Correcting for Misclassified Binary Regressors Using Instrumental Variables

Steven J. Haider*
Michigan State University
and
Melvin Stephens Jr.
University of Michigan and NBER

September 5, 2022

Abstract

Estimators that exploit an instrumental variable to correct for misclassification in a binary regressor typically assume that the misclassification rates are invariant across all values of the instrument. We show this assumption is invalid in routine empirical settings. We derive a new estimator which allows misclassification rates to vary across values of the instrumental variable. Our key identifying assumption, that the sum of misclassification rates remains constant across instrument values, follows from the empirical examples we present. We also show this assumption can be relaxed using moment inequalities that arise from our model. We demonstrate the usefulness of our estimator through Monte Carlo simulations and a re-analysis of the extent to which Medicaid eligibility crowds out other forms of health insurance. Correcting for measurement error substantially reduces estimates of crowd out and the extent to which Medicaid eligibility lowers the share of the uninsured.

Keywords: Measurement Error, Instrumental Variables, Medicaid

*We would like to thank Dan Black, Andreas Hagemann, Gary Solon, Jeff Wooldridge, and seminar participants at the University of Chicago Harris School, the University of Michigan Labor Lunch, Michigan State University, the Society of Labor Economists meetings, and the Michigan-Michigan State-Western University Labo(u)r Day Conference for helpful comments and suggestions. We thank Thomas Brosy, Cody Orr, and Brenden Timpe for excellent research assistance.

1 Introduction

It has long been recognized that measurement error is pervasive in applied economic research (e.g., Bound, Brown, and Mathiowetz 2001). It is well-known that standard instrumental variables (IV) estimation yields a consistent estimate when the measurement error in a regressor is "classical," i.e., uncorrelated with the true regressor. This approach does not work with a misclassified binary regressor because the measurement error is negatively correlated with the true regressor. Numerous methods are available to consistently estimate the impact of a misclassified binary regressor, many of which make use of an instrumental variable.[1]

A key identifying assumption in this literature is that the conditional probabilities of misclassifying the binary regressor, the "misclassification rates," are constant across values of the instrumental variable.[2] This seemingly innocuous assumption is, in fact, rather strong. For example, in the literature that examines the impact of Medicaid (e.g., Currie and Gruber 1996; Cutler and Gruber 1996; Gross and Notowidigdo 2011), the observed binary indicator for Medicaid eligibility is computed (primarily) by determining whether measured household income is below a given threshold (e.g., a state-specific threshold that depends on family size and structure) and changes in the threshold over time are used to construct an instrumental variable. Intuitively, those with true income that is relatively close to the threshold are more likely to have measured income fall on the wrong side of the threshold and, consequently, have their eligibility be misclassified. Because the share of households close to the threshold can vary as the threshold moves through the income distribution, the misclassification rate

[1] For example, see Card 1996; Kane, Rouse, and Staiger 1999; Black, Berger, and Scott 2000; Frazis and Lowenstein 2003; Mahajan 2006; Lewbel 2007; Hu 2008; Chen, Hu, and Lewbel 2008a,b; Battistin, De Nadai, and Sianesi 2014; DiTraglia and Garcia-Jimeno 2019; Calvi, Lewbel, and Tommasi 2017; and Yanagi 2019.

[2] Specifically, the misclassification rates are the probabilities of misclassifying the binary regressor conditional on the true value of the regressor.

of Medicaid eligibility can vary across values of the instrumental variable.

We refer to misclassification rates that vary with the instrumental variable as "varying misclassification" in contrast to the standard "fixed misclassification" assumption.[3] Solutions leveraging instruments for identification that assume fixed misclassification are inconsistent when binary regressors suffer from varying misclassification. In addition, under fixed misclassification, standard IV estimation overestimates the parameter of interest leading prior researchers to suggest using the IV estimator as an upper bound (Kane, Rouse, and Staiger 1999; Black, Berger, and Scott 2000).[4] In contrast, we show that with varying misclassification IV estimation can either overestimate or underestimate the parameter of interest. Because prior estimators are inconsistent and IV estimation does not yield an upper bound, it is important to develop new estimation strategies that account for varying misclassification.

In this paper, we present a method to consistently estimate the impact of a misclassified binary regressor when the misclassification rates vary across instrument values. Identification requires a discrete instrumental variable that takes on three or more values. Our key identifying assumption is that the sum of the misclassification rates is fixed, while allowing the underlying misclassification rates to vary with the instrument values. We present evidence from multiple empirical settings that illustrate misclassification rates do indeed vary while satisfying this identifying assumption. We also show how to relax this key assumption

[3]Although some earlier papers allow misclassification rates to vary with other observable characteristics, they do not allow the misclassification rates to vary with the instrumental variable. Recent exceptions are Ura (2018) and Yanagi (2019), which we discuss below. Additionally, our estimator can allow for misclassification that varies both with other observable characteristics *and* the instrumental variable.

[4]Ura (2018) develops an alternative upper bound based on differences in the joint distribution of the outcome and the misclassified regressor between groups with different values of the instrument. Jiang and Ding (2020) and Tommasi and Zhang (2020) also develop methods for constructing bounds in this setting. Nguimkeu, Denteh, and Tchernis (2019) show that the IV estimator will not be an upper bound in the case of endogenous misreporting. The Ordinary Least Squares (OLS) regression of an outcome on a misclassified binary regressor yields a corresponding lower bound (Aigner 1973; Bollinger 1996).

using moment inequalities. Using prior information on the extent to which the sum of the misclassification rates can vary across instrument values, we are able to construct bounds for the parameter of interest. In addition, it is straightforward to include covariates in our framework, and we can allow the misclassification rates to vary with the covariates.

We demonstrate the performance of our estimator both through simulations and an empirical example. Through a series of Monte Carlo simulations, we show the usefulness of our estimator under a variety of conditions when many alternative estimators are inconsistent. As an empirical example, we apply our estimator to the question of whether Medicaid eligibility, which is calculated by the researcher and likely mis-measured, crowds out the take-up of private health insurance. Re-examining the pioneering work by Cutler and Gruber (1996), we find that correcting for misclassification substantially lowers estimated crowd out and the extent to which non-insurance rates are reduced by Medicaid.

A variety of methodological approaches have been developed to estimate the impact of a mismeasured binary regressor. Our paper is most closely related to those which point identify the parameter of interest using an instrumental variable (Frazis and Lowenstein 2003; Mahajan 2006; Lewbel 2007; Hu 2008; DiTraglia and Garcia-Jimeno 2019); however, these methods require fixed misclassification. Another set of papers relies on the availability of multiple measures of the misclassified treatment (Card 1996; Kane, Rouse, and Staiger 1999; Black, Berger, and Scott 2000; Battistin, De Nadai, and Sianesi 2014). Two recent papers require both an instrumental variable and at least one additional variable that provides information on the measurement error to achieve identification (Calvi, Lewbel, and Tommasi 2017; Yanagi 2019), whereas our approach requires a single instrumental variable. Another set of papers utilizes restrictions on higher order moments (e.g., Chen, Hu, and Lewbel

2008a,b; DiTraglia and Garcia-Jimeno 2019), while our approach does not require such assumptions.

Two recent papers allow for varying misclassification. Ura (2018) bounds the impact of a mismeasured treatment allowing for general forms of misclassification. Yanagi (2019) point identifies the impact of a mismeasured binary regressor, but requires both an instrumental variable and an additional covariate that provides identifying information on the measurement error process. We point identify the impact of a mismeasured binary regressor in the presence of varying misclassification using only a single, discrete-valued instrument.

The paper proceeds as follows. In the next section, we provide motivating empirical examples to demonstrate that misclassification rates can vary with the instrumental variable. We then develop our main theoretical result. After discussing model estimation, including how to include moment inequalities when our key identifying assumption does not exactly hold, we provide a Monte Carlo simulation study to illustrate our main findings. We then use our estimator to re-examine Cutler and Gruber (1996), which estimates the extent to which Medicaid eligibility crowds out private health insurance. The final section concludes.

2 Examples of Varying Misclassification

To illustrate the importance of varying misclassification, we present examples from multiple data sources. We first use data from the March 1973 Current Population Survey (CPS) linked to tax returns filed with the Internal Revenue Service (IRS) (Social Security Administration 2005).[5] Because the IRS earnings are from tax filings which only contain a single, combined

[5]The data also contain Social Security (SS) Earnings records. We use the IRS earnings data because the available SS earnings records are top-coded at the annual ceiling for earnings subject to the SS tax, which censors earnings for nearly half of privately employed men. The IRS earnings data as well as self-reported

earnings measure for all household members, we restrict our sample to men and women whose marital status on their tax return is single. We further restrict the analysis to those who are privately employed, have positive IRS and CPS earnings, and have non-imputed CPS earnings, yielding 8,031 observations. We use sample weights that account for selection into the CPS and the match between the CPS and IRS records.[6]

The top panel of Figure 1 shows the probability density function of both (log) administrative and self-reported earnings in 1972. We treat IRS earnings as "true" earnings E^* and CPS earnings as the observed, mis-measured earnings E. The two distributions are quite similar throughout with the most noticeable differences in the middle of the distributions.

For this example, we construct a hypothetical program eligibility indicator, T^*, which equals one if true earnings E^* are at or below a threshold c and equals zero otherwise. At the threshold c_1, the leftmost vertical line in the top panel of Figure 1, 39.1 percent of individuals are assigned $T^* = 1$.[7] However, the researcher typically only observes the indicator T which is determined by whether E is at or below c. At c_1, $T = 1$ for 40.4 percent of individuals.

As discussed below, the misclassification rates, $\alpha_0 = P[T = 1|T^* = 0]$ and $\alpha_1 = P[T = 0|T^* = 1]$, are key inputs in deriving the bias of the OLS and IV estimators due to misclassification. At c_1, we calculate that $\alpha_0 = 0.068$ and $\alpha_1 = 0.072$. At c_2, the second threshold in the top panel of Figure 1, $P[T^* = 1]$ is 0.872 while $P[T = 1]$ is 0.877. However, the misclassification rates at c_2, $\alpha_0 = 0.149$ and $\alpha_1 = 0.016$, differ markedly from those at c_1.

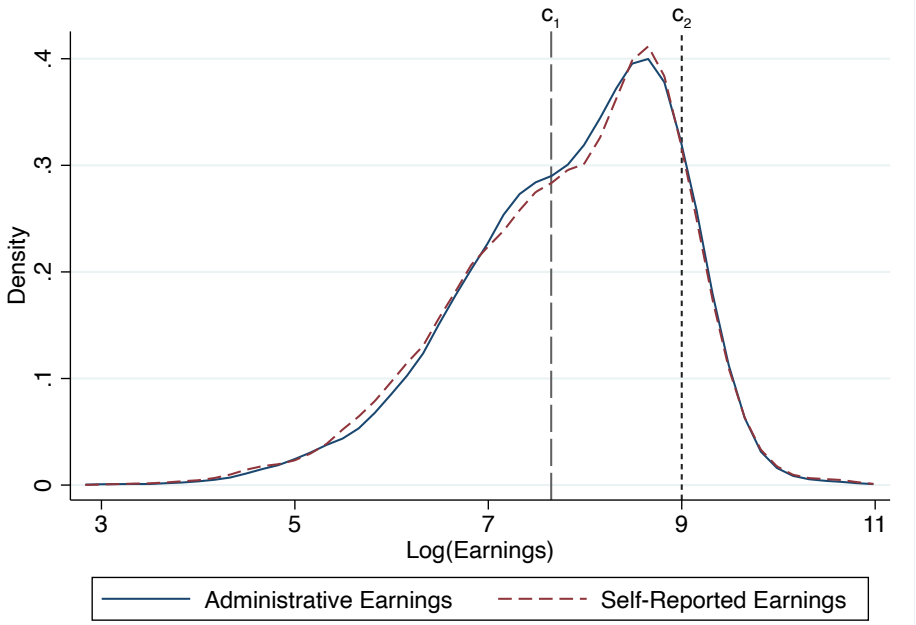
The bottom panel of Figure 1 plots the misclassification rates for numerous values of

earnings in the CPS are top-coded at \$50,000, which affects less than one-half of one percent of privately employed men in the data.

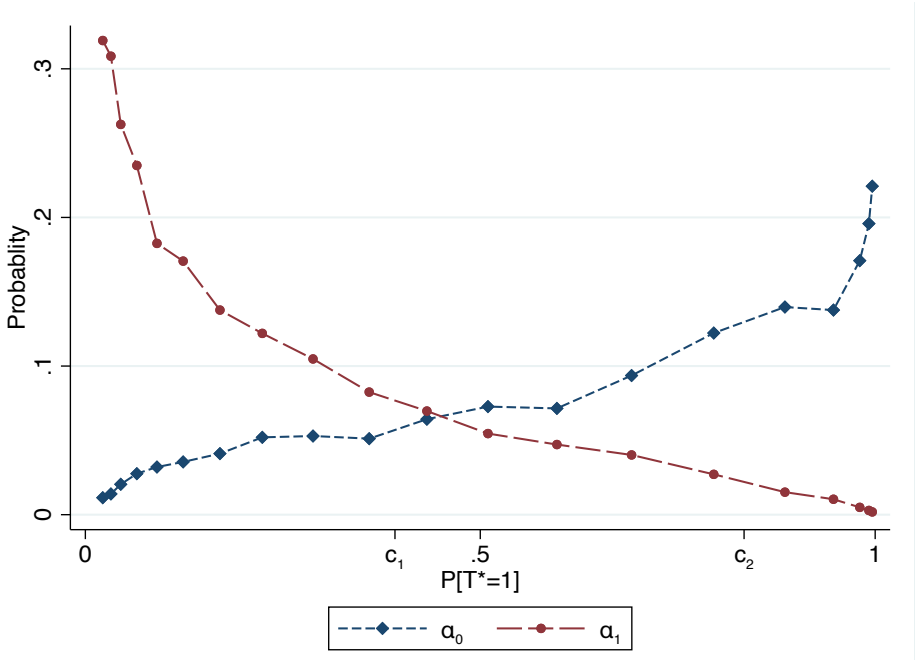
[6]More details on the linked data and selection criteria are provided in Appendix Section A.1.1.

[7]This threshold, $log(c_1) = 7.65$, is approximately \$2,100 which was the poverty line for a single, unrelated individual in 1972.

Figure 1: Misclassification Example - 1973 CPS-IRS Matched Earnings



(a) Administrative and Self-Reported Earnings PDFs



(b) Misclassification Rates

Notes: Both panels use data from the 1973 CPS-IRS matched earnings data. Details about the data are discussed in the text and in Appendix section A.1.1. Panel 1a plots the PDFs of both self-reported earnings and earnings reported on IRS tax returns from a sample of single men and women. Panel 1b reports misclassification rates; the details for constructing these rates can be found in Section 2.

c, placing $P[T^* = 1]$ on the horizontal axis. Notice that α_1 is decreasing in $P[T^* = 1]$. Intuitively, when a larger fraction of the population is eligible for the program ($T^* = 1$), the share of eligible individuals who are close to the eligibility threshold is smaller. For analogous reasons, α_0 increases with $P[T^* = 1]$.

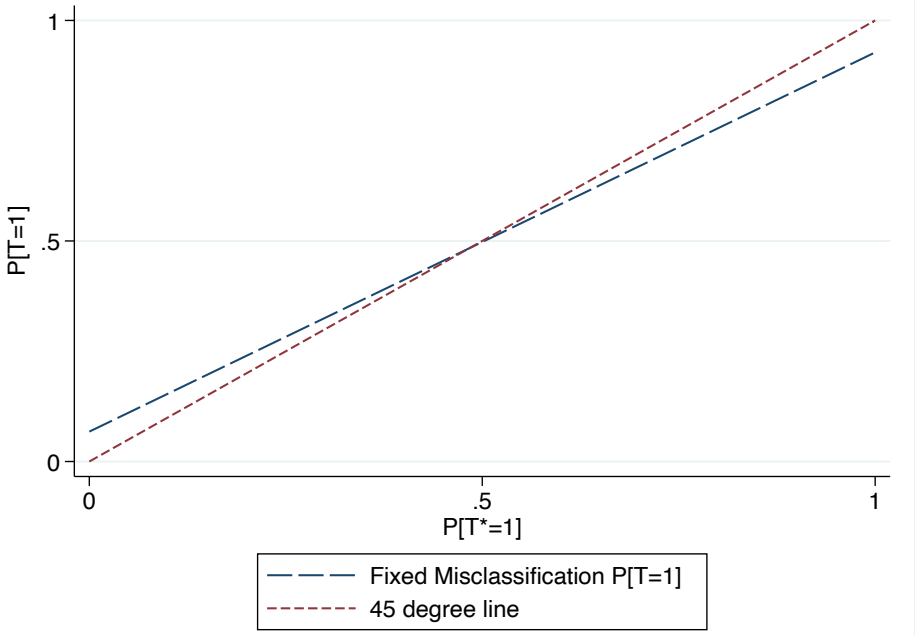
In a typical program eligibility setting, such as the Medicaid example mentioned in the Introduction, changes in the eligibility threshold are the basis for an instrumental variable. For example, Currie and Gruber (1996) treat the change in a state's eligibility threshold as an exogenous source of variation for the fraction of individuals in the state who have access to Medicaid. However, as with the example in Figure 1, the misclassification rates will systematically vary with these thresholds, which violates the fixed misclassification assumption.

The distinction between varying and fixed misclassification has important implications for interpreting IV estimation results. The IV estimator for the impact of T^* on Y when using Z as an instrument is $COV(Y, Z)/COV(T^*, Z)$, whereas when using T the IV estimator is $COV(Y, Z)/COV(T, Z)$. Note that the difference between these estimands is the denominators. In the case where Z is a binary instrument, the denominators are $COV(T^*, Z) = P[T^* = 1|Z = 1] - P[T^* = 1|Z = 0]$ and $COV(T, Z) = P[T = 1|Z = 1] - P[T = 1|Z = 0]$. Thus, the difference in the IV estimator using T versus T^* is seen by comparing the changes in $P[T = 1]$ and $P[T^* = 1]$ when the instrument changes values.

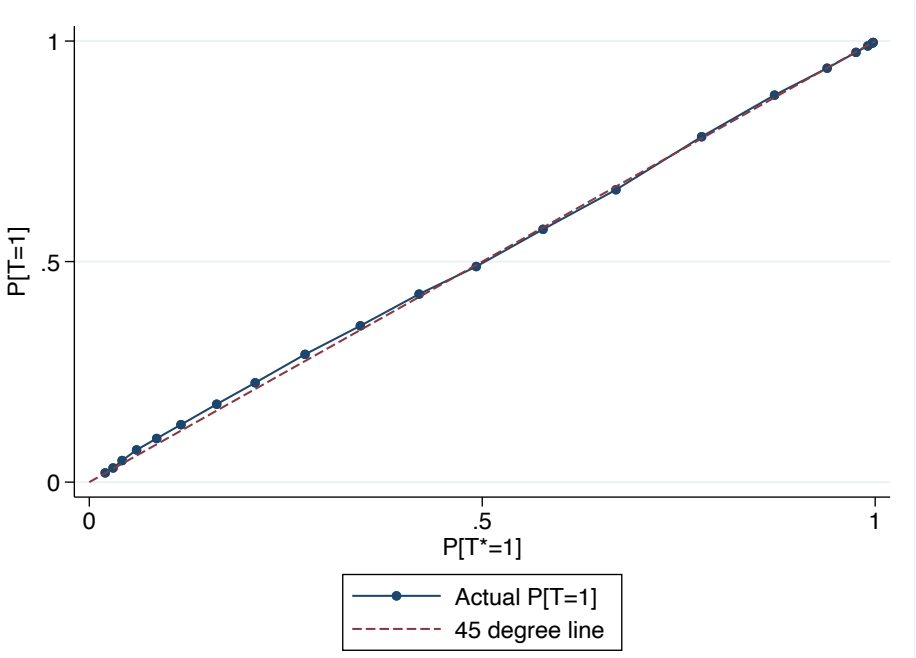
Figure 2 plots the relationship between $P[T = 1]$ and $P[T^* = 1]$ under the assumption of fixed misclassification (the line labeled "Hypothetical Fixed Misclassification" in the top panel) and for what is observed in the actual data (the line labeled "Actual $P[T = 1]$" in the bottom panel).[8] For the fixed misclassification case shown in the top panel, the slope of

[8]For the hypothetical fixed misclassification case, we use $\alpha_0 = 0.068$ and $\alpha_1 = 0.072$, which are the values corresponding the c_1 threshold shown in Figure 1.

Figure 2: Misclassification Example - 1973 CPS-IRS Matched Earnings - $P[T = 1]$ vs. $P[T^* = 1]$



(a) Using Hypothetical Fixed Misclassification Rates



(b) Using Actual Misclassification Rates

Notes: Panel 2a plots the hypothetical relationship between $P[T = 1]$ and $P[T^* = 1]$ under the assumption of fixed misclassification rates, where the α_0 and α_1 used to construct this figure correspond to the threshold c_1 shown in Figure 1. Panel 2b plots the actual relationship between $P[T = 1]$ and $P[T^* = 1]$ found in the 1973 CPS-IRS matched earnings data. Details about the data are discussed in the text and in Appendix section A.1.1.

the line is constant and always less than one (i.e., flatter than the 45 degree line).[9] Thus, a given change in a threshold c will always result in a smaller change in $P[T = 1]$ than the corresponding change in $P[T^* = 1]$ (i.e., $COV(T, Z) < COV(T^*, Z)$). As a result, when using T instead of T^*, the IV estimator will overestimate the impact of T^* on Y.[10]

However, as shown in the bottom panel of Figure 2, the actual relationship between $P[T = 1]$ and $P[T^* = 1]$ is non-linear. Around $P[T^* = 1] = 0.5$, the slope of this curve is less than one, similar to what occurs in the fixed misclassification case. On the other hand, at both small and large values of $P[T^* = 1]$, the change in $P[T = 1]$ *exceeds* the corresponding change in $P[T^* = 1]$. In these ranges, in contrast to the fixed misclassification case, applying the IV estimator to T will underestimate the impact of T^* on Y.

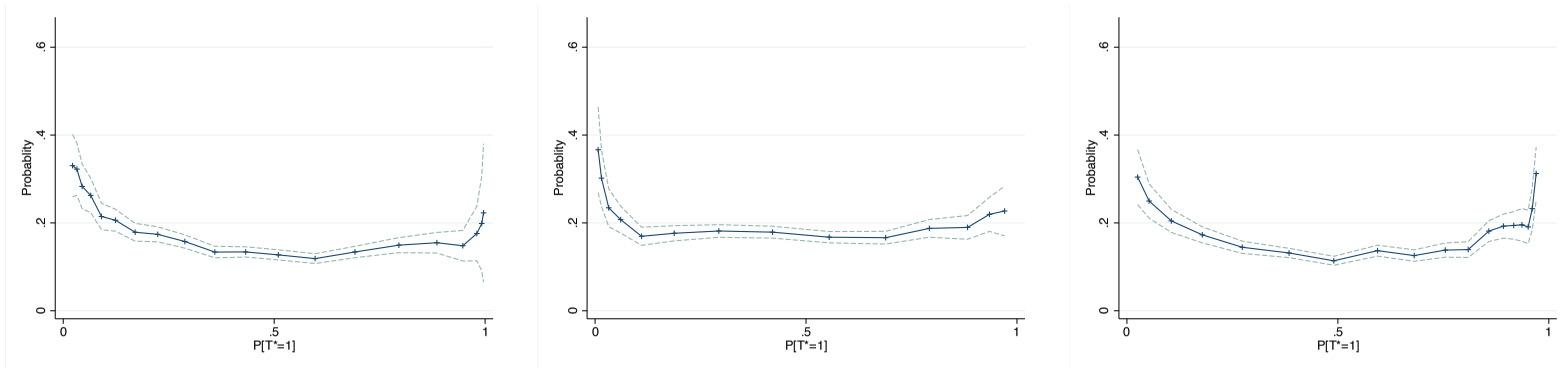
To foreshadow our key identifying assumption, Panel A of Figure 3 plots the sum of the misclassification rates along with bootstrapped 95% confidence intervals. As is readily apparent, the sum is relatively constant for much of the interior of the range of $P[T^* = 1]$. The remaining panels in Figure 3 provide analogous plots for the sum of the misclassification rates in additional examples where both survey and administrative data are available: wages in the January 1977 CPS (Panel D) and, using the 1999-2016 waves of the Continuous National Health and Nutrition Examination Survey (Continuous NHANES), height for males (Panel B) and females (Panel C) and weight for males (Panel E) and females (Panel F).[11] Across this range of outcomes, we find that the sum of the misclassification rates is roughly

[9] Since $P[T = 1] = \alpha_0 + (1 - \alpha_0 - \alpha_1)P[T^* = 1]$, the slope of this line will be less than one unless T is perfectly measured. In addition, under the standard assumption that $\alpha_0 + \alpha_1 < 1$, which is routinely made in this literature, T and T^* will be positively correlated and the slope of this line will exceed zero.

[10] For example, see Kane, Rouse, and Staiger 1999; Black, Berger, and Scott 2000; and Ura 2018.

[11] Appendix Section A.1.2 provides detailed descriptions of these data and how we construct the panels in Figure 3. Appendix Figures A1-A3 show that the plots of α_0 and α_1 for these additional examples move very similarly what we observe in Figure 1.

Figure 3: Misclassification Example - Sum of the Misclassification Rates



(a) 1973 CPS Earnings (b) Male Height (c) Male Weight

(d) 1977 CPS Wages (e) Female Height (f) Female Weight

Notes: The solid lines in the panels present the sum of the misclassification rates. The dashed lines in each panel are the bootstrapped 95% confidence intervals.

constant, especially when $P[T^* = 1]$ takes on values towards the middle of its range.

Overall, these examples yield three main implications. First, the misclassification rates vary with the instrument (here, the threshold), a finding at odds with the predominant assumption of fixed misclassification.[12] Second, changes in $P[T = 1]$ may be greater than, less than, or very close to changes in $P[T^* = 1]$ when moving between thresholds. As such, the IV estimator is no longer guaranteed to yield an upper bound for the true impact of T^* on the outcome of interest, in contrast to the fixed misclassification case. Third, the sum of the misclassification rates is fairly constant for a wide range of values of $P[T^* = 1]$.

3 The Model and Identification

The equation relating the outcome Y to the true, binary regressor $T^* = \{0, 1\}$ is

$$Y = \gamma + \beta T^* + \epsilon \tag{1}$$

However, instead of observing T^*, we observe T, which is subject to misclassification, i.e., T is also a binary variable in which either $T = T^*$ or $T = 1 - T^*$.[13]

Let Z be a discrete-valued instrumental variable with $J+1$ values, $j = 0, 1, \ldots, J$. Define the misclassification rates, α_{0j} and α_{1j}, $j = 0, 1, \ldots, J$, as $\alpha_{0j} = P[T = 1|T^* = 0, Z = j]$ and $\alpha_{1j} = P[T = 0|T^* = 1, Z = j]$, respectively. Defining $p_j^* = P[T^* = 1|Z = j]$, it follows that

[12] While our motivating examples focus on misclassification rates varying when a threshold varies, the results we present in the next section apply more generally to whenever misclassification rates vary with an instrument. See Bound, Brown, and Mathiowetz (2001) for an exhaustive review of validation studies for labor market outcomes, including for evidence of when misclassification rates vary.

[13] Our main result can be extended to a non-parametric regression framework such as that used in Mahajan (2006) and Lewbel (2007). We adopt a parametric regression framework for parsimony of notation and to accommodate covariates in a manner consistent with much applied work.

$$p_j = P[T = 1|Z = j] = (1 - \alpha_{1j})\, p_j^* + \alpha_{0j}\left(1 - p_j^*\right) = \alpha_{0j} + (1 - \alpha_{0j} - \alpha_{1j})\, p_j^*.$$

Assumption 1 *The following conditions are assumed to hold*

i. *There is a discrete instrumental variable, Z, with (at least) three values, i.e., $J \geq 2$*

ii. $E[Y|Z = j] = \gamma + \beta p_j^*,\ \forall j$

iii. $0 < p_j^* < 1,\ \forall j;\ p_j^* \neq p_k^*,\ \forall\ j \neq k$

iv. $E[Y|T, T^*, Z] = E[Y|T^*, Z]$

v. $\alpha_{0j} + \alpha_{1j} < 1,\ \forall j$

Assumption 1.ii is the usual IV exclusion condition and Assumption 1.iii is the usual IV relevance condition. The last two parts of Assumption 1 are standard in the misclassification literature (e.g., Frazis and Loewenstein 2003; Mahajan 2006; Lewbel 2007; DiTraglia and Garcia-Jimeno 2019). Assumption 1.iv implies "nondifferential" measurement error, i.e., no additional independent information is contained in the mis-measured regressor once the true value of the regressor is known. This assumption is analogous to the classical measurement error assumption for continuous variables.[14] Assumption 1.v requires that there is not "too much" misclassification. With this restriction, the OLS regression of Y on T using observations $Z = j$ will yield an estimated coefficient on T that has the same sign as β.

Assumption 2 *For all j, $E[Y|T^* = t, Z = j] = E[Y|T^* = t]$*

Assumption 2 requires the expected value of the outcome Y for a given value of the true regressor T^* to be the same across all values j of the instrument Z. The same assumption

[14] Recently, Nguimkeu, Denteh, and Tchernis (2019) allows for differential measurement error.

is made by Mahajan (2006) to identify the impact of a misclassified binary regressor with fixed misclassification. Similar assumptions are made elsewhere in this literature (e.g., Kane, Rouse, and Staiger 1999, Black, Berger, and Scott 2000; Frazis and Loewenstein 2003). This assumption also is proposed as a test of no selection bias in the literature on instrumental variables (Black et al. 2015; Brinch, Mogstad, and Wiswall 2017). While T^* is exogenous under Assumption 2, we relax this assumption below to admit some forms of endogeneity.

3.1 Identification

For a discrete instrumental variable Z that takes on exactly three discrete values, Assumption 1.ii yields three equations based on the exclusion restriction

$$E\left[Y|Z=j\right] = \gamma + \beta p_j^*,\ j = 0, 1, 2 \tag{2}$$

Assumption 2 yields two equations for any pair of instrument values, j and k

$$E\left[Y|T^*=1, Z=j\right] = E\left[Y|T^*=1, Z=k\right] \tag{3}$$

$$E\left[Y|T^*=0, Z=j\right] = E\left[Y|T^*=0, Z=k\right] \tag{4}$$

Equations (3) and (4) are based on unobserved quantities. For a given value of Z, the expected value of the outcome Y conditional on T is a weighted average of these unknown quantities, with the weights based on the misclassification rates. As shown in Appendix Section A.2, inverting these relationships yields the following expressions for the unknown

quantities in (3) and (4) in terms of observed quantities and the misclassification rates

$$E\left[Y|T^*=1,Z=j\right]=\frac{COV\left(Y,T|Z=j\right)}{p_j-\alpha_{0j}}+E\left[Y|Z=j\right] \tag{5}$$

$$E\left[Y|T^*=0,Z=j\right]=E\left[Y|Z=j\right]-\frac{COV\left(Y,T|Z=j\right)}{1-p_j-\alpha_{1j}} \tag{6}$$

where $COV\left(Y,T|Z=j\right)$ is the covariance of Y and T among those with $Z=j$.

Subtracting equation (4) from equation (3) and then inserting equations (5) and (6) into the resulting expression yields (see Appendix Section A.3)[15]

$$\frac{1}{(1-\alpha_{0j}-\alpha_{1j})}\cdot\frac{COV\left(Y,T|Z=j\right)}{p_j^*\left(1-p_j^*\right)}=\frac{1}{(1-\alpha_{0k}-\alpha_{1k})}\cdot\frac{COV\left(Y,T|Z=k\right)}{p_k^*\left(1-p_k^*\right)} \tag{7}$$

For a discrete instrument with three values, these substitutions provide five equations: three equations found in (2) and two additional equations by applying (7) to two pairs of instrument values. However, eleven unknown parameters appear in these five equations: β, γ, the three p_j^*, the three α_{0j}, and the three α_{1j}. Noting that misclassification rates only appear in (7) as the sum $\alpha_{0j}+\alpha_{1j}$, we can instead view the problem as containing five equations with eight parameters: β, γ, the three p_j^*, and the three sums $\alpha_{0j}+\alpha_{1j}$. Moving to an instrument with more discrete values will not solve this identification issue because each additional instrument value adds two equations but also two parameters: another p_j^* and another sum $\alpha_{0j}+\alpha_{1j}$. Instead, we make the following assumption regarding the misclassification rates.

Assumption 3 *The sum of the misclassification rates is constant, i.e.,* $\alpha_{0j}+\alpha_{1j}=\overline{\alpha},\ \forall j$.

[15]By first subtracting equation (4) from equation (3) yields $E\left[Y|T^*=1,Z=j\right]-E\left[Y|T^*=0,Z=j\right]=E\left[Y|T^*=1,Z=k\right]-E\left[Y|T^*=0,Z=k\right]$. An alternative approach, discussed in Section 3.2, is to replace Assumption 2 with this expression (as in Lewbel (2007)).

Assumption 3 allows both α_{0j} and α_{1j} to vary across instrument values j. Importantly, this assumption nests the typical fixed misclassification assumption, making our estimator applicable under the conditions assumed in previous papers. In addition, this assumption is consistent with the empirical examples in Section 2: α_{0j} and α_{1j} move in opposite directions as $P[T^* = 1]$ changes while their sum, $\alpha_{0j} + \alpha_{1j}$, is relatively constant for much of the range of $P[T^* = 1]$ for all six examples in Figure 3.

Equation (7) is greatly simplified under Assumption 3 because replacing $\alpha_{0j} + \alpha_{1j}$ with $\overline{\alpha}$ leads the misclassification rates to drop out of the expression to yield

$$\frac{COV\left(Y, T|Z = j\right)}{p_j^*\left(1 - p_j^*\right)} = \frac{COV\left(Y, T|Z = k\right)}{p_k^*\left(1 - p_k^*\right)} \tag{8}$$

Theorem 1 *The parameters β, γ, p_0^*, p_1^*, and p_2^*, and $\overline{\alpha}$ are identified under Assumptions 1, 2, and 3.*

We provide the detailed proof in Appendix Section A.4.[16] We note that the equations yield two solutions for β and Assumption 1.v allows us to select between the two options.

3.2 Relaxing the Exogeneity Assumption

The estimator derived above can be applied to settings with a weaker exogeneity assumption.

Assumption 4 *For all j, k, $E\left[Y|T^* = 1, Z = j\right] - E\left[Y|T^* = 0, Z = j\right] = E\left[Y|T^* = 1, Z = k\right] - E\left[Y|T^* = 0, Z = k\right]$*

[16] Under Assumptions 1 through 3, equation (2) holds for all $J + 1$ instrument values while the model generates J independent equations based on equation (8) using pairs of instrument values. If only a two-valued instrument were available, we would have three equations but four unknown parameters (β, γ, p_0^*, and p_1^*). However, an instrument with three values will generate five unique equations, three due to equation (2) and two due to equation (8), allowing us to identify the five parameters β, γ, p_0^*, p_1^*, and p_2^*.

In contrast to Assumption 2, Assumption 4 only requires the difference in the expected value of Y between the two values of T^* to be the same across all values of the instrument. A similar assumption is made by Lewbel (2007).[17] Assumption 4 also allows T^* to be endogenous under some conditions such as those found in Olsen's (1980) selection model (e.g., see Brinch, Mogstad, and Wiswall 2017).

Corollary 1.1 *The magnitude of β, but not the sign, is identified under Assumptions 1i.-iv., 3, and 4.*

Therefore, swapping Assumption 4 for Assumption 2 allows us to accommodate some forms of endogeneity. However, Assumption 1.v is not enough to sign β as in Theorem 1. We can use this result, however, to test the null hypothesis of there being no impact of T^* on Y.

We can sign β by assuming that the sign of the IV estimator does not change when using T instead of T^*.

Assumption 5 $sgn\left(p_j - p_k\right) = sgn\left(p_j^* - p_k^*\right) \ \forall \ j, k$

This assumption is consistent with Figure 2b because $P[T = 1]$ in an increasing function of $P[T^* = 1]$ and is sufficient to state our last corollary.

Corollary 1.2 *β, γ, p_0^*, p_1^*, and p_2^* are identified under Assumptions 1, 3, 4, and 5.*

Thus, by additionally invoking Assumption 5, we immediately sign β, although we still cannot identify $\overline{\alpha}$.[18]

[17] However, in addition to Assumption 4, Lewbel requires T^* to be (conditionally) exogenous.

[18] As shown in Appendix Section A.5, one can identify $\overline{\alpha}$ if, in addition to Assumptions 1, 3, and 4, it is assumed that the error term in equation (1) is homoskedastic.

4 Estimation and Inference

We first discuss the estimation of our model using Generalized Methods of Moments (GMM), and then discuss the inclusion of covariates. We also demonstrate how to replace the moments in (7) with inequality moments to relax Assumption 3. However, as we discuss in Appendix Section A.6.1, the GMM estimator will not always converge when β is small and/or the instruments are weak. As such, we briefly discuss how to estimate our model by incorporating bounds on β via inequality moments.

4.1 GMM Estimation

Equations in (2) and (8) are conditional on instrument values. It is straightforward to create unconditional moment conditions to use in GMM estimation. The unconditional moment analogous to equation (2) is

$$Z_j \cdot \left(Y - \gamma - \beta p_j^*\right) \tag{9}$$

where $\cdot$ is the element-by-element product operator, $Z_j = 1(Z = j)$, and $1(...)$ is the indicator function.[19]

The equality conditions in (8) are analogous to requiring $COV\left(Y, T|Z = j\right) / \left(p_j^* \left(1 - p_j^*\right)\right) = \phi$, where ϕ is an unknown constant. The corresponding unconditional moment is

$$Z_j \cdot \left[\frac{Y \cdot T - Y \cdot p_j}{p_j^* \left(1 - p_j^*\right)} - \phi\right] \tag{10}$$

because averaging the term $(Y \cdot T - Y \cdot p_j)$ is an estimator for $COV\left(Y, T|Z = j\right)$ since we are

[19] We suppress the i subscripts for the outcome Y and the indicator Z_j for expositional convenience.

conditioning on Z_j. These moments introduce additional parameters $p_j = P[T = 1|Z = j]$ that must be estimated for each instrument value.

Letting $W' = (Y, T, Z)$ (and continuing to suppress the i subscript), our unconditional moment conditions are thus given by $E[g(W, \boldsymbol{\delta})] = 0$, where

$$g(W, \boldsymbol{\delta}) = \begin{pmatrix} Z_j \cdot \left(Y - \gamma - \beta p_j^*\right) \\ Z_j \cdot \left[(Y \cdot T - Y \cdot p_j) / \left(p_j^* \left(1 - p_j^*\right)\right) - \phi\right] \\ Z_j \cdot (T - p_j) \end{pmatrix} \tag{11}$$

and $\boldsymbol{\delta} = (\beta, \gamma, p_j^*, p_j, \phi)$ is the vector of unknown parameters. When the instrument takes on three discrete values, this vector contains nine moment conditions because each row of (11) provides a moment for each instrument value. As $\boldsymbol{\delta}$ contains nine parameters, the system is exactly identified. Instruments with more than three discrete values will yield an over-identified system. Given our assumptions and our identification results, standard GMM results can be applied, providing us with consistency and asymptotic normality of $\hat{\boldsymbol{\delta}}$ (e.g., see Wooldridge 2010).

The moment conditions in (11) are applicable under either Assumption 2 or Assumption 4. If we invoke Assumption 5, then we can identify the sign of β with this set of moments, regardless of which exogeneity assumption we use. In addition, under Assumption 2, as shown in Appendix Section A.4, $COV\left(Y, T|Z = j\right) / \left(p_j^* \left(1 - p_j^*\right)\right) = \beta\left(1 - \overline{\alpha}\right)$. Thus, with Assumption 2, we can replace ϕ with $\beta\left(1 - \overline{\alpha}\right)$ in (11) and also estimate $\overline{\alpha}$.

4.2 Including Covariates

We present two approaches to including covariates. The first approach can incorporate both continuous and discrete covariates by assuming a standard parametric relationship between the outcome and the covariates. However, this approach requires that the model parameters do not vary with the covariates (although we continue to allow the misclassification rates to vary with the instrument values). The second approach allows for the misclassification rates to vary with the covariates as well as the instrument, but requires discrete covariates.

For the first approach to include covariates, we specify the equation of interest to be

$$Y = \gamma + \beta T^* + \boldsymbol{X}\boldsymbol{\psi} + \epsilon \tag{12}$$

where $\boldsymbol{X}$ is a vector of exogenous covariates (i.e., ϵ is uncorrelated with each element of $\boldsymbol{X}$) and classification error in T^* does not vary with $\boldsymbol{X}$ (e.g., $P[T = 1|T^* = 0, Z = j, \boldsymbol{X}] = P[T = 1|T^* = 0, Z = j]$). An advantage of assuming (12) is that it embodies specifications typically used in applied work, as illustrated by our empirical application in Section 6.

As shown in Appendix Section A.6.2, by updating the assumptions to condition on $\boldsymbol{X}$ and defining $\widetilde{Y} = Y - \boldsymbol{X}\boldsymbol{\psi}$, we derive moment conditions analogous to (2) and (8) to account for $\boldsymbol{X}$. The moment conditions analogous to equations (2) and (8) are

$$E[\widetilde{Y}|Z = j, \boldsymbol{X}] - \gamma - \beta p_j^* = 0 \tag{13}$$

$$\frac{COV\left(\widetilde{Y}, T|Z = j, \boldsymbol{X}\right)}{p_j^*\left(1 - p_j^*\right)} = \frac{COV\left(\widetilde{Y}, T|Z = k, \boldsymbol{X}\right)}{p_k^*\left(1 - p_k^*\right)} \tag{14}$$

For estimation, the moments corresponding to $p_j = P[T = 1|Z = j, \boldsymbol{X}]$ require the

additional assumption that p_j does not vary with $\boldsymbol{X}$. The empirical analogs of (13) and (14) can be implemented by replacing Y with $\widetilde{Y}$ in the first two lines in (11). Doing so requires knowledge of $\boldsymbol{\psi}$ which we can consistently pre-estimate by regressing Y on T and $\boldsymbol{X}$ while using Z to instrument for T (see Appendix Section A.6.2). Using the pre-estimated value of $\boldsymbol{\psi}$ to first compute $\widetilde{Y}$, we can apply GMM although the variances need to be adjusted for using pre-estimated parameters (Newey 1984) or the procedure can be bootstrapped.

Our second approach allows for the misclassification rates and the other parameters to vary with discrete covariates by expanding the estimating equations (e.g., Mahajan 2006, Lewbel 2007). For example, suppose that there is a dichotomous exogenous covariate V. We can allow the empirical moment conditions in (11) to depend on V, thereby allowing all parameters to vary with V. This expanded set of moments can be estimated using GMM. In addition, we can impose restrictions on these parameters, e.g., β does not vary with V, that can be readily tested. While this second approach for including covariates provides more flexibility, it requires variation in the instrument within each cell defined by the exogenous regressors. It is common in empirical work, however, for the instrument to vary across, but not within, cells (e.g., the instrument varies across states and years but not within a state at a point in time), which makes the former approach relevant for many applications.

4.3 Relaxing Assumption 3

Although the Figures in Section 2 indicate that Assumption 3 approximately holds for a wide range of $P[T^* = 1]$ values across multiple domains, it does not hold exactly and especially so when $P[T^* = 1]$ is close to either 0 or 1. However, we can relax Assumption 3 by replacing the moment equalities in (8) with moment inequalities as shown below.

Letting $\overline{\alpha}_j = \alpha_{0j} + \alpha_{1j}$, we can re-write (7) as

$$\left[\frac{COV\left(Y,T|Z=j\right)}{p_j^*\left(1-p_j^*\right)}\Big/\frac{COV\left(Y,T|Z=k\right)}{p_k^*\left(1-p_k^*\right)}\right] = R_{jk} \tag{15}$$

where $R_{jk} = \left(1-\overline{\alpha}_j\right)/\left(1-\overline{\alpha}_k\right)$. Assumption 3 imposes that $R_{jk} = 1$. Based on the data from Figure 3a, we see that $\overline{\alpha}_j$ ranges from 0.119 to 0.158 when $P[T^* = 1]$ lies in $[.29, .95]$. Within this range, the smallest value that R_{jk} can take on is $\left(1-0.158\right)/\left(1-0.119\right) \approx 0.95$.

Letting $\underline{R}_{jk}$ be the smallest value for R_{jk}, the corresponding inequality moment is

$$\left[\frac{COV\left(Y,T|Z=j\right)}{p_j^*\left(1-p_j^*\right)}\Big/\frac{COV\left(Y,T|Z=k\right)}{p_k^*\left(1-p_k^*\right)}\right] - \underline{R}_{jk} \geq 0 \tag{16}$$

Using the above example from Figure 3a, we would set $\underline{R}_{jk} = 0.95$ to implement (16).

The same data imply a corresponding upper bound for R_{jk}. However, the upper bound for R_{jk} is simply a lower bound for R_{kj}, where we switch the instrument values j and k. This observation yields a second moment inequality for each pair of instrument values

$$\left[\frac{COV\left(Y,T|Z=k\right)}{p_k^*\left(1-p_k^*\right)}\Big/\frac{COV\left(Y,T|Z=j\right)}{p_j^*\left(1-p_j^*\right)}\right] - \underline{R}_{kj} \geq 0 \tag{17}$$

While the discussion motivating this pair of moment inequalities does not take a stand on whether $\overline{\alpha}_j \geq \overline{\alpha}_k$ or $\overline{\alpha}_j \leq \overline{\alpha}_k$ (which leads to assuming $\underline{R}_{jk} = \underline{R}_{kj}$), we may choose to apply additional knowledge. For example, in Figure 3 we observe that $\overline{\alpha}_j$ tends to increase as we move away from $P[T^* = 1] = 0.5$ in either direction. Therefore, if $p_j = 0.2$ and $p_k = 0.4$, then we would suspect that $\overline{\alpha}_j \geq \overline{\alpha}_k$. We can impose this relationship by setting $\underline{R}_{jk}$ to some preselected value that is less than 1 and setting $\underline{R}_{kj}$ equal to 1.

It is important to note that, although $\underline{R}_{jk}$ and $\underline{R}_{kj}$ must be selected by the researcher for each pair of moment inequalities, this choice involves the relative level of misclassification across instrument values as opposed to the absolute level of misclassification. Such an assumption is likely to be appropriate when $\overline{\alpha}_j$ and $\overline{\alpha}_k$ fall into the same small range.

Estimation requires the combination of moments equalities and inequalities. The moment equalities in the first and third lines of (11) remain the same above. The second line of (11) is replaced with two sets of moments (in the case with three instrument values). First, we include moment equalites analogous to (10) for instrument values $Z = 0$ and 1 except replacing ϕ with ϕ_0 and ϕ_1, respectively, which yields estimates of $COV(Y, T|Z = j) / \left(p_j^* \left(1 - p_j^*\right)\right)$ for $j = 0, 1$. Second, we include six moment inequalities corresponding to (16) and (17) for all pairwise comparisons of instrument values. In these moment inequalities, we replace exactly one $COV(Y, T|Z = j) / \left(p_j^* \left(1 - p_j^*\right)\right)$ term with its corresponding ϕ_j parameter while conditioning on the other instrument value in order create unconditional moment inequalities. While inference can be performed using the moment inequality methods of Andrews and Soares (2010), their approach requires the joint confidence set for all parameters to be constructed. To focus on the marginal confidence set for β, we apply the methods found in Bugni, Canay, and Shi (2017).

4.4 Inference with Weak Identification

There is an important caveat to implementing GMM. While our estimator exists as long as $\beta \neq 0$ in the population, its sample analog may not exist when at least one of the following holds: i) β is close to zero, ii) there are small differences between p_j^* values, and iii) the sample size is relatively small; see Appendix Section A.6.1 for further details. Under these

conditions, the GMM estimator is weakly identified and may not converge in finite samples.[20] In an earlier working paper version of this paper (Haider and Stephens 2020), we demonstrate how to address this issue by constructing the marginal confidence set for β by combining bounds on β with moments from our model.[21]

5 Monte Carlo Simulations

We present Monte Carlo simulations to highlight multiple aspects of our estimator. For each simulation, we randomly assign observations to one of three instrument values, $j = 0, 1, 2$.[22] We generate a random variable $E^* \sim N(7.85, 1.18)$ and set T^* equal to one if E^* falls below an instrument value-specific cutoff which, for our baseline simulations, are chosen such that $p_0^* = 0.35$, $p_1^* = 0.50$, and $p_2^* = 0.65$.[23] We construct Y using equation (1) by setting $\gamma = \beta = 1$ and drawing the error term $\epsilon \sim N(0, \sigma_\epsilon^2)$.[24]

Finally, we construct the misclassified binary indicator T. For our baseline simulations, we base our misclassification rates on Figure 1b, but impose Assumption 3 that $\alpha_{0j} + \alpha_{1j} = \overline{\alpha}$. We set $\overline{\alpha} = 0.130$ and allow α_{0j} (and thus α_{1j}) to vary with Z where $\alpha_{00} = 0.055$, $\alpha_{01} = 0.070$, and $\alpha_{02} = 0.085$. We then draw the random variable $u \sim$ uniform$[0, 1]$. For observations with $T^* = 0$ and instrument value j, $T = 1$ if $u < \alpha_{0j}$ and $T = 0$ otherwise. For observations

[20] DiTraglia and Garcia-Jimeno (2019) note a similar issue with their estimator.

[21] There are a variety of choices for bounds on β. For lower bounds, see Aigner (1973) and Black, Berger, and Scott (2000). We develop an alternative lower bound using an instrument which, as shown in our working paper (Haider and Stephens 2020), is tighter than OLS. For upper bounds, see Bollinger (1996), Ura (2018), Jiang and Ding (2020), and Tommasi and Zhang (2020). Recall, as noted earlier, the IV estimator no longer yields an upper bound with varying misclassification.

[22] These were constructed from an underlying standard normal random variable where we assign $j = 0$ if the random variable is less than -0.3, $j = 2$ if the random variable is greater than or equal to 0.3, and $j = 1$ otherwise. The assignment probabilities are 0.382, 0.236, and 0.382, respectively.

[23] The distribution of E^* is matches the 1973 CPS/IRS administrative log earnings data from Section 2.

[24] For all simulations in the main text, the error term ϵ is assumed to have a standard deviation of 0.25.

with $T^* = 1$ and instrument value j, $T = 0$ if $u < \alpha_{1j}$ and $T = 1$ otherwise.

Table 1 illustrates the performance of our estimator in large samples ($N = 100,000$) as we vary the misclassification values across each column. In column (1) we impose "fixed misclassification," a special case of Assumption 3, using $\alpha_{0j} = 0.070$ and $\alpha_{1j} = 0.060$ for all j. Consistent with standard results, OLS using T yields an attenuated point estimate of 0.87 with a standard deviation (SD) across iterations of 0.002 while 2SLS yields an upwards biased estimate of 1.150 (SD=0.011) that aligns with the standard misclassification formula ($\beta/(1-\alpha_0-\alpha_1)$).[25] Our estimator is well-centered with a mean estimate of 1.003 (SD=0.033).

Table 1: Monte Carlo Simulations - Varying Misclassification

	(1)	(2)	(3)	(4)
Panel A: Data generating process				
α_{00}	0.070	0.055	0.035	0.110
α_{01}	0.070	0.070	0.070	0.140
α_{02}	0.070	0.085	0.105	0.170
$\alpha_{0j} + \alpha_{1j}$ $\forall j$	0.130	0.130	0.130	0.260
Panel B: Estimates of β				
OLS	0.870	0.877	0.886	0.750
	(.002)	(.002)	(.002)	(.003)
2SLS	1.150	1.032	0.907	1.065
	(.011)	(.009)	(.008)	(.012)
Our estimator	1.003	1.003	1.002	1.004
	(.033)	(.033)	(.033)	(.047)

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (11). Each simulation is based on 1,000 iterations, with a sample size of 100,000 observations. The mismeasured variables T are constructed so that misclassification rate α_{0j} and the sum of the misclassification rates $\alpha_{0j}+\alpha_{1j}$ for instrument value j match the values shown in the top panel. The values for the remaining parameters are discussed in Section 5. Panel B contains average estimates of β across iterations for the listed estimation methods. The standard deviations of the estimates across iterations are shown in parentheses.

We examine results under varying misclassification in the remainder of Table 1. In column

[25] With $\beta = 1$, $\alpha_0 = 0.070$, and $\alpha_1 = 0.060$, the 2SLS estimator converges to $1/(1-0.070-0.060) = 1.149$.

(2), the average 2SLS estimate of 1.032 (SD=0.009) substantially differs from the estimate with fixed misclassification in column (1).[26] Our estimator is centered on the true value (1.003 with SD=0.033). In column (3), we increase the rate at which the misclassification rates vary with p^*, but continue to maintain $\overline{\alpha} = .13$. We find that the average 2SLS estimate is *below* the true value of β (0.907 with SD=0.008). This result is consistent with our earlier analytical finding: under varying misclassification, the 2SLS estimator no longer provides an upper bound for β. Again, our estimator is well-centered (1.002 with SD=0.033). In column (4), we double the misclassification rates specified in column (2). Both the OLS and 2SLS estimates are further from the true value of 1 as compared to column (2) while our estimator remains well-centered (1.004 with SD=0.047).

Some papers in this literature impose assumptions on the distribution of the error term such as symmetry and/or homoskedasticity (e.g., Chen, Hu, Lewbel 2008a, 2008b; DiTraglia and Garcia-Jimeno 2019). By circumventing such assumptions, we can apply our estimator to binary outcomes as is frequently the case in applied settings such as the empirical example section 6. In Appendix Section A.7.1, we replicate the simulations of Table 1 with a binary outcome. The resulting takeaway is the same as above: OLS is inconsistent, 2SLS can be either downward or upward inconsistent, and our estimator remains well-centered.

Another advantage of our estimator is that it is consistent with certain forms of endogeneity, as discussed in Section 3.2. In Appendix Section A.7.1, we provide a series of simulations that build on those in Table 1, but allow for such endogeneity. With both misclassification and endogeneity present, neither OLS nor 2SLS systematically bound the true parameter value. However, in all cases, our estimator remains well-centered.

[26] Given the distribution of observations across the instrument values, the average misclassification rates for column (2) are the same as the fixed misclassification rates in column (1) (i.e., $\alpha_0 = 0.070$, and $\alpha_1 = 0.060$).

We also examine the performance of alternative estimators in Appendix Section A.7.1. We show that three estimators relying on fixed misclassification and a single instrumental variable (Frazis and Lowenstein 2003; Mahajan 2006; Lewbel 2007) are not well-centered in simulations with varying misclassification as examined in Table 1. Similarly, three estimators that rely on assumptions place on higher-order moments of the error term such as homoskedasticity and/or symmetry (Chen, Hu, and Lewbel 2008a, 2008b; DiTraglia and Garcia-Jimeno 2019) yield inconsistent estimates of β when we specify error distributions that deviate from these assumptions. In contrast, our estimator again remains well-centered.

6 Empirical Application: Medicaid and Crowd Out

We apply our estimator to examine whether Medicaid eligibility crowds out private health insurance coverage. In the pioneering work of Cutler and Gruber (1996), binary measures of health insurance coverage are regressed on a binary indicator for being eligible for Medicaid. During the years examined by Cutler and Gruber (1987 to 1992), a series of federal and state legislative decisions decoupled Medicaid eligibility from being tied to the receipt of Aid to Families with Dependent Children (AFDC) payments. We focus on the impact of the Medicaid expansions for children in our analysis. Cutler and Gruber estimate that the share of children eligible for Medicaid rose by 50 percent during this period from 18% to 27%.

Medicaid eligibility is a binary indicator that is computed based on federal and state-specific eligibility criteria along with household demographic and income information data found in the survey data. For a given set of demographic characteristics (including the number of family members, single or dual headed, age of child, and state of residence), eligibility

for Medicaid depends upon whether household income is below a particular threshold. Prior to these expansions, the eligibility threshold was determined by a state's AFDC income threshold. The federal expansions increased thresholds for children of certain ages and birth years and gave states the option to raise eligibility thresholds for non-AFDC recipients.[27]

Cutler and Gruber estimate specifications of the form

$$Y = \gamma + \beta ELIG + \boldsymbol{X}\boldsymbol{\psi} + \epsilon \tag{18}$$

where Y is a binary measure of insurance coverage (Medicaid, private, or uninsured) and $\boldsymbol{X}$ is a set of exogenous regressors, including household demographic variables and state and time period indicators.[28] $ELIG$ is the binary Medicaid eligibility indicator constructed using federal and state rules as noted above. Cutler and Gruber instrument for $ELIG$ given concerns about spurious correlations between computed eligibility and insurance coverage. They develop a "simulated eligibility" instrument that only exploits variation in Medicaid eligibility rules across states over time.[29]

We use data on 266,421 children from the CPS covering the same period as Cutler and Gruber.[30] However, we use an alternative instrumental variable to align our analysis with the examples from the earlier sections of this paper. Reported household income in the CPS

[27] See Currie and Gruber (1996) and Shore-Sheppard (2008) for more details of these Medicaid expansions.

[28] The vector $\boldsymbol{X}$ includes the number of people in the household and indicators for state of residence, calendar year, and child age as well as for male, white, household head type, and number of workers.

[29] Cutler and Gruber draw a random sample of 300 children for each age. The simulated Medicaid eligibility instrument for a child of age a in state s is the share of children in the age a random sample that are eligible when using the state s eligibility calculator. The process is repeated for each age in each state during each time period. For more details, see Cutler and Gruber (1996). This methodology, which isolates changes in programmatic rules to use as an instrument, has been used to study a variety of social programs. For recent examples, see Frean, Gruber, and Sommers (2017) and Brown, Kowalski, and Lurie (2020).

[30] We thank Lara Shore-Sheppard for generously providing her data and programs, which we use as the basis for our analysis, from her 2008 article.

is mismeasured, which leads Medicaid eligibility to be misclassified for some children. For a child of age a in a family of size n and structure f (single vs. dual headed) residing in state s during year t, we compute the highest income threshold under the prevailing federal and state Medicaid legislation that would allow them to be eligible for Medicaid. We convert this threshold to a percentage of the federal poverty line (FPL) for a family of size n in year t. As our estimator leverages a discrete instrument with three (or more) values, we construct an instrument by dividing observations into three groups based the Medicaid FPL threshold that determines their eligibility: thresholds less than 100% of the FPL, between 100% and 133% of the FPL, and greater than or equal to 133% of the FPL.

The results of our analysis are shown in Table 2. Each column corresponds to a different insurance coverage outcome: Medicaid, private, and uninsured. Row (A) presents the 2SLS estimates of β found in Table IV of Cutler and Gruber while row (B) shows our replication of their estimates. Our replication estimate for Medicaid coverage as the outcome is nearly identical to the Cutler and Gruber estimate, while our replications for the other two outcomes differ slightly from their estimates. Row (C) of Table 2 shows that the 2SLS estimates using our alternative, discrete-valued instrument are similar to our replication results, although the Medicaid and private coverage estimates are somewhat smaller in magnitude.[31]

As we noted above, many methods to address a misclassified binary regressor assume that the regressor is exogenous in the absence of measurement error and would not be applicable in this setting. Indeed, row (D) of Table 2 shows that the corresponding OLS estimates

[31]We use indicators for the discrete categories of FPL as separate instruments with $FPL < 100$ as the excluded category. The first stage parameter estimates for the two included instrument values, $100 \leq FPL < 133$ and $FPL \geq 133$ are 0.088 (SE=0.033) and 0.184 (0.003), respectively. Thus, as expected, higher state income thresholds, measured as a percentage of the federal poverty line, yield higher fractions of children that are eligible for Medicaid. The joint F-statistic for these instruments exceeds $2,500$.

Table 2: Empirical Application - Medicaid Eligibility and Health Insurance Crowd-Out

Outcome:	Medicaid (1)	Private (2)	Uninsured (3)
(A) 2SLS: Cutler and Gruber (1996)	0.235	-0.074	-0.119
	(.017)	(.021)	(.018)
(B) 2SLS: Replication	0.236	-0.101	-0.089
	(.016)	(.021)	(.018)
(C) 2SLS: Discrete instrument	0.186	-0.053	-0.091
	(.015)	(.019)	(.016)
(D) OLS	0.358	-0.403	0.084
	(.003)	(.003)	(.003)
(E) Our estimator with equality moments	0.151	-0.032	-0.037
	(.023)	(.025)	(.016)
(F) Our estimator with inequality moments			
$\underline{R}_{jk} = 0.95, \underline{R}_{kj} = .95 \ \forall j > k$	[0.11, 0.32]	[-0.56, -0.02]	[-0.08, -0.03]
$\underline{R}_{jk} = 1, \underline{R}_{kj} = 0.95 \ \forall j > k$	[0.12, 0.23]	[-0.65, -0.02]	[-0.07, -0.03]
$\underline{R}_{jk} = 1, \underline{R}_{kj} = 0.89 \ \forall j > k$	[0.11, 0.33]	[-0.79, -0.01]	[-0.07, -0.03]

Notes: This table reports estimates of β from equation (18). Standard errors robust to heteroskedasticity are reported in the first four rows while those in the fifth row are bootstrapped. All estimates use the CPS sampling weights. Each column uses a different insurance coverage outcome (Medicaid, Private, No Insurance). The CPS sample from 1988 to 1993 (calendar years 1987 to 1992) contains 266,421 children. The analysis in rows (A) and (B) use the Cutler and Gruber (1996) simulated instrument. The analysis in the rows (C), (E), and (F) use the discrete, three-valued instrument discussed in the text. The results in row (F) present the 95% confidence intervals using inequality moments with different choices for $\underline{R}_{jk}$ and $\underline{R}_{kj}$. The controls include the number of people and workers in the household and indicators for state of residence, calendar year, child age, and householder characteristics (male, white, and type).

are substantially larger in magnitude than the 2SLS estimates in the first two columns and, in the final column, is opposite in sign. These results suggest that $ELIG$ is endogenous, consistent with Cutler and Gruber's main motivation for using a 2SLS strategy.

We present results using our estimator with equality moments in row (E) of Table 2.[32] For all three outcomes, our estimates of β are the same sign but smaller in magnitude, substantially so for private insurance coverage and being uninsured, as the corresponding 2SLS

[32]With endogenous T^*, our estimator is identified under Corollary 1.2 after accounting for covariates as in Appendix Section A.6.2.

estimate. Thus, accounting for mis-measured Medicaid eligibility has important implications for estimates of the extent to which Medicaid eligibility impacts the use of health insurance.

Finally, we apply our inequality moment estimator (equations (16) and (17)) and present results for three sets of bounds. The first set of bounds are symmetric with $\underline{R}_{jk} = \underline{R}_{kj} = 0.95\ \forall j > k$ following from the discussion in section 4.3 based on Figure 3a. Next, noting that in the sample we observe $p_0 = 0.15$, $p_1 = 0.29$, and $p_2 = 0.44$ which, based on Figure 3a, suggests that $\overline{\alpha}_j \geq \overline{\alpha}_k,\ \forall j < k$, the second set of bounds are asymmetric with $\underline{R}_{jk} = 1, \underline{R}_{kj} = 0.95\ \forall j > k$. The third set of bounds are also asymmetric, $\underline{R}_{jk} = 1, \underline{R}_{kj} = 0.89\ \forall j > k$, to allow for larger differences in the misclassification rates between instrument values.[33]

The final row of Table 2 shows 95% confidence sets for β using our inequality moment estimator. The confidence sets when Medicaid is the outcome change as expected, getting larger when $\underline{R}_{jk}$ gets smaller (i.e., when the differences in the misclassification rates between instrument values are allowed to be larger) and when we use the asymmetric bounds.[34] Interestingly, the confidence sets for the no insurance outcome (column (3)) are relatively small and vary little across the three sets of bounds with the 2SLS point estimate falling outside of all these confidence sets. However, the confidence sets for the private insurance outcome (column (2)) are large, driven by a long tail in the confidence curve (see Appendix Section A.8). Overall, our findings suggest that allowing for realistic bounds on relative misclassification rates across instrument values can produce informative confidence sets.

[33] $\underline{R}_{kj} = 0.89$ is based on Figure 3a when $P[T^* = 1]$ lies in $[.09, .99]$.

[34] As a point of reference, when we use inequality bounds with $\underline{R}_{jk} = \underline{R}_{kj} = 1\ \forall j > k$ which imposes the same restriction as Assumption 3, the 95% confidence set for the Medicaid outcome is [0.13, 0.20]. See Appendix Section A.8 for additional discussion and results using our inequality moment estimator.

7 Conclusion

In this paper, we add a new dimension to estimation with a misclassified, binary regressor: allowing the misclassification rates to vary across values of the instrumental variable. We first show that such variability both arises in empirically relevant settings and overturns some key results relied upon in the previous literature. We derive a new estimator that matches the conditions found in these empirical settings. We extend our analytic results along several dimensions, including demonstrating how to include covariates and extending the main results to relax the exogeneity assumption. We demonstrate the usefulness of our estimator with Monte Carlo evidence and by applying our method to the analysis of the impact of Medicaid eligibility on private health insurance. Our reexamination of the pioneering work of Cutler and Gruber (1996) yields a notably smaller impact of Medicaid eligibility on reducing the share of children without insurance after accounting for misclassification error in Medicaid eligibility, their key regressor.

One appealing feature of our identifying assumption, $\alpha_{0j}+\alpha_{1j} = \overline{\alpha},\ \forall j$, is that it nests the standard fixed misclassification assumption. Alternative features of varying misclassification could be exploited in order to identify the model, such as parameterizing α_0 and/or α_1 as function(s) of p^*. Such an approach will almost certainly not nest the fixed misclassification model. Future work should explore the fruitfulness of alternative modeling choices for varying misclassification given the empirical results we present here.

8 References

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

Andrews, D. W. K. and Soares, G. (2010) "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78:119-157.

Angrist, J. D. and Krueger, A. B. (1999) "Empirical Strategies in Labor Economics," in Orley Ashenfelter and David Card, eds., Handbook of Labor Economics, Vol 3A, (North-Holland, Amsterdam) 1277-1366.

Battistin, E., De Nadai, M., and Sianesi, B. (2014) "Misreported Schooling, Multiple Measures and Returns to Educational Qualifcations," *Journal of Econometrics*, 181:136-150.

Black, D. A., Berger, M.C., and Scott, F.A. (2000) "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, 95(451): 739-748.

Black, D. A., Joo, J., LaLonde, R., Smith, J.A., and Taylor, E.J. (2015) "Simple Tests for Selection Bias: Learning More from Instrumental Variables," *IZA DP No. 9346*.

Brinch, C. N., Mogstad, M., and Wiswall, M. (2017) "Beyond LATE with a Discrete Instrument," *Journal of Political Economy*, 125(4): 985-1039.

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

Bound, J., Brown, C.C., and Mathiowetz, N. (2001) "Measurement error in survey data," in Handbook of Econometrics. Vol. 5, Elsevier, 3705-3843.

Brown, D. W., Kowalski, A.E., and Zurie, I. (2020) "Long-Term Impacts of Childhood Medi-

caid Expansions on Outcomes in Adulthood," *Review of Economic Studies*, 87(2):792-821.

Bugni, F. A., Canay, I.A., and Shi, X. (2017) "Inference for subvectors and other functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 8, 1-38.

Calvi, R., Lewbel, A., and Tommasi, D. (2017) "LATE With Mismeasured or Misspecified Treatment: An Application To Women's Empowerment in India"

Card, D. (1996) "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64(4):957-979.

Chen, X., Hu, Y. and Lewbel, A. (2008a) "A note on the closed-form identification of regression models with a mismeasured binary regressor," *Statistics & Probability Letters*, 78(12):1473-1479.

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

Currie, J. and Gruber, J. (1996) "Health Insurance Eligibility, Utilization of Medical Care, and Child Health," *Quarterly Journal of Economics*, 111(2):431-466.

Cutler, D, M. and Gruber, J. (1996) "Does Public Insurance Crowd out Private Insurance?" *Quarterly Journal of Economics*, 111(2):391-430.

DiTraglia, F. J. and Garcia-Jimeno, C. (2019) "Identifying the Effect of a Mis-classified, Binary, Endogenous Regressor," *Journal of Econometrics*, 209:376-390.

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

Frean, M., Gruber, J., and Sommers, B.D. (2017) "Premium subsidies, the mandate, and Medicaid Expansion: Coverage effects of the Affordable Care Act," *Journal of Health Economics*, 53:72-86.

Gross, T. and Notowidigdo, M. (2011) "Health Insurance and the Consumer Bankruptcy Decision: Evidence from Medicaid Expansions," *Journal of Public Economics*, 95(7-8):767-778.

Haider, S.J. and Stephens Jr., M. (2020)"Correcting for Misclassified Binary Regressors Using Instrumental Variables," *National Bureau of Economic Research WP No. 27797.*

Hu, Y. (2008) "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144:27-61.

Jiang, Z. and Ding, P. (2020): "Measurement Errors in the Binary Instrumental Variable Model," *Biometrika*, 107, 238-245.

Kane, T. J., Rouse, C.E., and Staiger, D. (1999) "Estimating Returns to Schooling When Schooling is Misreported," *National Bureau of Economic Research WP No. 7235.*

Klepper, S. (1988) "Bounding the Effects of Measurement Error In Regressions Involving Dichotomous Variables," *Journal of Econometrics*, 37:343-359.

Lewbel, A. (2007) "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75(2):537-551.

Mahajan, A. (2006) "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74(3):631-665.

Newey, W. K. (1984) "A Method of Moments Interpretation of Sequential Estimators," *Economics Letters*, 14:201-6.

Nguimkeu, P., Denteh, A. and Tchernis, R. (2019) "On the Estimation of Treatment Effects with Endogenous Misreporting," *Journal of Econometrics,* 208:487-506.

Olsen, R. J. (1980) "A Least Squares Correction for Selectivity Bias," *Econometrica*, 48(7): 1815-1820.

Shore-Sheppard, L. (2008) "Stemming the Tide? The Effect of Expanding Medicaid Eligibility on Health Insurance Coverage" *B.E. Journal of Economic Analysis and Policy: Advances* 8, 2.

Social Security Administration. Current Population Survey, 1973, and Social Security Records: Exact Match Data. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005-11-04.

Tommasi, D. and Zhang, L. (2020) "Bounding Program Benefits When Participation Is Misreported," IZA DP No. 13430.

Ura, T. (2018) "Heterogeneous Treatment Effects with Mismeasured Endogenous Treatment," *Quantitative Economics*, 9(3):1335-1370.

Yanagi, T. (2019) "Inference on Local Average Treatment Effects for Misclassified Treatment," *Econometric Reviews*, 38(8):938-960.

A Appendix: Supplementary Material

A.1 Details for Misclassification Examples

A.1.1 Dataset Descriptions

We use the March 1973 CPS that is linked both to Social Security earnings records and to tax returns filed with the IRS. The earnings data used in our examples refer to the prior calendar year, 1972. The available SS earnings records are top-coded at the annual ceiling for earnings subject to SS tax, which is $9,000 in 1972. Nearly half of privately employed men in the sample have top-coded SS earnings data. On the other hand, the IRS earnings records and self-reported earnings in the CPS are top-coded at $50,000, which affects less than one-half of one percent of privately employed men in the data.

A drawback to the available IRS earnings data is that these come directly from tax filings in which only the combined earnings of all household members is reported, not individual eanrings. To circumvent this issue, we restrict our sample to men and women who designate their household status as single on the tax return such that the IRS earnings amount should only reflect their own earnings. In addition, we restrict the analysis to men and women who are report to the CPS that they are privately employed, have positive IRS and CPS earnings, and have non-imputed CPS earnings. We also restrict the data to individuals in households that are deemed to be "good" matches (where the variable V1255 is not equal to 4) and to individuals with matches to available IRS data (the variable V1253 equals 0). Our final sample includes 8,031 single men and women. All our analyses are weighted using the "final" CPS-IRS-SSA STATS unit administrative weight (the variable V1264), which account for selection into the CPS and the match between the CPS and the administrative records.

We also use data from a special supplement to the January 1977 CPS. A subsample of survey respondents was asked additional questions regarding union status, earnings, and hours worked. These individuals were also asked to provide the name and address of their employer. Their employers were subsequently asked to provide information for the survey respondent, including the aforementioned variables. These data have been used previously to examine misclassification in union status (Freeman 1984; Card 1996) and measurement error in wages (Mellow and Sider 1983; Angrist and Krueger 1999). See Mellow and Sider (1983) for a more detailed data descxription. We thank David Card, Hank Farber, and Alan Krueger for making this data available. In particular, Alan Krueger graciously provided the code from Angrist and Krueger (1999) allowing us to employ the same sample restrictions.

We also use data from the 1999-2016 waves of the Continuous National Health and Nutrition Examination Survey, which is a survey of the health and diet of Americans conducted by the National Center for Health Statistics and is conducted on a two-year cycle. This survey replaced the prior National Health and Nutrition Examination Surveys which had previously been fielded on an idiosyncratic schedule. Households and sample persons are selected using a stratified, multistage sampling design for each cycle. A screener and basic questionnaire are completed as part of initial in-home visits. A randomizing computer algorithm selects a sample person from each household roster. Survey respondents complete a questionnaire covering demographic, dietary, socioeconomic, and health topics. During this in-home interview, respondents are also asked to provide their self-reported height (in inches) and weight (in pounds). Following the in-home interview, the selected sample persons make appointments to visit a Mobile Examination Center for a detailed physical examination about two weeks later. During this subsequent physical examination, the individual's height

(in centimeters) and weight (in kilograms) are measured by survey staff. All of our analyses are weighted to account for selection into the NHANES and completing both the in-home interview and the physical exam (the variable WTMEC2YR).

The final data set that we employ is for the empirical application of our estimator in which we focus on the impact of Medicaid eligibility on Medicaid take-up. We re-examine the analysis of Cutler and Gruber (1996) using programs and data kindly provided by Lara Shore-Sheppard. We use a sample which contains 266,421 children from the 1988-1993 CPS. Details of the construction of the dataset and variables can be found in Shore-Sheppard (2008). Key features of the legislative rules and changes include: income thresholds for receipt of Aid to Families with Dependent Children (AFDC) benefits which vary by state over time, the Omnibus Budget Reconciliation Act (OBRA) of 1989 which covered children ages six and under in families with incomes up to 133 percent of the FPL, OBRA 1990 which covered children born after September 30, 1983 with family incomes below 100 percent of the FPL, and the implementation and expansion of state optional programs.

A.1.2 Further Evidence on Misclassification

We find evidence of varying misclassification, comparable to what appears in Figures 1 and 2, when we construct hypothetical program eligibility measures using wage data from matched employer-employer reports in the January 1977 CPS. The top panel of Figure A1 shows that the misclassification rates are varying as the threshold moves through the wage distribution. Interestingly, the bottom panel of Figure A1 shows that changes in $P[T = 1]$ are nearly identical to changes in $P[T^* = 1]$, suggesting that IV using T rather than T^* will yield roughly consistent estimates throughout much of the wage distribution.

Figure A2 shows height results separately for men and women in the NHANES.[35] The misclassification rates for men, shown in the top left panel, vary with $P[T^* = 1]$ in ways that closely mirror those found in the previous two examples. Moreover, the relative movements in $P[T = 1]$ and $P[T^* = 1]$ are comparable to the first example in that slope of the line is less than one near the middle of the distribution but exceeds one towards the ends of the distribution. For women, as shown in the top right panel of Figure A2, α_0 exceeds α_1 much further to the left in the distribution. As shown in the bottom right panel, underreporting of women's height occurs, on average, throughout the distribution as we find that $P[T = 1]$ always exceeds $P[T^* = 1]$. The relationship between $P[T = 1]$ and $P[T^* = 1]$ moves much further from the 45 degree line than in the other examples, which implies much larger deviations between the change in $P[T = 1]$ and the change in $P[T^* = 1]$. At various points in the distribution, IV will generate larger underestimates and overestimates as compared to the previous examples.

The patterns found for weight in the NHANES, as shown in Figure A3, are broadly comparable to those for height with regards to the relationships between $P[T^* = 1]$ and the misclassifications and in that the relationship between $P[T = 1]$ and $P[T^* = 1]$ deviates further from the 45 degree line for women than for men.

[35]We limit the sample to men and women between the ages of 25 and 54, inclusive. For each outcome, we restrict the sample to individuals who have both self-reported and measured data. Our height samples have 11,004 men and 12,100 women while our weight sample has 10,694 men and 11,756 women. We did not include weight observations for the small fraction of individuals who were flagged for being weighed while wearing their clothing.

A.2 Derivation of Equations (5) and (6)

For a given value of Z, the expected value of the outcome when $T = 1$ is

$$\begin{aligned}
E[Y|T=1,Z=j] &= E[Y|T^*=0,T=1,Z=j]\cdot P[T^*=0|T=1,Z=j] \\
&\quad + E[Y|T^*=1,T=1,Z=j]\cdot P[T^*=1|T=1,Z=j] \\
&= E[Y|T^*=0,Z=j]\cdot P[T^*=0|T=1,Z=j] \\
&\quad + E[Y|T^*=1,Z=j]\cdot P[T^*=1|T=1,Z=j] \\
&= E[Y|T^*=0,Z=j]\cdot \frac{P[T=1|T^*=0,Z=j]\cdot P[T^*=0|Z=j]}{P[T=1|Z=j]} \\
&\quad + E[Y|T^*=1,Z=j]\cdot \frac{P[T=1|T^*=1,Z=j]\cdot P[T^*=1|Z=j]}{P[T=1|Z=j]} \\
&= E[Y|T^*=0,Z=j]\cdot \frac{\alpha_{0j}\left(1-p_j^*\right)}{p_j} + E[Y|T^*=1,Z=j]\cdot \frac{(1-\alpha_{1j})\,p_j^*}{p_j}
\end{aligned} \tag{19}$$

The second equality follows from the non-differential measurement error assumption and the third equality follows from Bayes Theorem. Similarly,

$$E[Y|T=0,Z=j] = E[Y|T^*=0,Z=j]\cdot \frac{(1-\alpha_{0j})\left(1-p_j^*\right)}{1-p_j} + E[Y|T^*=1,Z=j]\cdot \frac{\alpha_{1j}\cdot p_j^*}{1-p_j} \tag{20}$$

Sovling for $E[Y|T^*=1,Z=j]$ and $E[Y|T^*=0,Z=j]$ in terms of observed ($E[Y|T=0,Z=j]$, $E[Y|T=1,Z=j]$, and p_j) and unobserved quantities (α_{0j}, α_{1j}, and p_j^*) yields

$$\begin{aligned}
E[Y|T^*=1,Z=j] &= \frac{p_j E[Y|T=1,Z=j] - \alpha_{0j}E[Y|Z=j]}{p_j^*\left(1-\alpha_{0j}-\alpha_{1j}\right)} \\
E[Y|T^*=0,Z=j] &= \frac{(1-p_j)\,E[Y|T=0,Z=j] - \alpha_{1j}E[Y|Z=j]}{\left(1-p_j^*\right)\left(1-\alpha_{0j}-\alpha_{1j}\right)}
\end{aligned}$$

which use the substitution $E[Y|Z=j] = E[Y|T=1,Z=j]\,p_j + E[Y|T=0,Z=j]\,(1-p_j)$.

Noting that $p_j - \alpha_{0j} = (1-\alpha_{0j}-\alpha_{1j})\,p_j^*$, which follows immediately from the equation for p_j, we can simplify the expression for $E[Y|T^*=1,Z=j]$ to yield

$$\begin{aligned}
E[Y|T^*=1,Z=j] &= \frac{p_j E[Y|T=1,Z=j] - \alpha_{0j} E[Y|Z=j]}{p_j^*(1-\alpha_{0j}-\alpha_{1j})} \\
&= \frac{p_j E[Y|T=1,Z=j] - \alpha_{0j} E[Y|Z=j]}{p_j - \alpha_{0j}} \\
&= \frac{p_j E[Y|T=1,Z=j] - \alpha_{0j} E[Y|Z=j] + (p_j E[Y|Z=j] - p_j E[Y|Z=j])}{p_j - \alpha_{0j}} \\
&= \frac{p_j E[Y|T=1,Z=j] - p_j E[Y|Z=j] + (p_j - \alpha_{0j})\,E[Y|Z=j]}{p_j - \alpha_{0j}} \\
&= \frac{E[YT|Z=j] - p_j E[Y|Z=j]}{p_j - \alpha_{0j}} + E[Y|Z=j] \\
&= \frac{COV(Y,T|Z=j)}{p_j - \alpha_{0j}} + E[Y|Z=j] \qquad (21)
\end{aligned}$$

where we make use of the fact that

$$E[YT|Z=j] = p_j E[Y\cdot 1|T=1,Z=j] + (1-p_j)\,E[Y\cdot 0|T=0,Z=j] = p_j E[Y|T=1,Z=j]$$

The following expression follows similarly

$$E[Y|T^*=0,Z=j] = E[Y|Z=j] - \frac{COV(Y,T|Z=j)}{1-p_j-\alpha_{1j}} \qquad (22)$$

A.3 Derivation of Equation (7)

Subtracting equation (4) from equation (3) and then inserting equations (5) and (6) into the resulting expression yields

$$E\left[Y|T^*=1,Z=j\right]-E\left[Y|T^*=0,Z=j\right]=E\left[Y|T^*=1,Z=k\right]-E\left[Y|T^*=0,Z=k\right]$$

$$\frac{COV\left(Y,T|Z=j\right)}{p_j-\alpha_{0j}}+\frac{COV\left(Y,T|Z=j\right)}{1-p_j-\alpha_{1j}}=\frac{COV\left(Y,T|Z=k\right)}{p_k-\alpha_{0k}}+\frac{COV\left(Y,T|Z=k\right)}{1-p_k-\alpha_{1k}}$$

$$\frac{COV\left(Y,T|Z=j\right)}{\left(1-\alpha_{0j}-\alpha_{1j}\right)p_j^*}+\frac{COV\left(Y,T|Z=j\right)}{\left(1-\alpha_{0j}-\alpha_{1j}\right)\left(1-p_j^*\right)}=\frac{COV\left(Y,T|Z=k\right)}{\left(1-\alpha_{0k}-\alpha_{1k}\right)p_k^*}+\frac{COV\left(Y,T|Z=k\right)}{\left(1-\alpha_{0k}-\alpha_{1k}\right)\left(1-p_k^*\right)}$$

$$\frac{1}{\left(1-\alpha_{0j}-\alpha_{1j}\right)}\cdot\frac{COV\left(Y,T|Z=j\right)}{p_j^*\left(1-p_j^*\right)}=\frac{1}{\left(1-\alpha_{0k}-\alpha_{1k}\right)}\cdot\frac{COV\left(Y,T|Z=k\right)}{p_k^*\left(1-p_k^*\right)}$$

where we also make use of the definition of p_j for the following substitutions $p_j-\alpha_{0j}=\left(1-\alpha_{0j}-\alpha_{1j}\right)p_j^*$ and $1-p_j-\alpha_{1j}=\left(1-\alpha_{0j}-\alpha_{1j}\right)\left(1-p_j^*\right)$.

A.4 Proof of Theorem 1

Let the instrument Z take on three values, $j=0,1,2$. Re-writing (2) in terms of β and then equating this result for the cases $(j=1,k=0)$ and $(j=2,k=1)$ yields

$$\frac{E\left[Y|Z=1\right]-E\left[Y|Z=0\right]}{p_1^*-p_0^*}=\frac{E\left[Y|Z=2\right]-E\left[Y|Z=1\right]}{p_2^*-p_1^*}$$

Letting $\Delta_{jk}=E\left[Y|Z=j\right]-E\left[Y|Z=k\right]$, we can solve this expression for p_1^*

$$p_1^*=\frac{p_2^*\Delta_{10}+p_0^*\Delta_{21}}{\Delta_{21}+\Delta_{10}} \tag{23}$$

which we can use to derive an expression for $p_1^*(1-p_1^*)$

$$\begin{aligned}
p_1^*(1-p_1^*) &= \left(\frac{p_2^*\Delta_{10}+p_0^*\Delta_{21}}{\Delta_{21}+\Delta_{10}}\right)\left(\frac{\Delta_{21}+\Delta_{10}}{\Delta_{21}+\Delta_{10}}-\left(\frac{p_2^*\Delta_{10}+p_0^*\Delta_{21}}{\Delta_{21}+\Delta_{10}}\right)\right) \\
&= \left(\frac{p_2^*\Delta_{10}+p_0^*\Delta_{21}}{\Delta_{21}+\Delta_{10}}\right)\left(\frac{(1-p_2^*)\Delta_{10}+(1-p_0^*)\Delta_{21}}{\Delta_{21}+\Delta_{10}}\right) \\
&= \left(\frac{1}{\Delta_{20}}\right)^2\left(p_2^*(1-p_2^*)\Delta_{10}^2+p_0^*(1-p_0^*)\Delta_{21}^2+\Delta_{10}\Delta_{21}\left(p_2^*(1-p_0^*)+p_0^*(1-p_2^*)\right)\right)
\end{aligned} \tag{24}$$

where the last step uses $\Delta_{20}=\Delta_{21}+\Delta_{10}$. Next, we can re-write equation (8) for the case $j=1,k=0$ to yield

$$p_1^*(1-p_1^*) = p_0^*(1-p_0^*)\cdot\frac{C_1}{C_0} \tag{25}$$

where $C_j = COV(Y,T|Z=j)$. Inserting (24) into (25) and simplifying yields

$$\begin{aligned}
&\left(\frac{1}{\Delta_{20}}\right)^2\left(p_2^*(1-p_2^*)\Delta_{10}^2+p_0^*(1-p_0^*)\Delta_{21}^2+\Delta_{10}\Delta_{21}\left(p_2^*(1-p_0^*)+p_0^*(1-p_2^*)\right)\right) = p_0^*(1-p_0^*)\cdot\frac{C_1}{C_0} \\
&p_2^*(1-p_2^*)\Delta_{10}^2+p_0^*(1-p_0^*)\left(\Delta_{21}^2-\Delta_{20}^2\frac{C_1}{C_0}\right)+\Delta_{10}\Delta_{21}\left(p_2^*(1-p_0^*)+p_0^*(1-p_2^*)\right)=0
\end{aligned} \tag{26}$$

Similarly, we can re-write equation (8) for the case $(j=2,k=1)$

$$p_1^*(1-p_1^*) = p_2^*(1-p_2^*)\cdot\frac{C_1}{C_2} \tag{27}$$

which we can combine with (24) to yield

$$\begin{aligned}
&\left(\frac{1}{\Delta_{20}}\right)^2\left(p_2^*(1-p_2^*)\Delta_{10}^2+p_0^*(1-p_0^*)\Delta_{21}^2+\Delta_{10}\Delta_{21}\left(p_2^*(1-p_0^*)+p_0^*(1-p_2^*)\right)\right) = p_2^*(1-p_2^*)\cdot\frac{C_1}{C_2} \\
&p_2^*(1-p_2^*)\left(\Delta_{10}^2-\Delta_{20}^2\frac{C_1}{C_2}\right)+p_0^*(1-p_0^*)\Delta_{21}^2+\Delta_{10}\Delta_{21}\left(p_2^*(1-p_0^*)+p_0^*(1-p_2^*)\right)=0
\end{aligned} \tag{28}$$

To find the solutions to the remaining two unknowns, p_0^* and p_2^* given the two equations (26) and (28), we proceed in the following steps. First, we re-write these two equations as quadratics in p_0^*. We can then compute the resultant using the coefficients on p_0^* in the re-written equations. As these coefficients are only a function of p_2^*, the resultant also is only a function of p_2^*. Since the system of equations has a non-zero solutions if and only if the resultant equals zero, we set the resultant equal to zero and solve for roots of p_2^*. We then repeat the process by re-writing equations (26) and (28) as quadratics in p_2^* and solve that resultant for roots of p_0^*. Finally, we substitute all combinations of the roots of p_0^* and p_2^* into (26) and (28) to find the pair(s) of roots that satisfy these equations.

Re-writing both (26) and (28) as quadratics in p_0^* yields

$$\left(\Delta_{20}^2\frac{C_1}{C_0}-\Delta_{21}^2\right)p_0^{*2}+\left(\Delta_{21}^2-\Delta_{20}^2\frac{C_1}{C_0}+\Delta_{10}\Delta_{21}-2\Delta_{10}\Delta_{21}p_2^*\right)p_0^* + \left(\left(\Delta_{10}^2+\Delta_{10}\Delta_{21}\right)p_2^*-\Delta_{10}^2p_2^{*2}\right)=0 \tag{29}$$

$$-\Delta_{21}^2p_0^{*2}+\left(\Delta_{21}^2+\Delta_{10}\Delta_{21}-2\Delta_{10}\Delta_{21}p_2^*\right)p_0^* + \left(\left(\Delta_{10}^2-\Delta_{20}^2\frac{C_1}{C_2}+\Delta_{10}\Delta_{21}\right)p_2^*+\left(\Delta_{20}^2\frac{C_1}{C_2}-\Delta_{10}^2\right)p_2^{*2}\right)=0 \tag{30}$$

Denoting the coefficients of the system of quadratic equations formed by (29) and (30)

by

$$
\begin{aligned}
A_{00} &= \Delta_{20}^2 \frac{C_1}{C_0} - \Delta_{21}^2 \\
A_{01} &= \Delta_{21}^2 - \Delta_{20}^2 \frac{C_1}{C_0} + \Delta_{10}\Delta_{21} - 2\Delta_{10}\Delta_{21}p_2^* \\
A_{02} &= \left(\Delta_{10}^2 + \Delta_{10}\Delta_{21}\right) p_2^* - \Delta_{10}^2 p_2^{*2} \\
B_{00} &= -\Delta_{21}^2 \\
B_{01} &= \Delta_{21}^2 + \Delta_{10}\Delta_{21} - 2\Delta_{10}\Delta_{21}p_2^* \\
B_{02} &= \left(\Delta_{10}^2 - \Delta_{20}^2 \frac{C_1}{C_2} + \Delta_{10}\Delta_{21}\right) p_2^* + \left(\Delta_{20}^2 \frac{C_1}{C_2} - \Delta_{10}^2\right) p_2^{*2}
\end{aligned}
$$

we can form the Sylvester matrix, $S_{p_0^*}$, of this system of equations in p_0^*

$$
\begin{bmatrix}
A_{00} & A_{01} & A_{02} & 0 \\
0 & A_{00} & A_{01} & A_{02} \\
B_{00} & B_{01} & B_{02} & 0 \\
0 & B_{00} & B_{01} & B_{02}
\end{bmatrix}
$$

Since the resultant can be formed as the determinant of the Sylvester matrix, taking the

determinant of $S_{p_0^*}$ and setting it equal to zero yields

$$\frac{C_1^2\Delta_{20}^4\left(C_1^2\Delta_{20}^4+(C_2\Delta_{10}^2-C_0\Delta_{21}^2)^2-2C_1\Delta_{20}^2(C_2\Delta_{10}^2+C_0\Delta_{21}^2)\right)}{C_0^2C_2^2}\cdot p_2^{*4}$$
$$-\frac{2C_1^2\Delta_{20}^4\left(C_1^2\Delta_{20}^4+(C_2\Delta_{10}^2-C_0\Delta_{21}^2)^2-2C_1\Delta_{20}^2(C_2\Delta_{10}^2+C_0\Delta_{21}^2)\right)}{C_0^2C_2^2}\cdot p_2^{*3}$$
$$+\frac{1}{C_0^2C_2^2}\Big[C_1^2\Delta_{20}^4\left(C_0^2\Delta_{21}^4+C_1^2\Delta_{20}^4-C_0C_2\Delta_{10}\Delta_{21}^2(3\Delta_{20})+C_2^2\Delta_{10}^2(\Delta_{10}^2-\Delta_{10}\Delta_{21}-\Delta_{21}^2)\right.$$
$$\left.-C_1\Delta_{20}^2(C_2\Delta_{10}(2\Delta_{10}-\Delta_{21})+2C_0\Delta_{21}^2)\right)\Big]\cdot p_2^{*2}$$
$$+\frac{C_1^2\Delta_{10}\Delta_{21}\Delta_{20}^5(C_2\Delta_{10}+C_0\Delta_{21}-C_1\Delta_{20})}{C_0^2C_2}\cdot p_2^*=0 \tag{31}$$

Following an analogous process, we re-write (26) and (28) as quadratics in p_2^*, form the Sylvester matrix $S_{p_2^*}$, take its determinant and set it equal to zero to yield

$$\frac{C_1^2\Delta_{20}^4(C_1^2\Delta_{20}^4+(C_2\Delta_{10}^2-C_0\Delta_{21}^2)^2-2C_1\Delta_{20}^2(C_2\Delta_{10}^2+C_0\Delta_{21}^2))}{C_0^2C_2^2}\cdot p_0^{*4}$$
$$-\frac{2C_1^2\Delta_{20}^4\left(C_1^2\Delta_{20}^4+(C_2\Delta_{10}^2-C_0\Delta_{21}^2)^2-2C_1\Delta_{20}^2(C_2\Delta_{10}^2+C_0\Delta_{21}^2)\right)}{C_0^2C_2^2}\cdot p_0^{*3}$$
$$+\frac{1}{C_0^2C_2^2}\Big[C_1^2\Delta_{20}^4\left(C_2^2\Delta_{10}^4+C_1^2\Delta_{20}^4-C_0C_2\Delta_{10}^2\Delta_{21}(\Delta_{10}+3\Delta_{21})+C_0^2\Delta_{21}^2(-\Delta_{10}^2-\Delta_{10}\Delta_{21}+\Delta_{21}^2)\right.$$
$$\left.-C_1\Delta_{20}^2(2C_2\Delta_{10}^2+C_0\Delta_{21}(-\Delta_{10}+2\Delta_{21}))\right)\Big]\cdot p_0^{*2}$$
$$+\frac{C_1^2\Delta_{10}\Delta_{21}\Delta_{20}^5(C_2\Delta_{10}+C_0\Delta_{21}-C_1\Delta_{20})}{C_0C_2^2}\cdot p_0^*=0 \tag{32}$$

We find that (31) yields four roots of p_2^* and (32) yields four roots of p_0^*. We then use all combinations of these roots to determine which (p_0^*, p_2^*) root pairs satisfy (26) and (28).

Two of the four roots for both p_0^* and p_2^* are 0 and 1. When either p_0^* or p_2^* equals 0, the solution to the system of equations is $p_0^*=0$ and $p_2^*=0$. Similarly, when either p_0^* or p_2^* equals 1, the solution to the system of equations is $p_0^*=1$ and $p_2^*=1$. However, these two

pairs of solutions violate Assumption 1.iii that $0 < p_j^* < 1,\ \forall j$ and $p_j^* \neq p_k^*,\ j \neq k$. Thus, neither 0 nor 1 is a possible solution for either p_0^* or p_2^*.

The two pairs of roots which satisfy (26) and (28) for which $p_j^* \neq p_k^*,\ j \neq k$, along with corresponding solution for p_1^* computed by inserting each (p_0^*, p_2^*) pair into (23), are

$$\begin{aligned} p_0^* &= \frac{D_3 + \sqrt{D_3}\Big((C_1 - C_0)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{10}^2 \Big)}{2D_3} \\ p_1^* &= \frac{D_3 + \sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_1 - C_0)\,\Delta_{21}^2 \Big)}{2D_3} \\ p_2^* &= \frac{D_3 + \sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{21}^2 \Big)}{2D_3} \end{aligned} \tag{33}$$

and

$$\begin{aligned} p_0^* &= \frac{D_3 - \sqrt{D_3}\Big((C_1 - C_0)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{10}^2 \Big)}{2D_3} \\ p_1^* &= \frac{D_3 - \sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_1 - C_0)\,\Delta_{21}^2 \Big)}{2D_3} \\ p_2^* &= \frac{D_3 - \sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{21}^2 \Big)}{2D_3} \end{aligned} \tag{34}$$

where $D_0 = C_0\Delta_{21}^2$, $D_1 = C_1\Delta_{20}^2$, $D_2 = C_2\Delta_{10}^2$, and $D_3 = D_0^2 + D_1^2 + D_2^2 - 2D_0D_1 - 2D_0D_2 - 2D_1D_2$.

Both sets of solutions yield real numbers if $D_3 > 0$. We prove that is the case below in section A.4.1.

We can next solve for β by selecting any pair of p_j^* and substituting the appropriate values into equation (2). Without loss of generality, using the first solution found in equations (33) along with the fact that $\Delta_{20}^2 = (\Delta_{21} + \Delta_{10})^2 = \Delta_{21}^2 + \Delta_{10}^2 + 2\Delta_{21}\Delta_{10}$, we can first compute

$p_1^* - p_0^*$

$$
\begin{aligned}
p_1^* - p_0^* &= \frac{\sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_1 - C_0)\,\Delta_{21}^2 \Big)}{2D_3} - \frac{\sqrt{D_3}\Big((C_1 - C_0)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{10}^2 \Big)}{2D_3} \\
&= \frac{\sqrt{D_3}\Big((2C_2 - C_1 - C_0)\,\Delta_{10}^2 + (C_1 - C_0)\,(\Delta_{21}^2 - \Delta_{20}^2) \Big)}{2D_3} \\
&= \frac{\sqrt{D_3}\Big(2\,(C_2 - C_1)\,\Delta_{10}^2 + 2\,(C_0 - C_1)\,(\Delta_{21}\Delta_{10}) \Big)}{2D_3} \\
&= \frac{\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_0 - C_1)\,(\Delta_{21}\Delta_{10}) \Big)}{\sqrt{D_3}}
\end{aligned}
\tag{35}
$$

Thus, for the first solution found in equations (33), re-writing equation (2) to solve for β and then inserting the above result for $p_1^* - p_0^*$ yieids

$$
\begin{aligned}
\beta &= \frac{\Delta_{10}}{p_1^* - p_0^*} \\
&= \frac{\sqrt{D_3}\Delta_{10}}{\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_0 - C_1)\,(\Delta_{21}\Delta_{10}) \Big)} \\
&= \frac{\sqrt{D_3}}{\Big((C_2 - C_1)\,\Delta_{10} + (C_0 - C_1)\,\Delta_{21} \Big)}
\end{aligned}
\tag{36}
$$

Similarly, we can show that β corresponding to the second solution found in equations (34), is equal and oppose in sign from the result in equation (36), i.e.,

$$
\beta = -\frac{\sqrt{D_3}}{\Big((C_2 - C_1)\,\Delta_{10} + (C_0 - C_1)\,\Delta_{21} \Big)} \tag{37}
$$

To solve for $\overline{\alpha}$, notice that under the Assumption 2, an OLS regression of Y on T^* will yield a consistent estimate of β. Moreover, a regression of Y on T^* conditional on any value

taken on by the instrument will also consistently estimate β. Therefore,

$$\beta = \frac{COV\left(Y, T^* | Z = j\right)}{VAR\left(T^* | Z = j\right)} = \frac{COV\left(Y, T^* | Z = j\right)}{p_j^*\left(1 - p_j^*\right)} = \frac{COV\left(Y, T | Z = j\right)}{\left(1 - \alpha_{0j} - \alpha_{1j}\right) p_j^*\left(1 - p_j^*\right)} \tag{38}$$

where the last equality results from the fact that $COV\left(Y, T | Z = j\right) = \left(1 - \alpha_{0j} - \alpha_{1j}\right) COV\left(Y, T^* | Z = j\right)$ which we prove below in section A.4.2.

Re-writing (38) to solve for $\overline{\alpha} = \alpha_{0j} + \alpha_{1j}$ yields

$$1 - \overline{\alpha} = \frac{COV\left(Y, T | Z = j\right)}{\beta p_j^*\left(1 - p_j^*\right)} \tag{39}$$

Inspection of equations (33) and (34) reveals that the results for each p_j^* sum to one across the two solutions, i.e., p_j^* in (33) equals $1 - p_j^*$ in (34). Thus, $p_j^*\left(1 - p_j^*\right)$ is the same across both solutions as is the observed quantity $COV\left(Y, T | Z = j\right)$. Since the corresponding results for β are equal but opposite in sign across the two solutions, the right-hand side of (39) is positive for one solution and negative for another solution. Thus $\overline{\alpha} < 1$ for one solution and $\overline{\alpha} > 1$ for the other solution. Since Assumption 1.v requires $\overline{\alpha} < 1$, we can determine which of the two solutions matches this assumption and therefore determine the sign of β.

A.4.1 Proof that $D_3 > 0$

The solutions in equations (33) and (34) require $D_3 > 0$ in order to yield real numbers. Re-writing the expression for $p_1^*\left(1 - p_1^*\right)$, found in (24), making substitutions using equation

(8), and simplifying yields

$$
\begin{aligned}
p_1^*\left(1-p_1^*\right)\Delta_{20}^2 &= p_2^*\left(1-p_2^*\right)\Delta_{10}^2 + p_0^*\left(1-p_0^*\right)\Delta_{21}^2 + \Delta_{10}\Delta_{21}\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right) \\
\left[p_0^*\left(1-p_0^*\right)\cdot\frac{C_1}{C_0}\right]\Delta_{20}^2 &= \left[p_0^*\left(1-p_0^*\right)\cdot\frac{C_2}{C_0}\right]\Delta_{10}^2 + p_0^*\left(1-p_0^*\right)\Delta_{21}^2 + \Delta_{10}\Delta_{21}\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right) \\
C_1\Delta_{20}^2 &= C_2\Delta_{10}^2 + C_0\Delta_{21}^2 + C_0\Delta_{10}\Delta_{21}\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)}{p_0^*\left(1-p_0^*\right)} \\
D_1 - D_2 - D_0 &= C_0\Delta_{10}\Delta_{21}\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)}{p_0^*\left(1-p_0^*\right)} \qquad (40)
\end{aligned}
$$

Squaring both sides of equation (40) and simplifying yields

$$
\begin{aligned}
\left(D_1 - D_2 - D_0\right)^2 &= C_0^2\Delta_{10}^2\Delta_{21}^2\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)^2}{\left(p_0^*\left(1-p_0^*\right)\right)^2} \\
D_0^2 + D_1^2 + D_2^2 - 2D_0D_1 - 2D_1D_2 + 2D_0D_2 &= \left[\frac{C_0}{p_0^*\left(1-p_0^*\right)}\Delta_{10}^2\right]C_0\Delta_{21}^2\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)^2}{p_0^*\left(1-p_0^*\right)} \\
D_3 + 4D_0D_2 &= \left[\frac{C_2}{p_2^*\left(1-p_2^*\right)}\Delta_{10}^2\right]C_0\Delta_{21}^2\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)^2}{p_0^*\left(1-p_0^*\right)} \\
D_3 + 4D_0D_2 &= D_2D_0\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)^2}{p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)} \\
D_3 &= D_0D_2\left[\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)^2}{p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)} - 4\right] \qquad (41)
\end{aligned}
$$

Notice that C_0 and C_2 must always have the same sign in order to satisfy equation (8). Thus, $C_0C_2 > 0$ and, in turn, $D_0D_2 = C_0\Delta_{21}^2C_2\Delta_{10}^2 > 0$.[36] Therefore, $D_3 > 0$ as long as the term in brackets on the right hand side of (41) is positive. To confirm this condition holds

[36] Rule out $C_j = 0\ \forall\ j$.

notice that

$$\begin{aligned}
\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)^2}{p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)}-4>0 \\
\frac{\left(p_2^*\left(1-p_0^*\right)\right)^2+\left(p_0^*\left(1-p_2^*\right)\right)^2+2p_2^*\left(1-p_0^*\right)p_0^*\left(1-p_2^*\right)}{p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)}>4 \\
\frac{\left(p_2^*\left(1-p_0^*\right)\right)^2+\left(p_0^*\left(1-p_2^*\right)\right)^2}{p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)}+2>4 \\
\left(p_2^*\left(1-p_0^*\right)\right)^2+\left(p_0^*\left(1-p_2^*\right)\right)^2>2p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right) \\
\left(p_2^*\left(1-p_0^*\right)\right)^2-2p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)+\left(p_0^*\left(1-p_2^*\right)\right)^2>0 \\
\left[\left(p_2^*\left(1-p_0^*\right)\right)-\left(p_0^*\left(1-p_2^*\right)\right)\right]^2>0 \\
\left[p_2^*-p_0^*\right]^2>0 \\
\frac{\Delta_{20}^2}{\beta^2}>0 \qquad (42)
\end{aligned}$$

As long as $\beta \neq 0$, this condition is satisfied and, thus, $D_3 > 0$.

A.4.2 Proof that $COV\left(W, T|Z=j\right) = \left(1-\alpha_{0j}-\alpha_{1j}\right) COV\left(W, T^*|Z=j\right)$

For a given variable, W, the covariance between W and T^* among those observations where $Z = j$ simplifies to

$$\begin{aligned}
COV\left(W, T^*|Z=j\right) &= E\left[WT^*|Z=j\right] - E\left[W|Z=j\right]E\left[T^*|Z=j\right] \\
&= p_j^* E\left[W|T^*=1, Z=j\right] - E\left[W|Z=j\right]p_j^* \\
&= \big(E\left[W|T^*=1, Z=j\right] - E\left[W|Z=j\right]\big)p_j^* \qquad (43)
\end{aligned}$$

where we make use of the fact that

$$E\left[WT^*|Z=j\right] = p_j^* E\left[W\cdot 1|T^*=1,Z=j\right] + \left(1-p_j^*\right) E\left[W\cdot 0|T^*=0,Z=j\right] = p_j^* E\left[W|T^*=1,Z=j\right]$$

Similarly, for a given variable, W, the covariance between W and T among those observations where $Z=j$ simplifies to

$$COV\left(W,T|Z=j\right) = \left(E\left[W|T=1,Z=j\right] - E\left[W|Z=j\right]\right)p_j \tag{44}$$

Notice that replacing Y with W in equation (19) and extending the assumption of non-differential measurement error to W yields an expression for $E\left[W|T=1,Z=j\right]$. Inserting this term into equation (44) yields

$$\begin{aligned} COV\left(W,T|Z=j\right) &= \Bigg(E\left[W|T^*=0,Z=j\right]\cdot\frac{\alpha_{0j}\left(1-p_j^*\right)}{p_j} + E\left[W|T^*=1,Z=j\right]\cdot\frac{\left(1-\alpha_{1j}\right)p_j^*}{p_j} \\ &\quad - E\left[W|Z=j\right]\Bigg)p_j \\ &= E\left[W|T^*=0,Z=j\right]\cdot\alpha_{0j}\left(1-p_j^*\right) + E\left[W|T^*=1,Z=j\right]\cdot\left(1-\alpha_{1j}\right)p_j^* \\ &\quad - E\left[W|Z=j\right]p_j \end{aligned} \tag{45}$$

Finally, adding zero in the form of $\alpha_{0j}E[W|Z=j] - \alpha_{0j}E[W|Z=j]$ yields

$$\begin{aligned}
COV(W,T|Z=j) &= E[W|T^*=0,Z=j]\cdot\alpha_{0j}\left(1-p_j^*\right) + E[W|T^*=1,Z=j]\cdot(1-\alpha_{1j})\,p_j^* \\
&\quad - \alpha_{0j}E[W|Z=j] \\
&\quad - \left(E[W|Z=j]\,p_j - \alpha_{0j}E[W|Z=j]\right) \\
&= E[W|T^*=0,Z=j]\cdot\alpha_{0j}\left(1-p_j^*\right) + E[W|T^*=1,Z=j]\cdot(1-\alpha_{1j})\,p_j^* \\
&\quad - \alpha_{0j}\Big(E[W|T^*=0,Z=j]\left(1-p_j^*\right) + E[W|T^*=1,Z=j]\,p_j^*\Big) \\
&\quad - E[W|Z=j]\,(p_j-\alpha_{0j}) \\
&= E[W|T^*=1,Z=j]\cdot(1-\alpha_{0j}-\alpha_{1j})\,p_j^* \\
&\quad - E[W|Z=j]\cdot(1-\alpha_{0j}-\alpha_{1j})\,p_j^* \\
&= \left(E[W|T^*=1,Z=j] - E[W|Z=j]\right)\cdot(1-\alpha_{0j}-\alpha_{1j})\,p_j^* \\
&= (1-\alpha_{0j}-\alpha_{1j})\,COV(W,T^*|Z=j) \qquad (46)
\end{aligned}$$

A.5 Identification of $\overline{\alpha}$ Under Assumption 4 Using Homoskedasticity

One possibility for identifying $\overline{\alpha}$ when using Assumption 4 is to impose restrictions on the higher order moments. In particular, we can require that the error term, ϵ, in equation (1) is homoskedastic, i.e., $E[\epsilon^2|Z=j] = E[\epsilon^2]$.

Following equation (1), the expected value of the square of Y when $Z = j$ is

$$\begin{aligned} E\left[Y^2|Z=j\right] &= E\left[\left(\gamma+\beta T^*+\epsilon\right)^2|Z=j\right] \\ &= E\left[\gamma^2+\beta^2\left(T^*\right)^2+\epsilon^2+2\beta T^*\gamma+2\gamma\epsilon+2\beta T^*\epsilon|Z=j\right] \\ &= \beta^2 E\left[\left(T^*\right)^2|Z=j\right]+2\beta E\left[\left(\gamma+\epsilon\right)T^*|Z=j\right]+E\left[\gamma^2+2\gamma\epsilon+\epsilon^2|Z=j\right] \end{aligned} \tag{47}$$

The first term in (47) simplifies using $\beta^2 E\left[\left(T^*\right)^2|Z=j\right] = \beta^2 p_j^*$. The third term in (47) can be simplified using $E\left[\gamma^2+2\gamma\epsilon+\epsilon^2|Z=j\right] = \gamma^2+2\gamma E\left[\epsilon|Z=j\right]+E\left[\epsilon^2|Z=j\right] = \gamma^2+E\left[\epsilon^2\right]$ where the second equality follows both from the fact that $E\left[\epsilon|Z=j\right]$ and by the homoskedasticity assumption.

To find an expression for the second term in (47), we begin with right-hand side of (6) multiplied by $-\left(1-p_j^*\right)$. Simplifying this expression yields

$$\begin{aligned} \left(1-p_j^*\right)\left[\frac{COV\left(Y,T|Z=j\right)}{1-p_j-\alpha_{1j}}-E\left[Y|Z=j\right]\right] &= \frac{COV\left(Y,T|Z=j\right)}{1-\alpha_{0j}-\alpha_{1j}}-\left(1-p_j^*\right)E\left[Y|Z=j\right] \\ &= COV\left(Y,T^*|Z=j\right)-\left(1-p_j^*\right)E\left[Y|Z=j\right] \\ &= \left(E\left[YT^*|Z=j\right]-E\left[Y|Z=j\right]p_j^*\right)-\left(1-p_j^*\right)E\left[Y|Z=j\right] \\ &= E\left[\left(\gamma+\beta T^*+\epsilon\right)T^*|Z=j\right]-E\left[Y|Z=j\right] \\ &= E\left[\left(\gamma+\epsilon\right)T^*|Z=j\right]+E\left[\beta\left(T^*\right)^2|Z=j\right]-E\left[Y|Z=j\right] \\ &= E\left[\left(\gamma+\epsilon\right)T^*|Z=j\right]+\beta p_j^*-\left(\gamma+\beta p_j^*\right) \\ &= E\left[\left(\gamma+\epsilon\right)T^*|Z=j\right]-\gamma \end{aligned} \tag{48}$$

where the first equality uses the fact that $1-p_j-\alpha_{1j} = \left(1-\alpha_{0j}-\alpha_{1j}\right)\left(1-p_j^*\right)$ and the

second equality uses the result $COV\left(Y,T|Z=j\right)=\left(1-\alpha_{0j}-\alpha_{1j}\right)COV\left(Y,T^{*}|Z=j\right)$.

Substituting these results back into (47) yields

$$E\left[Y^{2}|Z=j\right]=\beta^{2}p_{j}^{*}+2\beta\left[\frac{COV\left(Y,T|Z=j\right)}{1-\alpha_{0j}-\alpha_{1j}}-\left(1-p_{j}^{*}\right)E\left[Y|Z=j\right]+\gamma\right]+\gamma^{2}+E\left[\epsilon^{2}\right] \tag{49}$$

Differencing (49) for any two values of the instrument, j and k, yields

$$\begin{aligned}E\left[Y^{2}|Z=j\right]-E\left[Y^{2}|Z=k\right]&=\beta^{2}\left(p_{j}^{*}-p_{k}^{*}\right)+2\beta\left[\frac{COV\left(Y,T|Z=j\right)}{1-\alpha_{0j}-\alpha_{1j}}-\frac{COV\left(Y,T|Z=k\right)}{1-\alpha_{0k}-\alpha_{1k}}\right]\\&\quad-2\beta\left[\left(1-p_{j}^{*}\right)E\left[Y|Z=j\right]-\left(1-p_{k}^{*}\right)E\left[Y|Z=k\right]\right]\\&=\beta\left(E\left[Y|Z=j\right]-E\left[Y|Z=k\right]\right)\\&\quad+2\beta\left[\frac{COV\left(Y,T|Z=j\right)}{1-\alpha_{0j}-\alpha_{1j}}-\frac{COV\left(Y,T|Z=k\right)}{1-\alpha_{0k}-\alpha_{1k}}\right]\\&\quad-2\beta\left[\left(1-p_{j}^{*}\right)E\left[Y|Z=j\right]-\left(1-p_{k}^{*}\right)E\left[Y|Z=k\right]\right]\end{aligned} \tag{50}$$

where the second equality arises by using equation (2) to adjust the first term of (50). Again invoking Assumption 3 (i.e., $\overline{\alpha}=\alpha_{0j}+\alpha_{1j}$) and solving for $1-\overline{\alpha}$ yields

$$1-\overline{\alpha}=\frac{2\beta\left(COV\left(Y,T|Z=j\right)-COV\left(Y,T|Z=k\right)\right)}{\left(E\left[Y^{2}|Z=j\right]-E\left[Y^{2}|Z=k\right]\right)+\beta\left\{\Delta_{jk}+2\left(E\left[Y|Z=j\right]p_{j}^{*}-E\left[Y|Z=k\right]p_{k}^{*}\right)\right\}} \tag{51}$$

where $\Delta_{jk}=E\left[Y|Z=j\right]-E\left[Y|Z=k\right]$.

To see that (51) implies that $\overline{\alpha}<1$ for one solution and $\overline{\alpha}>1$ for the other solution, let $\{\beta,p_{j}^{*},p_{k}^{*}\}$ be the estimators for the first solution while $\{\widetilde{\beta},\widetilde{p}_{j}^{*},\widetilde{p}_{k}^{*}\}$ are the corresponding estimators for the second solution. Recall from (36) and (37) that $\widetilde{\beta}=-\beta$ and, since the estimators for a p_{j}^{*} sum to one across the two solutions, $\widetilde{p}_{j}^{*}=1-p_{j}^{*}$ for each j.

Inserting these results into the second term in the denominator of (51) yields

$$\begin{aligned}\widetilde{\beta}\left\{\Delta_{jk}+2\left(E\left[Y|Z=j\right]\widetilde{p}_j^*-E\left[Y|Z=k\right]\widetilde{p}_k^*\right)\right\} &= -\beta\left\{E\left[Y|Z=j\right]-E\left[Y|Z=k\right]\right. \\ &\quad \left.+\,2\left(E\left[Y|Z=j\right]\left(1-p_j^*\right)-E\left[Y|Z=k\right]\left(1-p_k^*\right)\right)\right\} \\ &= -\beta\left\{-\Delta_{jk}-2\left(E\left[Y|Z=j\right]p_j^*-E\left[Y|Z=k\right]p_k^*\right)\right\} \\ &= \beta\left\{\Delta_{jk}+2\left(E\left[Y|Z=j\right]p_j^*-E\left[Y|Z=k\right]p_k^*\right)\right\} \end{aligned} \tag{52}$$

Thus, the second term in the denominator of (51) is the same for both solutions and, since $E\left[Y^2|Z=j\right]-E\left[Y^2|Z=k\right]$ does not vary across solutions, we see the entire denominator does not vary across the two solutions. The sign of the numerator, however, will vary across the two solutions since β is of equal magnitude but opposite in sign across the two solutions. Thus, $\overline{\alpha}<1$ for one solution and $\overline{\alpha}>1$ for the other solution.

A.6 Estimation and Inference Details

A.6.1 The Potential for Weak Identification

Inspection of the equation for β found in (37) shows that it yields a real solution as long as the term $D3$ is non-negative. As is further shown in the proof of Theorem 1 in Appendix Section A.4, $D3$ must be strictly positive as the solution to the system of equations for the p_j^* terms have $D3$ in the denominator. Although $D3$ is positive in the population, except when $\beta=0$ (see Section A.4.1), it may be negative when replaced with its sample analog. Such situations are more likely to occur when at least one of the following holds: i) β is close to zero, ii) there are small differences between p_j^* values, and iii) the sample size is

relatively small. Under these conditions, the GMM estimator is weakly identified and may not converge in finite samples. DiTraglia and Garcia-Jimeno (2019) note a similar issue with their estimator.

A.6.2 Including Covariates

For our first approach for including covariates, we specify the equation for Y as $Y = \gamma + \beta T^* + \boldsymbol{X}\boldsymbol{\psi} + \epsilon$ as shown in equation (12), where $\boldsymbol{X}$ is a row vector of K regressors and $\boldsymbol{\psi}$ is a column vector of K parameters. Correspondingly modifying the Assumptions 1, 2, and 4 to account for the additional regressor(s) yields

Assumption 1′.ii: $E[Y|Z = j, \boldsymbol{X}] = \gamma + \beta p_j^* + E[\boldsymbol{X}\boldsymbol{\psi}|Z = j, \boldsymbol{X}],\ \forall j$

Assumption 1′.iv: $E\left[Y|T, T^*, Z, \boldsymbol{X}\right] = E\left[Y|T^*, Z, \boldsymbol{X}\right]$

Assumption 2′: $E\left[Y|T^* = t, Z = j, \boldsymbol{X}\right] = E\left[Y|T^* = t, \boldsymbol{X}\right],\ \forall j$

Assumption 4′: $E\left[Y|T^* = 1, Z = j, \boldsymbol{X}\right] - E\left[Y|T^* = 0, Z = j, \boldsymbol{X}\right] = E\left[Y|T^* = 1, Z = k, \boldsymbol{X}\right] - E\left[Y|T^* = 0, Z = k, \boldsymbol{X}\right]\ \ \forall\ j, k$

Assumption 1′.ii is analogous to assuming that each element in $\boldsymbol{X}$ is uncorrelated with the error ϵ. In addition, notice that $p_j^* = E[T^*|Z = j, \boldsymbol{X}]$. Re-writing the equation for Assumption 1′.ii, the analogous moment conditions to (2) are

$$E[Y|Z = j, \boldsymbol{X}] - E[\boldsymbol{X}\boldsymbol{\psi}|Z = j, \boldsymbol{X}] = E[\widetilde{Y}|Z = j, \boldsymbol{X}] = \gamma + \beta p_j^*,\ \forall j \tag{53}$$

where $\widetilde{Y} = Y - \boldsymbol{X}\boldsymbol{\psi}$.

We next derive the equations found in Appendix Section A.2, extending the results to explicitly condition on $\boldsymbol{X}$. Adjusting equation (19) to account for $\boldsymbol{X}$ yields

$$E\left[Y|T=1,Z=j,\boldsymbol{X}\right]=E\left[Y|T^{*}=0,Z=j,\boldsymbol{X}\right]\cdot\frac{\alpha_{0j}\left(1-p_{j}^{*}\right)}{p_{j}}+E\left[Y|T^{*}=1,Z=j,\boldsymbol{X}\right]\cdot\frac{\left(1-\alpha_{1j}\right)p_{j}^{*}}{p_{j}} \tag{54}$$

To include $\boldsymbol{X}$, we have assumed that $\alpha_{0j}=P\left[T=1|T^{*}=0,Z=j,\boldsymbol{X}\right]$, $\alpha_{1j}=P\left[T=0|T^{*}=1,Z=j,\boldsymbol{X}\right]$, and $p_{j}=P\left[T=1|Z=j,\boldsymbol{X}\right]$ do not depend on the covariates (our second approach to including covariates relaxes these assumptions).

Similarly, adjusting equation (20) to account for $\boldsymbol{X}$ yields

$$E\left[Y|T=0,Z=j,\boldsymbol{X}\right]=E\left[Y|T^{*}=0,Z=j,\boldsymbol{X}\right]\cdot\frac{\left(1-\alpha_{0j}\right)\left(1-p_{j}^{*}\right)}{1-p_{j}}+E\left[Y|T^{*}=1,Z=j,\boldsymbol{X}\right]\cdot\frac{\alpha_{1j}\cdot p_{j}^{*}}{1-p_{j}} \tag{55}$$

Using these last two expressions we can solve for the following unobserved quantities

$$E\left[Y|T^{*}=1,Z=j,\boldsymbol{X}\right]=\frac{COV\left(Y,T|Z=j,\boldsymbol{X}\right)}{p_{j}-\alpha_{0j}}+E\left[Y|Z=j,\boldsymbol{X}\right] \tag{56}$$

$$E\left[Y|T^{*}=0,Z=j,\boldsymbol{X}\right]=E\left[Y|Z=j,\boldsymbol{X}\right]-\frac{COV\left(Y,T|Z=j,\boldsymbol{X}\right)}{1-p_{j}-\alpha_{1j}} \tag{57}$$

Recalling that $\widetilde{Y}=Y-\boldsymbol{X}\boldsymbol{\psi}$, we can next write

$$COV\left(\widetilde{Y},T|Z=j,\boldsymbol{X}\right)=COV\left(Y,T|Z=j,\boldsymbol{X}\right)-COV\left(\boldsymbol{X}\boldsymbol{\psi},T|Z=j,\boldsymbol{X}\right) \tag{58}$$

$$=COV\left(Y,T|Z=j,\boldsymbol{X}\right) \tag{59}$$

where the second term on the right hand side equals zero since we are conditioning on $\boldsymbol{X}$. Using this result to substitute for $COV\left(Y,T|Z=j,\boldsymbol{X}\right)$ in equations (56) and (57), we can then use either Assumption $2'$ or Assumption $4'$ to derive equations analogous to (8)

$$\frac{COV\left(\widetilde{Y},T|Z=j,\boldsymbol{X}\right)}{p_j^*\left(1-p_j^*\right)}=\frac{COV\left(\widetilde{Y},T|Z=k,\boldsymbol{X}\right)}{p_k^*\left(1-p_k^*\right)} \tag{60}$$

As in the text, we eliminate the misclassification rates from the system of moment conditions by assuming that the sum of the misclassification rates is constant across instrument values.

Thus, when using this first approach to including covariates, we end up with moments that are identical to those in the text, but substituting $\widetilde{Y}$ for Y. We can pre-estimate $\boldsymbol{\psi}$ by regressing Y on T and $\boldsymbol{X}$ while using Z as an instrument for T. The proof that $\boldsymbol{\psi}$ can be consistently estimated with 2SLS is provided in Proposition 3 in Fraziz and Loewenstein (2003). With this consistent estimate of $\boldsymbol{\psi}$, we can then use GMM to estimate the set of moment conditions found in (11), except replacing Y with $\widetilde{Y}$.

Our second approach to including covariates, which is discussed in the text, involves estimating the moments in equation (8) separately for different cells defined by the covariates in $\boldsymbol{X}$. This approach can allow all the parameters to vary across cells or restrict some parameters (e.g., β) to be the same across all cells.

A.7 Additional Simulation Results

Table A1 illustrates that the GMM estimator is affected by weak identification when β is (relatively) close to zero and/or the sample size is small (see Section 4.4).[37] The table

[37] Another situation that can lead to weak identification is when there are small differences between the p_j^* values. We do not explore this aspect of weak identification in the simulations reported here.

presents GMM estimates for multiple combinations of β and the sample size N, using 1,000 iterations for each combination. Panel A of Table A1 shows the fraction of iterations where GMM successfully converges.[38] Consistent with the conditions for weak identification, the GMM estimator is less likely to converge as β approaches zero and/or as the sample size decreases. When $N = 1,000$, GMM fails to converge for at least some fraction of iterations for every value of β. However, when $N = 100,000$ GMM always converges except where $\beta = 0$, the parameter value for which our estimator is not identified.

Even when GMM converges, the remaining Panels of Table A1 show that weak identification yields imprecise estimates and raises inference concerns. As shown in Panel B, when GMM converges, the average point estimate is too large when $N = 1,000$ while it is well-centered when $N = 100,000$. Moreover, as shown in Panel C, GMM yields relatively larger standard errors when identification is weak, but as the sample size increases, the GMM estimates are quite precise. Finally, in Panel D, the coverage rates for the 95% confidence intervals are as expected when $N = 100,000$, whereas coverage is incomplete when $N = 1,000$ in spite of the substantially wider confidence intervals. See Section A.7.2 for evidence that the patterns in Table A1 remain when we vary $\overline{\alpha}$.

In the text and the next two appendix subsections, we present "large sample" simulation results with $N = 100,000$ and $\beta = 1$ to demonstrate the performance of our estimator in the absence of the weak identification concerns. We focus on how our estimator and the Two Stage Least Squares (2SLS) estimator perform under different assumptions regarding misclassification error.[39] We also examine the robustness of our estimator when we relax

[38] All GMM simulations are completed in MatLab. We obtain starting values using fminunc and its default settings. We obtain final parameter values using fminsearch, setting the search tolerance at 10^{-7} and the maximum iterations at 20,000.

[39] We use 2SLS as we use indicators for all but one of the discrete values of Z as excluded instruments.

Assumption 3 (i.e., that $\alpha_{0j} + \alpha_{1j}$ is constant) in a manner that mimics the examples from Section 2 and when allowing T^* to be endogenous (as consistent with Assumption 4). In the working paper version of this paper (Haider and Stephens 2020), we present "small sample" results with $N = 1,000$ in which we examine inference when identification is weak.

A.7.1 "Large Sample" Simulation Results Discussed in Paper

In Table A2, we replicate the simulations of Table 1, but do so with a binary outcome. To accommodate a binary outcome, we specify that $Y = 1(\gamma + \beta T^* > \epsilon)$, where 1(.) is the indicator function, $\gamma = 0.25$, $\beta = 0.5$, and $\epsilon \sim$ uniform$[0, 1]$. All of the conclusions remain the same.

We allow T^* to be endogenous in Table A3. These simulations are making use of such a form specified in Olsen (1980). See Brinch, Mogstad, and Wiswall (2017) for a recent application that leverages endogeneity of this type. Specifically, the error term ϵ is specified to have two components. The endogenous component is $v = \Phi(E^*)$, where Φ is the standard normal cumulative distribution function so that $v \sim$ uniform$[0, 1]$. The exogenous component is $e \sim N(0, \sigma_\epsilon^2(1 - \rho^2))$, where ρ is the correlation between ϵ and v. With this structure,

$$\epsilon = e - \rho(v - .5)\sigma_\epsilon / \sqrt{1/12} \tag{61}$$

We hold σ_ϵ^2 as specified in Table 1 while ρ is 0.250 for columns (1) and (3) and 0.500 for columns (2) and (4).

For the first two columns of the table, we start with the basic specification used in column (2) of Table 1 where, in the absence of a correlation between ϵ and T^*, the OLS estimate is

less than β while the 2SLS estimate exceeds β. Notice that as this correlation, ρ, increases from 0.25 to 0.5 as we move between columns (1) and (2), the OLS estimate changes from being downward inconsistent to upward inconsistent. In both columns, the 2SLS estimate is identical to that found in column (2) of Table 1, and thus remains upward inconsistent because of the nature of the varying misclassification. We repeat this same exercise in columns (3) and (4), but instead base it on the specification of column (3) of Table 1. The basic results are similar: OLS can be downward or upward inconsistent depending on the amount of endogeneity, but in this case, 2SLS is downward inconsistent due to the nature of the varying misclassification. Thus, with both misclassification and endogeneity present, neither OLS nor 2SLS systematically bound the true parameter value. However, in all cases, our estimator remains consistent.

As we have discussed throughout, other papers have proposed methods to estimate the impact of a mismeasured binary regressor. In Table A4, we show three estimators that rely on fixed misclassification and a single instrumental variable (Frazis and Lowenstein 2003; Mahajan 2006; Lewbel 2007) are not well-centered in simulations that have varying misclassification. However, three estimators that instead rely on higher-order moments of the error term (Chen, Hu, and Lewbel 2008a, b; DiTraglia and Garcia-Jimeno 2019) are well-centered (see column (2) of Table A4) as the errors used in our simulations (which are normally distributed) satisfy the error term restrictions required by these estimators. However, in the final three columns where we specify error distributions that deviate from homoskedasticity and/or symmetry (as would be the case with binary outcomes), we find that these alternative methods yield estimates of β that systematically deviate from its true value. In contrast, our estimator remains well-centered in all cases.

In Table A4, we examine how other proposed estimators perform in the face of misclassification rates that vary. Columns (1) and (2) of Table A4 use exactly the same data generating process as columns (1) and (2) of Table 1, respectively. All of the estimators are well-centered in column (1) when the underlying misclassification rates are fixed. When we allow the underlying misclassifications rates to vary (but continue to assume that the sum is fixed) in column (2), the three estimators that rely on fixed misclassification and an instrumental variable (Frazis and Lowenstein 2003, Mahajan 2006, and Lewbel 2007) are not well-centered. The three estimators that rely on higher-order moments of the error term (Chen, Hu, and Lewbel 2008a; Chen, Hu, and Lewbel 2008b; DiTraglia and Garcia-Jimeno 2019) continue to be well-centered because the error term specification (normal errors) for the simulations in Table 1 are consistent with the higher order error term requirements of these estimators.

To examine the sensitivity of the estimators to these distributional assumptions, columns (3) through (5) of Table A4 are created to be identical to column (2), but differ in their specification for the error term. Specifically, consider the expanded DGP

$$Y = \gamma + \beta T^* + \epsilon_0(1 - T^*) + \epsilon_1 T^* \tag{62}$$

As in the main text, the standard deviation of ϵ is 0.25, and the expanded error term is

- Column (1): $\epsilon_0 = \epsilon_1 = \epsilon$
- Column (2): $\epsilon_0 = \epsilon_1 = \epsilon$
- Column (3): $\epsilon_0 = \epsilon_n$; $\epsilon_1 = 2.5\epsilon$

- Column (4): $\epsilon_0 = \epsilon_1 = |2.5\epsilon| - \overline{|2.5\epsilon|}$
- Column (5): $\epsilon_0 = |\epsilon| - \overline{|\epsilon|}$; $\epsilon_1 = |2.5\epsilon| - \overline{|2.5\epsilon|}$

Chen, Hu, and Lewbel (2008a) and DiTraglia and Garcia-Jimeno (2019) assume homoskedastic errors, so these estimators are not well-centered in column (3) that specifies a heteroskedastic error term.[40] Chen, Hu, and Lewbel (2008b) assumes that the error distribution has a third moment that is equal to zero, so this estimator starts is less well-centered in column (4) that specifies the error term to be a half-normal distribution. All three estimators are poorly centered in the final column that specifies a heteroskedastic and non-symmetric error term. For all columns in Table A4, our estimator is well-centered.

A.7.2 Further "Large Sample" Simulations Results

Tables A6-A8 shows large sample results when varying α and β. All of the general patterns remain.

A.8 Additional Empirical Results Using Moment Inequalities

In this subsection, we present additional empirical results that are relevant to the moment inequality results in the text.

Row (A) of Table A9 shows the implied confidence intervals based on the primary results in the text that make use of equality moments, presented in Row (E) of Table 2. These results are based on bootstrapped standard errors to account for the preliminary estimation

[40] Chen, Hu, and Lewbel (2008a) require that the errors are homoskedastic with respect to T^* while DiTraglia and Garcia-Jimeno (2019) require homoskedasticity with respect to Z. In addition, both estimators make a constant skewness assumption, with Chen, Hu, and Lewbel (2008a) requiring it with respect to T^* and with DiTraglia and Garcia-Jimeno (2019) requiring it with respect to Z.

of the coefficients on the other covariates in the model.

All of our results based on inequality moments ignore the preliminary estimation of the coefficients on the other covariates in the model. Row (B) of Table A9 shows confidence intervals based on equality moments and analytic standard errors that ignore the preliminary estimation. As expected, the standard errors are smaller than those in Row (A).

Rows (C) - (F) use inequality moments and the methods described in Bugni, Canay, and Shi (2017) to obtain the confidence sets on β based on different values of R_{jk}. Row (C) sets $R_{jk} = R_{kj} = 1 \ \forall j, k$, which effectively turns the inequality moments into equality moments. The confidence sets are reasonably close to Row (B), except for the private insurance outcome. This finding for the private insurance outcome is discussed more below.

Rows (D) - (F) present the inequality moments using the values in the text (see Row (E) of Table 2) and adds in one additional specification: $\underline{R}_{jk} = 1, \underline{R}_{kj} = .89 \ \forall j > k$. The patterns are as expected: loosening the inequality bounds tends to widen the confidence sets.

The simple confidence set presentation of these findings overlooks the fact that, with our reliance on a simulation-based test statistic, the probability of rejection can exhibit noise.[41] Moreover, not only does the confidence region need not be symmetric around the point estimate (as is clear from the table of results), it not need be symmetric around any point.

To show transparently what the broader acceptance/rejection regions look like in our empirical example, Figure A5 presents the percentile of the test statistic as compared to its empirical distribution based on the null that is being evaluated. In other words, we are simply graphing 1 minus the p-value of the null listed on the x-axis. Each panel shows two sets of results based on inequality moments, the first that mimics our equality moment

[41] For the confidence regions, we defined "rejection" to occur once the p-value dipped below 0.05 for two successive ordinates on a 0.01 grid.

results (i.e., $R_{jk} = R_{kj} = 1 \ \forall j, k$) and the second that loosens the inequality moments (i.e., $R_{jk} = R_{kj} = 0.95 \ \forall j, k$). Panels A and B are based on a 0.01 grid, except for being on a 0.001 grid around 0.01 of the parameter estimate based on the equality moments. Panel C is based on a 0.001 grid, except for being on a 0.0001 grid around 0.001 of the parameter estimate based on the equality moments.

All three panels mostly deliver the expected results. First, the graphs reach zero around the point estimates based on the equality moments. Second, the graphs based on weaker inequality moments (i.e., $R_{jk} = R_{kj} = 0.95 \ \forall j, k$) are outside of those based on the stricter inequality moments (i.e., $R_{jk} = R_{kj} = 1 \ \forall j, k$). Third, the graphs for Medicaid (Panel A) and No Insurance (Panel B) produce confidence sets that plausibly line up with the equality moment results: The confidence intervals based on the tight inequality moments (i.e., $R_{jk} = R_{kj} = 1 \ \forall j, k$) line up with what is delivered from the equality moments and the confidence sets based on the looser inequality moments (i.e., $R_{jk} = R_{kj} = 0.95 \ \forall j, k$) widen a little bit.

The one set of results in Table A9 and Figure A5 that deviate from this general pattern are those for the lower bound for the Private Insurance outcome in that the confidence set becomes much wider. The graph shows why this is happening. Like all of the outcomes, the curves tend to flatten out and becomes noisier as one moves away from zero. For the Private Insurance outcome, it flattens out before 0.95. It simply appears that this outcome is less-well identified as compared to our other two outcomes.

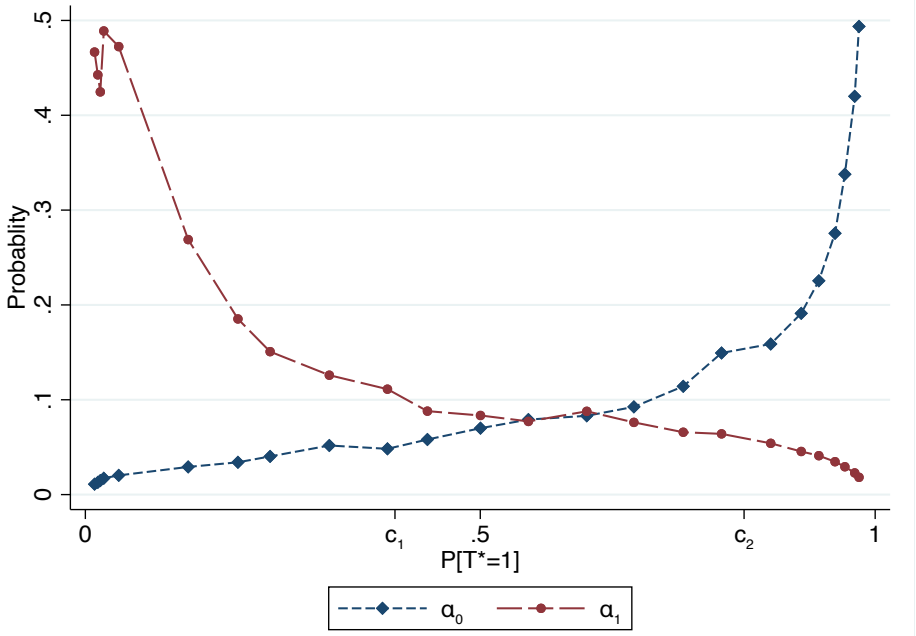
A.9 Supplementary Material References

Freeman, R. B. (1984) "Longitudinal Analysis of the Effects of Trade Unions," *Journal of Labor Economics*, 2(1):1-26.

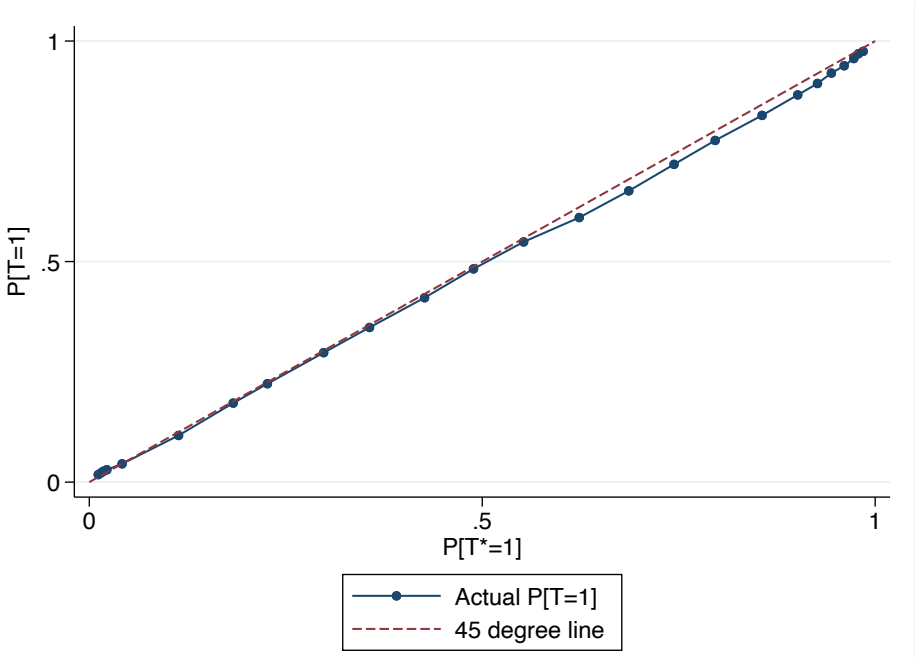
Kreider, B., Pepper, J.V., Gundersen, C. and Jolliffe, D. (2012) "Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation Is Endogenous and Misreported," *Journal of the American Statistical Association,* 107:499, 958-975.

Mellow, W. and Sider, H. (1983) "Accuracy of Response in Labor Market Surveys: Evidence and Implications," *Journal of Labor Economics*, 1(4):331-344.

Figure A1: Misclassification Example - 1977 January CPS Matched Employer-Employee Wages



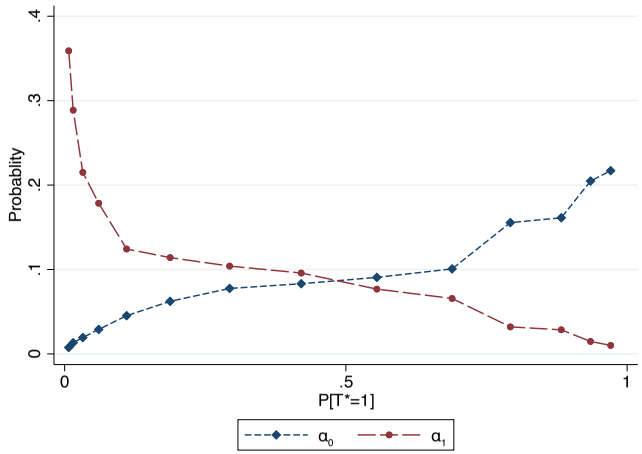
(a) Misclassification Rates



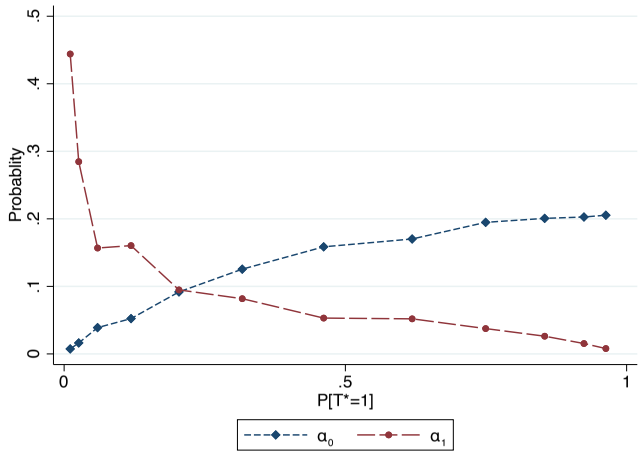
(b) $P[T = 1]$ vs. $P[T^* = 1]$

Notes: Both panels use data from the January 1977 CPS matched employer-employee wage data. Details about the data are discussed in the text and in Appendix section A.1.1. Panel A1a reports misclassification rates, with the details for constructing these rates found in Section 2. Panel A1b plots the relationship between $P[T = 1]$ and $P[T^* = 1]$.

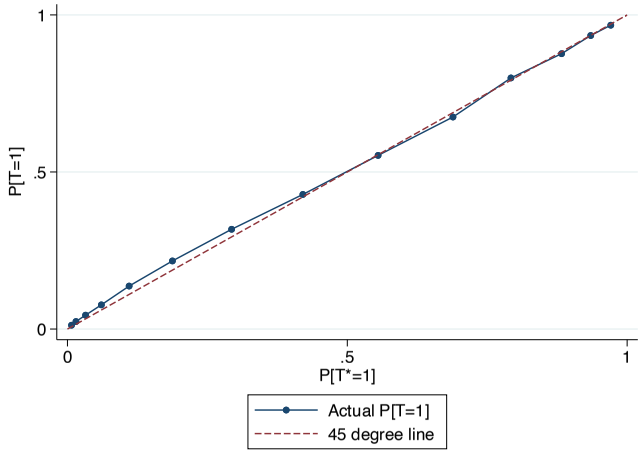
Figure A2: Misclassification Example - Continuous NHANES Height



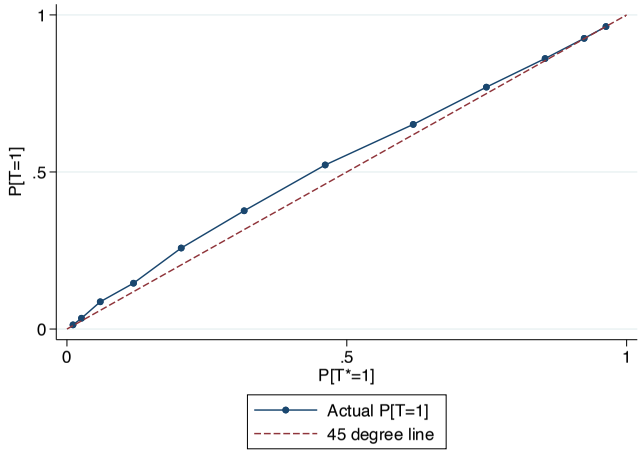
(a) Male Height Misclassification Rates



(b) Female Height Misclassification Rates



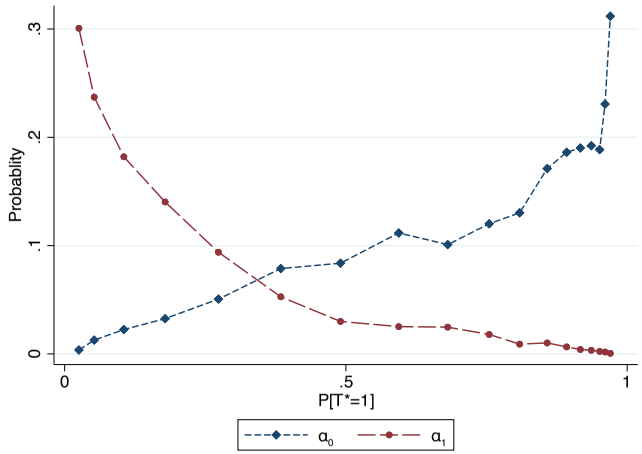
(c) Male Height $P[T = 1]$ vs. $P[T^* = 1]$



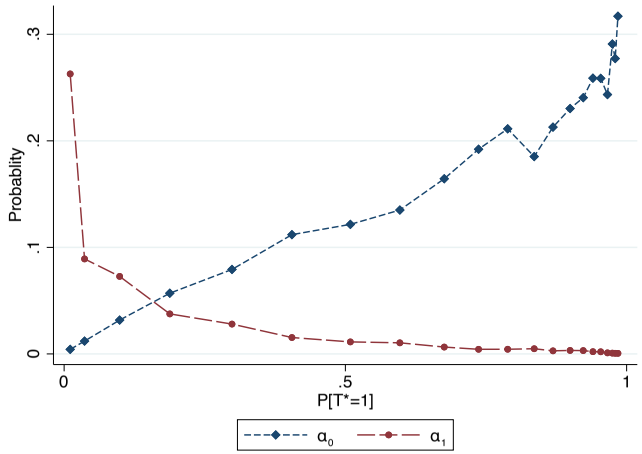
(d) Female Height $P[T = 1]$ vs. $P[T^* = 1]$

Notes: The panels use both self-reported and measured height from Continuous NHANES data. Details about the data are discussed in the text and in Appendix section A.1.1. Panels A2a and A2b report misclassification rates for men and women, respectively, with the details for constructing these rates found in Section 2. Panels A2c and A2d plots the relationship between $P[T = 1]$ and $P[T^* = 1]$ for men and women, respectively.

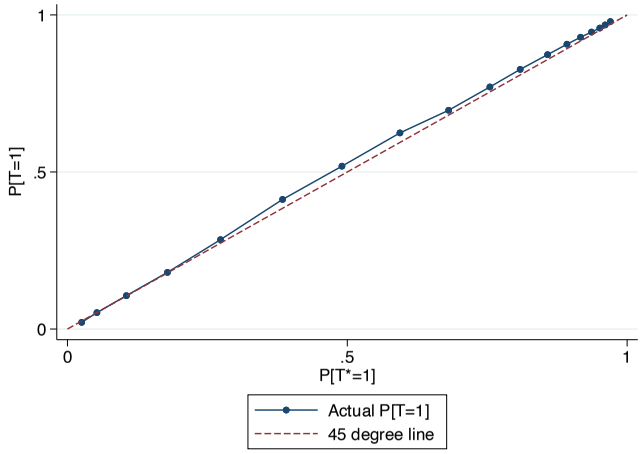
Figure A3: Misclassification Example - Continuous NHANES Weight



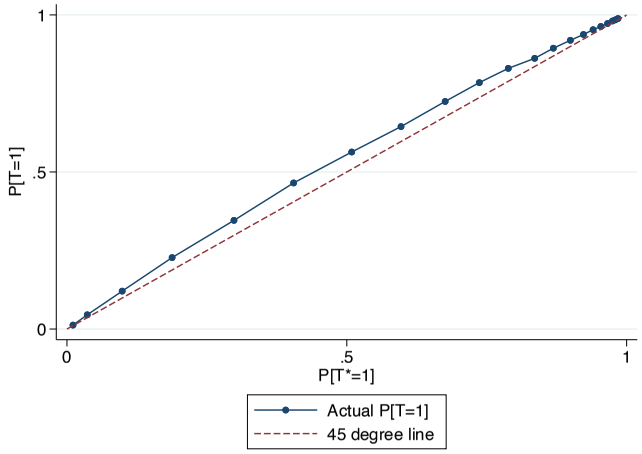
(a) Male Weight Misclassification Rates



(b) Female Weight Misclassification Rates



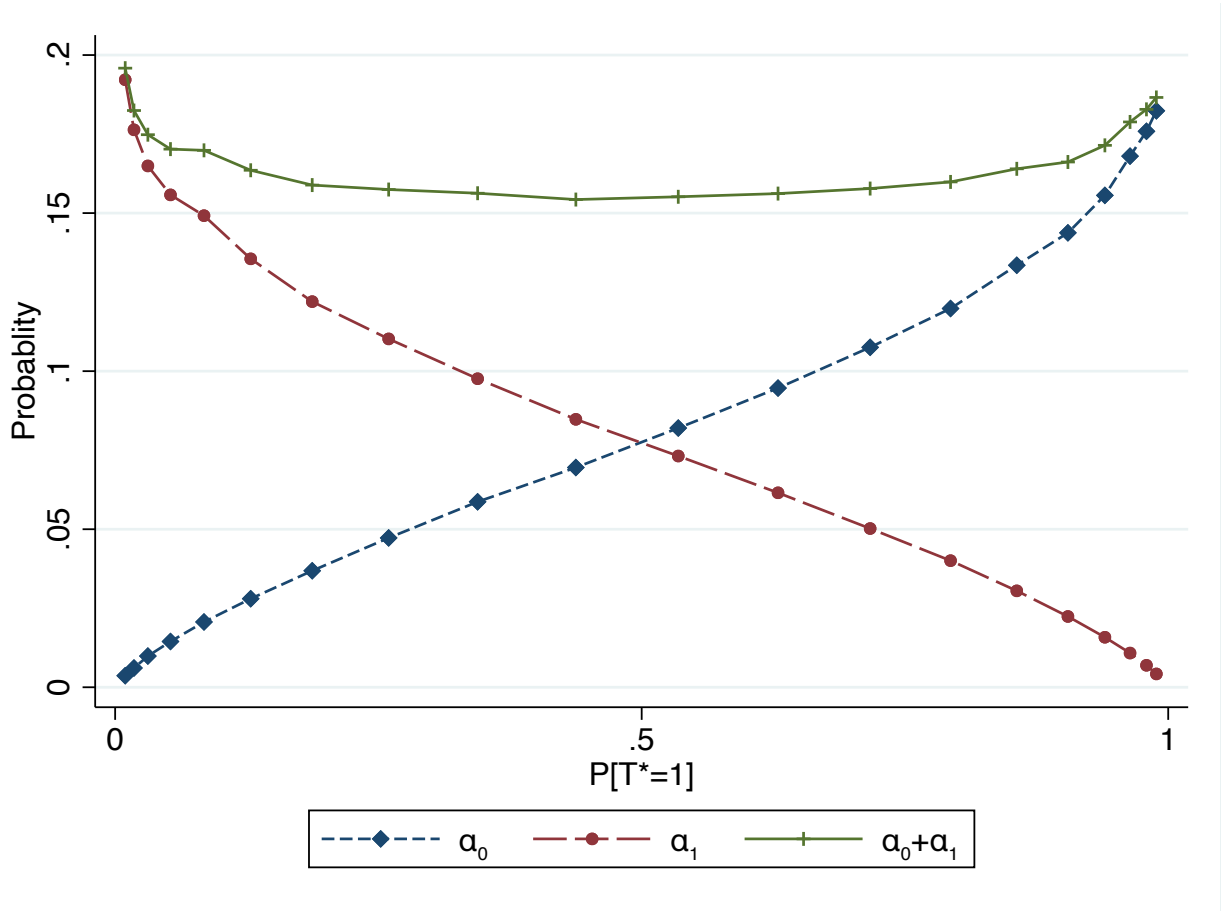
(c) Male Weight $P[T = 1]$ vs. $P[T^* = 1]$



(d) Female Weight $P[T = 1]$ vs. $P[T^* = 1]$

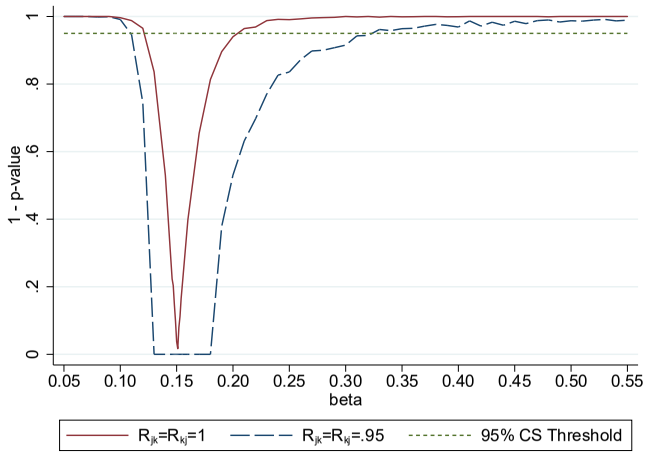
Notes: The panels use both self-reported and measured weight from Continuous NHANES data. Details about the data are discussed in the text and in Appendix section A.1.1. Panels A3a and A3b report misclassification rates for men and women, respectively, with the details for constructing these rates found in Section 2. Panels A3c and A3d plots the relationship between $P[T = 1]$ and $P[T^* = 1]$ for men and women, respectively.

Figure A4: Simulated Misclassification Rates Using a Mis-Measured Index

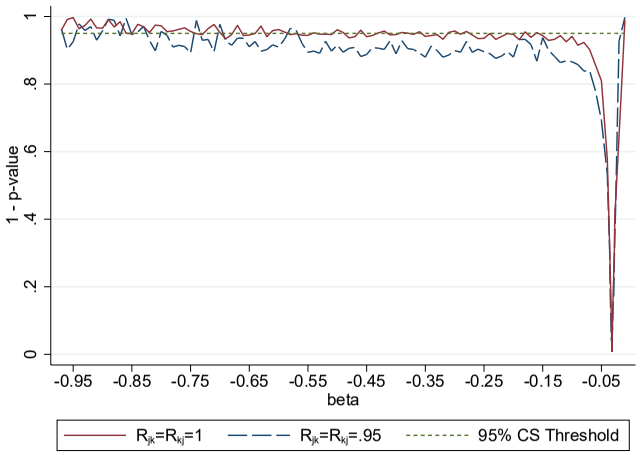


Notes: To construct this figure, we generate the mis-measured index E by adding a mean zero measurement error ν to the random variable E^*. For any given threshold c, we construct the binary indicators $T^* = 1(E^* \leq c)$ and $T = 1(E \leq c)$ and the corresponding misclassification rates as discussed in the text. This figure shows these values average over 500,000 iterations.

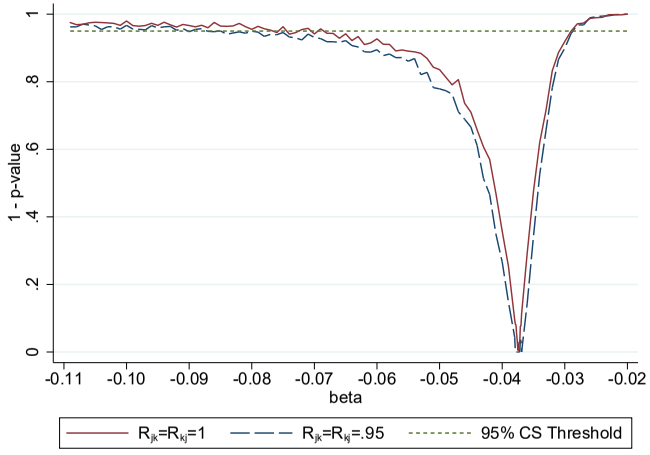
Figure A5: Confidence Curves for β - Empirical Example with Inequality Moments



(a) Medicaid



(b) Private Insurance



(c) No Insurance

Notes: The lines in each panel present 1 minus the p-value of the null listed on the x-axis. Each panel presents results for a different outcome with results for Medicaid, private insurance, and no insurance shown in panels A, B, and C, respectively. Each panel shows two sets of results based on inequality moments: the solid line uses bound values (i$R_{jk} = R_{kj} = 1 \; \forall j, k$) that mimic our moment equality restrictions while the long dashed line uses symmetric bound values discussed in the text ($R_{jk} = R_{kj} = 0.95 \; \forall j, k$). The horizontal dashed line shows the threshold for the 95% confidence set. Panels A and B use null hypotheses (shown in the x-axis) that are based on a 0.01 grid, except for being on a 0.001 grid around 0.01 of the parameter estimate based on the equality moments. Panel C is based on a 0.001 grid, except for being on a 0.0001 grid around 0.001 of the parameter estimate based on the equality moments.

Table A1: Monte Carlo Simulations - GMM Estimation

N	0.0	0.25	0.5	1.0	1.5	2.0
Panel A: Proportion GMM Converges						
1,000	0.970	0.844	0.885	0.940	0.969	0.978
10,000	0.969	0.979	0.999	1.000	1.000	1.000
100,000	0.951	1.000	1.000	1.000	1.000	1.000
Panel B: Mean of β if Converges						
1,000	-0.009	0.755	0.984	1.427	2.021	2.567
10,000	0.048	0.385	0.539	1.017	1.517	2.018
100,000	0.006	0.256	0.504	1.003	1.502	2.001
Panel C: Mean of SE of β if Converges						
1,000	1350	636.8	1346	15.04	15.06	11.71
10,000	1170	33.50	0.300	0.113	0.137	0.168
100,000	495.7	0.026	0.026	0.032	0.040	0.049
Panel D: Coverage of True β if Converges						
1,000	0.992	0.906	0.884	0.911	0.918	0.921
10,000	0.995	0.892	0.925	0.925	0.943	0.946
100,000	1.000	0.948	0.954	0.954	0.951	0.952

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (11). Each simulation uses 1,000 iterations. The numbers within each of the four panels correspond to a different simulation, where the sample size N and the parameter of interest β vary across simulations. The values for the remaining parameters are discussed in Section 5. Panel A reports the proportion of times that the GMM estimator converges across iterations. Panel B reports the average of β across the iterations where the estimator converges. Panel C reports the average standard error of β across the iterations where the estimator converges. Panel D reports the coverage for β across the iterations where the estimator converges.

Table A2: Monte Carlo Simulations - Varying Misclassification with a Binary Outcome

	(1)	(2)	(3)	(4)
Panel A: Data generating process				
α_{00}	0.070	0.055	0.035	0.110
α_{01}	0.070	0.070	0.070	0.140
α_{02}	0.070	0.085	0.105	0.170
$\alpha_{00} + \alpha_{10}$	0.130	0.130	0.130	0.260
$\alpha_{01} + \alpha_{11}$	0.130	0.130	0.130	0.260
$\alpha_{02} + \alpha_{12}$	0.130	0.130	0.130	0.260
Panel B: Estimates of β				
OLS	0.435	0.438	0.443	0.377
	(.003)	(.003)	(.003)	(.003)
2SLS	0.574	0.515	0.453	0.532
	(.013)	(.011)	(.010)	(.012)
Our estimator	0.506	0.506	0.505	0.509
	(.044)	(.043)	(.043)	(.055)

Notes: This table reports Monte Carlo simulations that are structured identically to those in Table 1, but modifies the DGP so that the outcome variable is binary. We specify that $Y = 1(\gamma + \beta T^* > \epsilon)$, where $1(.)$ is the indicator function, $\gamma = 0.25$, $\beta = 0.5$, and $\epsilon \sim uniform(0, 1)$. All other parts of the DGP and the simulation are identical to that in Table 1.

Table A3: Monte Carlo Simulations - Varying Misclassification with Endogenous T^*

	(1)	(2)	(3)	(4)
Panel A: Data generating process				
α_{00}	0.055	0.055	0.035	0.035
α_{01}	0.070	0.070	0.070	0.070
α_{02}	0.085	0.085	0.105	0.105
$\alpha_{0j} + \alpha_{1j} \ \ \forall j$	0.130	0.130	0.130	0.130
Error correlation (ρ)	0.250	0.500	0.250	0.500
Panel B: Estimates of β				
OLS	0.965	1.052	0.974	1.062
	(.002)	(.002)	(.002)	(.002)
2SLS	1.032	1.032	0.907	0.907
	(.010)	(.010)	(.008)	(.008)
Our estimator	1.002	1.002	1.002	1.002
	(.031)	(.029)	(.031)	(.029)

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (11). Each simulation is based on 1,000 iterations, with a sample size of 100,000 observations. The mis-measured variables T in columns (1) and (2) match column (2) of Table 1 and in columns (3) and (4) match column (3) of Table 1. The values for the remaining parameters are discussed in Section 5. All four columns specify that T^* is endogenous following Olsen (1980), with error correlation ρ; see footnote 35 for a complete description. Panel B contains average estimates of β across iterations for the listed estimation methods. The standard deviations of the estimates across iterations are shown in parentheses.

Table A4: Monte Carlo Simulations - Comparing to Other Estimators

	(1)	(2)	(3)	(4)	(5)
Panel A: Data generating process					
α_{00}	0.070	0.055	0.055	0.055	0.055
α_{01}	0.070	0.070	0.070	0.070	0.070
α_{02}	0.070	0.085	0.085	0.085	0.085
$\alpha_{00} + \alpha_{10}$	0.130	0.130	0.130	0.130	0.130
$\alpha_{01} + \alpha_{11}$	0.130	0.130	0.130	0.130	0.130
$\alpha_{02} + \alpha_{12}$	0.130	0.130	0.130	0.130	0.130
Homoskedastic errors?	Yes	Yes	No	Yes	No
Symmetric errors?	Yes	Yes	Yes	No	No
Panel B: Estimates of β					
Our estimator	1.003	1.003	1.003	1.002	1.000
	(.033)	(.033)	(.051)	(.042)	(.034)
Frazis and Lowenstein (2003)	1.000	0.951	0.951	0.951	0.951
	(.005)	(.005)	(.007)	(.006)	(.004)
Mahajan (2006)	1.000	0.951	0.951	0.951	0.951
	(.005)	(.005)	(.007)	(.006)	(.005)
Lewbel (2007)	0.999	0.896	0.897	0.896	0.899
	(.038)	(.034)	(.049)	(.041)	(.035)
Chen, Hu, and Lewbel (2008a)	1.000	1.000	1.150	1.000	0.964
	(.002)	(.002)	(.004)	(.003)	(.002)
Chen, Hu, and Lewbel (2008b)	1.000	1.000	1.027	0.981	0.854
	(.002)	(.002)	(.014)	(.003)	(.009)
DiTraglia and Garcia-Jimeno (2019)	1.001	1.001	1.150	1.000	0.965
	(.009)	(.008)	(.016)	(.013)	(.011)

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (11). Each simulation is based on 1,000 iterations, with a sample size of 100,000 observations. The mismeasured variables T are constructed so that misclassification rate α_{0j} and the sum of the misclassification rates $\alpha_{0j} + \alpha_{1j}$ for instrument value j match the values shown in the top panel. The values for the remaining parameters are discussed in Section 5. See Appendix section A.7.2 for a complete description of the data generating processes specified for each column. Panel B contains average estimates of β across iterations for the listed estimation method. The standard deviation of the estimates across iterations are shown in parentheses.

Table A5: Monte Carlo Simulations - Varying Misclassification

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Data generating process						
α_{00}	0.070	0.055	0.035	0.110	0.059	0.014
α_{01}	0.070	0.070	0.070	0.140	0.078	0.039
α_{02}	0.070	0.085	0.105	0.170	0.097	0.059
$\alpha_{00} + \alpha_{10}$	0.130	0.130	0.130	0.260	0.156	0.173
$\alpha_{01} + \alpha_{11}$	0.130	0.130	0.130	0.260	0.155	0.159
$\alpha_{02} + \alpha_{12}$	0.130	0.130	0.130	0.260	0.156	0.156
Panel B: Estimates of β						
OLS	0.870	0.877	0.886	0.750	0.853	0.842
	(.002)	(.002)	(.002)	(.003)	(.002)	(.003)
2SLS	1.150	1.032	0.907	1.065	1.030	1.004
	(.011)	(.009)	(.008)	(.012)	(.009)	(.008)
Our estimator	1.003	1.003	1.002	1.004	0.998	1.012
	(.033)	(.033)	(.033)	(.047)	(.033)	(.052)

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (11). Each simulation is based on 1,000 iterations, with a sample size of 100,000 observations. The mismeasured variables T are constructed so that misclassification rate α_{0j} and the sum of the misclassification rates $\alpha_{0j} + \alpha_{1j}$ for instrument value j match the values shown in the top panel. The values for the remaining parameters are discussed in Section 5. Panel B contains average estimates of β across iterations for the listed estimation methods. The standard deviations of the estimates across iterations are shown in parentheses.

Table A6: Monte Carlo Simulations - Varying α and β, $N = 1,000$

$\alpha_{0j} + \alpha_{0j}$	β=0.0	β=0.25	β=0.5	β=1.0	β=1.5	β=2.0
Panel A: Proportion GMM Converges						
0.00	0.964	0.847	0.907	0.988	0.997	1.000
0.13a	0.970	0.844	0.885	0.940	0.969	0.978
0.13b	0.964	0.838	0.897	0.944	0.971	0.967
0.26a	0.963	0.825	0.863	0.916	0.925	0.930
0.26b	0.968	0.848	0.869	0.914	0.926	0.937
0.39a	0.960	0.827	0.833	0.865	0.899	0.914
0.39b	0.974	0.828	0.843	0.891	0.895	0.905
Panel B: Mean of β if Converges						
0.00	0.003	0.679	0.883	1.187	1.586	2.064
0.13a	-0.010	0.755	0.984	1.427	2.021	2.567
0.13b	-0.001	0.896	1.000	1.496	2.098	2.559
0.26a	0.013	0.825	1.131	1.812	2.385	2.990
0.26b	-0.020	0.917	1.135	1.732	2.412	2.941
0.39a	-0.060	0.779	1.288	1.911	2.689	3.337
0.39b	-0.020	0.886	1.128	1.915	2.666	3.385
Panel C: Mean of SE of β if Converges						
0.00	6347	252.8	212.1	3.158	0.480	0.425
0.13a	1350	636.8	1346	15.04	15.06	11.71
0.13b	3011	3305	97.77	21.78	20.66	15.54
0.26a	3914	1017	412.7	232.7	66.74	46.41
0.26b	590.2	2499	1557	77.15	51.98	47.38
0.39a	2892	343.3	1362	183.3	207.7	87.83
0.39b	3655	5147	281.4	92.79	129.7	103.2

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (11). Each simulation uses 1,000 iterations. The numbers within each of the three panels correspond to a different simulation, varying the parameters β and $\alpha_{0j} + \alpha_{1j}$ and keeping the sample size fixed at $N = 1,000$. Each $\alpha_{0j} + \alpha_{1j}$ value has two specifications for α_{0j}, with α_{1j} then being set to equal the designated sum. The α_{0j} values for 0.13a are (0.055, 0.070, 0.085) and for 0.13b are (0.035, 0.070, 0.105), following columns (2) and (3) of Table 2; the values for 0.26a and 0.26b multiply these values by 2, respectively, and the values for 0.39a and 0.39b multiply these values by 3, respectively. The values for the remaining parameters are discussed in Section 5. Panel A reports the proportion of times that the GMM estimator converges across iterations. Panel B reports the average of β across the iterations where the estimator converges. Panel C reports the average standard error of β across the iterations where the estimator converges.

Table A7: Monte Carlo Simulations - Varying α and β, $N = 10,000$

$\alpha_{0j} + \alpha_{0j}$	β=0.0	β=0.25	β=0.5	β=1.0	β=1.5	β=2.0
Panel A: Proportion GMM Converges						
0.00	0.963	0.985	1.000	1.000	1.000	1.000
0.13a	0.960	0.979	0.999	1.000	1.000	1.000
0.13b	0.965	0.973	1.000	1.000	1.000	1.000
0.26a	0.969	0.952	0.997	0.999	1.000	1.000
0.26b	0.971	0.956	0.992	0.999	0.999	1.000
0.39a	0.968	0.928	0.986	0.995	0.995	0.998
0.39b	0.975	0.937	0.982	0.994	0.996	0.995
Panel B: Mean of β if Converges						
0.00	0.005	0.330	0.515	1.006	1.504	2.003
0.13a	0.048	0.385	0.539	1.017	1.517	2.018
0.13b	-0.010	0.357	0.537	1.017	1.516	2.017
0.26a	-0.001	0.414	0.566	1.042	1.571	2.059
0.26b	-0.001	0.412	0.551	1.039	1.553	2.062
0.39a	-0.001	0.432	0.653	1.130	1.647	2.217
0.39b	-0.001	0.437	0.624	1.109	1.672	2.208
Panel C: Mean of SE of β if Converges						
0.00	2962	20.06	0.082	0.066	0.065	0.065
0.13a	1170	33.50	0.300	0.113	0.137	0.168
0.13b	1045	14.46	0.24	0.113	0.138	0.168
0.26a	518.7	58.63	0.687	0.205	1.168	0.302
0.26b	649.6	48.01	0.244	0.186	0.264	0.315
0.39a	842.6	14.38	9.299	1.414	0.786	1.618
0.39b	329.6	13.57	3.114	0.763	1.730	0.985

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (11). Each simulation uses 1,000 iterations. The numbers within each of the three panels correspond to a different simulation, varying the parameters β and $\alpha_{0j}+\alpha_{1j}$ and keeping the sample size fixed at $N = 10,000$. Each $\alpha_{0j} + \alpha_{1j}$ value has two specifications for α_{0j}, with α_{1j} then being set to equal the designated sum. The α_{0j} values for 0.13a are (0.055, 0.070, 0.085) and for 0.13b are (0.035, 0.070, 0.105), following columns (2) and (3) of Table 2; the values for 0.26a and 0.26b multiply these values by 2, respectively, and the values for 0.39a and 0.39b multiply these values by 3, respectively. The values for the remaining parameters are discussed in Section 5. Panel A reports the proportion of times that the GMM estimator converges across iterations. Panel B reports the average of β across the iterations where the estimator converges. Panel C reports the average standard error of β across the iterations where the estimator converges.

Table A8: Monte Carlo Simulations - Varying α and β, $N = 100,000$

$\alpha_{0j} + \alpha_{0j}$	β=0.0	β=0.25	β=0.5	β=1.0	β=1.5	β=2.0
Panel A: Proportion GMM Converges						
0.00	0.963	1.000	1.000	1.000	1.000	1.000
0.13a	0.951	1.000	1.000	1.000	1.000	1.000
0.13b	0.976	1.000	1.000	1.000	1.000	1.000
0.26a	0.959	1.000	1.000	1.000	1.000	1.000
0.26b	0.974	1.000	1.000	1.000	1.000	1.000
0.39a	0.953	1.000	1.000	0.999	1.000	1.000
0.39b	0.965	1.000	1.000	1.000	1.000	1.000
Panel B: Mean of β if Converges						
0.00	0.007	0.254	0.503	1.002	1.502	2.002
0.13a	0.006	0.256	0.504	1.003	1.502	2.001
0.13b	0.003	0.256	0.504	1.002	1.502	2.001
0.26a	0.001	0.258	0.505	1.004	1.504	2.003
0.26b	0.020	0.258	0.505	1.005	1.504	2.004
0.39a	0.003	0.268	0.509	1.009	1.510	2.027
0.39b	-0.001	0.265	0.509	1.008	1.511	2.016
Panel C: Mean of SE of β if Converges						
0.00	255	0.022	0.020	0.020	0.020	0.020
0.13a	496	0.026	0.026	0.032	0.040	0.049
0.13b	337	0.026	0.026	0.032	0.040	0.050
0.26a	552	0.033	0.033	0.045	0.059	0.075
0.26b	479	0.033	0.033	0.045	0.060	0.076
0.39a	492	0.178	0.044	0.062	0.084	0.121
0.39b	209	0.050	0.044	0.061	0.085	0.110

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (11). Each simulation uses 1,000 iterations. The numbers within each of the three panels correspond to a different simulation, varying the parameters β and $\alpha_{0j}+\alpha_{1j}$ and keeping the sample size fixed at $N = 100,000$. Each $\alpha_{0j} + \alpha_{1j}$ value has two specifications for α_{0j}, with α_{1j} then being set to equal the designated sum. The α_{0j} values for 0.13a are (0.055, 0.070, 0.085) and for 0.13b are (0.035, 0.070, 0.105), following columns (2) and (3) of Table 2; the values for 0.26a and 0.26b multiply these values by 2, respectively, and the values for 0.39a and 0.39b multiply these values by 3, respectively. The values for the remaining parameters are discussed in Section 5. Panel A reports the proportion of times that the GMM estimator converges across iterations. Panel B reports the average of β across the iterations where the estimator converges. Panel C reports the average standard error of β across the iterations where the estimator converges.

Table A9: Further Empirical Results Based on Moment Inequalities

Outcome:	Medicaid (1)	Private (2)	Uninsured (3)
(A) Equality moments, bootstrapped	[0.11, 0.20]	[-0.08, 0.02]	[-0.07, -0.01]
(B) Equality moments, analytic	[0.12, 0.18]	[-0.05, -0.01]	[-0.04, -0.02]
(C) $\underline{R}_{jk} = 1, \underline{R}_{kj} = 1 \ \forall j > k$	[0.13, 0.20]	[-0.29, -0.02]	[-0.07, -0.03]
(D) $\underline{R}_{jk} = .95, \underline{R}_{kj} = .95 \ \forall j > k$	[0.11, 0.32]	[-0.56, -0.02]	[-0.08, -0.03]
(E) $\underline{R}_{jk} = 1, \underline{R}_{kj} = .95 \ \forall j > k$	[0.12, 0.23]	[-0.65, -0.02]	[-0.07, -0.03]
(F) $\underline{R}_{jk} = 1, \underline{R}_{kj} = .89 \ \forall j > k$	[0.11, 0.33]	[-0.79, -0.02]	[-0.07, -0.03]

Notes: This table reports estimates of β from equation (18). The controls include the number of people and workers in the household and indicators for state of residence, calendar year, child age, and householder characteristics (male, white, and type).