1]$. The true outcome Y^* is generated by $Y^* = \mathbb{1}\{\epsilon \leq X'\beta_0\}$, where the true parameter is $\beta_0 = [1; 1.5]$. Similarly, we consider two specifications of error term ϵ: a standard normal distribution $\mathcal{N}(0, 1)$ and a Cauchy distribution $Cauchy(0, 0.5)$.

Instrument W is uniformly distributed over the set $\{1, 2, 3, 4, 5\}$. The reporting variables M_y follows Bernoulli(α_y) distribution, where α_y depends on instrument W in the following way:

$$\alpha_1 = 0.9, \quad \alpha_0 = 1 - \frac{1}{1 + 0.3W^2}.$$

Therefore, the monotonicity condition in Assumption 6 on instrument W is satisfied. The observed outcome Y is given as $Y = M_1 \cdot Y^* + (1 - M_0) \cdot (1 - Y^*)$. The number of hypercubes is $k = \{20, 30, 40\}$ for the sample size $N = \{500, 1000, 2000\}$, respectively. The repetition number is $B = 300$.

Table 3 presents the performance of the three methods for β_0^2 under different sample sizes and specifications of ϵ. The results show that the semiparametric method in this paper can perform uniformly better than the two other methods regardless of the sample size and distributional assumptions on ϵ, and this pattern becomes more significant when the sample size increases. In summary, the simulation results show the robustness advantage of the two approaches in this paper.

Table 3: Performance Comparisons for $\hat{\beta}^2$

Design	$\hat{\beta}^l_{semi}$		$\hat{\beta}^u_{semi}$		$\hat{\beta}_{par}$		$\hat{\beta}_{mse}$	
	rMSE	MAD	rMSE	MAD	rMSE	MAD	rMSE	MAD
	$N = 500$							
Normal	0.405	0.269	0.411	0.234	1.623	0.674	0.905	0.416
Cauchy	0.455	0.269	0.451	0.269	1.965	0.909	0.933	0.414
	$N = 1000$							
Normal	0.302	0.183	0.322	0.176	1.357	0.609	0.956	0.461
Cauchy	0.310	0.219	0.305	0.183	1.694	0.625	0.721	0.406
	$N = 2000$							
Normal	0.235	0.168	0.297	0.151	0.855	0.458	0.861	0.454
Cauchy	0.282	0.219	0.318	0.168	1.292	0.449	0.666	0.419

6 Empirical Illustration

As an empirical illustration, we apply our methods to analyze educational attainment using a binary choice model with potential misreporting. The dataset we use is drawn from the National Longitudinal Surveys in 1976 (NLSY76), which is also used in Card (1995) to estimate returns to education. This survey data contains 3613 individuals' self-reported information including educational experiences and family backgrounds. The objective is to explore how people's characteristics affect the probability of them attaining a college degree. However, there may be misreporting in self-reports of educational attainment in this data, which could severely bias the estimation results. The two approaches in this paper allow for misreporting in the binary outcome and we apply the first approach to study how people's observed characteristics affect the likelihood of attending a college.

In this application, the reported outcome Y is whether an individual reports attending a college which may be subject to misreporting. Instrument Z is whether an individual grew up near a four-year college (college proximity). This instrument affects people's true decision of attending college, but may not affect their misreporting behaviors. We also include two other covariates in the binary choice model: parents' average education X_1 and whether an individual is black X_2. The following table shows the summary statistics of all variables.

Table 4: Summary Statistics

	Y	X_1	X_2	Z
min	0	0	0	0
max	1	18	1	1
mean	0.268	10.173	0.230	0.678
std	0.443	2.786	0.421	0.467

We implement the first approach in this paper by using instrument Z. The full vector of covariate is $X = [1; X_1; X_2; Z]$, and the corresponding coefficient is denoted as $\beta_0 = [\beta_0^0, \beta_0^1, \beta_0^2, \beta_0^3]$. For the method in this paper, the coefficient β_0 can be only identified up to a constant. The coefficient of instrument Z is normalized as one since people tend to be more likely to attend a college when they live closer to a college. For comparison, we also display the results of the parametric approach described in Section 5.

Table 5: Estimation Results

	$\hat{\beta}^0$	$\hat{\beta}^1$	$\hat{\beta}^2$	$\hat{\beta}^3$
this paper	[-1.667, -0.879]	[0.030, 0.091]	[-1, -0.333]	1
parametric	-32.577	-23.368	-10.582	-27.823

Table 5 presents the estimation results for the coefficients in the binary choice model. The method in this paper shows a positive sign of parents' education and a negative sign of being black for educational attainment. However, the parametric method shows negative signs for all coefficients, which seems inconsistent with economic intuition. Parents' education and living closer to a college are likely to increase the chance of attending a college instead of decreasing the chance. The results show that misspecifications in either parametric assumptions or misreporting processes may lead to opposite signs of the coefficients.

7 Extension

7.1 Two Instruments

Section 3.1 and 3.2 provide bounds for the true conditional probability $p^*(x)$ when there is only one instrument available. This section studies the joint identifying power of the two instruments (Z, W). The observed variables are $(X, W, Y) = (\tilde{X}, Z, W, Y)$ when two instruments are available. We adjust previous assumptions in Section 3.1 and 3.2 slightly to accommodate the availability of the two instruments.

Assumption 8. *(1) Exclusion: for any $\tilde{x} \in \tilde{\mathcal{X}}, z \in \mathcal{Z}, w \in \mathcal{W}$, and $y \in \{0, 1\}$,*

$$\Pr(Y^* = 1 \mid x, w) = \Pr(Y^* = 1 \mid x) = p^*(x),$$
$$\Pr(Y = 1 - y \mid Y^* = y, x, w) = \Pr(Y = 1 - y \mid Y^* = y, \tilde{x}, w).$$

(2) Degree of misreporting: for any $\tilde{x} \in \tilde{\mathcal{X}}$ and $w \in \mathcal{W}$,

$$\Pr(Y = 0 \mid Y^* = 1, \tilde{x}, w) + \Pr(Y = 0 \mid Y^* = 1, \tilde{x}, w) \leq 1.$$

(3) Monotonicity & Relevance: for any $\tilde{x} \in \tilde{\mathcal{X}}$, $w_1 > w_2 \in \mathcal{W}$, and $y \in \{0, 1\}$,

$$\Pr(Y = 1 - y \mid Y^* = y, \tilde{x}, w_1) \leq \Pr(Y = 1 - y \mid Y^* = y, \tilde{x}, w_2),$$

and there exists $k \in \{0, 1\}$ such that the above inequality is strict.

(4) Relevance: for any $\tilde{x} \in \tilde{\mathcal{X}}$, there exists $z_1 \neq z_2 \in \mathcal{Z}$ such that $p^*(\tilde{x}, z_1) \neq p^*(\tilde{x}, z_2)$.

Assumption 8 summarizes all assumptions for the two instruments (Z, W) in previous sections. Assumption (1) states exclusion restrictions for the two instruments, which requires that instrument Z does not affect misreporting probabilities and instrument W does not affect the true probability. Assumption (2) requires the sum of the two-sided misreporting probabilities to be smaller than one. Assumption (3) adds one relevance restriction for instrument W so that W at least affects the misreporting probability for one group $Y^* = y$ strictly. Assumption (4) is the relevance condition for instrument Z, but the direction of how the true probability is affected by instrument Z is not restricted so that instrument Z can either increase or decrease the true probability. The relevance condition of instrument Z can guarantee that the supremum and infimum of the reported probability over Z are bounded away from one and zero respectively. Therefore, the boundary condition in previous sections is no longer needed in this section.

Under the above assumptions, we can use joint variation in the two instruments to derive bounds for misreporting probabilities and the true probability. The joint variation can bound misreporting probabilities via a new channel and can thus provide more informative results than simply taking intersections over the bounds derived using each instrument separately.

We consider that instrument W is bounded and let w_m denote the maximum value of instrument W. Next, we establish bounds for the misreporting probabilities evaluated at w_m by using the two instruments jointly. Under the monotonicity condition of instrument W in Assumption 8 (iii), the misreporting probability at w_m

is the smallest misreporting probability. The bounds for misreporting probabilities evaluated at other values of W can be established similarly, which will lead to the same identification result for the true probability. Therefore, we focus on the results for the smallest misreporting probabilities.

The next lemma derives bounds on misreporting probabilities evaluated at $W = w_m$.

Lemma 1. *Under Assumption 8, the misreporting probability* $\Pr(Y = 1 - y \mid Y^* = y, \tilde{x}, w_m)$ *can be bounded as:* $[0, U_{\alpha_y}(\tilde{x}, w_m)]$ *for any* $\tilde{x} \in \tilde{\mathcal{X}}$ *and* $y \in \{0, 1\}$*, where*

$$U_{\alpha_1}(\tilde{x}, w_m) = 1 - \sup_{z \in \mathcal{Z}, w < w_m \in \mathcal{W}} \left\{ \frac{q_1(\tilde{x}, w_m, w) p_W(\tilde{x}, z_1, w) - p_W(\tilde{x}, z_1, w_m)}{q_1(\tilde{x}, w_m, w) - 1}, p_W(\tilde{x}, z, w_m) \right\},$$

$$U_{\alpha_0}(\tilde{x}, w_m) = \inf_{z \in \mathcal{Z}, w < w_m \in \mathcal{W}} \left\{ \frac{q_1(\tilde{x}, w_m, w) p_W(\tilde{x}, z_1, w) - p_W(\tilde{x}, z_1, w_m)}{q_1(\tilde{x}, w_m, w) - 1}, p_W(\tilde{x}, z, w_m) \right\},$$

$$q_1(\tilde{x}, w_m, w) = \frac{p_W(\tilde{x}, z_1, w_m) - p_W(\tilde{x}, z_2, w_m)}{p_W(\tilde{x}, z_1, w) - p_W(\tilde{x}, z_2, w)}.$$

Lemma 1 characterizes the lower and upper bound for the misreporting probabilities evaluated at $W = w_m$ by using two instruments jointly. The lower bounds for the smallest misreporting probabilities are zero, since we cannot rule out the possibility of no misreporting.

The upper bounds display the joint identifying power of the two instruments. From the definition of $U_{\alpha_y}(\tilde{x}, w_m)$, we can see that it uses information from all possible values of the two instruments (Z, W). The term $p_W(\tilde{x}, z, w_m)$ shows the identifying power of instrument Z, and the other term involving $q_1(\tilde{x}, w_m, w)$ uses joint information of the two instruments which is more informative compared to only using instrument W. The main idea is that the joint variation in the two instruments can build relationships between misreporting probabilities at different values of W, which can further bound the misreporting probabilities.

Given bounds on misreporting probabilities, the next proposition characterizes the identification result for the true probability $p^*(x)$.

Proposition 3. *Under Assumption 8, the true conditional choice probability* $p^*(x)$ *is* *bounded as* $p^*(x) = [L_3(x), U_3(x)]$ *for any* $x \in \mathcal{X}$*, where*

$$L_3(x) = \frac{p_W(x, w_m) - U_{\alpha_0}(\tilde{x}, w_m)}{1 - U_{\alpha_0}(\tilde{x}, w_m)}, \qquad U_3(x) = \frac{p_W(x, w_m)}{1 - U_{\alpha_1}(\tilde{x}, w_m)}.$$

And the above bounds are sharp when instrument W *is binary.*

Proposition 3 establishes bounds for the true probability by using two instruments jointly, and the bounds have exhausted all information from assumptions and observed data when instrument W is binary. From the definition of the bounds, the lower bound $L_3(x)$ decreases with respect to the bound $U_{\alpha_0}(\tilde{x}, w_m)$ on misreporting probabilities, and the upper bound $U_3(x)$ increases with respect to $U_{\alpha_1}(\tilde{x}, w_m)$. Therefore a smaller bound $U_{\alpha_y}(\tilde{x}, w_m)$ for misreporting probabilities would imply tighter bounds for the true conditional probability. As discussed, the upper bound $U_{\alpha_y}(\tilde{x}, w_m)$ on misreporting probabilities shown in Lemma 1 would be smaller than the one by only using one instrument. Therefore, Proposition 3 derives more informative bounds for $p^*(x)$ by using the two instruments jointly.

7.2 One-Sided Misreporting

This section studies identification of the true probability $p^*(x)$ under one-sided misreporting. One-sided misreporting refers to where only one group with the true outcome $Y^* = y$ misreport but the other group with $Y^* = 1 - y$ always report the truth. This assumption has practical applications, for example, Nguimkeu, Denteh, and Tchernis (2019) study the application of participation in the food stamp program and show that the overreporting probability is small. The assumption about which group has no misreporting depends on the specific application, so we provide results for both cases.

The next proposition provides bounds for the true probability $p^*(x)$ under one-sided misreporting by using one of the two instruments (Z, W) respectively.

Proposition 4. *(1) Under Assumptions 2-4, the sharp bounds for the true probability*

$p^*(x)$ *are characterized as*

$$p^*(x) \in \begin{cases} [p(x), U_1(x)] & \textit{when } \Pr(Y = 1 \mid Y^* = 0, \tilde{x}) = 0, \\ [L_1(x), p(x)] & \textit{when } \Pr(Y = 0 \mid Y^* = 1, \tilde{x}) = 0. \end{cases}$$

(2) Under Assumptions 5-7, the sharp bounds for the true probability $p^*(x)$ *are characterized as*

$$p^*(x) \in \begin{cases} \left[\sup_{w \in \mathcal{W}} p_W(x, w), U_2(x)\right] & \textit{when } \Pr(Y = 1 \mid Y^* = 0, x, w) = 0, \\ \left[L_2(x), \inf_{w \in \mathcal{W}} p_W(x, w)\right] & \textit{when } \Pr(Y = 0 \mid Y^* = 1, x, w) = 0. \end{cases}$$

Proposition 4 provides sharp identification results for $p^*(x)$ under different scenarios of one-sided misreporting by using one of the two instruments respectively. Compared to the results in Proposition 1 and 2 allowing for two-sided misreporting, the assumption of one-sided misreporting can shrink the bounds by either increasing the lower bound or decreasing the upper bound. While one of the bounds for the true probability still keeps the same under one-sided misreporting. It is because even allowing for two-sided misreporting, one of the bounds will be achieved when there is no misreporting for one group which coincides with the one-sided misreporting assumption.

8 Conclusion

This paper provides semiparametric identification of a binary choice model when the binary dependent variable is subject to misreporting. We introduce two different approaches by exploiting different instrumental variables respectively. The first approach uses an instrument that only affects the true outcome but not the misreporting probabilities. The second approach uses an instrument that only influences misreporting probabilities in a monotone way but does not affect the true outcome. Our approaches can allow for unknown distributions of errors and a flexible misreporting process. We characterize conditional moment inequalities based on the identification results. This approach is shown via simulations to perform more robustly than the parametric method concerning distributional assumptions and misreporting process. In the extension, we also explore the joint identifying power of the two instruments together and study identification under one-sided misreporting.

Our approaches do not impose any parametric models for the misreporting and allow the misreporting probabilities to depend on covariates. In future research, it is worthwhile to explore how additional assumptions such as a parametric model for misreporting or homogeneous misreporting probabilities can tighten the identified set derived in this paper. Moreover, our paper focuses on misreporting in the binary dependent variable, while assuming that the covariate is perfectly observed. It still needs substantial future work to investigate whether and how our methods can be applied to the case where both the dependent variable and covariates are subject to misreporting.

A Appendix

We first introduce the notation for conditional misreporting probabilities. Define the conditional misreporting probability $\alpha_y(x)$ for people with $Y^* = y$ as follows: for $y \in \{0, 1\}$,

$$\alpha_y(x) = \Pr(Y = 1 - y \mid Y^* = y, X = x).$$

With a slight abuse of notation, we use the same function name α_y when it is conditional on different covariates to avoid the complexity of introducing more notation. As such, the function α_y is defined conditional on $\tilde{x}$ in Section 3.1, conditional on (x, w) in Section 3.2, and conditional on $(\tilde{x}, w)$ in Section 7.1.

A.1 Proof of Proposition 1

Proof. The covariate X is divided into two parts: $X = (\tilde{X}, Z)$. Under the exclusion restriction of instrument Z (Assumption 2), we know that the conditional misreporting probability only depends on covariate $\tilde{x}$ so it is denoted as $\alpha_y(\tilde{x})$ for $y \in \{0, 1\}$.

The proof of Proposition 1 comprises two steps: the first step is to bound the misreporting probabilities $\alpha_0(\tilde{x}), \alpha_1(\tilde{x})$ using variation in instrument Z. The second step is to bound the true probability $p^*(x)$.

Step 1: bound the misreporting probabilities $\alpha_y(\tilde{x})$. The reported probability $p(x) = \Pr(Y = 1 \mid x)$ comes from two parts: people with $Y^* = 1$ and report the truth, as well as people with $Y^* = 0$ and misreport. Then under Assumption 2, the reported probability $p(x)$ can be expressed as follows:

$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)].$$

By combining the common term $p^*(x)$, the above equation can be written as

$$[1 - \alpha_0(\tilde{x}) - \alpha_1(\tilde{x})]p^*(x) = p(x) - \alpha_0(\tilde{x}).$$

Assumption 3 (degree of misreporting) implies that $1 - \alpha_0(\tilde{x}) - \alpha_1(\tilde{x}) \in [0, 1]$. The fact that the true conditional probability satisfies $p^*(x) \in [0, 1]$ can bound the misreporting probability $\alpha_y(\tilde{x})$ as follows:

$$\begin{aligned} p^*(x) \geq 0 \ \forall z &\Longrightarrow 0 \leq \alpha_0(\tilde{x}) \leq p(x), \\ p^*(x) \leq 1 \ \forall z &\Longrightarrow 0 \leq \alpha_1(\tilde{x}) \leq 1 - p(x). \end{aligned}$$

Since the misreporting probability $\alpha_y(\tilde{x})$ does not depend on z, we can take the smallest upper bound over z:

$$\begin{aligned} &0 \leq \alpha_0(\tilde{x}) \leq \inf_{z \in \mathcal{Z}} p(\tilde{x}, z), \\ &0 \leq \alpha_1(\tilde{x}) \leq \sup_{z \in \mathcal{Z}} \{1 - p(\tilde{x}, z)\} = 1 - \sup_{z \in \mathcal{Z}} p(\tilde{x}, z). \end{aligned}$$

Step 2: bound the true probability $p^*(x)$. Now we revisit the equation for the reported probability $p(x)$:

$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)]. \tag{5}$$

Given bounds on misreporting probabilities $\alpha_0(\tilde{x}), \alpha_1(\tilde{x})$ derived in the first step, equation (5) leads to the following inequalities:

$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)] \geq \sup_{z \in \mathcal{Z}} p(\tilde{x}, z)p^*(x),$$
$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)] \leq p^*(x) + \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)[1 - p^*(x)].$$

Under Assumption 4 (boundary condition), we know that the infimum and supremum of the reported probability $p(x)$ is bounded away from one and zero respectively. Then the true probability $p^*(x)$ can be bounded as follows:

$$\frac{p(x) - \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)}{1 - \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)} \leq p^*(x) \leq \frac{p(x)}{\sup_{z \in \mathcal{Z}} p(\tilde{x}, z)}.$$

Now we need to prove the sharpness of the above bounds. It can be proved by showing that the lower bound and upper bound can be achieved. The idea is to show that given the lower bound and upper bound, we can construct misreporting probabilities and the true probability which match the reported probability $p(x)$ and satisfy Assumptions 2-4.

We first look at the upper bound. The misreporting probability is constructed as $\alpha_1(\tilde{x}) = 1 - \sup_{z \in \mathcal{Z}} p(\tilde{x}, z)$, $\alpha_0(\tilde{x}) = 0$, and the true probability is constructed as $p^*(x) = \frac{p(x)}{\sup_{z \in \mathcal{Z}} p(\tilde{x}, z)}$. It can be verified that this construction matches the reported probability $p(x)$ and satisfies assumptions. Similarly the lower bound can be achieved when $\alpha_1(\tilde{x}) = 0$, $\alpha_0(\tilde{x}) = \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)$, and $p^*(x) = \frac{p(x) - \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)}{1 - \inf_{z \in \mathcal{Z}} p(\tilde{x}, z)}$.

□

A.2 Proof of Proposition 2

Proof. In this part, the misreporting probability depends on the covariate (x, w) so it is denoted as $\alpha_y(x, w)$. The strategy is similar to the proof in Section A.1: we first establish bounds for the misreporting probability $\alpha_y(x, w)$ using instrument W and then derive bounds for the true probability $p^*(x)$.

Under Assumption 5 (exclusion) for instrument W, the reported probability $p_W(x, w)$

can be expressed as

$$p_W(x,w) = [1-\alpha_1(x,w)]p^*(x) + \alpha_0(x,w)[1-p^*(x)].$$

Given the fact that the true probability satisfies $p^*(x) \in [0,1]$ and the sum of two-sided misreporting probabilities is smaller than one (Assumption 7), the misreporting probabilities are bounded as

$$\begin{aligned} p^*(x) \geq 0 &\implies 0 \leq \alpha_0(x,w) \leq p_W(x,w), \\ p^*(x) \leq 1 &\implies 0 \leq \alpha_1(x,w) \leq 1 - p_W(x,w). \end{aligned} \tag{6}$$

However the misreporting probability $\alpha_y(x,w)$ also depends on instrument W so that the above results are not informative for the true probability. Next we use the monotonicity assumption of instrument W to further bound the misreporting probabilities. Under Assumption 6 (monotonicity), the following holds for any $\tilde{w} < w \in \mathcal{W}$:

$$\begin{aligned} \alpha_0(x,w) &\leq \alpha_0(x,\tilde{w}) \leq p_W(x,\tilde{w}), \\ \alpha_1(x,w) &\leq \alpha_1(x,\tilde{w}) \leq 1 - p_W(x,\tilde{w}). \end{aligned} \tag{7}$$

Then the misreporting probability $\alpha_y(w)$ at each w can be further bounded by taking infimum over all upper bounds of misreporting probabilities evaluated at $\tilde{w} < w$:

$$\begin{aligned} 0 &\leq \alpha_0(x,w) \leq \inf_{\tilde{w} \leq w} p_W(x,\tilde{w}), \\ 0 &\leq \alpha_1(x,w) \leq 1 - \sup_{\tilde{w} \leq w} p_W(x,\tilde{w}). \end{aligned}$$

Now we are ready to derive bounds on the true probability by the reported probability $p_W(x,w)$. Given bounds on the misreporting probability $\alpha_y(x,w)$, it has the following implication for each w:

$$\begin{aligned} p_W(x,w) &= [1-\alpha_1(x,w)]p^*(x) + \alpha_0(x,w)[1-p^*(x)] \geq \sup_{\tilde{w} \leq w} p_W(x,\tilde{w})p^*(x), \\ p_W(x,w) &= [1-\alpha_1(x,w)]p^*(x) + \alpha_0(x,w)[1-p^*(x)] \leq p^*(x) + \inf_{\tilde{w} \leq w} p_W(x,\tilde{w})[1-p^*(x)]. \end{aligned}$$

By Assumption 7 (boundary condition), the reported probability is bounded away from zero and one. Then the true probability $p^*(x)$ can be bounded as follows:

$$\frac{p_W(x,w) - \inf\limits_{\tilde{w}\le w} p_W(x,\tilde{w})}{1 - \inf\limits_{\tilde{w}\le w} p_W(x,\tilde{w})} \le p^*(x) \le \frac{p_W(x,w)}{\sup\limits_{\tilde{w}\le w} p_W(x,\tilde{w})}.$$

Since the above bounds hold for any w and the true probability does not depend on w, we can take intersections over all possible values of w:

$$\sup_{w\in\mathcal{W}} \left\{ \frac{p_W(x,w) - \inf\limits_{\tilde{w}\le w\in\mathcal{W}} p_W(x,\tilde{w})}{1 - \inf\limits_{\tilde{w}\le w\in\mathcal{W}} p_W(x,\tilde{w})} \right\} \le p^*(x) \le \inf_{w\in\mathcal{W}} \left\{ \frac{p_W(x,w)}{\sup\limits_{\tilde{w}\le w\in\mathcal{W}} p_W(x,\tilde{w})} \right\}.$$

In the end, we need to show that the above bounds are sharp when instrument $W \in \{w_1, w_2\}$ only takes two values with $w_1 > w_2$. When instrument W only takes two values, the upper bound becomes $p^*(x) = \frac{p_W(x,w_1)}{\sup_{w\in\mathcal{W}} p_W(x,w)}$. It can be achieved when the misreporting probability satisfies $\alpha_1(x,w_1) = 1 - \sup\limits_{w\in\mathcal{W}} p_W(x,w), \alpha_0(x,w_1) = 0$ as well as $\alpha_1(x,w_2) = 1 - p_W(x,w_2), \alpha_0(x,w_2) = p_W(x,w_2)$. It can be verified that this construction satisfies Assumptions 6-7 and matches the reported probability.

The lower bound for the true probability is $p^*(x) = \frac{p_W(x,w_1) - \inf_{w\in\mathcal{W}} p_W(x,w)}{1-\inf_{w\in\mathcal{W}} p_W(x,w)}$. It can be achieved when $\alpha_1(x,w_1) = 0, \alpha_0(x,w_1) = \inf_{w\in\mathcal{W}} p_W(x,w)$ and $\alpha_1(x,w_2) = 1 - \alpha_0(x,w_2), \alpha_0(x,w_2) = p_W(x,w_2)$. This construction satisfies Assumptions 6-7 and matches the reported probability.

□

A.3 Proof of Theorem 1

Proof. Theorem 1 contain two results: one is that $B_{I,t}$ is an identified set for β_0 under corresponding assumptions. This has already been shown in the main text. The second result remaining to be shown is that the identified set $B_{I,t}$ is sharp when the bounds $[L_t(x), U_t(x)]$ for $p^*(x)$ are sharp.

To prove the sharpness of the identified set $B_{I,t}$, we need to show that for any β satisfying condition (4), there exits a data generating process that satisfies our assumptions and matches the reported probability.

The reported probability consists of the true probability and misreporting probabilities, where the true probability is generated from model (1). The proof proceeds in two steps. In the first step, we construct a conditional distribution of $\epsilon \mid x$ satisfying Assumption 1 such that the true probability derived from model (1) is equal to any value between the lower bound $L_t(x)$ and upper bound $U_t(x)$. Then given the true probability, we need to construct misreporting probabilities such that the observed reported probability is generated. The second step has been shown in Propositions 1-2 when we prove the sharpness of the bounds $[L_t(x), U_t(x)]$.

So now we need to prove the first step which is about constructing a conditional distribution $F_{\epsilon|X}(\cdot \mid x)$ of $\epsilon \mid x$ such that the true probability is equal to any value between the lower bound $L_t(x)$ and the upper bound $U_t(x)$. The only restriction for the distribution $F_{\epsilon|X}(\cdot \mid x)$ is Assumption 1, which is equivalent to the following condition: for any x,

$$F_{\epsilon|x}(0 \mid x) = 0.5.$$

For any candidate β in the identified set $B_{I,t}$, we discuss three cases for the sign of the covariate index: $x'\beta > 0$, $x'\beta = 0$, and $x'\beta < 0$. When $x'\beta > 0$, condition (4) implies that $U_t(x) - 0.5 \geq 0$. Then the conditional distribution is constructed as

$$F_{\epsilon|x}(c \mid x) = \begin{cases} 0.5 & \text{when } c = 0, \\ U_t(x) & \text{when } c = x'\beta > 0. \end{cases}$$

The above construction satisfies the weak monotonicity since $U_t(x) \geq 0.5$, so it satisfies the requirement for a distribution. Actually in this construction, $U_t(x)$ can be replaced with any number h satisfying $h \in [0.5, U_t(x)]$. Under the above construction, the true probability is given as:

$$p^*(x) = \Pr(\epsilon_i \leq x'\beta \mid x) = U_t(x) \in [L_t(x), U_t(x)].$$

When $x'\beta = 0$, then we only require $F_{\epsilon|x}(0 \mid x) = 0.5$. Condition (4) implies that $U_t(x) - 0.5 \geq 0$ and $L_t(x) - 0.5 \leq 0$. So the true probability satisfies $p^*(x) = 0.5 \in [L_t(x), U_t(x)]$.

When $x'\beta < 0$, condition (4) implies that $L_t(x) - 0.5 \leq 0$. Then the conditional distribution can be constructed similarly by replacing $U_t(x)$ with $L_t(x)$.

□

A.4 Proof of Lemma 1

Proof. In this part, the misreporting probability $\alpha_y(\tilde{x}, w)$ depends on $(\tilde{x}, w)$. We suppress the covariate $\tilde{X}$ in this proof to simplify notation. Under the exclusion restrictions imposed on the two instruments in Assumption 8, the reported probability $p_W(z, w)$ can be expressed as follows:

$$p_W(z, w) = [1 - \alpha_1(w)]p^*(z) + \alpha_0(w)[1 - p^*(z)]. \tag{8}$$

Under Assumption 8, instrument Z only affects the true probability $p^*(z)$ and instrument W only affects the misreporting probabilities $\alpha_y(w)$. We first derive bounds on misreporting probabilities $\alpha_y(w)$ and then establish bounds for the true probability based on equation (8).

We first look at identification for the misreporting probabilities $\alpha_y(w)$. Since the true conditional probability satisfies $p^*(z) \in [0, 1]$ for any z, we can bound $\alpha_y(w)$ following similar arguments to the proofs in Section A.1 and A.2:

$$\alpha_0(w) \in [0, \inf_{z \in \mathcal{Z}} p_W(z, w)], \qquad \alpha_1(w) \in [0, 1 - \sup_{z \in \mathcal{Z}} p_W(z, w)]. \tag{9}$$

Next, we use joint variation in the two instruments to build relationships between misreporting probabilities evaluated at different values of W. This relationship can further bound the misreporting probabilities. We fix instrument $Z = z$ and look at the reported probability evaluated at two different values $w_1 \neq w_2 \in \mathcal{W}$:

$$\begin{aligned} p_W(z, w_1) &= [1 - \alpha_1(w_1)]p^*(z) + \alpha_0(w_1)[1 - p^*(z)], \\ p_W(z, w_2) &= [1 - \alpha_1(w_2)]p^*(z) + \alpha_0(w_2)[1 - p^*(z)]. \end{aligned}$$

From condition (9), we know that $1-\alpha_1(w)-\alpha_0(w) \geq \sup_{z \in \mathcal{Z}} p_W(z, w) - \inf_{z \in \mathcal{Z}} p_W(z, w) > 0$ by the relevance condition of instrument Z. The above two equations both contain the common term $p^*(z)$, and canceling out the same term $p^*(z)$ has the following implication:

$$p_W(z,w_1) = \frac{1-\alpha_1(w_1)-\alpha_0(w_1)}{1-\alpha_1(w_2)-\alpha_0(w_2)}[p_W(z,w_2)-\alpha_0(w_2)] + \alpha_0(w_1).$$

Let $A_1(w_1,w_2) = \frac{1-\alpha_1(w_1)-\alpha_0(w_1)}{1-\alpha_1(w_2)-\alpha_0(w_2)}$ and $A_0(w_1,w_2) = \alpha_0(w_2)A_1(w_1,w_2) - \alpha_0(w_1)$. Then the above equation can be rewritten as follows for $z \in \{z_1, z_2\}$:

$$\begin{aligned}
p_W(z_1,w_1) &= A_1(w_1,w_2)p_W(z_1,w_2) - A_0(w_1,w_2), \\
p_W(z_2,w_1) &= A_1(w_1,w_2)p_W(z_2,w_2) - A_0(w_1,w_2).
\end{aligned}$$

The two equations can jointly identify $A_1(w_1,w_2)$ and $A_0(w_1,w_2)$ as long as the equations are not collinear. By the relevance condition of instrument Z, we know that $p_W(z_1,w) - p_W(z_2,w) = [1-\alpha_1(w)-\alpha_0(w)][p^*(z_1)-p^*(z_2)] \neq 0$ so the two equations are not collinear. Then $A_1(w_1,w_2)$ and $A_0(w_1,w_2)$ can be identified as follows:

$$\begin{aligned}
A_1(w_1,w_2) &= \frac{p_W(z_1,w_1)-p_W(z_2,w_1)}{p_W(z_1,w_2)-p_W(z_2,w_2)} \equiv q_1(w_1,w_2), \\
A_0(w_1,w_2) &= q_1(w_1,w_2)p_W(z,w_2) - p_W(z,w_1) \equiv q_0(w_1,w_2) \quad \forall z.
\end{aligned}$$

According to the definition of $A_1(w_1,w_2)$ and $A_0(w_1,w_2)$, they build relationships between misreporting probabilities and this relationship can be used to further bound misreporting probabilities. From the definition of $A_1(w_1,w_2)$ and $A_0(w_1,w_2)$, the following holds for $\alpha_y(w)$:

$$\begin{aligned}
1-\alpha_1(w_1) &= [1-\alpha_1(w_2)]q_1(w_1,w_2) - q_0(w_1,w_2), \\
\alpha_0(w_1) &= \alpha_0(w_2)q_1(w_1,w_2) - q_0(w_1,w_2).
\end{aligned} \tag{10}$$

By the monotonicity condition of instrument W in Assumption 8 and bounds on misreporting probabilities in condition (9), the following condition summarizes restrictions on misreporting probabilities:

$$\begin{aligned}
1 \geq 1-\alpha_1(w_1) &\geq 1-\alpha_1(w_2), \qquad 0 \leq \alpha_0(w_1) \leq \alpha_0(w_2), \\
1-\alpha_1(w) &\geq \sup_{z\in\mathcal{Z}} p_W(z,w), \qquad \alpha_0(w) \leq \inf_{z\in\mathcal{Z}} p_W(z,w).
\end{aligned} \tag{11}$$

Given the above restrictions on $\alpha_y(w)$ and equation (10), we can derive bounds on misreporting probabilities $\alpha_y(w_1)$. In order to derive explicit bounds on $\alpha_y(w_1)$, we also need to discuss the value $q_1(w_1, w_2)$ and $q_0(w_1, w_2)$. By the monotonicity condition of instrument W, the following holds:

$$
\begin{aligned}
q_1(w_1, w_2) &= \frac{1 - \alpha_1(w_1) - \alpha_0(w_1)}{1 - \alpha_1(w_2) - \alpha_0(w_2)} > 1, \\
q_0(w_1, w_2) &= \alpha_0(w_2)q_1(w_1, w_2) - \alpha_0(w_1) \geq \alpha_0(w_1)[q_1(w_1, w_2) - 1] \geq 0, \\
q_1(w_1, w_2) - q_0(w_1, w_2) - 1 &= q_1(w_1, w_2) - \alpha_1(w_1) - [1 - \alpha_1(w_2)]q_1(w_1, w_2) \\
&= \alpha_1(w_2)q_1(w_1, w_2) - \alpha_1(w_1) \geq 0.
\end{aligned}
$$

Then conditions (10) together with the restrictions (11) leads to bounds for $\alpha_y(w_1)$ for any w_1, w_2:

$$
\begin{aligned}
&\max\left\{\frac{q_0(w_1, w_2)}{q_1(w_1, w_2) - 1}, \sup_{z\in\mathcal{Z}} p_W(z, w_1)\right\} \leq 1 - \alpha_1(w_1) \leq 1, \\
&\qquad\qquad 0 \leq \alpha_0(w_1) \leq \min\left\{\frac{q_0(w_1, w_2)}{q_1(w_1, w_2) - 1}, \inf_{z\in\mathcal{Z}} p_W(z, w_1)\right\}.
\end{aligned}
$$

The above bounds hold for any $w_2 < w_1$ and any z. Therefore we can derive bounds for $\alpha_y(w_m)$ by taking intersections over bounds derived from all possible values of $z, w < w_m$ which leads to:

$$
\begin{aligned}
0 \leq \alpha_1(w_m) &\leq 1 - \sup_{z\in\mathcal{Z}, w<w_m\in\mathcal{W}} \left\{\frac{q_0(w_m, w)}{q_1(w_m, w) - 1}, p_W(z, w_m)\right\} = U_{\alpha_1}(w_m), \\
0 \leq \alpha_0(w_m) &\leq \inf_{z\in\mathcal{Z}, w<w_m\in\mathcal{W}} \left\{\frac{q_0(w_m, w)}{q_1(w_m, w) - 1}, p_W(z, w_m)\right\} = U_{\alpha_0}(w_m).
\end{aligned}
$$

□

A.5 Proof of Proposition 3

Proof. We look at the equation for the reported probability conditional on (x, w_m):

$$
p_W(x, w_m) = [1 - \alpha_1(\tilde{x}, w_m)]p^*(x) + \alpha_0(\tilde{x}, w_m)[1 - p^*(x)]. \tag{12}
$$

Given bounds on misreporting probabilities derived in Lemma 1, it implies the following conditions:

$$
\begin{aligned}
&p_W(x, w_m) \leq p^*(x) + U_{\alpha_0}(\tilde{x}, w_m)[1 - p^*(x)], \\
&p_W(x, w_m) \geq [1 - U_{\alpha_1}(\tilde{x}, w_m)]p^*(x).
\end{aligned}
$$

First we can show that the upper bounds for $U_{\alpha_1}(\tilde{x}, w_m)$ and $U_{\alpha_0}(\tilde{x}, w_m)$ are strictly smaller than one. It can be shown by proving that $\sup_z p(\tilde{x}, z, w) > 0$ and $\inf_z p_W(\tilde{x}, z, w) < 1$ for any $(\tilde{x}, w)$.

We show it by contradiction. If the supremum of $p_W(\tilde{x}, z, w)$ over z is zero which implies that $p_W(\tilde{x}, z, w) = 0$ for all z. However we know that there exists z_1, z_2 such that $p_W(\tilde{x}, z_1, w) - p_W(\tilde{x}, z_2, w) = [1 - \alpha_1(\tilde{x}, w) - \alpha_0(\tilde{x}, w)][p^*(\tilde{x}, z_1) - p^*(\tilde{x}, z_2)] \neq 0$ by the relevance condition of instrument Z and the degree of misreporting assumptions in Assumption 8. Therefore $p_W(\tilde{x}, z, w)$ cannot be zero for all z and the supremum of it over z is strictly larger than zero. Similarly we can conclude that $\inf_z p_W(\tilde{x}, z, w) < 1$ and the upper bounds $U_{\alpha_y}(\tilde{x}, w_m)$ are strictly smaller than one.

Then the true probability $p^*(x)$ can be bounded as follows:

$$
\frac{p_W(x, w_m) - U_{\alpha_0}(\tilde{x}, w_m)}{1 - U_{\alpha_0}(\tilde{x}, w_m)} \leq p^*(x) \leq \frac{p_W(x, w_m)}{1 - U_{\alpha_1}(\tilde{x}, w_m)}.
$$

Now we need to show that the above bounds are sharp when instrument W only takes two values: $W \in \{w_1, w_2\}$ with $w_1 > w_2$ so that $w_m = w_1$. We prove it by constructing misreporting probabilities and the true probability such that they match with the reported probability and also satisfy Assumption 8.

We first show that the upper bound $p^*(x) = \frac{p_W(x, w_1)}{1 - U_{\alpha_1}(\tilde{x}, w_1)}$ can be achieved. The misreporting probabilities $\alpha_y(\tilde{x}, w)$ are constructed as: $\alpha_1(\tilde{x}, w_1) = U_{\alpha_1}(\tilde{x}, w_1)$, $\alpha_0(\tilde{x}, w_1) = 0$, $\alpha_1(\tilde{x}, w_2) = 1 - \frac{1 - U_{\alpha_1}(\tilde{x}, w_1) + q_0(\tilde{x}, w_1, w_2)}{q_1(\tilde{x}, w_1, w_2)}$, $\alpha_0(\tilde{x}, w_2) = \frac{q_0(\tilde{x}, w_1, w_2)}{q_1(\tilde{x}, w_1, w_2)}$. It can be verified that they match with the reported probability $p_W(x, w)$ for any $x, w \in \{w_1, w_2\}$ under some algebra.

Next we show that misreporting probabilities we construct satisfy Assumption 8 which assumes monotonicity and the degree of misreporting. Given the upper bound $U_{\alpha_1}(\tilde{x}, w_1)$ is strictly smaller than one, then the misreporting probabilities constructed above satisfy the degree of misreporting: $\alpha_0(\tilde{x}, w) + \alpha_1(\tilde{x}, w) < 1$ for any $\tilde{x}, w \in \{w_1, w_2\}$.

The monotonicity condition for $\alpha_0(\tilde{x}, w)$ is satisfied since $\alpha_0(\tilde{x}, w_2) \geq 0 = \alpha_0(\tilde{x}, w_1)$, so we only need to show the monotonicity condition for $\alpha_1(\tilde{x}, w)$. We look at the difference of the two misreporting probabilities multiplied by $q_1(\tilde{x}, w_1, w)$:

$$
\begin{aligned}
& q_1(\tilde{x}, w_1, w_2)[\alpha_1(\tilde{x}, w_1) - \alpha_1(\tilde{x}, w_2)] \\
= & -[1 - U_{\alpha_1}(\tilde{x}, w_1)][q_1(\tilde{x}, w_1, w_2) - 1] + q_0(\tilde{x}, w_1, w_2) \\
\leq & -\frac{q_0(\tilde{x}, w_1, w_2)}{q_1(\tilde{x}, w_1, w_2) - 1}[q_1(\tilde{x}, w_1, w_2) - 1] + q_0(\tilde{x}, w_1, w_2) \leq 0.
\end{aligned}
$$

Therefore the monotonicity condition is also satisfied.

Lastly we need to show that the lower bound $p^*(x) = \frac{p_W(x, w_1) - U_{\alpha_0}(\tilde{x}, w_1)}{1 - U_{\alpha_0}(\tilde{x}, w_1)}$ can be achieved. The misreporting probabilities are constructed as: $\alpha_1(\tilde{x}, w_1) = 0$, $\alpha_0(\tilde{x}, w_1) = U_{\alpha_0}(\tilde{x}, w_1)$, $\alpha_1(\tilde{x}, w_2) = 1 - \frac{1 + q_0(\tilde{x}, w_1, w_2)}{q_1(\tilde{x}, w_1, w_2)}$, and $\alpha_0(\tilde{x}, w_2) = \frac{U_{\alpha_0}(\tilde{x}, w_1) + q_0(\tilde{x}, w_1, w_2)}{q_1(\tilde{x}, w_1, w_2)}$. Similarly it can be shown that they satisfy Assumption 8 and match with the reported probability $p_W(x, w)$ for any (x, w).

□

A.6 Proof of Proposition 4

Proof. We start with the identification result by using instrument Z under Assumptions 2-4 and consider the case $\Pr(Y = 1 \mid Y^* = 0, \tilde{x}) = 0$. The analysis applies to other cases. The upper bound $U_1(x)$ can be derived similarly as Proposition 1. The lower bound can be established as follows:

$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) \leq p^*(x).$$

Now we show that the lower bound is sharp. We can construct misreporting probabilities $\alpha_1(\tilde{x}) = 0$ and true probability $p^*(x) = p(x)$ such that they match with the reported probability $p(x)$. The analysis also applies to the case where $\Pr(Y = 0 \mid Y^* = 1, \tilde{x}) = 0$.

Next consider that we have instrument W satisfying Assumptions 5-7, and assume that $\Pr(Y = 1 \mid Y^* = 0, \tilde{x}) = 0$. The upper bound is the same with the result in Proposition 2. Under the exclusion restriction, the reported probability can be characterized as

$$p_W(x, w) = [1 - \alpha_1(x, w)]p^*(x) \leq p^*(x).$$

Then the true probability can be bounded by all values of W:

$$p^*(x) \geq \sup_{w \in \mathcal{W}} p_W(x, w).$$

The lower bound can be achieved when $1 - \alpha_1(x, w) = \frac{p(x,w)}{\sup_{w \in \mathcal{W}} p_W(x,w)}$ and $p^*(x) = \sup_{w \in \mathcal{W}} p_W(x, w)$, which matches with the reported probability. The analysis is similar for the case $\Pr(Y = 0 \mid Y^* = 1, \tilde{x}) = 0$.

□

References

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

D. Andrews, W. Kim, and X. Shi. Commands for testing conditional moment inequalities and equalities. *Stata Journal*, 17:56–72, 03 2017.

D. W. Andrews and X. Shi. Inference based on conditional moment inequalities. *Econometrica*, 81(2):609–666, 2013.

E. Battistin, M. De Nadai, and B. Sianesi. Misreported schooling, multiple measures and returns to educational qualifications. *Journal of Econometrics*, 181(2):136–150, 2014.

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

C. R. Bollinger and M. H. David. Modeling discrete choice with response error: Food stamp participation. *Journal of the American Statistical Association*, 92(439):827–835, 1997.

C. R. Bollinger and M. H. David. Estimation with response error and nonresponse: food-stamp participation in the sipp. *Journal of Business & Economic Statistics*, 19(2):129–141, 2001.

R. Calvi, A. Lewbel, and D. Tommasi. Late with mismeasured or misspecified treatment: An application to women's empowerment in India. *ECARES (Free University of Brussels) Working Paper*, 27, 2017.

D. Card. Using geographic variation in college proximity to estimate the return to schooling, aspects of labour market behaviour: essays in honour of john vanderkamp. ed. *LN Christofides, EK Grant, and R. Swidinsky*, 1995.

V. Chernozhukov, H. Hong, and E. Tamer. Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica*, 75(5):1243–1284, 2007.

V. Chernozhukov, S. Lee, and A. M. Rosen. Intersection bounds: estimation and inference. *Econometrica*, 81(2):667–737, 2013.

M. Davern, J. A. Klerman, J. Ziegenfuss, V. Lynch, and G. Greenberg. 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(4):219–240, 2009.

F. J. DiTraglia and C. Garcia-Jimeno. Identifying the effect of a mis-classified, binary, endogenous regressor. *Journal of Econometrics*, 209(2):376–390, 2019.

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

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

Y. Hu. Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, 144(1): 27–61, 2008.

T. J. Kane, C. E. Rouse, and D. O. Staiger. Estimating returns to schooling when schooling is misreported, 1999.

K. i. Kim. Set estimation and inference with models characterized by conditional moment inequalities. *Working Paper*, 2008.

B. Kreider and J. V. Pepper. Disability and employment: Reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association*, 102 (478):432–441, 2007.

B. Kreider, J. V. Pepper, C. Gundersen, and D. Jolliffe. 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, 2012.

A. Lewbel. Identification of the binary choice model with misclassification. *Econometric Theory*, 16(4):603–609, 2000.

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

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

C. F. Manski. Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics*, 27(3):313–333, 1985.

B. D. Meyer and N. Mittag. Misclassification in binary choice models. *Journal of Econometrics*, 200(2):295–311, 2017.

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

B. D. Meyer, N. Mittag, and R. M. George. Errors in survey reporting and imputation and their effects on estimates of food stamp program participation. *Journal of Human Resources*, pages 0818–9704R2, 2020.

F. Molinari. Partial identification of probability distributions with misclassified data. *Journal of Econometrics*, 144(1):81–117, 2008.

P. Nguimkeu, A. Denteh, and R. Tchernis. On the estimation of treatment effects with endogenous misreporting. *Journal of Econometrics*, 208(2):487–506, 2019.

T. Ura. Heterogeneous treatment effects with mismeasured endogenous treatment. *Quantitative Economics*, 9(3):1335–1370, 2018.