Partial Identification of Binary Choice Models with Misreported Outcomes*

Orville Mondal† and Rui Wang‡

June 9, 2024

Abstract

This paper studies partial identification of binary choice models with misreported dependent variables. We propose two distinct approaches by exploiting different discrete 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 influences misreporting probabilities monotonically while having no effect on the true dependent variable. Moreover, we derive identification results under additional restrictions on misreporting, including one-sided misreporting and bounded/monotone misreporting probabilities. We use simulations to demonstrate the robust performance of our approaches and apply the method to study educational attainment.

Keywords: Misreporting, Binary Choice Models, Instrumental Variables, Partial Identification, Exclusion Restriction

JEL classification: C01, C14, C26.

†Economist, Bates White Economic Consulting. Email: orville.dm@gmail.com

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

1 Introduction

This paper provides partial identification of binary choice models when the 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 several 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.

Numerous studies have examined the bias introduced by misreporting across various 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 significant 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. Firstly, misreporting in a binary variable involves non-classical measurement errors, as the measurement error is always negatively correlated with the true outcome. Moreover, misreporting stem from unobserved incentives among respondents; 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 introduces two different approaches to partially identify binary choice models with a potentially misreported dependent variable, under a median independence condition. With potential misreporting in the dependent variable, conventional approaches for binary choice models do not apply, as the true dependent variable is not observed, and the conditional expectation of the true dependent variable is not identified. Our identification strategy derives bounds for the conditional expectation of the true outcome given covariates, by exploiting variation in different discrete instruments. Given the bounds, we derive identifying restrictions for the model co-

efficient based on the relationship between the model coefficient and the conditional expectation of the true dependent variable.

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, such as job training program, this instrument can be randomly assigned eligibility for the program. This instrument affects the true participation of the program, but it is unlikely to affect misreporting, given its random nature. The second approach uses an instrument that only affects misreporting probabilities monotonically, but does not influence the true outcome. Examples of such a variable could include interview-relevant variables such as an interviewer's evaluation score of respondents' accuracy or interview styles in survey data, including in-person, phone, or email interviews. Individuals are more likely to provide truthful responses during in-person interviews than during email interviews.

Our approaches can allow for a wide range of misreporting processes. Firstly, we allow for heterogeneous misreporting probabilities, which can 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. Furthermore, we do not assume a parametric model for the misreporting process (such as a linear index model) and permit arbitrary dependence between the true outcome and the misreporting process.

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, and demonstrate its robust performance concerning 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 explore the identifying power of each instrument under additional restrictions on misreporting, including one-sided misreporting, bounded misreporting probabilities, and monotone misreporting probabilities. These restrictions can potentially provide more informative bounds for misreporting probabilities and thus tighten the bounds for the conditional expectation of the true outcome. Furthermore, we examine the joint identifying power of the two instrumental variables together. The two instruments jointly provide a new channel for identification, thus leading to more informative results compared to intersecting the bounds obtained by using each instrument separately.

1.1 Related Literature

This paper directly contributes to the line of literature on binary choice models with misreported dependent variables. See Chen, Hong, and Nekipelov (2011) for a comprehensive survey of nonlinear models with measurement errors, including binary choice models. The literature on binary choice models dates back to Chamberlain (1980), Manski (1985), and Manski (1987). When the dependent variable is misreported, Hausman, Abrevaya, and Scott-Morton (1998) proposes a modified maximum likelihood estimator for binary choice models to correct potential misreporting, and Abrevaya and Hausman (1999) explores a more general linear index model. These studies assume homogeneous misreporting probabilities, where misreporting rates are constant regardless of values of covariates. Bollinger and David (1997) allows misreporting rates to depend on covariates in a Probit model, imposing parametric assumptions for both the true binary choice model and misreporting processes. Lewbel (2000) studies semiparametric identification of binary choice models 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) proposes different parametric estimators, relying on a parametric model for misreporting processes or the availability of validation data. Our paper allows for heterogeneous misreporting probabilities, which can depend on covariates in an arbitrary way. Furthermore, we explore the identifying power of discrete instruments (along with additional restrictions on misreporting) and the results are valid even if the instrument only take two values.

Our work also relates to a large body of literature on various models with misreported regressors. Mahajan (2003) investigates misclassified regressors in binary choice models. Several studies explore regression models with misreported regressors under homogeneous misreporting probabilities, including Aigner (1973), Bollinger (1996), Frazis and Loewenstein (2003). Mahajan (2006), Lewbel (2007), DiTraglia and Garcia-Jimeno (2019) allow misreporting probabilities to be covariates dependent and achieve identification by using a binary instrument for the true regressor. Chen, Hong, and Tamer (2005), Chen and Hu (2006), and Chen, Hu, and Lewbel (2008) provide nonparametric identification without instrumental variables. They achieve identification by exploiting auxiliary data, two samples, and higher-order moments, respectively. Additionally, Hu (2008) and Molinari (2008) explore misclassification in a general discrete regressor, and Hu and Schennach (2008) and Hu, Schennach, and Shiu (2022) extend the study to a continuous regressor with nonclassical measurement errors. These papers investigate different frameworks with misreported regressors,

and thus their assumptions and identification approaches/results differ from those in our paper. In a more closely related study, Nguimkeu, Denteh, and Tchernis (2019) uses two instruments jointly for identification–one for the true regressor and the other for misreporting. They obtain point identification under one-sided misreporting and parametric structures on misreporting, while our paper allows for two-sided misreporting and nonparametric structures on misreporting processes.

The above literature focuses on homogeneous effects of the regressor given covariates in a regression setting. Numerous papers study 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, auxiliary administrative data to restrict misreporting errors, or an instrumental variable related to the true treatment. This literature also studies different framework and requires different assumptions to identify heterogeneous treatment effects from our paper. Additionally, these works exploit the instrument for treatment, while we also explore the identifying power of the instrument for misreporting. More differently, Horowitz and Manski (1995), Kreider and Pepper (2008), and Kreider and Pepper (2011) study identification with corrupted data under a mixing framework on data errors.

The rest of this paper is organized as follows. Section 2 presents the binary choice model with potential misreporting. Section 3 describes the identification strategy using different instruments. Section 4 characterizes conditional moment inequalities for the model coefficient and Section 5 examines the finite sample performance via simulations. Section 6 studies the application of educational attainment. Section 7 explores various extensions and Section 8 concludes.

2   Model

The analysis studies partial identification of binary choice models with potential misreporting (or misclassification) in the binary dependent variable. Let $Y_i^* \in \{0,1\}$ denote the true binary dependent variable, and $Y_i \in \{0,1\}$ denote the observed variable which may be subject to misreporting. Let $X_i \in \mathcal{X}$ denotes a vector of observed covariate, which is relevant to the true variable Y_i^* and can also affect misreporting probabilities. We study the following binary choice model:

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

The objective is to identify 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 median independence assumption on the unobserved term ϵ_i.

Assumption 1 (Median Independence). $\text{Med}(\epsilon_i \mid X_i) = 0$.

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:

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

where $\text{sgn}\{x\} = \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$ denotes the sign function. However, the identification result in condition (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 based on condition (2).

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 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:

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

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 present our identification approaches to establish 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. Therefore, it is essential to distinguish between these two components to identify the true probability from the reported probability.

We provide two different approaches to identify the true probability $p^*(x)$ by exploiting different exclusion restrictions. In the first approach, we use a discrete instrument $Z_i \in \mathcal{Z} := \{z_1, z_2, ..., z_k\}$ that only affects the true probability $p^*(x)$

but does not affect misreporting probabilities. The other approach uses a discrete instrument $W_i \in \mathcal{W} := \{w_1, w_2, ..., w_l\}$ that only affects misreporting probabilities but not the true probability $p^*(x)$. We first study the identifying power of each individual instrument and then discuss how the two instruments 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.
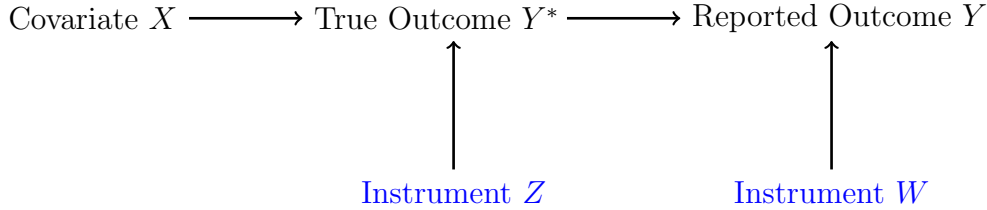
Figure 1: Relationship among Variables

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 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.

3.1 Instrument Z

This section studies the identifying power of instrument $Z \in \mathcal{Z} := \{z_1, z_2, ..., z_k\}$ that only affects the true choice probability but does not influence misreporting probabilities. Instrument Z affects the true variable Y^* 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 process, so it only affects the reported probability by shifting the true probability. When the true outcome is participation in social programs, such as a job training program, one example of instrument Z could be randomly assigned eligibility for the

8

program. The random assignment will affect true participation but not influence misreporting, given its random nature. Studies, e.g., Mahajan (2006), Ura (2018), and DiTraglia and Garcia-Jimeno (2019), provide more examples of 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, ensuring 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 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. Moreover, our identification analysis only requires the degree of misreporting to be known, and the analysis is applicable when the sum of two-sided misreporting rates is larger than one.

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

Assumption 4 is a boundary condition for the reported probability. This assumption is relatively weak, requiring that the supremum of the reported probability is bounded away from zero and the infimum of the reported probability is bounded away from one. It can be satisfied when 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*

$$L_1(x) = \frac{p(x) - \underline{p}_z(\tilde{x})}{1 - \underline{p}_z(\tilde{x})}, \qquad U_1(x) = \frac{p(x)}{\bar{p}_z(\tilde{x})}.$$

where $\underline{p}_z(\tilde{x}) := \min\limits_{z \in \mathcal{Z}} p(\tilde{x}, z)$ and $\bar{p}_z(\tilde{x}) := \max\limits_{z \in \mathcal{Z}} p(\tilde{x}, z)$.

Proposition 1 characterizes partial identification for the true probability $p^*(x)$ using variation in instrument Z. According to the definition of the bounds, we know

that $L_1(x), U_1(x) \in [0,1]$. The identification strategy is to exploit the exclusion restriction of instrument Z. Through variation in the reported probability $p(x)$ under different values of z, we can bound misreporting probabilities and thus derive bounds for the true probability $p^*(x)$.

Proposition 1 establishes sharpness of the results, showing that the bounds are the best possible given the assumptions and data. The identifying power of instrument Z depends on the range of the reported probability $p(\tilde{x}, z)$ as one varies z: tighter bounds for the true probability are achieved with a larger range of the reported probability. A larger variation in the reported probability can yield smaller bounds for misreporting probabilities, leading to tighter bounds on the true probability.

3.2   Instrument W

This section introduces an alternative method for identifying the true probability using instrument $W \in \mathcal{W} := \{w_1, w_2, ..., w_l\}$. 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 an additional variable related to misreporting processes but excluded from the true binary choice model. Our approach does not impose any parametric models for misreporting processes, allowing instrument W 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 presents assumptions on instrument 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 dependent variable Y^*. Examples for instrument W could include interview-related variables, such as interviewers' assessments of respondents' accuracy or different interview styles like phone interviews and in-person interviews. These variables are unlikely to affect the true dependent variable but are related to respondents' probabilities of reporting the truth.

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. For instance, when the instrument is interviewers' evaluations of respondents' accuracy, it is natural that misreporting probabilities are smaller with higher evaluations. In the case of interview styles, individuals are more likely to report the truth during an in-person interview than a phone interview. It is also worth noting that Assumption 6 is a weak monotonicity condition, as it only requires monotonicity for the average probability but allows for potential violations in certain 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*

$$L_2(x) = \max_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w) - \underline{p}_w(x, w)}{1 - \underline{p}_w(x, w)} \right\}, \quad U_2(x) = \min_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w)}{\bar{p}_w(x, w)} \right\},$$

where $\underline{p}_w(x, w) := \min_{\tilde{w} \leq w} p_W(x, \tilde{w})$ *and* $\bar{p}_w(x, w) := \max_{\tilde{w} \leq w} p_W(x, \tilde{w})$. *Moreover, the above bounds are sharp when instrument* W *is binary.*

Proposition 2 derives partial identification by using a distinct instrument that only monotonically influences misreporting probabilities but not the true probability. Furthermore, it demonstrates that this result exhausts all possible information when the instrument W is binary. The bounds, as per their definitions, satisfy that $L_2(x), U_2(x) \in [0, 1]$. Additionally, the results in Proposition 2 imply that $U_2(x) \geq L_2(x)$ for all x, providing testable implications for our assumptions.

11

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 intersecting all bounds derived from any value of instrument W, as demonstrated 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, we have $\beta_0 \in B_{I,k}$, where the identified set $B_{I,k}$ consists of all $\beta \in \mathcal{B}$ that satisfies the following condition for any $x \in \mathcal{X}$:*

$$
\begin{aligned}
x'\beta \geq 0 &\implies U_k(x) - 0.5 \geq 0; \\
x'\beta \leq 0 &\implies L_k(x) - 0.5 \leq 0,
\end{aligned}
\tag{4}
$$

where $k = 1$ under Assumptions 2-4 and $k = 2$ under Assumptions 5-7.

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 $B_{I,k}$ depends on the informativeness of the bounds, and tighter bounds lead to a smaller identified set for β_0. The 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 and allow for heterogeneous misreporting probabilities that can depend on covariates arbitrarily.

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 $(L_k(x), U_k(x))$ 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 $k = 1$, and a similar analysis applies to the case $k = 2$. When $k = 1$, we define $(l_1(x), u_1(x))$ as follows:

$$l_1(x) = (L_1(x) - 0.5)(1 - \underline{p}_z(\tilde{x})) = p(x) - 0.5\underline{p}_z(\tilde{x}) - 0.5,$$
$$u_1(x) = (U_1(x) - 0.5)\bar{p}_z(\tilde{x}) = p(x) - 0.5\bar{p}_z(\tilde{x}).$$

We look at the identifying restrictions with nonnegative covariate index $x'\beta_0 \geq 0$ 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:

$$E[m_1^1(X, Y, \beta_0) \mid X = x] \equiv E[X'\beta_0 \mathbb{1}\{X'\beta_0 \geq 0\}(Y - 0.5\bar{p}_z(\tilde{x})) \mid X = x]$$
$$= x'\beta_0 \mathbb{1}\{x'\beta_0 \geq 0\}u_1(x) \geq 0.$$

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:

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

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:

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

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

$$l_2(x) = \max_{w \in \mathcal{W}} \left\{ p_W(x, w) - 0.5\underline{p}_w(x, w) - 0.5 \right\},$$
$$u_2(x) = \min_{w \in \mathcal{W}} \left\{ p_W(x, w) - 0.5\bar{p}_w(x, w) \right\}.$$

The identifying restriction with nonnegative covariate index in condition (4) has

the following implication: for any x,

$$x'\beta_0 \geq 0 \implies u_2(x) \geq 0 \iff p_W(x, w) - 0.5\bar{p}_w(x, w) \geq 0 \quad \forall w.$$

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

$$E[m_2^1(X, W, Y, \beta_0) \mid X = x, W = w]$$
$$= E[X'\beta_0 \mathbb{1}\{X_i'\beta_0 \geq 0\}(Y - 0.5\bar{p}_w(x, w)) \mid X = x, W = w] \geq 0 \quad \forall x, w.$$

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

$$g_2(x, w, \beta_0) = \begin{cases} E[X'\beta_0 \mathbb{1}\{X'\beta_0 \geq 0\}(Y - 0.5\bar{p}_w(x, w)) \mid X = x, W = w], \\ E[X'\beta_0 \mathbb{1}\{X'\beta_0 \leq 0\}(Y - 0.5\underline{p}_w(x, w) - 0.5) \mid X = x, W = w]. \end{cases}$$

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

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

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

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

Theorem 2 characterizes the identified set $B_{I,k}$ using conditional moment inequalities. The literature has developed various methods to conduct inference for conditional moment inequalities such as Chernozhukov, Hong, and Tamer (2007), 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 $B_{I,k}$ for $k \in \{1, 2\}$ via Monte Carlo simulations. To better evaluate our approach, we also implement the maximum likelihood estimation approach proposed in Hausman, Abrevaya, and Scott-Morton (1998) for comparison, referred to as the HAS approach. Their method accounts for potential misreporting, but assumes a distributional as-

sumption on error term and homogeneous (constant) misreporting probabilities. The simulation results demonstrate the robustness of our approach concerning heterogeneous misreporting probabilities and parametric assumptions.

We adopt the approach in Chernozhukov, Hong, and Tamer (2007) and Andrews and Shi (2013) to estimate the identified set $B_{I,k}$. 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. The number of hypercubes is $\{30, 40, 50\}$ for the sample size $n \in \{1000, 2000, 4000\}$. 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

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

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 $\hat{\beta}_{semi}^l$ and $\hat{\beta}_{semi}^u$ denote the estimated lower and upper bound of the model coefficient in our method.

5.1 Instrument Z

This section studies the identification results using instrument Z. The DGP 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 $[-1, 1]$, 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 $Cauchy(0, 0.5)$.

The reported outcome Y is given by $Y = M_1 \cdot Y^* + (1 - M_0) \cdot (1 - Y^*)$, where $M_y \in \{0, 1\}$ denotes the reporting variable and $M_y = 0$ represents misreporting for any $y \in \{0, 1\}$. We allow for heterogeneous misreporting probabilities, depending on the value of covariate $\tilde{X}$:

$$\Pr(M_1 = 0 \mid \tilde{x}) = 0.1 - 0.1\tilde{x}, \quad \Pr(M_0 = 0 \mid \tilde{x}) = 0.3 + 0.1\tilde{x}.$$

In this DGP, it is clear that Assumptions 2 and 3 on the exclusion and degrees of misreporting are satisfied. The sample size is $n \in \{500, 1000, 2000\}$ and repetition number is $B = 300$.

To compare the performance of the two 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, along with the parametric estimator $\hat{\beta}_{HAS}$ which assumes constant misreporting probabilities and a normal distribution of the error term. Let β^k denote the kth element of the parameter β. We normalize the first element of the parameter to one for our method: $\beta^1_0 = 1$.

Tables 1 and 2 display the performance of the two methods under different specifications of the error term ϵ and different sample sizes. The results illustrate that our approach uniformly performs well across various error term specifications. When the distribution of ϵ is correctly specified, the HAS method shows reasonable performance but does not necessarily outperform our method, as the constant misreporting probabilities assumption still remains misspecified. Moreover, the HAS estimator exhibits significant bias when the distributional assumption is also misspecified (under the Cauchy design), while our approach has robust performance under different designs.

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

Design	$\hat{\beta}^l_{semi}$		$\hat{\beta}^u_{semi}$		$\hat{\beta}_{HAS}$	
	rMSE	MAD	rMSE	MAD	rMSE	MAD
			$n = 500$			
Normal	0.482	0.338	0.468	0.338	0.475	0.265
Cauchy	0.518	0.338	0.500	0.338	1.115	0.407
			$n = 1000$			
Normal	0.434	0.338	0.408	0.268	0.302	0.245
Cauchy	0.385	0.268	0.356	0.237	0.703	0.349
			$n = 2000$			
Normal	0.363	0.338	0.331	0.237	0.242	0.180
Cauchy	0.376	0.268	0.323	0.237	0.438	0.274

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

Design	$\hat{\beta}^l_{semi}$		$\hat{\beta}^u_{semi}$		$\hat{\beta}_{HAS}$	
	rMSE	MAD	rMSE	MAD	rMSE	MAD
			$n = 500$			
Normal	0.519	0.369	0.527	0.439	0.587	0.563
Cauchy	0.544	0.439	0.566	0.439	0.899	0.573
			$n = 1000$			
Normal	0.427	0.338	0.449	0.338	0.602	0.617
Cauchy	0.452	0.338	0.485	0.439	0.640	0.472
			$n = 2000$			
Normal	0.376	0.237	0.391	0.338	0.596	0.583
Cauchy	0.351	0.237	0.393	0.338	0.479	0.416

5.2 Instrument W

This section examines the finite sample performance of the identified set using instrument W. The DGP 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 reported outcome Y is given by $Y = M_1 \cdot Y^* + (1 - M_0) \cdot (1 - Y^*)$, where $M_y \in \{0, 1\}$ and $M_y = 0$ denotes misreporting. The misreporting probability $\Pr(M_y = 0 \mid \tilde{x}, w)$ depends on covariate $\tilde{X}$ and instrument W as follows:

$$\Pr(M_1 = 0 \mid \tilde{x}, w) = 0.1 - 0.1\tilde{x}, \quad \Pr(M_0 = 0 \mid \tilde{x}, w) = \frac{1}{1 + 0.3w^2}.$$

In this specification, it can be verified that the monotonicity condition in Assumption 6 on instrument W and the degree of misreporting probabilities in Assumption 7 are satisfied. The sample size is $n \in \{500, 1000, 2000\}$ and repetition number is $B = 300$.

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

Design	$\hat{\beta}^l_{semi}$		$\hat{\beta}^u_{semi}$		$\hat{\beta}_{HAS}$	
	rMSE	MAD	rMSE	MAD	rMSE	MAD
			$n = 500$			
Normal	0.567	0.338	0.561	0.338	2.342	1.735
Cauchy	0.598	0.439	0.587	0.338	2.966	3.500
			$n = 1000$			
Normal	0.423	0.237	0.412	0.237	2.244	1.456
Cauchy	0.391	0.237	0.372	0.237	2.739	3.081
			$n = 2000$			
Normal	0.363	0.237	0.351	0.167	1.935	1.128
Cauchy	0.341	0.237	0.334	0.237	2.736	3.098

Similarly, Table 3 presents the performance of β_0^2 under different sample sizes and

18

specifications of ϵ. The results show that our approach consistently outperforms the HAS approach across all specifications. The HAS method exhibits a significant bias even when the distribution of ϵ is correctly specified, as the range of misreporting probabilities (under different values of $(\tilde{x}, w)$) is very large in this DGP and the constant misreporting probabilities assumption is seriously misspecified. The bias of HAS method becomes larger when the distributional assumption is also misspecified under the Cauchy design. In summary, the simulation results demonstrate the robust performance of our two approaches concerning flexible misreporting processes and distributional assumptions.

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.

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
s.d.	0.443	2.786	0.421	0.467

We adopt the semiparametric binary choice model with instrument Z to study how individuals' observed characteristics affect the likelihood of attending a college. In addition to accommodating flexible misreporting processes, our method is robust to distributional assumptions. 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 our semiparametric binary choice model, the coefficient β_0 can be only identified up to a constant. Therefore, the coefficient of instrument Z is normalized to one since people tend to be more likely to attend a college when they live closer to a college.

Table 5: Application: 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
HAS	-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 HAS 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 misreporting processes or distributional assumptions may lead to opposite signs of the coefficients.

7   Extensions

This section explores additional restrictions on misreporting probabilities including one-sided misreporting, bounded misreporting probabilities, and monotone misreporting probabilities. Moreover, Section 7.2 studies the joint identifying power of exploiting two instruments together.

7.1   Additional Restrictions on Misreporting

The identification analysis in Section 2 can be also combined with other additional information on misreporting probabilities to further tighten the bounds for the true probabilities $p^*(x)$. Section 7.1.1-7.1.3 studies one-sided misreporting, bounded misreporting probabilities, as well as monotone misreporting probabilities.

7.1.1 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, while the other group with $Y^* = 1 - y$ always report the truth. This assumption has practical applications and has been used in previous studies, such as Nguimkeu, Denteh, and Tchernis (2019). In their paper, they investigate the participation in the food stamp program and provide evidence for a small overreporting probability. The assumption regarding which group has no misreporting depends on the specific application, and we present 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 3. *(1) Under Assumptions 2-4, the sharp bounds for the true probability $p^*(x)$ are characterized as*

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

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

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

Proposition 3 provides sharp identification results for $p^*(x)$ under different scenarios of one-sided misreporting, using one of the two instruments, respectively. The one-sided misreporting assumption ensures accurate data from one group and establishes the misreporting probability to be zero for this group. In comparison to the results in Proposition 1 and 2, which allow for two-sided misreporting, Proposition 3 consistently provides tighter bounds on $p^*(x)$ with either larger lower bound or smaller upper bound.[1]

7.1.2 Bounded Misreporting Probabilities

This section considers that the misreporting probabilities are bounded by known numbers. This restriction may come from information such as previous studies on

---

[1]When instrument W is available, the one-sided misreporting assumption implies the monotonicity of $p_W(x, w)$ in w, and the upper bound $U_2(x)$ becomes one. This implication can serve as a testable implication for the one-sided misreporting assumption.

misreporting, auxiliary administrative data, or theoretical models that provide insights into the potential range of misreporting probabilities.

Assumption 8. *The misreporting probabilities satisfy the following condition: for* $y \in \{1, 2\}$ *and any* (x, w),

$$\Pr(Y = 1 - y \mid Y^* = y, x, w) \leq \bar{\alpha}_y,$$

where $\bar{\alpha}_y \in [0, 1]$ *is a known constant.*

Assumption 8 may stem from empirical results in other research or auxiliary information, such as administrative data from other samples. For simplicity, Assumption 8 adopts uniform bounds for misreporting probabilities regardless of values of (x, w). Our approach can be also adjusted to allow $\bar{\alpha}_y$ to depend on (x, w). When instrument W is not available, then the above conditional misreporting probability can be adjusted by only conditioning on the covariate x.

Proposition 4. *(1) Under Assumptions 2-4 & 8, the true probability* $p^*(x)$ *can be bounded as*

$$p^*(x) \in \left[\frac{p(x) - \min\left\{ \underline{p}(\tilde{x}), \bar{\alpha}_0 \right\}}{1 - \min\left\{ \underline{p}(\tilde{x}), \bar{\alpha}_0 \right\}}, \frac{p(x)}{\max\left\{ \bar{p}(\tilde{x}), 1 - \bar{\alpha}_1 \right\}} \right];$$

(2) Under Assumptions 5-7 & 8, the true probability $p^*(x)$ *can be bounded as*

$$p^*(x) \in \left[\max_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w) - \min\left\{ \underline{p}_w(x, w), \bar{\alpha}_0 \right\}}{1 - \min\left\{ \underline{p}_w(x, w), \bar{\alpha}_0 \right\}} \right\}, \min_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w)}{\max\left\{ \bar{p}_w(x, w), 1 - \bar{\alpha}_1 \right\}} \right\} \right].$$

Proposition 4 derives partial identification for $p^*(x)$ through additional bounds on misreporting probabilities, with the identifying power depending on the value of the bounds $\bar{\alpha}_y$. A smaller value of $\bar{\alpha}_y$ results in tighter bounds for $p^*(x)$, and $\bar{\alpha}_y = 0$ corresponds to no misreporting. It is also possible that this additional restriction may not provide any information if the value of the bound $\bar{\alpha}_y$ is larger than that derived by using each instrument.

7.1.3 Monotone Misreporting Probabilities

This section explores monotone misreporting probabilities, where the misreporting probability of one group is smaller than that of the other group. The direction of monotonicity depends on specific applications. To illustrate, we study the following monotonicity on misreporting probabilities.

Assumption 9. *The misreporting probabilities satisfy the following condition: for any (x, w),*

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

Assumption 9 says that the misreporting probability for people with $Y^* = 0$ is smaller than those with $Y^* = 1$. This assumption is applicable for scenarios where Y^* represents participation in social assistance programs or whether one is smoking or not. It is documented in the literature that people who did not participate in social program are more likely to report the truth than people who participated. The direction of the monotonicity can be reversed in some applications, such as Y^* represents educational attainment. Our identification methods can accommodate both cases, where the direction of the monotonicity is required to be known.

Proposition 5. *(1) Under Assumptions 2-4 & 9, the true probability $p^*(x)$ can be bounded as*

$$p^*(x) \in \left[\frac{p(x) - \min\left\{\underline{p}_z(\tilde{x}), 1 - \bar{p}_z(\tilde{x})\right\}}{1 - \min\left\{\underline{p}_z(\tilde{x}), 1 - \bar{p}_z(\tilde{x})\right\}}, \frac{p(x)}{\bar{p}(\tilde{x})} \right];$$

(2) Under Assumptions 5-7 & 9, the true probability $p^*(x)$ can be bounded as

$$p^*(x) \in \left[\max_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w) - \min\left\{\underline{p}_w(x, w), 1 - \bar{p}_w(x, w)\right\}}{1 - \min\left\{\underline{p}_w(x, w), 1 - \bar{p}_w(x, w)\right\}} \right\}, \min_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w)}{\bar{p}_w(x, w)} \right\} \right].$$

Proposition 5 shows that Assumption 9 can further increase the lower bound for $p^*(x)$ with each instrument. The idea is that under the monotone misreporting rates assumption, the misreporting probability for group $Y^* = 0$ can also be bounded by the upper bound of the misreporting probability for the other group $Y^* = 1$, yielding more informative results for the true probability $p^*(x)$. Symmetrically, when assuming a smaller misreporting rate for group $Y^* = 1$ compared to $Y^* = 0$, the upper bound for $p^*(x)$ can be tightened.

7.2 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 10. *(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 10 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 through a new channel, thereby providing more informative results than simply taking intersections over the bounds derived using each instrument separately.

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 10 (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,

24

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 10, 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*

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

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

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

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 demonstrate the joint identifying power of the two instruments. From the definition of $U_{\alpha_y}(\tilde{x}, w_m)$, it uses variation from both 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, providing more information compared to using only instrument W. The main idea is that the joint variation in the two instruments imposes additional restrictions 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 6. *Under Assumption 10, 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 6 establishes bounds for the true probability by using two instruments jointly, and these bounds have exhausted all possible 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 6 derives more informative bounds for $p^*(x)$ by using the two instruments jointly.

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 discrete 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. Moreover, we explore additional restrictions on misreporting probabilities and the joint identifying power of the two instruments.

Our approach allows for flexible misreporting processes in the sense that we do not impose any parametric model for misreporting processes and allow for heterogeneous misreporting probabilities. It would be interesting to explore how additional parametric structures on misreporting can tighten the bounds for the conditional expectation of the true dependent variable. Furthermore, we focus on binary choice models in this paper, while the approach may be applied more broadly. It still requires substantial future work to investigate how the method can be applied in other models with potential misreporting, such as ordered and multinomial choice models with misreported dependent variables.

References

J. Abrevaya and J. A. Hausman. Semiparametric estimation with mismeasured dependent variables: an application to duration models for unemployment spells. *Annales d'Economie et de Statistique*, pages 243–275, 1999.

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.

G. Chamberlain. Analysis of covariance with qualitative data. *The review of economic studies*, 47(1):225–238, 1980.

X. Chen and Y. Hu. Identification and inference of nonlinear models using two samples with aribrary measurement errors. 2006.

X. Chen, H. Hong, and E. Tamer. Measurement error models with auxiliary data. *The Review of Economic Studies*, 72(2):343–366, 2005.

X. Chen, Y. Hu, and A. Lewbel. Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments. *Economics Letters*, 100(3):381–384, 2008.

X. Chen, H. Hong, and D. Nekipelov. Nonlinear models of measurement errors. *Journal of Economic Literature*, 49(4):901–937, 2011.

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.

J. L. Horowitz and C. F. Manski. Identification and robustness with contaminated and corrupted data. *Econometrica: Journal of the Econometric Society*, pages 281–302, 1995.

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.

Y. Hu and S. M. Schennach. Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216, 2008.

Y. Hu, S. Schennach, and J.-L. Shiu. Identification of nonparametric monotonic regression models with continuous nonclassical measurement errors. *Journal of Econometrics*, 226(2):269–294, 2022.

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

B. Kreider and J. Pepper. Inferring disability status from corrupt data. *Journal of Applied Econometrics*, 23(3):329–349, 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 and J. V. Pepper. Identification of expected outcomes in a data error mixing model with multiplicative mean independence. *Journal of Business & Economic Statistics*, 29(1):49–60, 2011.

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. Misclassified regressors in binary choice models. *Unpublished manuscript, Stanford University*, 2003.

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.

C. F. Manski. Semiparametric analysis of random effects linear models from binary panel data. *Econometrica: Journal of the Econometric Society*, pages 357–362, 1987.

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.

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 7.2.

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:

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

30

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:

$$p^*(x) \geq 0 \; \forall z \implies 0 \leq \alpha_0(\tilde{x}) \leq p(x),$$
$$p^*(x) \leq 1 \; \forall z \implies 0 \leq \alpha_1(\tilde{x}) \leq 1 - p(x).$$

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

$$0 \leq \alpha_0(\tilde{x}) \leq \min_{z \in \mathcal{Z}} p(\tilde{x}, z) = \underline{p}_z(\tilde{x}),$$
$$0 \leq \alpha_1(\tilde{x}) \leq 1 - \max_{z \in \mathcal{Z}} p(\tilde{x}, z) = 1 - \bar{p}_z(\tilde{x}).$$

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:

$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)] \geq \bar{p}_z(\tilde{x})p^*(x),$$
$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)] \leq p^*(x) + \underline{p}_z(\tilde{x})[1 - p^*(x)].$$

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:

$$\frac{p(x) - \underline{p}_z(\tilde{x})}{1 - \underline{p}_z(\tilde{x})} \leq p^*(x) \leq \frac{p(x)}{\bar{p}_z(\tilde{x})}.$$

Now we prove the sharpness of the above bounds. It can be proved by showing that the lower bound and upper bound can be achieved under some DGP. The idea is to show that given the lower bound or 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 $\alpha_1(\tilde{x}) = 1 - \bar{p}_z(\tilde{x})$, $\alpha_0(\tilde{x}) = 0$, and the true probability is constructed as $p^*(x) = \frac{p(x)}{\bar{p}_z(\tilde{x})}$.

It can be verified that this construction matches the reported probability $p(x)$ and satisfies assumptions. Similarly the lower bound can be achieved when $\alpha_1(\tilde{x}) = 0$, $\alpha_0(\tilde{x}) = \underline{p}_z(\tilde{x})$, and $p^*(x) = \frac{p(x) - \underline{p}_z(\tilde{x})}{1 - \underline{p}_z(\tilde{x})}$.

$\square$

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 $\tilde{w} < w$:

$$0 \leq \alpha_0(x, w) \leq \min_{\tilde{w} \leq w} p_W(x, \tilde{w}) = \underline{p}_w(x, w),$$

$$0 \leq \alpha_1(x, w) \leq 1 - \max_{\tilde{w} \leq w} p_W(x, \tilde{w}) = 1 - \bar{p}_w(x, w).$$

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:

$$p_W(x, w) = [1 - \alpha_1(x, w)]p^*(x) + \alpha_0(x, w)[1 - p^*(x)] \geq \bar{p}_w(x, w)p^*(x),$$

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

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:

$$\frac{p_W(x, w) - \underline{p}_w(x, w)}{1 - \underline{p}_w(x, w)} \leq p^*(x) \leq \frac{p_W(x, w)}{\bar{p}_w(x, w)}.$$

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:

$$\max_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w) - \underline{p}_w(x, w)}{1 - \underline{p}_w(x, w)} \right\} \leq p^*(x) \leq \min_{w \in \mathcal{W}} \left\{ \frac{p_W(x, w)}{\bar{p}_w(x, w)} \right\}.$$

Now we prove 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 $p^*(x) = \min\left\{ \frac{p_W(x, w_1)}{\bar{p}_w(x, w_1)}, 1 \right\} = \frac{p_W(x, w_1)}{\bar{p}_w(x, w_1)}$. This bound can be achieved when the misreporting probability satisfies $\alpha_1(x, w_1) = 1 - \bar{p}_w(x, w_1), \alpha_0(x, w_1) = 0$ 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 $p_W(x, w)$.

The lower bound for the true probability is $p^*(x) = \max\left\{ \frac{p_W(x, w_1) - \underline{p}_w(x, w_1)}{1 - \underline{p}_w(x, w_1)}, 0 \right\} = \frac{p_W(x, w_1) - \underline{p}_w(x, w_1)}{1 - \underline{p}_w(x, w_1)}$. It can be achieved when $\alpha_1(x, w_1) = 0, \alpha_0(x, w_1) = \underline{p}_w(x, w_1)$ 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.

$\square$

A.3   Proof of Proposition 3

Proof. We focus on one type of one-sided misreporting where $\alpha_0(\tilde{x}) = 0$ or $\alpha_0(\tilde{x}, w) = 0$, and the same analysis applies to the other case.

When instrument Z is available, the upper bound $U_1(x)$ remains the same and it is sharp. It can be achieved when $\alpha_1(\tilde{x}) = 1 - \bar{p}_z(\tilde{x}), \alpha_0(\tilde{x}) = 0$, and the true probability is $p^*(x) = \frac{p(x)}{\bar{p}_z(\tilde{x})} = U_1(x)$. 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 as $\alpha_1(\tilde{x}) = 0, \alpha_0(\tilde{x}) = 0$ and the true probability as $p^*(x) = p(x)$ such that they match with the reported probability $p(x)$.

When instrument W is available, one-sided misreporting assumption implies the monotonicity of $p_W(x, w)$ in w under the monotonicity assumption of W in Assumption 6, and the upper bound should equal one $U_2(x) = 1$. This can be achieved when $1 - \alpha_1(x, w) = p_W(x, w)$ and $p^*(x) = 1$.

For the lower bound, the following restriction holds under the exclusion restriction:

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

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

$$p^*(x) \geq \max_w p_W(x, w).$$

The sharpness of the lower bound can be shown by constructing misreporting probabilities and the true probability as: $1 - \alpha_1(x, w) = \frac{p(x,w)}{\max_w p_W(x,w)}$ and $p^*(x) = \max_w p_W(x, w)$. This construciton satisfies all assumptions and matches with the reported probability.

$\square$

A.4 Proof of Propositions 4 and 5

Proof. The proofs of Propositions 4 and 5 are similar to Appendix A.1-A.2. We focus on the scenario where instrument Z is available, and the analysis for instrument W is omitted since the idea is the same.

We first study Proposition 4 where misreporting probabilities are bounded. As shown in A.1, under Assumptions 2-4, the misreporting probability $\alpha_y(\tilde{x})$ can be bounded below:

$$\alpha_0(\tilde{x}) \in [0, \underline{p}_z(\tilde{x})] \quad \alpha_1(\tilde{x}) \in [0, 1 - \bar{p}_z(\tilde{x})].$$

Combining Assumption 8, we can tighten the bounds for $\alpha_y(\tilde{x})$:

$$\alpha_0(\tilde{x}) \in [0, \min\{\underline{p}_z(\tilde{x}), \bar{\alpha}_0\}] \quad \alpha_1(\tilde{x}) \in [0, 1 - \max\{\bar{p}_z(\tilde{x}), 1 - \bar{\alpha}_1\}].$$

Recall that the reported probability $p(x)$ is given as

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

yielding the following inequalities:

$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)] \geq \max\{\bar{p}_z(\tilde{x}), 1 - \bar{\alpha}_1\}p^*(x),$$
$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)] \leq p^*(x) + \min\{\underline{p}_z(\tilde{x}), \bar{\alpha}_0\}[1 - p^*(x)].$$

Therefore, the true choice probability $p^*(x)$ can be bounded as

$$\frac{p(x) - \min\left\{\underline{p}(\tilde{x}), \bar{\alpha}_0\right\}}{1 - \min\left\{\underline{p}(\tilde{x}), \bar{\alpha}_0\right\}} \leq p^*(x) \leq \frac{p(x)}{\max\left\{\bar{p}(\tilde{x}), 1 - \bar{\alpha}_1\right\}}.$$

In Proposition 5 with monotone misreporting probabilities $\alpha_0(\tilde{x}) \leq \alpha_1(\tilde{x})$, the misreporting probabilities can be further bounded as

$$\alpha_0(\tilde{x}) \in [0, \min\{\underline{p}_z(\tilde{x}), 1 - \bar{p}_z(\tilde{x})\}] \quad \alpha_1(\tilde{x}) \in [0, 1 - \bar{p}_z(\tilde{x})].$$

Under similar arguments, we have

$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)] \geq \bar{p}_z(\tilde{x})p^*(x),$$
$$p(x) = [1 - \alpha_1(\tilde{x})]p^*(x) + \alpha_0(\tilde{x})[1 - p^*(x)] \leq p^*(x) + \min\{\underline{p}_z(\tilde{x}), 1 - \bar{p}_z(\tilde{x})\}[1 - p^*(x)].$$

Therefore, misreporting probabilities can be bounded as

$$\frac{p(x) - \min\left\{\underline{p}_z(\tilde{x}), 1 - \bar{p}_z(\tilde{x})\right\}}{1 - \min\left\{\underline{p}_z(\tilde{x}), 1 - \bar{p}_z(\tilde{x})\right\}} \leq p^*(x) \leq \frac{p(x)}{\bar{p}(\tilde{x})}.$$

$\square$

A.5 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 10, 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 10, 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:

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

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 $w_1 \neq w_2 \in \mathcal{W}$:

$$p_W(z, w_1) = [1 - \alpha_1(w_1)]p^*(z) + \alpha_0(w_1)[1 - p^*(z)],$$
$$p_W(z, w_2) = [1 - \alpha_1(w_2)]p^*(z) + \alpha_0(w_2)[1 - p^*(z)].$$

From condition (9), we know that $1 - \alpha_1(w) - \alpha_0(w) \geq \max\limits_{z \in \mathcal{Z}} p_W(z, w) - \min\limits_{z \in \mathcal{Z}} p_W(z, w) > 0$ 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 $z \in \{z_1, z_2\}$:

$$p_W(z_1, w_1) = A_1(w_1, w_2)p_W(z_1, w_2) - A_0(w_1, w_2),$$
$$p_W(z_2, w_1) = A_1(w_1, w_2)p_W(z_2, w_2) - A_0(w_1, w_2).$$

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:

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

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 $\alpha_y(w)$:

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

By the monotonicity condition of instrument W in Assumption 10 and bounds on misreporting probabilities in condition (9), the following condition summarizes restrictions on misreporting probabilities:

$$1 \geq 1 - \alpha_1(w_1) \geq 1 - \alpha_1(w_2), \qquad 0 \leq \alpha_0(w_1) \leq \alpha_0(w_2),$$
$$1 - \alpha_1(w) \geq \max\limits_{z \in \mathcal{Z}} p_W(z, w), \qquad \alpha_0(w) \leq \min\limits_{z \in \mathcal{Z}} p_W(z, w). \tag{11}$$

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:

$$q_1(w_1, w_2) = \frac{1 - \alpha_1(w_1) - \alpha_0(w_1)}{1 - \alpha_1(w_2) - \alpha_0(w_2)} > 1,$$

$$q_0(w_1, w_2) = \alpha_0(w_2)q_1(w_1, w_2) - \alpha_0(w_1) \geq \alpha_0(w_1)[q_1(w_1, w_2) - 1] \geq 0,$$

$$q_1(w_1, w_2) - q_0(w_1, w_2) - 1 = q_1(w_1, w_2) - \alpha_1(w_1) - [1 - \alpha_1(w_2)]q_1(w_1, w_2)$$
$$= \alpha_1(w_2)q_1(w_1, w_2) - \alpha_1(w_1) \geq 0.$$

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

$$\max\left\{\frac{q_0(w_1, w_2)}{q_1(w_1, w_2) - 1}, \max_{z \in \mathcal{Z}} p_W(z, w_1)\right\} \leq 1 - \alpha_1(w_1) \leq 1,$$

$$0 \leq \alpha_0(w_1) \leq \min\left\{\frac{q_0(w_1, w_2)}{q_1(w_1, w_2) - 1}, \min_{z \in \mathcal{Z}} p_W(z, w_1)\right\}.$$

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:

$$0 \leq \alpha_1(w_m) \leq 1 - \max_{z \in \mathcal{Z}, w < w_m \in \mathcal{W}}\left\{\frac{q_0(w_m, w)}{q_1(w_m, w) - 1}, p_W(z, w_m)\right\} = U_{\alpha_1}(w_m),$$

$$0 \leq \alpha_0(w_m) \leq \min_{z \in \mathcal{Z}, w < w_m \in \mathcal{W}}\left\{\frac{q_0(w_m, w)}{q_1(w_m, w) - 1}, p_W(z, w_m)\right\} = U_{\alpha_0}(w_m).$$

$$\square$$

A.6 Proof of Proposition 6

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:

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

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 $\max_z p(\tilde{x}, z, w) > 0$ and

$\min_z p_W(\tilde{x}, z, w) < 1$ for any $(\tilde{x}, w)$.

We prove 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 10. 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 $\min_z p_W(\tilde{x}, z, w) < 1$ 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 10.

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 verrity that misreporting probabilities we construct satisfy Assumption 10 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 $q_1(\tilde{x}, w_1, w)$:

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

Therefore the monotonicity condition is also satisfied.

Lastly, we need to prove 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 obtained. 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 $\alpha_0(\tilde{x}, w_2) = \frac{U_{\alpha_0}(\tilde{x}, w_1) + q_0(\tilde{x}, w_1, w_2)}{q_1(\tilde{x}, w_1, w_2)}$. Simi-

larly it can be shown that they satisfy Assumption 10 and match with the reported probability $p_W(x, w)$ for any (x, w).

$\square$