Bounding program benefits when participation is misreported: Estimation and inference with Stata

Bounding Program Benefits When Participation is Misreported

Denni Tommasi[1] and Lina Zhang*[2]

[1]University of Bologna, IZA and CDES

[2]University of Amsterdam and Tinbergen Institute

October 11, 2023

Abstract

Instrumental variables (IV) are commonly used to estimate treatment effects in case of noncompliance. However, program participation is often misreported in survey data and standard techniques are not sufficient to point identify and consistently estimate the effects of interest. In this paper, we show that the identifiable IV estimand that ignores treatment misclassification is a weighted average of local average treatment effects with weights that can also be negative. This is troublesome because it may fail to deliver a correct causal interpretation, and this is true even if all the weights are non-negative. Therefore, we provide three IV strategies to bound the program benefits when both noncompliance and misreporting are present. We demonstrate the gain of identification power achieved by leveraging multiple exogenous variations when discrete or multiple-discrete IVs are available. At last, we use our new Stata command, `ivbounds`, to study the benefits of participating in the 401(k) pension plan on savings.

JEL Codes: C14, C21, C26, C35, C51.

Keywords: heterogeneous treatment effects, causality, binary treatment, endogeneity, measurement error, discrete or multiple-discrete instruments, weighted average of LATEs, program evaluation.

*Tommasi: Department of Economics, University of Bologna. Piazza Scaravilli 2, Bologna, 40126, Italy. E-mail: `denni.tommasi@unibo.it`. Zhang: Amsterdam School of Economics, University of Amsterdam. Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. E-mail: `l.zhang5@uva.nl`. We are grateful to the editor (Elie Tamer), associate editor, and two anonymous referees for their detailed and constructive comments. This is a substantial revision of a paper that circulated with the same title. We have benefited from discussions with Isaiah Andrews, Giuseppe Cavaliere, Peng Ding, Luca Fanelli, David Frazier, Martin Huber, Kosuke Imai, Zhichao Jiang, Desire Kedagni, Umair Khalil, Arthur Lewbel, Charles Manski, Francesca Molinari, Akanksha Negi, Whitney Newey, Didier Nibbering, Tatsushi Oka, John Pepper, Alessandro Tarozzi, Takuya Ura and Kaspar Wuthrich, participants at the Italian Congress of Econometrics and Empirical Economics, and seminar participants at the University of Bologna for valuable comments. Hai Duong provided excellent research assistance. Tommasi acknowledges financial support from SEED fund of Monash University. The `ivbounds` Stata command is available from the Statistical Software Components (SSC) repository. See Lin, Tommasi, and Zhang (2021) for explanations on how to use the command. All remaining errors are ours.

1 Introduction

The instrumental variables (IV) method is commonly used to estimate treatment effects in case of noncompliance (Athey and Imbens, 2017). Standard approaches for the identification and inference of causal parameters require that the treatment variable is correctly measured. In program evaluation, misreporting (misclassification) of key variables due to the "desire to shorten the time spent on the interview, the stigma of program participation, the sensitivity of income information, or changes in the characteristics of those who receive transfers" is an increasing problem for social scientists (Meyer et al., 2015, p. 219). Since participation is usually binary, attempting to evaluate the benefits of a program using standard IV techniques would lead to biased estimates (Kreider, 2010; Millimet, 2011).[1] In this paper, we focus on the IV estimand that captures the average causal effect for compliers (Imbens and Angrist, 1994), and develop IV methods that can be used to measure the benefits of a binary program when both noncompliance and misreporting of treatment are present.

We aim to address the challenge posed by endogenous (differential) treatment misclassification within a comprehensive framework that accounts for (i) endogenous treatment selection, (ii) heterogeneous treatment effects, and (iii) binary, discrete, or multiple-discrete IVs. When the IV is binary, our target parameter is the local average treatment effect (LATE); with discrete or multiple-discrete IVs, our target parameter is the weighted average of LATEs (WLATE).[2] Recent works have developed methods to deal with misclassified binary treatment variables.[3] When confronted with endogenous (differential) misclassification of the treatment, Kreider et al. (2012) established an approach that necessitates auxiliary data to partially identify the average treatment effect. Nguimkeu et al. (2018) achieved point identification of the target parameter(s) by assuming homogeneous treatment effects and a parametric model. Whereas, Ura (2018) focused on heterogeneous treatment effects and proposed a partial identification strategy for the LATE using a binary IV.

Our analysis proceeds in three steps. Firstly, we delve into the mismeasured IV estimand, which overlooks treatment misclassification. We discover that this estimand is a weighted average of LATEs with potentially negative weights. This poses a significant challenge as the sign of the identifiable effect may differ from that of the underlying LATEs. To tackle this, we characterize the bias of the mismeasured IV estimand and establish its link with the true IV estimand. This connection is mediated by a novel parameter defined in terms of misclassification probabilities, which can be utilized to approximate the potential extent of bias in estimated program benefits. Additionally, we provide a straightforward sufficient condition that ensures the nonnegative weights of the mismeasured IV estimand. These findings complement the work of Chalak (2017), who also addressed the problem of nonnegative weights in an IV framework with accurate treatment observations but potentially mismeasured instruments.

Secondly, we demonstrate that even when the mismeasured IV estimand has nonnegative weights, it fails to accurately interpret the true causal effects of a program. To address this issue, we propose partial identification methods to study both LATE and WLATE. In this regard, we improve the bound for LATE

[1] In a classical measurement error scenario, IVs correct for endogeneity and measurement error of the treatment simultaneously. However, measurement error is always nonclassical for a binary treatment, because of the negative correlation between the true treatment and the error term.

[2] IVs with a support larger than two values are commonly used in empirical research. According to Mogstad et al. (2021), more than half of the empirical papers using IVs and published in top journals in the last 20 years "make use of multiple instrumental variables for a single treatment." In practice, WLATE is often used by applied researchers to combine multiple IVs for estimation efficiency. In heterogeneous treatment effect settings, WLATE has a causal interpretation as a positively weighted average of LATEs, which has been emphasized in the literature as an attractive property.

[3] Our paper aligns with a well-established literature concerned with treatment misclassification. This is thoroughly acknowledged in Appendix A.

of Ura (2018) to accommodate instruments with support expanding from binary to discrete to multiple-discrete. Compared to the bound for LATE in a binary IV setting, we enhance the identification power by leveraging multiple exogenous variations that capture the distributional effects of the discrete IV(s) on observable variables. Moreover, we propose two strategies to establish bounds for the WLATE: first, by leveraging the bounds of LATEs, and second, by leveraging the bounds of a newly introduced estimand that captures the local average of treatment misclassifications (LATMs) for compliers. Furthermore, we provide sufficient conditions that ensure sharp bounds for both LATEs and LATMs.

Finally, we formalize a third partial identification strategy to combine external information about the extent of misclassification to obtain tighter bounds or even a point estimate.[4] Our strategy improves upon others that rely on auxiliary data because (i) external information are only used to narrow our bounds (Kreider and Pepper, 2007; Kreider et al., 2012), (ii) we do not need to assume an exogenous treatment (Imai and Yamamoto, 2010; Battistin and Sianesi, 2011), nor to observe who has missing treatment (Molinari, 2010), and (iii) treatment misclassification in our case can be endogenous (differential). When the practitioner has a credible approximation of the misclassification rates, our proposed methods, using external information, can generate a point estimate that is less biased compared to the naive IV estimate that ignores treatment misclassification. Hence, our method has a bias reduction property.

Building upon the core principle of partial identification, which explores "what can be learned about the parameter of interest given the data and model assumptions" (Kline and Tamer, 2022), we aim to provide useful tools to study both LATE and WLATE. Applied researchers can make their own decisions about which target parameter (LATE, WLATE, or both) they want to pursue in their specific contexts and research objectives. In cases where the causal effects for compliers do not capture the desired effect of interest, LATE or WLATE can be employed to extrapolate to causal effects for a wider population. Notably, recent works by Mogstad et al. (2018, 2020) explore the use of multiple IVs to study policy-relevant treatment effects (PRTEs) utilizing the marginal treatment effect (MTE) approach. However, these papers assume that the treatment variable is accurately measured, highlighting the significance of our approach as a starting point for investigating PRTEs in the presence of treatment misclassification. In contrast, Acerenza et al. (2021), Acerenza (2021), and Possebom (2021) conduct bounding analyses for MTEs and PRTEs when the treatment variable suffers from misclassification. Our paper can be viewed as a valuable complement to these recent works.

The remainder of the paper is organized as follows. Section 2 introduces the model setup and the limitations of the standard IV approach with misclassified treatment. Section 3 develops the main partial identification results. Section 4 applies our dedicated Stata command, `ivbounds`, to reassess the benefits of participating in the 401(k) pension plan on savings in the US. Concluding remarks are in Section 5. Due to space constraints, we relegate the proofs for results in the main text to Appendix B and C, extensions of our findings to settings with multiple treatment proxies and covariates to Appendix D and E, and Monte Carlo simulations to Appendix F.

[4]Misclassification rates of program participation are increasingly accessible in various survey data. For example, Meyer et al. (2020) found false negative rates of 23%, 35%, and 50% for participating in a food stamp program, Supplemental Nutrition Assistance Program (SNAP), in the CPS, ACS, and Survey of Income and Program Participation (SIPP), respectively.

2 Setup and Limitations of the Standard IV Approach

In this section, we describe our model setup and show the limitations of the standard IV approach when the treatment variable is contaminated by measurement error. This leads to a simple relationship between the true and mismeasured effect, which can be captured by a summary statistic of the misclassification probabilities.

2.1 True effect

Let D be the true binary treatment variable that affects the outcome of interest. Throughout the paper, we assume that D is *not* observed. Let Z be a $h \times 1$ vector of discrete instruments. Let $\Omega_Z = \{z_0, z_1, ..., z_K\}$ be the support of Z with $z_k \in \mathbb{R}^h$. Denote $D_k \in \{0,1\}$, for $k = 0, 1, ..., K$, as the potential treatment corresponding to possible realization z_k of Z. By definition,

$$D = \sum_{k=0}^{K} 1[Z = z_k] D_k,$$

where $1[\cdot]$ denotes the indicator function. Denote by $\Pr(z_k) = \mathbb{E}(D|Z = z_k)$ the propensity score. Let Y be an observed outcome of interest and let Y_d be the potential outcome with $d \in \{0,1\}$ for possible realization of D. Denote by $\Omega_Y \subset \mathbb{R}$ the support of Y, Y_1 and Y_0. Then,

$$Y = DY_1 + (1-D)Y_0.$$

A common way to exploit multiple instruments is to introduce a scalar function $g : \Omega_Z \mapsto \mathbb{R}$, for example, $g(z)$ can be an estimate of $\Pr(z)$ or other known functions.[5]

Assumption 2.1. *Y, D and Z satisfy the standard Imbens and Angrist (1994) assumptions:*

(i) (i.i.d.) $(Y_1, Y_0, \{D_k\}_{k=0}^K, Z)$ are independent and identically distributed across all individuals and have finite first and second moments;

(ii) (Unconfoundedness) $Z \perp (Y_1, Y_0, \{D_k\}_{k=0}^K)$ and $Pr(z) = \mathbb{E}(D|Z = z)$ for $z \in \Omega_Z$ is a nontrivial function of z; $0 < \pi_k = Pr(Z = z_k) < 1$, $k = 0, 1, ..., K$;

(iii) (First stage) $Cov(D, g(Z)) \neq 0$;

(iv) (Monotonicity) For any $z_l, z_w \in \Omega_Z$, with probability one, either $D_l \geq D_w$ for all individuals, or $D_l \leq D_w$ for all individuals. Furthermore, for all $z_l, z_w \in \Omega_Z$, either $Pr(z_l) \leq Pr(z_w)$ implies $g(z_l) \leq g(z_w)$, or $Pr(z_l) \leq Pr(z_w)$ implies $g(z_l) \geq g(z_w)$.

The monotonicity assumption is satisfied if no subjects respond in the opposite way to their instrument assignment status (no defiers).[6] Throughout the paper, we denote compliers $(D_{k-1} = 0, D_k = 1)$ as C_k.

[5] If Z is a scalar binary or discrete instrument, we can simply set $g(z) = z$. If Z includes multiple instruments, $g(z)$ can be set as, for example, an estimate of $E[Y|Z = z]$ or of $\Pr(T = 1|Z = z)$ for $z \in \Omega_Z$, where T represents a proxy of the true treatment and will be introduced later.

[6] When there is more than one instrument, Mogstad et al. (2020, 2021) point out that the monotonicity assumption is satisfied if the selection into treatment is homogeneous. In the presence of multiple IVs, it may be possible to extend the analysis in this paper via replacing the monotonicity condition with a weaker "partial monotonicity" condition proposed by Mogstad et al. (2021), where the monotonicity is satisfied for each instrument separately. We leave it to future research.

If D was observed, under Assumption 2.1, the Imbens and Angrist (1994)'s weighted average of local average treatment effect (WLATE) would be identified by the IV estimand:

$$\alpha^{IV} := \frac{\text{Cov}(Y, g(Z))}{\text{Cov}(D, g(Z))} = \frac{\mathbb{E}[(Y - \mathbb{E}(Y))(g(Z) - \mathbb{E}[g(Z)])]}{\mathbb{E}[(D - \mathbb{E}(D))(g(Z) - \mathbb{E}[g(Z)])]} = \sum_{k=1}^{K} \gamma_k^{IV} \alpha_{k,k-1}, \tag{1}$$

where $\gamma_k^{IV} := \frac{\Pr(C_k)\sum_{l=k}^{K} \pi_l (g(z_l) - \mathbb{E}[g(Z)])}{\sum_{m=1}^{K} \Pr(C_m)\sum_{l=m}^{K} \pi_l (g(z_l) - \mathbb{E}[g(Z)])}$ are the weights, $\Pr(C_k) = \Pr(z_k) - \Pr(z_{k-1})$ and $\alpha_{k,k-1} := \mathbb{E}[Y_1 - Y_0 | C_k]$ is the local average treatment effect (LATE) for compliers C_k. The weights $\{\gamma_k^{IV}\}_{k=1}^{K}$ are nonnegative and $\sum_{k=1}^{K} \gamma_k^{IV} = 1$. The IV estimand α^{IV} has a useful causal interpretation, because it assigns nonnegative weights to all the LATEs, so that it will be positive (or negative) if all of the underlying LATEs are positive (or negative). However, since in practice we do not observe D, we cannot implement this standard approach.

2.2 Mismeasured effect

Instead of D, suppose we can observe a binary treatment indicator T, which could be a proxy for D, or could correspond to reported values of D that are misclassified for some observations. Define $T_d \in \{0, 1\}$ as the potential observed treatment with $d \in \{0, 1\}$ for possible realization of D. Then by definition:

$$T = DT_1 + (1 - D)T_0.$$

The variables T_0 and T_1 can be used to indicate whether the treatment is misclassified or not: if $T_0 = 1$, then a true $D = 0$ is misclassified as treated (false positive), and if $T_1 = 0$, then a true $D = 1$ is misclassified as untreated (false negative).

Assumption 2.2. *The treatment indicator T is such that the following conditions are satisfied:*

(i) (Extended unconfoundedness) $Z \perp (Y_1, Y_0, \{D_k\}_{k=0}^{K}, T_1, T_0)$;

(ii) (Extended first stage) $Cov(T, g(Z)) \neq 0$.

Assumption 2.2-(i) combines the LATE unconfoundedness assumption that $Z \perp \left(Y_1, Y_0, \{D_k\}_{k=0}^{K}\right)$ with the assumption that the instruments are also independent of the treatment measurement errors, and hence of (T_1, T_0).[7] Assumption 2.2-(ii) is used to ensure that the identifiable estimand from observable data is well-defined and it is a minimal relevance condition. These assumptions do not impose any restriction on the type of misclassification error. Thus, we allow both exogenous (non-differential) misclassification error, in the sense that $(T_0, T_1) \perp (Y_1, Y_0, \{D_k\}_{k=0}^{K})$, such as clerical errors, as well as endogenous (differential) misclassification error, in the sense that $(T_0, T_1) \not\perp (Y_1, Y_0, \{D_k\}_{k=0}^{K})$, such as misreporting due to fear of stigma.

Using the proxy T in place of D leads to the identification of a new parameter, which is useful to characterize. Let $p_{d,k} = \mathbb{E}(T_d \mid C_k)$ for $d \in \{0, 1\}$ and $k = 1, 2, ..., K$. By definition, $p_{1,k}$ is the probability

[7] The assumption of independence between the instrument and treatment misclassification errors is frequently employed in recent literature investigating self-selection into treatment using a misclassified treatment variable (Ura, 2018; Calvi et al., 2021; Tommasi and Zhang, 2022; Acerenza et al., 2021; Acerenza, 2021). However, scholars such as Bound et al. (2001), Haider and Stephens Jr. (2019), and Possebom (2021) have raised concerns about its plausibility across different empirical contexts, where misclassification rates may vary across values of the instrument. Later, we discuss the validity of this assumption in the context of our empirical illustration. We thank an anonymous referee for pointing this out.

that compliers C_k would have their treatment correctly observed if they were treated. In contrast, $p_{0,k}$ is the probability that compliers C_k would have their treatment incorrectly observed if they were untreated.

Theorem 2.1. *Let Assumption 2.1 and 2.2 hold. Then, the mismeasured IV estimand*

$$\alpha^{Mis} := \frac{Cov(Y, g(Z))}{Cov(T, g(Z))} = \frac{\mathbb{E}[(Y - \mathbb{E}(Y))(g(Z) - \mathbb{E}[g(Z)])]}{\mathbb{E}[(T - \mathbb{E}(T))(g(Z) - \mathbb{E}[g(Z)])]} = \sum_{k=1}^{K} \gamma_k^{Mis} \alpha_{k,k-1}, \tag{2}$$

where $\gamma_k^{Mis} := \frac{Pr(C_k)\sum_{l=k}^{K} \pi_l (g(z_l) - \mathbb{E}[g(Z)])}{\sum_{m=1}^{K} (p_{1,m} - p_{0,m}) Pr(C_m) \sum_{l=m}^{K} \pi_l (g(z_l) - \mathbb{E}[g(Z)])}$ *are the weights for compliers* C_k.

Proof of Theorem 2.1. See Appendix B.1. □

Intuitively, α^{Mis} denotes the identifiable estimand if we ignore the misclassification error and use the mismeasured treatment indicator T in place of the true treatment D. We can see that α^{Mis} is also a weighted average of LATEs, but, unlike α^{IV}, its weights may be negative. If the misclassification is particularly severe, it is possible that $p_{1,k} - p_{0,k} < 0$ and α^{Mis} may become negative (positive) even if the true treatment effects are positive (negative) for all compliers in the population. In contrast, if the treatment indicator contains sufficient information about the true treatment, that is, if $0 < p_{1,k} - p_{0,k} \leq 1$ for all complier groups, then the weight γ_k^{Mis} is nonnegative and α^{Mis} can sign α^{IV}. That said, the summation $\sum_{l=k}^{K} \gamma_k^{Mis}$ is greater than one because each γ_k^{Mis} is biased upward by the misclassification error in the denominator, leading to overestimation $|\alpha^{Mis}| \geq |\alpha^{IV}|$. A sufficient condition for $\alpha^{Mis} = \alpha^{IV}$ is that $p_{1,k} = 1$ and $p_{0,k} = 0$ for all k (no misclassification error).

2.3 Relationship between the true and mismeasured effect

Denote by $\Delta p_k = p_{1,k} - p_{0,k}$ the difference between misclassification probabilities for complier group C_k. We refer to Δp_k as the local average of treatment misclassification (LATM)

$$\text{LATM} = \Delta p_k = \mathbb{E}[T_1 - T_0 | C_k], \tag{3}$$

because it is analogous to the LATE if we replace $Y_1 - Y_0$ by $T_1 - T_0$. There is a simple relationship between α^{IV} and α^{Mis} which can be captured by a weighted average of the LATMs.

Corollary 2.1. *Let Assumption 2.1 and 2.2 hold and, without loss of generality, assume* $\gamma_k^{IV} \neq 0$ *and* $\gamma_k^{Mis} \neq 0$ *for* $\forall k$. *Then, there exists a summary statistic* ξ *such that:*

$$\alpha^{Mis} = \sum_{k=1}^{K} \gamma_k^{IV} \alpha_{k,k-1} \times \frac{\gamma_k^{Mis}}{\gamma_k^{IV}} \implies \alpha^{IV} = \xi \alpha^{Mis} \tag{4}$$

where the ratio $\xi = \gamma_k^{IV} / \gamma_k^{Mis} = \sum_{k=1}^{K} \gamma_k^{IV} \Delta p_k$.

Proof of Corollary 2.1. See Appendix B.2. □

The parameter ξ is a constant across k, with absolute value less than or equal to one, and is unobserved in practice. Denote the weighted average probability of false negative as $w^n = 1 - \sum_{k=1}^{K} \gamma_k^{IV} p_{1,k}$ (this is

the probability of treated individuals misclassified as untreated) and false positive as $w^p = \sum_{k=1}^{K} \gamma_k^{IV} p_{0,k}$ (this is the probability of untreated individuals misclassified as treated). Then, by definition,

$$\xi = 1 - w^n - w^p \implies \alpha^{IV} = (1 - w^n - w^p)\alpha^{Mis} \tag{5}$$

which makes clear that ξ can be interpreted as an overall measure of how severe the treatment misclassification is. When there is no misclassification ($w^n = w^p = 0$), $\xi = 1$ and $\alpha^{Mis} = \alpha^{IV}$. As misclassification worsens ($w^n > 0, w^p > 0$), the value of ξ falls and the bias in α^{Mis} becomes increasingly severe. When $0 < \xi < 1$, α^{Mis} and α^{IV} are of the same sign, while α^{Mis} inflates the effect. When $\xi < 0$, α^{Mis} and α^{IV} have opposite signs.

A similar link between the causal and the identifiable parameter has been established in the literature under a variety of different conditions.[8] Our main contribution with respect to the previous works is twofold. First, we demonstrate that ξ can be used to derive a sufficient condition under which α^{Mis} can sign α^{IV}. Second, since the probabilities of false positive and false negative are increasingly available to practitioners through administrative datasets or validations studies, we improve upon the literature by showing that, using those information, we can approximate the relative bias $\frac{\alpha^{Mis} - \alpha^{IV}}{\alpha^{IV}} = 1/\xi - 1$ via ξ.

Assumption 2.3 (Informative Treatment Proxy)**.** *For all* $k = 1, 2, ..., K$, $Pr(T = d|C_k, D = d) > Pr(T = d|C_k, D = 1 - d)$, $d = \{0, 1\}$.

Assumption 2.3 requires that T is an informative proxy of the actual treatment status, and it does not impose further restrictions regarding the dependence of misclassification errors on the potential outcomes. One sufficient condition for it is that $\max\{\Pr(T = 1|C_k, D = 0), \Pr(T = 0|C_k, D = 1)\} < 1/2$ for all k, meaning that the observations of T are more accurate than pure guesses about the true treatment. Similar restrictions are widely invoked in the measurement error literature (e.g., Hausman et al., 1998). This assumption is sufficient to ensure that α^{Mis} and α^{IV} have the same sign.

Corollary 2.2. *Under Assumption 2.1, 2.2 and 2.3, we have*

(i) $Pr(T = 1|Z = z_l) \leq Pr(T = 1|Z = z_w)$ *implies* $Pr(z_l) \leq Pr(z_w)$ *for* $\forall z_l, z_w \in \Omega_Z$;

(ii) $\gamma_k^{Mis} \geq 0$ *for all* k *and* $sign(\alpha^{Mis}) = sign(\alpha^{IV})$.

Proof of Corollary 2.2. See Appendix B.3. □

Corollary 2.2 says that if the proxy T is informative, we can reveal the direction of the effect of the instrumental variable(s) Z on the true treatment status D by using the naive propensity score $\Pr(T = 1|Z)$. This is still possible even though the magnitude of the propensity score $\Pr(z) = \Pr(D = 1|Z = z)$ cannot be recovered from the observed data. Hereafter, we assume the elements $\{z_0, z_1, ..., z_K\}$ of the support Ω_Z follow an *ascending order*, in the sense that $\forall l, w \in \{0, 1, ..., K\}$, $l < w$ implies $\Pr(z_l) \leq \Pr(z_w)$. Such an ascending order assures that the weights in the true effect α^{IV} are nonnegative. More importantly, the informative proxy T rules out negative weights in α^{Mis} since all LATMs are strictly positive, hence α^{Mis} and α^{IV} have the same sign.

[8] First, differently from Frazis and Loewenstein (2003) and Stephens Jr and Unayama (2019), we establish a link between the causal and identifiable parameter in a heterogeneous treatment effect framework, while they assume homogeneous treatment effects. Second, Lewbel (2007) and Battistin and Sianesi (2011) assume an exogenous treatment, which is not required in our context. Third, we generalizes the B-LATE (for Biased LATE) estimator of Calvi, Lewbel, and Tommasi (2021) and its link to the true effect to a discrete and multiple-discrete-instrument setting. All these papers, including ours, are related to one another and benefited from the result by Hausman et al. (1998).

Table 1: Bias of α^{Mis} relative to α^{IV} for different misclassification probabilities

Bias $= (1/\xi - 1) \times \alpha^{IV}$						
$w^n \downarrow$ | $w^p \rightarrow$	0	0.05	0.10	0.20	0.30	0.40
0.00	0.000	0.053	0.111	0.250	0.429	0.667
0.05	0.053	0.111	0.176	0.333	0.538	0.818
0.10	0.111	0.176	0.250	0.429	0.667	1.000
0.20	0.250	0.333	0.429	0.667	1.000	1.500
0.30	0.429	0.538	0.667	1.000	1.500	2.333
0.40	0.667	0.818	1.000	1.500	2.333	4.000

Notes: Each cell reports $(1/\xi - 1) \times \alpha^{IV}$ for different values of w^n (false negative) and w^p (false positive). α^{IV} is set to 1.

Finally, a practitioner can use relationship (5) to approximate the possible level of bias of the estimated benefits of a program for different misclassification probabilities.[9] In Table 1 we report the difference in the values between α^{Mis} and α^{IV} for different values of w^n and w^p. The true effect α^{IV} is normalized to 1. Hence, if the sample contains, e.g., 10% false negative and 5% false positive, this table tells the practitioner that the estimated effect α^{Mis} is approximately 17.6% larger than the (unknown) true effect. Results in Table 1 confirm that even with nonnegative weights, α^{Mis} still overestimates (in absolute value) α^{IV}, and the overestimation can be severe even with small or moderate misclassification rates.

3 Partial Identification Strategies

This section proceeds in three acts. First, we present tractable outer sets of the LATEs and the LATMs, and provide sufficient conditions for the sharp identified sets. Second, we develop two strategies to partially identify α^{IV} based on the LATEs and the LATMs, respectively. Finally, we show how to use external information regarding the extent of the misclassification probabilities to obtain tighter bounds of α^{IV} or, under certain conditions, to obtain its point estimate. All of our results remain valid when conditioning on covariates (see Appendix E for more details).

3.1 Bounds of the LATEs and the LATMs

Bounding the probability of compliers. To bound the probability of compliers, we use the concept of total variation (TV) distance. For any generic random variable (or vector) A and $z_k, z_{k-1} \in \Omega_Z$, the TV distance is a L^1 distance between the two conditional distribution functions $f_{A|Z=z_k}$ and $f_{A|Z=z_{k-1}}$, defined as follows:

$$TV_{A,k} = \frac{1}{2} \int |f_{A|Z=z_k}(a) - f_{A|Z=z_{k-1}}(a)| d\mu_A(a),$$

where μ_A denotes a dominating measure for the distribution of A. If A is discrete, the integral is replaced by summation across all possible values of A. The $TV_{A,k}$ captures the extent of the distributional effect of Z on A, when Z changes from z_{k-1} to z_k. For example, if we set $A = Y$, then $TV_{Y,k}$ is the distribution

[9] Notice that a practitioner does not need to know the values of $p_{1,k}$ and $p_{0,k}$ for all k, to be able to approximate the value of ξ. This is because, in practice, the type of information that is increasingly reported in the data is the overall misclassification probabilities, w^n and/or w^p. These are the only information actually required to approximate ξ.

version of the "intention-to-treat" effect.

Lemma 3.1. *Let Assumption 2.1-(ii) to (iv), 2.2-(i) and 2.3 hold. We have that, for* $\forall k = 1, 2, ..., K$*:*

$$TV_{(Y,T),k} \leq Pr(C_k) \leq 1 - \sum_{k' \neq k} TV_{(Y,T),k'}.$$

Proof of Lemma 3.1. See Appendix C.1. □

Using a binary IV, Ura (2018, Lemma 3) shows that the probability of compliers can be bounded from below by the TV distance and from above by one. In Lemma 3.1, our lower bound, $TV_{(Y,T),k}$, is the same to the one obtained using a binary instrument. However, when the instrument(s) are discrete, our upper bound, $1 - \sum_{k' \neq k} TV_{(Y,T),k'}$, is novel as it improves the upper bound of Ura (2018) and gains identification power by incorporating multiple exogenous variations captured by the TV distances of other complier groups. The magnitudes of the two bounds in Lemma 3.1 depend on the strength of the instrument(s). For example, if the change of Z from z_{k-1} to z_k causes no distributional variation in Y and T, the lower bound reduces to 0. Similarly, if no distributional variation is triggered by the change of Z from $z_{k'-1}$ to $z_{k'}$ for all $k' \neq k$, the upper bound increases to 1.

Bounds of the LATEs. Let $\mathbf{P}$ be an arbitrary data generating process (DGP) of (Y, T, Z). Denote the class of DGPs of $\mathbf{P}$ as $\mathscr{P}_0$, then we have $\mathbf{P} \in \mathscr{P}_0$. Denote Θ to be the parameter space of α^{IV} and of all $\alpha_{k,k-1}$.[10] For example, $\Theta = [-1, 1]$ if outcome Y is binary. For $A = \{Y, T\}$, denote $\Delta_k \mathbb{E}(A|Z) = \mathbb{E}(A|Z = z_k) - \mathbb{E}(A|Z = z_{k-1})$. Theorem 1 in Imbens and Angrist (1994) says that under Assumption 2.1 in our paper, we have:

$$\Delta_k \mathbb{E}(Y|Z) = \alpha_{k,k-1} \Pr(C_k). \tag{6}$$

Multiplying both sides of (6) by $\alpha_{k,k-1}$, we obtain that:

$$\alpha_{k,k-1} \Delta_k \mathbb{E}(Y|Z) = \alpha_{k,k-1}^2 \Pr(C_k) \geq 0. \tag{7}$$

Moreover, by applying Lemma 3.1 to the absolute value of (6), we have:

$$|\Delta_k \mathbb{E}(Y|Z)| \leq |\alpha_{k,k-1}| \left[1 - \sum_{k' \neq k} TV_{(Y,T),k'} \right], \tag{8}$$

$$|\Delta_k \mathbb{E}(Y|Z)| \geq |\alpha_{k,k-1}| TV_{(Y,T),k}. \tag{9}$$

Thus, under Assumptions 2.1, 2.2 and 2.3, each LATE ($\alpha_{k,k-1}$) satisfies the inequalities (7)-(9). Inequality (7) indicates that the sign of $\alpha_{k,k-1}$ is identified by $\Delta_k \mathbb{E}(Y|Z)$ whenever $\Pr(C_k)$ is nonzero. In addition, when $\Delta_k \mathbb{E}(Y|Z) \neq 0$, inequalities (8) and (9) give the lower and upper bounds of $|\alpha_{k,k-1}|$, respectively.

Denote the set characterized by (7)-(9) as $\Theta_k^{\alpha}(\mathbf{P}) \subset \Theta$. In the next Lemma, we derive explicit expressions for $\Theta_k^{\alpha}(\mathbf{P})$ and provide sufficient conditions under which $\Theta_k^{\alpha}(\mathbf{P})$ is a sharp identified set of $\alpha_{k,k-1}$.

[10] The parameter space for $\alpha_{k,k-1}$ may be different for each k. However, we ignore this possibility for notational simplicity.

Lemma 3.2. *Let Assumption 2.1-(ii)-(iv), 2.2-(i) and 2.3 hold. Then, for* $\forall k = 1, 2, ..., K$*:*

(i) If $TV_{(Y,T),k} = 0$*, then* $\Theta_k^{\alpha}(\mathbf{P}) = \Theta$*. Whereas if* $TV_{(Y,T),k} > 0$*, then:*

$$\Theta_k^{\alpha}(\mathbf{P}) = \begin{cases} \left[\frac{\Delta_k \mathbb{E}(Y|Z)}{1 - \sum_{k' \neq k} TV_{(Y,T),k'}}, \frac{\Delta_k \mathbb{E}(Y|Z)}{TV_{(Y,T),k}} \right], & \text{if } \Delta_k \mathbb{E}(Y|Z) > 0, \\ \{0\}, & \text{if } \Delta_k \mathbb{E}(Y|Z) = 0, \\ \left[\frac{\Delta_k \mathbb{E}(Y|Z)}{TV_{(Y,T),k}}, \frac{\Delta_k \mathbb{E}(Y|Z)}{1 - \sum_{k' \neq k} TV_{(Y,T),k'}} \right], & \text{if } \Delta_k \mathbb{E}(Y|Z) < 0; \end{cases} \tag{10}$$

(ii) If $\max_{0 \leq m \leq K} TV_{(Y,T),m} = 0$*, then* $\Theta_k^{\alpha}(\mathbf{P}) = \Theta$ *is the sharp identified set of* $\alpha_{k,k-1}$*. Whereas, if* $TV_{(Y,T),k} > 0$ *and* $TV_{(Y,T),k'} = 0$ *for all* $k' \neq k$*, then* $\Theta_k^{\alpha}(\mathbf{P})$ *in (10) is the sharp identified set of* $\alpha_{k,k-1}$*.*

Proof of Lemma 3.2. See Appendix C.2. □

Lemma 3.2 (i) shows that, if $TV_{(Y,T),k} = 0$, then no useful information about how the instrument's value changing from z_{k-1} to z_k affects the treatment can be extracted from the observable data. If this is the case, then we fail to exclude any values from the parameter space of the LATE, Θ. Once $TV_{(Y,T),k} > 0$, the instrument has nontrivial identification power, and analytic bounds can be derived for the LATE. Importantly, $\Theta_k^{\alpha}(\mathbf{P})$ can be seen as an informative bound for LATE, because it excludes the intention-to-treat effect $\Delta_k \mathbb{E}(Y|Z)$ and the naive Wald estimand $\Delta_k \mathbb{E}(Y|Z)/\Delta_k \mathbb{E}(T|Z)$,[11] which are the two trivial bounds of the LATE. Moreover, if the TV distance is nonzero only when Z changes from z_{k-1} to z_k, then $\Theta_k^{\alpha}(\mathbf{P})$ is the sharp identified set and it reduces to the identified set of Ura (2018). This is intuitive because $TV_{(Y,T),k'} = 0$ for all $k' \neq k$ implies that the multiple TV distances generated from the discrete IV(s) are essentially equivalent to that generated from a binary IV.

Bounds of the LATMs. We are interested in bounding Δp_k because it plays a crucial role in connecting α^{Mis} to the object of interest, α^{IV}. Similar arguments for (7)-(9) can be applied to obtain the inequalities (11)-(13) below, satisfied by each Δp_k:

$$\Delta p_k \Delta_k \mathbb{E}(T|Z) \geq 0, \tag{11}$$

$$|\Delta_k \mathbb{E}(T|Z)| \leq |\Delta p_k| \left[1 - \sum_{k' \neq k} TV_{(Y,T),k'} \right], \tag{12}$$

$$|\Delta_k \mathbb{E}(T|Z)| \geq |\Delta p_k| TV_{(Y,T),k}. \tag{13}$$

Denote the set characterized by (11)-(13) as $\Theta_k^{p}(\mathbf{P})$. The Lemma below gives analytic bounds of Δp_k, as well as sufficient conditions for the sharp identified set.

Lemma 3.3. *Let Assumption 2.1-(ii)-(iv), 2.2-(i) and 2.3 hold. For* $\forall k = 1, 2, ..., K$*,*

[11] See Lemma C.1 in Appendix C.2.

(i) If $TV_{(Y,T),k} = 0$*, then* $\Theta_k^p(\mathbf{P}) = [-1,1]$*. Whereas, if* $TV_{(Y,T),k} > 0$*, then:*

$$\Theta_k^p(\mathbf{P}) = \begin{cases} \left[\frac{\Delta_k \mathbb{E}(T|Z)}{1-\sum_{k'\neq k} TV_{(Y,T),k'}}, \frac{\Delta_k \mathbb{E}(T|Z)}{TV_{(Y,T),k}}\right], & \text{if } \Delta_k \mathbb{E}(T|Z) > 0, \\ \{0\}, & \text{if } \Delta_k \mathbb{E}(T|Z) = 0, \\ \left[\frac{\Delta_k \mathbb{E}(T|Z)}{TV_{(Y,T),k}}, \frac{\Delta_k \mathbb{E}(T|Z)}{1-\sum_{k'\neq k} TV_{(Y,T),k'}}\right], & \text{if } \Delta_k \mathbb{E}(T|Z) < 0; \end{cases} \tag{14}$$

(ii) If $\max_{0\leq m\leq K} TV_{(Y,T),m} = 0$*, then* $\Theta_k^p(\mathbf{P}) = [-1,1]$ *is the sharp identified set of* Δp_k*. Whereas, if* $TV_{(Y,T),k} > 0$ *and* $TV_{(Y,T),k'} = 0$ *for all* $k' \neq k$*, then* $\Theta_k^p(\mathbf{P})$ *in* (14) *is the sharp identified set of* Δp_k*.*

Proof of Lemma 3.3. See Appendix C.3. □

As shown in Lemma 3.3, the sign and an analytic bound for Δp_k can be obtained as long as $TV_{(Y,T),k} > 0$.[12] It is also clear that, in order to partially identify Δp_k, we do not need any prior or external information about how severely the treatment proxy T is contaminated by measurement error.

When Bounds of the LATEs and LATMs are Sharp. Lemma 3.2 and 3.3 establish that the bounds of the LATE ($\alpha_{k,k-1}$) and LATM (Δp_k) are sharp if the TV distance is nonzero only when Z changes from z_{k-1} to z_k. For more general cases where more than one TV distances are nonzero, the sharpness result is not established. However, in this scenario, the outer sets $\Theta_k^\alpha(\mathbf{P})$ and $\Theta_k^p(\mathbf{P})$ still possess desirable properties. First, the bound for LATE excludes the intention-to-treat effect and the Wald estimand. Second, the bound for $\alpha_{k,k-1}$ is tighter than the bound provided by Ura (2018) using only two values z_{k-1} and z_k of Z. Third, since an outer set is always a superset of the sharp identified set, inference based on the outer set is conservative yet valid. Finally, if $TV_{(Y,T),k} = 1 - \sum_{k'\neq k} TV_{(Y,T),k'}$, then $\Theta_k^\alpha(\mathbf{P})$ and $\Theta_k^p(\mathbf{P})$ reduce to a point so that $\alpha_{k,k-1}$ and Δp_k are both point identified. This is the case if two conditions are satisfied simultaneously (see the proof of Lemma 3.1): (i) there is no misclassification ($T_1 = 1, T_0 = 0$), and (ii) there are no always takers and no never takers. The bounds of LATEs and LATMs will be tighter in cases that are "closer" to this extreme case.

3.2 Bounds of α^{IV}

Recall that the estimand α^{IV} is a weighted average of LATEs with nonnegative weights summing up to one. Hence, our first partial identification result of α^{IV} can be obtained from the union of bounds of LATEs given in Lemma 3.2.

Theorem 3.1 (First Strategy). *Let Assumption 2.1, 2.2, and 2.3 hold. Denote* $\Theta^\alpha(\mathbf{P}) = \bigcup_{k\in\{1,2,\ldots,K\}} \Theta_k^\alpha(\mathbf{P})$*. Then we have* $\alpha^{IV} \in \Theta^\alpha(\mathbf{P})$*.*

Proof of Theorem 3.1. See Appendix C.4. □

Note that the set $\Theta^\alpha(\mathbf{P})$ might be uninformative about the sign of α^{IV} if two LATEs, say $\alpha_{k,k-1}$ and $\alpha_{k',k'-1}$, have opposite signs. Fortunately, because we can recover the sign of all the LATEs from the

[12] Since, without loss of generality, we assume $\{z_0, z_1, ..., z_K\}$ follow the ascending order which can be identified by Corollary 2.2, then it is clear that $\Delta_k \mathbb{E}(T|Z) \geq 0$ for all k. For the sake of completeness, in Lemma 3.3 we still present the result for the case $\Delta_k \mathbb{E}(T|Z) < 0$.

observed data (Lemma 3.2), we are able to recover the sign of α^{IV} as long as all the LATEs stand on the same side of zero.

Our second strategy is built upon the relation between α^{IV} and α^{Mis}, and the bounds of LATMs. Recall from Corollary 2.1 that $\alpha^{IV} = \xi\alpha^{Mis}$, where $\xi = \sum_{k=1}^{K} \gamma_k^{IV} \Delta p_k$. Again, due to the nonnegative weights, γ_k^{IV}, summing up to one, our second partial identification result of α^{IV} is characterized using the identifiable estimand α^{Mis} and the union of bounds of LATMs.

Theorem 3.2 (Second Strategy). *Let Assumption 2.1, 2.2, and 2.3 hold. Denote* $\Theta^p(\mathbf{P}) = \left\{\alpha^{Mis} \times \Delta p : \ \Delta p \in \bigcup_{k=1,2,\ldots,K} \Theta_k^p(\mathbf{P})\right\}$*, where* Δp *represents any generic value in the union* $\bigcup_{k=1,2,\ldots,K} \Theta_k^p(\mathbf{P})$*. Then we have* $\alpha^{IV} \in \Theta^p(\mathbf{P})$*.*

Proof of Theorem 3.2. See Appendix C.5. □

If $\alpha^{Mis} = 0$, α^{IV} is point identified as zero. In addition, we can identify the sign of α^{IV} as long as all the LATMs stand on the same side of zero, which is actually guaranteed by Assumption 2.3.[13] The two partial identification strategies make distinct contributions to the identification of α^{IV} because they are based on different sources of information. We discuss and compare their relative performance in Appendix C.8.

When Bounds of α^{IV} are Sharp. For both strategies, we can distinguish two cases. First, if the instrument is binary, α^{IV} is just the LATE and α^{Mis} reduces to:

$$\alpha^{Mis} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[T|Z=1] - \mathbb{E}[T|Z=0]} = \frac{\mathbb{E}[Y_1 - Y_0|D_1 = 1, D_0 = 0]}{p_1 - p_0}.$$

Then, the bounds of α^{IV} in Theorem 3.1 and 3.2 will be identical, sharp, and also coincide with the sharp identified set of LATE in Ura (2018). This is because $K = 1$ and $\sum_{k' \neq k} TV_{(Y,T),k'}$ degenerates to zero. Second, if the instrument(s) are discrete or multiple-discrete, and more than one total variation distances are nonzero, the sharpness of the bounds of LATEs and LATMs are not established. This means that the bounds of α^{IV} in the above two strategies may not be sharp identified sets but are valid outer sets. The outer set is still considered useful in practice, since it may be sufficient to answer important empirical questions, such as whether the treatment effect is positive or negative, and what is the possible range of program benefits (see Molinari, 2020, for more details). As discussed above, when all the LATEs and all the LATMs have the same sign, our outer sets $\Theta^{\alpha}(\mathbf{P})$ and $\Theta^{p}(\mathbf{P})$ are meaningful in revealing the sign of α^{IV} and its possible range.

3.3 Bounds of α^{IV} Using External Information

Administrative records of program receipts are not easily accessible to all researchers. Hence, when working on a standard survey dataset, often times we cannot know whose individual treatment is misclassified. However, an increasing number of studies report the average extent of misreporting for a wide range of programs. This information often comes in the form of possible values of average false negative probability (w^n) and/or false positive probability (w^p) in the sample. Validation studies or repeated measurements of the same individual can also provide valuable information. Suppose the practitioner has some prior or

[13] For both Strategy 1 and 2, we provide analytical expressions for the bounds of α^{IV} when all LATEs or all LATMs have the same sign in Appendix C.7.

external information about the range of $\xi = 1 - w^n - w^p$. Then we can utilize this information to further tighten the bounds using the fact that $\alpha^{IV} = \xi\alpha^{Mis}$.

Theorem 3.3 (**Third Strategy**). *Let Assumption 2.1 and 2.2 hold. Suppose there exist two known constants* $\underline{\xi} \leq \overline{\xi}$ *and* $\underline{\xi}, \overline{\xi} \in [0,1]$, *such that* $\underline{\xi} \leq \xi \leq \overline{\xi}$.

(i) If $\alpha^{Mis} \geq 0$, *denote* $\Theta^{\xi}(\mathbf{P}) = \left[\underline{\xi}\alpha^{Mis}, \overline{\xi}\alpha^{Mis}\right]$. *Then,* $\alpha^{IV} \geq 0$ *and* $\alpha^{IV} \in \Theta^{\xi}(\mathbf{P})$.

(ii) If $\alpha^{Mis} < 0$, *denote* $\Theta^{\xi}(\mathbf{P}) = \left[\overline{\xi}\alpha^{Mis}, \underline{\xi}\alpha^{Mis}\right]$. *Then,* $\alpha^{IV} < 0$ *and* $\alpha^{IV} \in \Theta^{\xi}(\mathbf{P})$.

Proof of Theorem 3.3. See Appendix C.6. □

The constants $\underline{\xi}$ and $\overline{\xi}$ are two bounds of the weighted average of LATMs, ξ. If researchers are certain that every value in the interval $[\underline{\xi}, \overline{\xi}]$ can be the true value of ξ, then $\Theta^{\xi}(\mathbf{P})$ is the sharp identified set of α^{IV}. By using these extra information, the set $\Theta^{\xi}(\mathbf{P})$ will be at least as good as that in Theorem 3.2 (the second strategy). This is because one could always set $\underline{\xi}$ and $\overline{\xi}$ as the ending points of $\bigcup_{k=1,2,\ldots,K} \Theta_k^p(\mathbf{P})$. Therefore, compared to the first two partial identification strategies, which are based purely on the observable data, by following our third strategy one can further tighten the bounds of α^{IV} and obtain (potentially) significant improvements.[14] In empirical illustration section, we provide guidance regarding how to incorporate external information about the misclassification error to find $[\underline{\xi}, \overline{\xi}]$.

Point estimate. From Theorem 3.3, two sets of conditions suffice to obtain tighter bounds. Firstly, having $\underline{\xi}$ close to 1 means less overall misclassification. At the extreme, when $\underline{\xi} = 1$, we have no misclassification error at all ($w^n = w^p = 0$), hence we can achieve point identification of $\alpha^{IV} = \alpha^{Mis}$. Secondly, having $\underline{\xi}$ and ξ close to each other indicates more accurate knowledge of the overall misclassification probabilities, which produces a narrower bound as well. At the extreme, when $\underline{\xi} = \overline{\xi} = \xi$, we can also achieve point identification of $\alpha^{IV} = \xi\alpha^{Mis}$. Notice that, in application, the constants $\underline{\xi}$ and $\overline{\xi}$ are going to be two approximations of the bounds of the misclassification probabilities. Hence, if the practitioner set $\underline{\xi} = \overline{\xi} = \xi$, the point estimate $\xi\alpha^{Mis}$ is likely biased with respect to α^{IV}, unless ξ is the exact value of misclassification rate. If $\underline{\xi}$ and $\overline{\xi}$ are credible approximations, then our approach, in general, can be used as a bias reduction method with respect to a naïve IV estimator, α^{Mis}.

4 Empirical Illustration

In this section, we utilize our dedicated Stata command, `ivbounds`, to investigate the impact of participating in the 401(k) pension plan on savings in the U.S. To conduct statistical inference, we construct confidence intervals for the bounds using a two-step bootstrap procedure proposed by Chernozhukov, Chetverikov, and Kato (2019). This procedure guarantees both uniform and asymptotic size control, which are crucial for meaningful empirical analyses. For the sake of clarity and due to space constraints, we leave detailed explanations of the inference procedure to a compendium paper (Lin, Tommasi, and Zhang, 2021), and Monte Carlo simulations that verify the finite sample performance of our partial identification strategies to Appendix F.

[14] Since the bounds of α^{IV} obtained under different partial identification strategies can be different, in principle, we can intersect $\Theta^{\alpha}(\mathbf{P})$, $\Theta^{p}(\mathbf{P})$ and $\Theta^{\xi}(\mathbf{P})$ (if available) to take advantage of all identifying information and achieve a tighter bound: $\alpha^{IV} \in \Theta^{\alpha}(\mathbf{P}) \bigcap \Theta^{p}(\mathbf{P}) \bigcap \Theta^{\xi}(\mathbf{P})$. For intersection bounds, it has been noticed that their sample analog estimators are systematically biased (e.g. Manski and Pepper, 2000, 2009; Kreider and Pepper, 2007; Chernozhukov et al., 2013). We defer a comprehensive investigation of this issue to future research.

4.1 The effect of the 401(k) pension plan on savings

The 401(k) pension plan is one of the most popular defined contribution (DC) retirement plans in the US. It aims at increasing financial savings through the tax deductibility of contributions to retirement accounts. We study its effects using data from the SIPP round of 1991 following the construction Abadie (2003). The resulting sample size is 9,275. Table G.1 in appendix reports the summary statistics of the main variables used in the empirical analysis. Although the effects of this plan have been examined elsewhere (e.g., Abadie, 2003; Ura, 2018), the application contains all the ingredients to demonstrate the full extent of the usefulness of the approach proposed in our paper.

First, the participation to the program is binary and notoriously misreported in survey data. Second, the eligibility to the pension plan, which is provided only to workers in firms offering the plan, is arguably a valid instrument (e.g., Poterba et al., 1995). Third, the eligibility can be interacted with the year of introduction of the plan, which yields a discrete instrument that accounts for the duration of the exposure to the plan.[15] Fourth, credible information on treatment misclassification probabilities are available from the literature and can be incorporated in estimation. Fifth, although our main theoretical results hold without covariates, including covariates is almost always crucial in application. In our specific context, for the instrument to be valid, it is really important to condition on family income and age. Hence, this specific application gives us also the opportunity to show the performance of our proposed suggestions to incorporate covariates in a realistic context. Finally, the fact that the application is well known makes it easier for us to evaluate our results in light of the existing literature.

In measuring the benefits of the 401(k) pension plan, a researcher would face two main difficulties: endogenous participation in the plan and misreporting of participation. The first problem may arise due to unobserved differences in saving behaviors. That is, participants in the plan might save more in general than those who do not participate. Hence, a comparison of accumulated financial assets between participants and nonparticipants is likely to yield a positive bias of the true effect of the program. If this was the only problem in the data, a practitioner could just use the eligibility to the plan as a valid instrument and perform inference on the causal parameter as in Abadie (2003). However, the contemporaneous presence of the second problem makes the task difficult. Misreporting in this context may arise because individuals find it difficult to remember or understand their pension plan, leading to reporting error. Indeed, Dushi and Iams (2010) documented that in the SIPP, over 17% of participants in the 401(k) pension plan self-report as nonparticipants (false negative) and almost 10% of nonparticipants self-report as participants (false positive). Understanding plan benefits is relevant for the economic well-being of future retirees because these plans are important for retirement income security. This is the economic motivation underlying our efforts.

[15] In both the binary and discrete instrument cases, our empirical analysis relies on the assumption of independence between the instrument and misclassification errors (Assumption 2.2-(i)). In our specific context, the binary instrument reflects the firm's decision to introduce the 401(k) plan, while the treatment signifies the worker's choice to participate in the same plan. To justify the validity of Assumption 2.2-(i), we must assume that the misreporting of pension plan participation by workers, influenced by their awareness (or lack thereof) of the plan, is reasonably uncorrelated with the firm's decision, after adjusting for workers' income levels. Similarly, the discrete instrument relies on data concerning the introduction of the new pension plan. To justify the validity of Assumption 2.2-(i) in this case, we need to assume that, given the worker's age, misreporting is reasonably unrelated to their knowledge of the pension plan. However, we acknowledge a valid concern raised by a referee: in our context, treatment misclassification rates may vary across different values of the instrument (Haider and Stephens Jr., 2019; Possebom, 2021). As such, we caution against interpreting our findings as definitive causal effects. Instead, our empirical application primarily serves to illustrate a novel theoretical tool applied in a real-world setting.

4.2 Results

Given the three main theoretical contributions of the paper, we aim to answer the following questions: (i) What is the likely bias of the estimated program benefits if we do not account for treatment misclassification? (ii) In case of a binary instrument, how do the bounds of the program benefits shrink by incorporating external information on misclassification probabilities? How do they compare with the results of the existing literature? (iii) In case of a discrete instrument, how are the bounds compared to a naive approach that does not account for treatment misclassification? How do these bounds shrink by incorporating external information on misclassification probabilities? We provide comprehensive answers to each of these questions.

To begin with, researchers can employ the available information on treatment misclassification probabilities and utilize our newly established relationship between the true and mismeasured treatment effect, as represented by Equation (5). By doing so, they can approximate the potential level of biases in the benefits of the 401(k) plan. In our specific case, assuming $w^n \approx 17\%$ and $w^p \approx 10\%$, we conclude that the estimated (mismeasured) treatment effect reported in the literature is likely biased (upward) by approximately $\frac{\alpha^{Mis}-\alpha^{IV}}{\alpha^{IV}} = \frac{w^n+w^p}{1-w^n-w^p} = 37\%$.

We then proceed by estimating the program benefits using a binary instrument, the eligibility to the pension plan. In this case, our target parameter is the unconditional IV estimand $\alpha^{IV} := \mathbb{E}[Y_1 - Y_0|C_1] = \mathbb{E}\{\mathbb{E}[Y_1 - Y_0|X, C_1]|C_1\} = \frac{\mathbb{E}[\mathbb{E}(Y|X,Z=1)-\mathbb{E}(Y|X,Z=0)]}{\mathbb{E}[\mathbb{E}(D|X,Z=1)-\mathbb{E}(D|X,Z=0)]}$,[16] where $C_1 = \{D_0 = 0, D_1 = 1\}$ and covariates in X include family income, age, age squared, marital status, and family size. Panel A of Table 2 reports the results. Column (1) reports the conventional 2SLS estimate (assuming homogenous treatment effect) as shown in Column (3) of Table 2 by Abadie (2003). This represents a biased point estimate because it ignores the potential treatment misclassification. The effect is statistically significant and says that participating in the 401(k) plan increases the total financial assets by roughly $9,400, with a 95% confidence interval of $5,300–13,500. Column (2) reports the nonparametric estimate of the mismeasured treatment effect, $\alpha^{Mis} := \frac{\mathbb{E}[\mathbb{E}(Y|X,Z=1)-\mathbb{E}(Y|X,Z=0)]}{\mathbb{E}[\mathbb{E}(T|X,Z=1)-\mathbb{E}(T|X,Z=0)]}$ to incorporate covariates and treatment effect heterogeneity, while ignoring the potential treatment misclassification.[17] Next, Column (3) displays the 95% CI for the bound of unconditional IV estimand α^{IV} from Ura (2018) which accounts for both treatment effect heterogeneity and treatment misclassification error. This is our benchmark result from the literature, to which we compare the performance of our partial identification strategies.

Columns (4)–(8) report the 95% CI of our partial identification strategies under different assumptions about the availability of misclassification probabilities. The estimation error of the nuisance parameters $\pi(X) = \Pr(Z = 1|X)$ and α^{Mis} are taken into account following the inference process presented in Lin, Tommasi, and Zhang (2021), where their confidence intervals are obtained by nonparametric bootstrapping. Column (4) assumes no information about the misclassification probabilities is available. Since the IV is binary, Strategy 1 and 2 coincide and are equivalent to the method developed by Ura (2018).[18]

Whereas, Columns (5)-(8) display the estimates of the bound for α^{IV} under Strategy 3, $\hat{\alpha}^{Mis} * [\underline{\xi}, \overline{\xi}]$

[16] The last equality follows from Theorem 1 of Frölich (2007). See more detailed explanations in Appendix E.1.

[17] In the binary IV case, Abadie (2003) and Frölich (2007) show that, for $\pi(X) := \Pr(Z = 1|X)$, $\mathbb{E}\left[\frac{Z-\pi(X)}{\pi(X)(1-\pi(X))}Q\right] = \mathbb{E}[\mathbb{E}(Q|X, Z = 1) - \mathbb{E}(Q|X, Z = 0)]$ holds for any random variable Q. Thus, $\hat{\alpha}^{Mis}$ in Panel A Column (2) is calculated as the sample analogue of $\mathbb{E}\left[\frac{Z-\pi(X)}{\pi(X)(1-\pi(X))}Y\right] / \mathbb{E}\left[\frac{Z-\pi(X)}{\pi(X)(1-\pi(X))}T\right]$, where $\pi(X)$ is estimated via a linear probability model. The confidence interval of $\hat{\alpha}^{Mis}$ is computed using a nonparametric bootstrap method.

[18] In the binary IV case with covariates, the computation of our CI using Strategy 1 and 2 is identical to that described in Section 4 of Ura (2018). The differences between the results in Column (3) and (4) in Panel A arise from different random samples in the two-step multiplier bootstrap method.

Table 2: Empirical Illustration

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Binary instrument									
		α^{Mis}		**Target parameter: unconditional LATE α^{IV}**					
		2SLS Abadie (2003)	nonpara.	Ura (2018)	Strategy 1 ≡ 2	Strategy 3			
						appr. w^n	appr. w^p	bounds w^n and w^p	appr. w^n and w^p
	est.	9.4	16.3			(10.8, 13.5)	(6.5, 13.0)	(11.9, 13.0)	11.9
	95% CI	(5.3, 13.5)	(7.1, 25.5)	(4.4, 28.3)	(4.3, 27.8)	(4.7, 21.2)	(2.8, 20.4)	(5.2, 20.4)	(5.2, 18.6)
Panel B: Discrete instrument									
			$\alpha^{Mis}(e \in A_s)$	**Target parameter: conditional WLATE $\alpha^{IV}(e \in A_s)$**					
			nonpara.	Strategy 1	Strategy 2	Strategy 3			
						appr. w^n	appr. w^p	bounds w^n and w^p	appr. w^n and w^p
Stratum 1		est.	21.8			(14.4, 18.1)	(8.7, 17.4)	(15.9, 17.4)	15.9
		95% CI	(16.7, 27.6)	(2.5, 42.4)	(2.9, 29.4)	(11.0, 23.0)	(6.7, 22.1)	(12.2, 22.1)	(12.2, 20.2)
Stratum 2		est.	23.1			(15.2, 19.2)	(9.2, 18.5)	(16.9, 18.5)	16.9
		95% CI	(19.2, 27.0)	(2.3, 70.1)	(4.6, 28.2)	(12.7, 22.4)	(7.7, 21.6)	(14.0, 21.6)	(14.0, 19.7)
Stratum 3		est.	54.5			(36.0, 45.2)	(21.8, 43.6)	(39.8, 43.6)	39.8
		95% CI	(44.9, 64.0)	(19.2, 120.9)	(15.5, 68.2)	(29.6, 53.2)	(17.9, 51.2)	(32.8, 51.2)	(32.8, 46.7)
Panel C: Discrete instrument									
			$\alpha^{Mis}_{k,k-1}(e \in A_s)$	**Target parameter: conditional LATE $\alpha_{k,k-1}(e \in A_s)$**					
			nonpara.	Ura (2018)	Lemma 3.2	Strategy 3			
						appr. w^n	appr. w^p	bounds w^n and w^p	appr. w^n and w^p
Stratum 1	$k=1$	est.	49.4			(32.6, 41.0)	(19.7, 39.5)	(36.0, 39.5)	36.0
		95% CI	(33.6, 65.1)	(3.9, 44.1)	(8.0, 42.4)	(22.2, 54.0)	(13.4, 52.1)	(24.5, 52.1)	(24.5, 47.5)
	$k=2$	est.	11.3			(7.5, 9.4)	(4.5, 9.0)	(8.2, 9.0)	8.2
		95% CI	(3.6, 19.1)	(1.0, 22.4)	(2.5, 19.4)	(2.4, 15.8)	(1.4, 15.3)	(2.6, 15.3)	(2.6, 13.9)
Stratum 2	$k=1$	est.	63.6			(42.0, 52.8)	(25.5, 50.9)	(46.5, 50.9)	46.5
		95% CI	(45.2, 82.1)	(7.1, 75.3)	(12.5, 70.1)	(29.8, 68.1)	(18.1, 65.7)	(33.0, 65.7)	(33.0, 59.9)
	$k=2$	est.	12.7			(8.4, 10.5)	(5.1, 10.2)	(9.3, 10.2)	9.3
		95% CI	(6.4, 19.0)	(1.9, 21.7)	(2.3, 22.0)	(4.2, 15.8)	(2.6, 15.2)	(4.7, 15.2)	(4.7, 13.9)
Stratum 3	$k=1$	est.	132.8			(87.7, 110.2)	(53.1, 106.3)	(97.0, 106.3)	97.0
		95% CI	(99.6, 166.1)	(17.9, 113.8)	(40.6, 120.9)	(65.7, 137.7)	(39.8, 132.9)	(72.7, 132.9)	(72.7, 121.3)
	$k=2$	est.	43.0			(28.4, 35.7)	(17.2, 34.4)	(31.4, 34.4)	31.4
		95% CI	(30.6, 55.5)	(14.1, 61.5)	(19.2, 59.7)	(20.2, 46.1)	(12.2, 44.4)	(22.3, 44.4)	(22.3, 40.5)

Notes: Results in this Table are in 1,000$ units. Point or bound estimate is in the first row and 95% CI is in the second row. Panel A reports the results using a binary IV. Panel B and C report illustrative results using a discrete IV. In Panel A, Column (1) reports the conventional 2SLS estimate as shown in Column (3) of Table 2 by Abadie (2003). Column (2) reports the nonparametric estimate of α^{Mis} taking unobserved heterogeneity into account. Column (3) reports the 95% CI of the LATE as shown in Table 2 by Ura (2018). Columns (4)-(7) report the results of our partial identification strategies under different assumptions regarding the misclassification probabilities. Finally, Column (8) delivers a point estimate of the effect and its 95% CI. In Panel B, our target parameter is the conditional IV estimand $\alpha^{IV}(e \in A_s)$ for each stratum A_s and $s = 1, 2, 3$. We stratify the samples into three strata based on their estimated $e = e(X) = \Pr(T = 1|X)$. Column (2) reports the nonparametric estimates of $\alpha^{Mis}(e \in A_s)$. Columns (3)-(7) report the results using our partial identification strategies under different assumptions regarding the misclassification probabilities. Finally, Column (8) delivers a point estimate of the effect and its 95% CI. In Panel C, our target parameter is the conditional LATE estimand $\alpha_{k,k-1}(e \in A_s)$ with $k = 1, 2$ for each stratum A_s and $s = 1, 2, 3$. Column (2) reports the estimates of $\alpha^{Mis}_{k,k-1}(e \in A_s)$ which is the mismeasured LATE ignoring the treatment misclassification. Column (3) reports the 95% CI of $\alpha_{k,k-1}(e \in A_s)$ obtained by applying the method of Ura (2018) to the subsample with IV taking two values, $k-1, k$. Column (4) reports the results obtained by applying our Lemma 3.2 to the whole sample with IV taking all three values. Columns (5)-(7) deliver partial identification results of $\alpha_{k,k-1}(e \in A_s)$ using our Strategy 3 under different assumptions regarding the misclassification probabilities. Finally, Column (8) presents a point estimate of the LATE and its 95% CI.

with $\hat{\alpha}^{Mis}$ from Column (2), and its 95% CIs, using external information about the misclassification probabilities. In particular, we take into account the fact that $\hat{\alpha}^{Mis}$ is likely upward biased by approximately 37% (see Column (8)). We impose two common restrictions: (i) the probability of false positive w^p and

false negative w^n are both less than 50%,[19] and (ii) people are more prone to under-report rather than over-report due to the incomplete awareness so that $w^p \leq w^n$. We consider four cases:

- Case 1. Column (5) assumes that we know only an approximation of the probability of false negative ($w^n = 17\%$). Then, $\xi \leq 1 - w^n$ because w^p is nonnegative, and $\xi \geq 1 - 2w^n$ because $w^p \leq w^n$. In this case, we can set $\xi \in [1 - 2w^n, 1 - w^n] = [66\%, 83\%]$.

- Case 2. Column (6) assumes that we know only an approximation of the probability of false positive ($w^p = 10\%$). Then, $\xi \leq 1 - 2w^p$ because $w^p \leq w^n$ and $\xi \geq 0.5 - w^p$ because $w^n \leq 50\%$. In this case, we can set $\xi \in [0.5 - w^p, 1 - 2w^p] = [40\%, 80\%]$.

- Case 3. Column (7) assumes that we know the bound of w^n, where $w^p \leq w^n \leq \overline{w}^n$ with $\overline{w}^n = 17\%$ and $w^p = 10\%$. In this case, $\xi \in [1 - \overline{w}^n - w^p, 1 - 2 * w^p] = [73\%, 80\%]$.

- Case 4. Column (8) assumes that both w^n and w^p are approximately known, as in this application $w^n \approx 17\%$ and $w^p \approx 10\%$ thus $\xi = 1 - w^n - w^p \approx 73\%$. Then, the upward bias is $\frac{\alpha^{Mis} - \alpha^{IV}}{\alpha^{IV}} = \frac{1-\xi}{\xi} \approx 37\%$ and our approach can deliver a point estimate of the effect $\hat{\alpha}^{IV} = \hat{\alpha}^{Mis} * \xi \approx \hat{\alpha}^{Mis} * 73\%$.

As one can see, using our partial identification strategy in the presence of external information, as described in Columns (5) to (7), we obtain bounds of the true benefits of the program that can be between 25% and 36% narrower compared to those depicted in Column (3).[20] Notice that, the point estimate in Column (8) is likely to be biased since we only have approximations of the misclassification probabilities. However, if the true false positive and the true false negative rates are close to 17% and 10%, our estimate is closer to the true (unknown) effect than the value reported in Column (2), which ignores treatment misclassification.[21]

Next, we illustrate the performance of our method when the instrument is discrete by interacting the eligibility for the 401(k) plan and the duration of exposure to the plan. The duration of exposure is defined as how many years one has been exposed to the 401(k) program, which became active in 1981. Those with less than 10 years of exposure were 15 to 24 years old in 1981. Those with at least 10 years of exposure were 25 or older in 1981. The discrete instrument takes the value $Z = 0$ if an individual is not eligible and has been exposed for less than 10 years, $Z = 1$ they are eligible and have less than 10 years of exposure or are ineligible and have at least 10 years of exposure, and $Z = 2$ if they are eligible and have at least 10 years of exposure. Naturally, with this instrument the ascending order requirement is satisfied. Also in this case, the instrumental variable plausibly satisfies the exclusion restriction after conditioning on covariates X.

When the instrument is discrete (or multiple-discrete), we follow the method of stratification matching adopted by Dehejia and Wahba (1999) and Battistin and Sianesi (2011) to incorporate covariates. Denote

[19] This is the sufficient condition for Assumption 2.3. Otherwise the misclassification is too problematic and this dataset should probably be abandoned.

[20] In Panel A, the width of the 95% confidence interval (CI) in Column (3) is 23.9. However, the 95% CI in Column (5) is narrower, with a width of 16.5, representing a 31% reduction. Similarly, the 95% CI in Column (6) has a width of 17.6, which is 25% narrower, while in Column (7), the width is 15.2, reflecting a 36% reduction. In Column (8), where both approximations of w^n and w^p are available, we obtain a point estimate and achieve the most significant reduction in the width of the bounds. These findings suggest that having more accurate external information regarding the probabilities of misclassification leads to more informative bounds on the true benefits.

[21] The reliability of the external information used to approximate the misclassification probabilities is crucial in determining the accuracy of our proposed estimators obtained using Strategy 3. Strategy 3 is advantageous as long as the source of the external information is deemed informative. We thank an anonymous referee for pointing this out.

$e = e(X) = \Pr(T = 1|X)$ as the probability of self-reported participation. We stratify the sample into three strata A_1, A_2, A_3 based on their estimated $e(X)$. Each stratum consists of one-third of the samples, where samples in stratum 1 have the smallest $e(X)$, and samples in stratum 3 have the largest $e(X)$. We proceed with the estimation within each stratum. Our target parameter in Panel B is the conditional IV estimand $\alpha^{IV}(e \in A_s) = \frac{\text{Cov}(Y,Z|e\in A_s)}{\text{Cov}(D,Z|e\in A_s)}$, where A_s represents one of the S user-specified strata.[22] Column (2) of Panel B reports the estimates for $\alpha^{Mis}(e \in A_s) = \frac{\text{Cov}(Y,Z|e\in A_s)}{\text{Cov}(T,Z|e\in A_s)}$, which indicate that, without accounting for treatment misclassification, $\alpha^{Mis}(e \in A_s)$ overestimates the true effect of the program. Indeed, as one can notice, the left-end points of each confidence interval for $\alpha^{IV}(e \in A_s)$ in Columns (3)-(8) are much smaller than the left-end point of the confidence interval of $\alpha^{Mis}(e \in A_s)$ in Column (2). More importantly, in Columns (5)-(8), the right-end points of the confidence interval are also much smaller than the right-end point of the confidence interval and even the point estimate (in most of the cases) of $\alpha^{Mis}(e \in A_s)$ in Column (2). The same results hold for the binary IV case in Panel A. Similar to the binary instrument case, in the discrete instrument case, and for each stratum, the more informative external information is incorporated in estimation, the better is the performance of Strategy 3 in terms of tightest of the bounds.[23]

Our target parameter in Panel C is the conditional LATE estimand in each stratum, defined as: $\alpha_{k,k-1}(e \in A_s) = \frac{\mathbb{E}[Y|Z=z_k,e\in A_s]-\mathbb{E}[Y|Z=z_{k-1},e\in A_s]}{\mathbb{E}[D|Z=z_k,e\in A_s]-\mathbb{E}[D|Z=z_{k-1},e\in A_s]}$ for $k = 1, 2$ and $s = 1, 2, 3$. Column (2) reports the estimates for $\alpha^{Mis}_{k,k-1}(e \in A_s)$, which use the mismeasured treatment T and ignore the treatment misclassification. Column (3) reports the 95% CI of $\alpha_{k,k-1}(e \in A_s)$ obtained by applying the method of Ura (2018) to the subsample with IV taking two values, $k-1$ and k. Column (4) reports the results obtained by applying our Lemma 3.2 to the entire sample. Columns (5)-(8) present the bound or point estimates, $\hat{\alpha}^{Mis}_{k,k-1}(e \in A_s) \times [\underline{\xi}, \xi]$, with $\hat{\alpha}^{Mis}_{k,k-1}(e \in A_s)$ from Column (2) and the values of $\underline{\xi}$ and ξ computed using external information about w^n and w^p as described in Cases 1 to 4 in the binary IV example. When comparing the bounds in Columns (3) and (4), we can see that, first, the lower bounds obtained using our method are much larger than those obtained using the method of Ura (2018), and second, the upper bounds produced by these two methods are comparable. These two findings align with the theoretical results in Lemma 3.2 and provide confirmation of the identification gains achieved by incorporating a discrete IV in bounding individual LATEs.

Finally, two remarks are in order for Panel A and Panel B. First, the intersection bounds from all three strategies give the bounds using Strategy 3 and the approximations of both w^p and w^n in Column (8). This is because it dominates other bounds and thus is the preferred one. Second, we conduct a sensitivity analysis to evaluate the effects of treatment misclassification probabilities across a wide range of values for w^n and w^p, in order to provide more comprehensive illustration of the bounding analysis using our Strategy 3. Due to space limitation, the results are presented in Appendix G.1. In summary, the accuracy of the prior information about w^n and w^p generally results in narrower confidence intervals.

[22] Detailed discussions and results regarding our partial identification strategies with covariates are given in Appendix E. In Appendix E.1, we discuss the reasons for choosing the conditional IV estimand as the target parameter in the case of discrete or multiple-discrete IV(s) with covariates.

[23] Notice that the results in Panel A and B are not directly comparable for two reasons. First, in Panel A we use the full information of covariates, whereas in Panel B we use only discretized information about the covariates via dummies indicating $e(X) \in A_s$. Second, since the instruments in Panel A and B generate different complier groups, the effects for one complier group need not be comparable to effects for others.

5 Conclusion

This paper provides useful tools for applied researchers to study the local average treatment effect (LATE) and the weighted average of LATEs (WLATE) using binary, discrete, or multiple-discrete IVs, when the binary treatment is mismeasured. We demonstrate the limitations of the standard LATE approach and introduce a parameter based on misclassification rates to approximate the potential bias in the estimable program benefits. We find that the mismeasured effect is a weighted average of LATEs with potentially negative weights, resulting in overestimation of the true effect even with non-negative weights. Partial identification results are established for both LATE and WLATE. Moreover, we provide a way to tighten the bounds using external information on misclassification probabilities. Finally, we illustrate the potentials of our approach in the context of the 401(k) pension plan in the US. We conclude that (i) the naive estimates of program benefits (ignoring measurement error) are overestimated approximately by 37%, and (ii) our estimated bounds of the true program benefits can be up to 36% narrower in width than comparable results in the literature.

Our work is positioned between the literature which examines the consequences of using an error-laden proxy on the causal interpretation of commonly used IV estimands (e.g. Chalak, 2017), and the literature which offers a means to learn the possible values of target parameters under treatment misclassification using partial identification strategies (e.g. Ura, 2018). Our method offers at least three applications for practitioners. First, it can be used as the leading identification strategy in any setting where the practitioner knows that the endogenous binary treatment is not well measured. Second, it can be used as the leading robustness check if misreporting is only suspected. Third, it can assess the sensitivity of program benefits under different assumptions of the misclassification probabilities. Although our method is primarily motivated by (and directed to practitioners in) the program evaluation literature, it is not limited to applications within this context. It can be applied to any setting where the endogenous binary treatment is contaminated by endogenous measurement error, and the researcher considers LATE or WLATE the relevant parameter(s) for evaluating the policy change.

References

ABADIE, A. (2003): "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 113, 231–263. [14], [15], [16]

ACERENZA, S. (2021): "Partial Identification of Marginal Treatment Effects with discrete instruments and misreported treatment," *arXiv preprint arXiv:2110.06285*. [3], [5]

ACERENZA, S., K. BAN, AND D. KÉDAGNI (2021): "Marginal Treatment Effects with Misclassified Treatment," Tech. rep. [3], [5]

ATHEY, S. AND G. IMBENS (2017): "Chapter 3 - The econometrics of randomized experiments," in *Handbook of Field Experiments*, ed. by A. V. Banerjee and E. Duflo, North-Holland, vol. 1 of *Handbook of Economic Field Experiments*, 73 – 140. [2]

BATTISTIN, E. AND B. SIANESI (2011): "Misclassified treatment status and treatment effects: An application to returns to education in the United Kingdom," *Review of Economics and Statistics*, 93, 495–509. [3], [7], [17]

BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): "Measurement error in survey data," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, vol. 5, chap. 59, 3705–3843, 1 ed. [5]

CALVI, R., A. LEWBEL, AND D. TOMMASI (2021): "LATE With Missing or Mismeasured Treatment," *Journal of Business & Economic Statistics*. [5], [7]

CHALAK, K. (2017): "Instrumental variables methods with heterogeneity and mismeasured instruments," *Econometric Theory*, 33, 69–104. [2], [19]

CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2019): "Inference on causal and structural parameters using many moment inequalities," *The Review of Economic Studies*, 86, 1867–1900. [13]

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection bounds: Estimation and inference," *Econometrica*, 81, 667–737. [13]

DEHEJIA, R. H. AND S. WAHBA (1999): "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," *Journal of the American statistical Association*, 94, 1053–1062. [17]

DUSHI, I. AND H. M. IAMS (2010): "The impact of response error on participation rates and contributions to defined contribution pension plans," *Social Security Bulletin*, 70, 45—60. [14]

FRAZIS, H. AND M. A. LOEWENSTEIN (2003): "Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables," *Journal of Econometrics*, 117, 151 – 178. [7]

FRÖLICH, M. (2007): "Nonparametric IV estimation of local average treatment effects with covariates," *Journal of Econometrics*, 139, 35–75. [15]

HAIDER, S. AND M. STEPHENS JR. (2019): "Correcting for Misclassified Binary Regressors Using Instrumental Variables," Tech. rep., Working Paper. [5], [14]

HAUSMAN, J., J. ABREVAYA, AND F. SCOTT-MORTON (1998): "Misclassification of the dependent variable in a discrete-response setting," *Journal of Econometrics*, 87, 239 – 269. [7]

IMAI, K. AND T. YAMAMOTO (2010): "Causal Inference with Differential Measurement Error: Nonparametric Identification and Sensitivity Analysis," *American Journal of Political Science*, 54, 543–560. [3]

IMBENS, G. W. AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [2], [4], [5], [9]

KLINE, B. AND E. TAMER (2022): "Recent developments in partial identification," . [3]

KREIDER, B. (2010): "Regression coefficient identification decay in the presence of infrequent classification errors," *The Review of Economics and Statistics*, 92, 1017–1023. [2]

KREIDER, B. AND J. V. PEPPER (2007): "Disability and employment: Reevaluating the evidence in light of reporting errors," *Journal of the American Statistical Association*, 102, 432–441. [3], [13]

KREIDER, B., J. V. PEPPER, C. GUNDERSEN, AND D. JOLLIFFE (2012): "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported," *Journal of the American Statistical Association*, 107, 958–975. [2], [3]

LEWBEL, A. (2007): "Estimation of average treatment effects with misclassification," *Econometrica*, 75, 537–551. [7]

LIN, A., D. TOMMASI, AND L. ZHANG (2021): "Bounding Program Benefits when Participation is Misreported: Estimation and Inference with Stata," Tech. rep., Working paper. [1], [13], [15]

MANSKI, C. F. AND J. V. PEPPER (2000): "Monotone instrumental variables: With an application to the returns to schooling," *Econometrica*, 68, 997–1010. [13]

——— (2009): "More on monotone instrumental variables," *The Econometrics Journal*, 12, S200–S216. [13]

MEYER, B. D., N. MITTAG, AND R. M. GEORGE (2020): "Errors in survey reporting and imputation and their effects on estimates of food stamp program participation," *Journal of Human Resources*, 0818–9704R2. [3]

MEYER, B. D., W. K. C. MOK, AND J. X. SULLIVAN (2015): "Household Surveys in Crisis," *Journal of Economic Perspectives*, 29, 199–226. [2]

MILLIMET, D. (2011): "The elephant in the corner: a cautionary tale about measurement error in treatment effects models," in *Missing Data Methods: Cross-Sectional Methods and Applications. In: Advances in Econometrics*, Emerald Group Publishing Limited, vol. 27, 1–39, 1 ed. [2]

MOGSTAD, M., A. SANTOS, AND A. TORGOVITSKY (2018): "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 86, 1589–1619. [3]

MOGSTAD, M., A. TORGOVITSKY, AND C. R. WALTERS (2020): "Policy evaluation with multiple instrumental variables," Tech. rep., National Bureau of Economic Research. [3], [4]

——— (2021): "The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables," *American Economic Review*, 111, 3663–98. [2], [4]

MOLINARI, F. (2010): "Missing Treatments," *Journal of Business & Economic Statistics*, 28, 82–95. [3]

——— (2020): "Microeconometrics with partial identification," *Handbook of econometrics*, 7, 355–486. [12]

NGUIMKEU, P., A. DENTEH, AND R. TCHERNIS (2018): "On the estimation of treatment effects with endogenous misreporting," *Journal of Econometrics*. [2]

POSSEBOM, V. (2021): "Crime and Mismeasured Punishment: Marginal Treatment Effect with Misclassification," . [3], [5], [14]

POTERBA, J. M., S. F. VENTI, AND D. A. WISE (1995): "Do 401(k) contributions crowd out other personal saving?" *Journal of Public Economics*, 58, 1 – 32. [14]

STEPHENS JR, M. AND T. UNAYAMA (2019): "Estimating the impacts of program benefits: Using instrumental variables with underreported and imputed data," *Review of Economics and Statistics*, 101, 468–475. [7]

TOMMASI, D. AND L. ZHANG (2022): "Identifying Program Benefits When Participation Is Misreported," *IZA Discussion Paper*. [5]

URA, T. (2018): "Heterogeneous treatment effects with mismeasured endogenous treatment," *Quantitative Economics*, 9, 1335–1370. [2], [3], [5], [9], [10], [11], [12], [14], [15], [16], [18], [19]