Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://economics.sas.upenn.edu/pier

PIER Working Paper 15-040

"On Mis-measured Binary Regressors: New Results and Some Comments on the Literature"

By

Francis J. DiTraglia and Camilo Garcia-Jimeno

http://ssrn.com/abstract=2687151

On Mis-measured Binary Regressors: New Results and Some Comments on the Literature*

Francis J. DiTraglia Camilo García-Jimeno

University of Pennsylvania

This Version: November 24, 2015, First Version: October 31, 2015

Abstract

This paper studies the use of a discrete instrumental variable to identify the causal effect of an endogenous, mis-measured, binary treatment. We begin by showing that the only existing identification result for this case, which appears in Mahajan (2006), is incorrect. As such, identification in this model remains an open question. We first prove that the treatment effect is unidentified based on conditional first-moment information, regardless of the number of values that the instrument may take. We go on to derive a novel partial identification result based on conditional second moments that can be used to test for the presence of mis-classification and to construct simple and informative bounds for the treatment effect. In certain special cases, we can in fact obtain point identification of the treatment effect based on second moment information alone. When this is not possible, we show that adding conditional third moment information point identifies the treatment effect and the measurement error process.

Keywords: Instrumental variables, Measurement error, Endogeneity, Binary regressor, Partial Identification

JEL Codes: C10, C18, C25, C26

*We thank Daron Acemoglu, Manuel Arellano, Kristy Buzard, Xu Cheng, Bernardo da Silveira, Bo Honoré, Yuya Takahashi, and seminar participants at Chicago Booth and Princeton for valuable comments and suggestions.

1 Introduction

Many treatments of interest in applied work are binary. To take a particularly prominent example, consider treatment status in a randomized controlled trial. Even if the randomization is pristine, which yields a valid binary instrument (the offer of treatment), subjects may select into treatment based on unobservables, and given the many real-world complications that arise in the field, measurement error may be an important concern. This paper studies the use of a discrete instrumental variable to identify the causal effect of an endogenous, mis-measured, binary treatment in a model with additively separable errors. Specifically, we consider the following model

$$y = h(T^*, \mathbf{x}) + \varepsilon \tag{1.1}$$

where $T^* \in \{0, 1\}$ is a mis-measured, endogenous treatment, $\mathbf{x}$ is a vector of exogenous controls, and ε is a mean-zero error. Since T^* is potentially endogenous, $\mathbb{E}[\varepsilon|T^*, \mathbf{x}]$ may not be zero. Our goal is to non-parametrically estimate the average treatment effect (ATE) function

$$\tau(\mathbf{x}) = h(1, \mathbf{x}) - h(0, \mathbf{x}). \tag{1.2}$$

using a single discrete instrumental variable $z \in \{z_k\}_{k=1}^K$. We assume throughout that z is a relevant instrument for T^*, in other words

$$\mathbb{P}(T^* = 1|z_j, \mathbf{x}) \neq \mathbb{P}(T^* = 1|z_k, \mathbf{x}), \quad \forall k \neq j. \tag{1.3}$$

While the structural relationship involves T^*, we observe only a noisy measure T, polluted by non-differential measurement error. In particular, we assume that

$$\mathbb{P}(T = 1|T^* = 0, z, \mathbf{x}) \;\; = \;\; \alpha_0(\mathbf{x}) \tag{1.4}$$

$$\mathbb{P}(T = 0|T^* = 1, z, \mathbf{x}) \;\; = \;\; \alpha_1(\mathbf{x}) \tag{1.5}$$

where the mis-classification error rates $\alpha_0(\mathbf{x})$ and $\alpha_1(\mathbf{x})$ can depend on $\mathbf{x}$ but not z, and additionally that, conditional on true treatment status, observed treatment status provides no additional information about the error term. In other words, we assume that

$$\mathbb{E}[\varepsilon|T^*, T, z, \mathbf{x}] = \mathbb{E}[\varepsilon|T^*, z, \mathbf{x}]. \tag{1.6}$$

Although a relevant case for applied work, the setting we consider here has received little attention in the literature. The only existing result for the case of an endogenous treatment appears in an important paper by Mahajan (2006), who is primarily concerned with the case of an exogenous treatment. As we show below, Mahajan's identification result for the endogenous treatment case is incorrect. As far as we are aware, this leaves the problem considered in this paper completely unsolved.

We begin by showing that the proof in Appendix A.2 of Mahajan (2006) leads to a contradiction. Throughout his paper, Mahajan (2006) maintains an assumption (Assumption 4) which he calls the "Dependency Condition." This assumption requires that the instrumental variable be relevant, namely that it generates variation in true treatment status. When extending his result for an exogenous treatment to the more general case of an endogenous one, however, he must impose an additional condition on the model (Equation 11), which turns out to violate the Dependency Condition. Since one cannot impose the condition in Equation 11 of Mahajan (2006), we go on to study the prospects for identification in this model more broadly. We consider two possibilities. First, since Mahajan's identification results require only a binary instrument, we borrow an idea from Lewbel (2007) and explore whether expanding the support of the instrument yields identification based on moment equations similar to those used by Mahajan (2006). While allowing the instrument to take on additional values does increase the number of available moment conditions, we show that these moments cannot point identify the treatment effect, regardless of how many (finite) values the instrument takes on.

We then consider a new source of identifying information that arises from imposing stronger assumptions on the instrumental variable. If the instrument is not merely mean independent but in fact *statistically independent* of the regression error term, as in a randomized controlled trial or a true natural experiment, additional moment conditions become available. We show that adding a conditional second moment independence assumption on the instrument identifies the *difference* of mis-classification rates $\alpha_1(\mathbf{x}) - \alpha_0(\mathbf{x})$. Because these rates must equal each other when there is no mis-classification error, our result can be used to test a necessary condition for the absence of measurement error. It can also be used to construct simple and informative partial identification bounds for the treatment effect. When one of the mis-classification rates is known, this identifies the treatment effect. More generally, however, this is not the case. We go on to show that a conditional third moment independence assumption on the instrument point identifies both $\alpha_0(\mathbf{x})$ and $\alpha_1(\mathbf{x})$ and hence the ATE function

$\tau(\mathbf{x})$. Both our point identification and partial identification results require only a binary instrument, and lead to simple, closed-form method of moments estimators that should be straightforward to apply in practice.

The remainder of this paper is organized as follows. In section 2 we discuss the literature in relation to the problem considered here. Section 3 introduces notation and assumptions, and presents our main results. Section 4 concludes. All proofs appear in the Appendix.

2 Related Literature

Measurement error is a pervasive feature of economic data, motivating a long tradition of measurement error modelling in econometrics. The textbook case considers a continuous regressor (treatment) subject to classical measurement error in a linear model. In this setting, the measurement error is assumed to be unrelated to the true, unobserved, value of the treatment of interest. Regardless of whether this unobserved treatment is exogenous or endogenous, a single valid instrument suffices to identify its effect. When an instrument is unavailable, Lewbel (1997) shows that higher moment assumptions can be used to construct one, provided that the mis-measured treatment is exogenous. When it is endogenous, Lewbel (2012) uses a heteroskedasticity assumption to obtain identification.

Departures from the linear, classical measurement error setting pose serious identification challenges. One strand of the literature considers relaxing the assumption of linearity while maintaining that of classical measurement error. Schennach (2004), for example, uses repeated measures of each mis-measured treatment to obtain identification, while Schennach (2007) uses an instrumental variable. Both papers consider the case of exogenous treatments.[1] More recently, Song et al. (2015) rely on a repeated measure of the mis-measured treatment and the existence of a set of additional regressors, conditional upon which the treatment of interest is unrelated to the unobservables, to obtain identification. Another strand of the literature considers relaxing the assumption of classical measurement error, by allowing the measurement error to be related to the true value of the unobserved treatment. Chen et al. (2005) obtain identification in a general class of moment condition models with mis-measured data

[1] For comprehensive reviews of the challenges of addressing measurement error in non-linear models, see Chen et al. (2011) and Schennach (2013).

by relying on the existence of an auxiliary dataset from which they can estimate the measurement error process. In contrast, Hu and Shennach (2008) and Song (2015) rely on an instrumental variable and an additional conditional location assumption on the measurement error distribution. More recently, Hu et al. (2015) use a continuous instrument to identify the ratio of partial effects of two continuous regressors, one measured with error, in a linear single index model.

Many treatments of interest in economics, however, are binary, and in this case classical measurement error is impossible. Because a true 1 can only be mis-measured as a 0 and a true 0 can only be mis-measured as a 1, the measurement error must be *negatively* correlated with the true treatment status (Aigner, 1973; Bollinger, 1996). For this reason, even in a textbook linear model, the instrumental variables estimator can only remove the effect of endogeneity, not that of measurement error (Frazis and Loewenstein, 2003). Measurement error in a discrete variable is usually called mis-classification.[2] The simplest form of mis-classification is so-called *non-differential* measurement error. In this case, conditional on true treatment status, and possibly a set of exogenous covariates, the measurement error is assumed to be unrelated to all other variables in the system.

A number of papers have studied this problem without the use of instrumental variables under the assumption that the mis-measured binary treatment is exogenous. The first to address this problem was Aigner (1973), who characterized the asymptotic bias of the OLS estimator in this setting, and proposed a technique for correcting it using outside information on the mis-classification process. Another early contribution by Bollinger (1996) provides partial identification bounds. More recently, Chen et al. (2008a) use higher moment assumptions to obtain identification in a linear regression model, and Chen et al. (2008b) extend these results to the non-parametric setting. van Hasselt and Bollinger (2012) and Bollinger and van Hasselt (2015) provide additional partial identification results.

Continuing under the assumption of an exogenous treatment, a number of other papers in the literature have considered the identifying power of an instrumental variable, or something like one. Black et al. (2000) and Kane et al. (1999) more-or-less simultaneously pointed out that when *two* alternative measures of treatment are available, both subject to non-differential measurement error, a non-linear GMM estimator can be used to recover the treatment effect. In essence, one measure serves as an instrument

[2]For general results on the partial identification of discrete probability distributions using misclassified observations, see Molinari (2008).

for the other although the estimator is quite different from IV.[3] Subsequently, Frazis and Loewenstein (2003) correctly note that an instrumental variable can take the place of one of the measures of treatment in a linear model with an exogenous treatment, allowing one to implement a variant of the GMM estimator proposed by Black et al. (2000) and Kane et al. (1999). However, as we will show below, the assumptions required to obtain this result are stronger than Frazis and Loewenstein (2003) appear to realize: the usual IV assumption that the instrument is mean independent of the regression error is insufficient for identification.

Mahajan (2006) extends the results of Black et al. (2000) and Kane et al. (1999) to a more general nonparametric regression setting using a binary instrument in place of one of the treatment measures. Although unaware of Frazis and Loewenstein (2003), Mahajan (2006) makes the correct assumption over the instrument and treatment to guarantee identification of the conditional mean function. When the treatment is in fact exogenous, this coincides with the treatment effect. Hu (2008) derives related results when the mis-classified discrete regressor may take on more than two values. Lewbel (2007) provides an identification result for the same model as Mahajan (2006) under different assumptions. In particular, the variable that plays the role of the "instrument" need not satisfy the exclusion restriction provided that it does not interact with the treatment and takes on at least three distinct values.

Much less is known about the case in which a binary, or discrete, treatment is not only mis-measured but endogenous. Frazis and Loewenstein (2003) briefly discuss the prospects for identification in this setting. Although they do not provide a formal proof they argue, in the context of their parametric linear model, that the treatment effect is unlikely to be identified unless one is willing to impose strong and somewhat unnatural conditions.[4] The first paper to provide a formal result for this case is Mahajan (2006). He extends his main result to the case of an endogenous treatment, providing an explicit proof of identification under the usual IV assumption in a model with additively separable errors. As we show below, however, Mahajan's proof is incorrect.

[3] Ignoring covariates, the observable moments in this case are the joint probability distribution of the two binary treatment measures and the conditional means of the outcome variable given the two measures. Although the system is highly non-linear, it can be manipulated to yield an explicit solution for the treatment effect provided that the true treatment is exogenous.

[4] For example, one could consider using the results of Hausman et al. (1998), who study regressions with a mis-classified, discrete *outcome* variable, as a first-stage in an IV setting. In principle, this approach would fully identify the mis-classification error process. Using these results, however, requires either an explicit, nonlinear, parametric model for the first stage, or an identification at infinity argument.

The results we derive here most closely relate to the setting considered in Mahajan (2006) in that we study non-parametric identification of the effect of a binary, endogenous treatment, using a discrete instrument. Unlike Mahajan (2006) we consider and indeed show the necessity of using higher-moment information to identify the causal effect of interest. Unlike Kreider et al. (2012), who partially identify the effects of food stamps on health outcomes of children under weak measurement error assumptions, we do not rely on auxiliary data. Unlike Shiu (2015), who considers a sample selection model with a discrete, mis-measured, endogenous regressor, we do not rely on a parametric assumption about the form of the first-stage. Finally unlike Ura (2015), who studies local average treatment effects under very general forms of mis-classification but presents only partial identification results, we point identify an average treatment effect under non-differential measurement error. Moreover, unlike the identification strategies from the existing literature described above, we do not rely upon continuity of the instrument, a large support condition, or restrictions on the relationship between the true, unmeasured treatment and its observed surrogate, subject to the condition that the measurement error process is non-differential.

3 Main Results

3.1 Notation and Basic Properties of the Model

Consider the model described in Equations 1.1–1.6. Our arguments below, like those of Mahajan (2006) and Lewbel (2007), proceed by holding the exogenous covariates *fixed* at some level $\mathbf{x}_a$. As such, there is no loss of generality from suppressing dependence on $\mathbf{x}$ in our notation. It should be understood throughout that any conditioning statements are evaluated at $\mathbf{x} = \mathbf{x}_a$. To this end let $c = h(0, \mathbf{x}_a)$ and define $\beta = h(1, \mathbf{x}_a) - h(0, \mathbf{x}_a)$. Using this notation, Equation 1.1 can be re-expressed as a simple linear model, namely

$$y = \beta T^* + u \tag{3.1}$$

where we define $u = c + \varepsilon$, an error term that need not be mean zero. We maintain throughout that $\beta \neq 0$. If $\beta = 0$ then there is no meaningful sense in which there is "mis-classification" since T^* is irrelevant for y. Because the probability limit of the usual Wald IV estimator in this model is proportional to β, as we will discuss below, this condition can be directly assessed from the data.

	$z=1$	$z=2$	$\dots$	$z=K$
$T=0$	$\bar{y}_{01}$ / p_{01}	$\bar{y}_{02}$ / p_{02}	$\dots$	$\bar{y}_{0K}$ / p_{0K}
$T=1$	$\bar{y}_{11}$ / p_{11}	$\bar{y}_{12}$ / p_{12}	$\dots$	$\bar{y}_{1K}$ / p_{1K}

Table 1: Observables, using the shorthand $p_{0k} = q_k(1-p_k)$ and $p_{1k} = q_k p_k$.

From the perspective of non-parametric identification, the observables in this problem are the conditional distribution of y given (T, z), the conditional probabilities of T given z, and the marginal probabilities of z. For now, following the existing literature, we will restrict attention to the conditional mean of y. Below we consider using higher moments of y. Let $\bar{y}_{t,k}$ denote $\mathbb{E}[y|T=t, z=z_k]$, let p_k denote $\mathbb{P}(T=1|z=z_k)$ and let q_k denote $\mathbb{P}(z=z_k)$. Table 1 depicts the observable first moments for this problem.

The observed cell means $\bar{y}_{tk}$ depend on a number of unobservable parameters which we now define. Let m^*_{tk} denote the conditional mean of u given $T^*=t$ and $z=z_k$, $\mathbb{E}[u|T^*=t, z=z_k]$, and let p^*_k denote $\mathbb{P}(T^*=1|z=z_k)$. These quantities are depicted in Table 2. By the Law of Total Probability and the definitions of p_k and p^*_k,

$$
\begin{aligned}
p_k &= \mathbb{P}(T=1|z=z_k, T^*=0)(1-p^*_k) + \mathbb{P}(T=1|z=z_k, T^*=1)p^*_k \\
&= \alpha_0(1-p^*_k) + (1-\alpha_1)p^*_k
\end{aligned}
$$

since the misclassification probabilities do not depend on z by Equations 1.4–1.5. Rearranging,

$$
p^*_k = \frac{p_k - \alpha_0}{1-\alpha_0-\alpha_1}, \quad 1 - p^*_k = \frac{1-p_k-\alpha_1}{1-\alpha_0-\alpha_1}. \tag{3.2}
$$

Equation 3.2 implies that p^*_k is observable given knowledge of α_0 and α_1, since p_k is observable. Note that for these equations to be meaningful, we require that $\alpha_0 + \alpha_1 \neq 1$. Indeed, the existing literature (Black et al., 2000; Frazis and Loewenstein, 2003; Kane et al., 1999; Lewbel, 2007; Mahajan, 2006) imposes the stronger condition that $\alpha_0 + \alpha_1 < 1$, namely that the measurement error is not so severe that $1 - T$ is a better predictor of T^* than T is, and vice-versa. In the absence of this assumption the treatment effect would only be identified up to sign. Our identification result, presented below, will require that $\alpha_0 + \alpha_1 < 1$ whereas our partial identification result will not.

A key assumption below will be the conditional mean independence of the error

	$z=1$	$z=2$	$\dots$	$z=K$
$T^*=0$	m^*_{01} / p^*_{01}	m^*_{02} / p^*_{02}	$\dots$	m^*_{0K} / p^*_{0K}
$T^*=1$	m^*_{11} / p^*_{11}	m^*_{12} / p^*_{12}	$\dots$	m^*_{1K} / p^*_{1K}

Table 2: Unobservables, using the shorthand $p^*_{0k} = q_k(1-p^*_k)$ and $p^*_{1k} = q_k p^*_k$.

term and instrument, in other words $\mathbb{E}\left[\varepsilon|z\right] = 0$. Since we have defined $u = c + \varepsilon$, this assumption can be expressed in terms of m^*_{tk} as

$$(1-p^*_k)m^*_{0k} + p^*_k m^*_{1k} = c \tag{3.3}$$

for all $k = 1, \dots, K$. This restriction imposes that a particular weighted sum over the rows of a given column of Table 2 takes the same value *across* columns.

3.2 Mahajan's Approach

Here we show that Mahajan's proof of identification for an endogenous treatment is incorrect. The problem is subtle so we give his argument in full detail. We continue to supress dependence on the exogenous covariates $\mathbf{x}$.

The first step of Mahajan's argument is to show that if one could recover the conditional mean function of y given T^*, then a valid and relevant binary instrument would suffice to identify the treatment effect.

Assumption 1 (Mahajan A2)**.** *Suppose that* $y = c + \beta T^* + \varepsilon$ *where*

(i) $\mathbb{E}[\varepsilon|z] = 0$

(ii) $\mathbb{P}(T^* = 1|z_k) \neq \mathbb{P}(T^* = 1|z_\ell)$ *for all* $k \neq \ell$

(iii) $\mathbb{P}(T = 1|T^* = 0, z) = \alpha_0$, $\mathbb{P}(T = 0|T^* = 1, z) = \alpha_1$

(iv) $\alpha_0 + \alpha_1 \neq 1$

(v) $\beta \neq 0$

Lemma 1 (Mahajan A2)**.** *Under Assumption 1, knowledge of the mis-classification error rates* α_0, α_1 *suffices to identify* β*.*

In his Theorem 1, Mahajan (2006) proves that α_0, α_1 can in fact be identified under the following assumptions.[5]

Assumption 2 (Mahajan A1)**.** *Define* $\nu = y - \mathbb{E}[y|T^*]$ *so that by construction we have* $\mathbb{E}[\nu|T^*] = 0$*. Assume that*

(i) $\mathbb{E}[\nu|T^*, T, z] = 0$.[6]

(ii) $\mathbb{P}(T^* = 1|z_k) \neq \mathbb{P}(T^* = 1|z_\ell)$ *for all* $k \neq \ell$

(iii) $\mathbb{P}(T = 1|T^* = 0, z) = \alpha_0$, $\mathbb{P}(T = 0|T^* = 1, z) = \alpha_1$

(iv) $\alpha_0 + \alpha_1 < 1$

(v) $\mathbb{E}[y|T^* = 0] \neq \mathbb{E}[y|T^* = 1]$

Lemma 2 (Mahajan Theorem 1)**.** *Under Assumptions 2, the error rates* α_0, α_1 *are identified as is the conditional mean function* $\mathbb{E}[y|T^*]$*.*

Notice that the identification of the error rates in Lemma 2 does not depend on the interpretation of the conditional mean function $\mathbb{E}[y|T^*]$. If T^* is an exogenous treatment, the conditional mean coincides with the treatment effect; if it is endogenous, this is not the case. Either way, the meaning of α_0, α_1 is unchanged: these parameters simply characterize the mis-classification process. Based on this observation, Mahajan (2006) claims that he can rely on Lemma 2 to identify α_0, α_1 and thus the causal effect β when the treatment is endogenous via Lemma 1. To do this, he must build a bridge between Assumption 1 and Assumption 2 that allows T^* to be endogenous. Mahajan (2006) does this by imposing one additional assumption: Equation 11 in his paper.

Assumption 3 (Mahajan Equation 11)**.** *Let* $y = c + \beta T^* + \varepsilon$ *where* $\mathbb{E}[\varepsilon|T^*]$ *may not be zero and suppose that*

$$\mathbb{E}[\varepsilon|T^*, T, z] = \mathbb{E}[\varepsilon|T^*].$$

Lemma 3. *Suppose that* $y = c + \beta T^* + \varepsilon$ *where* $E[\varepsilon|z] = 0$ *and define the unobserved projection error* $\nu = y - \mathbb{E}[y|T^*]$*. Then Assumption 3 implies that* $E[\nu|T^*, T, z] = 0$*, which is Assumption 2(i).*

[5] Technically, one additional assumption is required, namely that the conditional mean of y given T^* and any covariates would be identified if T^* were observed.

[6] This is Mahajan's Equation (I).

To summarize, Mahajan's claim is equivalent to the proposition that under Assumptions 1(i), 2(ii)–(v), and 3, β is identified even if T^* is endogenous. although Lemmas 1, 2 and 3 are all correct, Mahajan's claim is not.[7] While Assumption 3 does guarantee that Assumption 2(i) holds, when combined with Assumption 1(i) it also implies that 2(ii) fails if T^* is endogenous. The failure of Assumption 2(ii) in turn leads to a division by zero in the solution to the linear system following Mahajan's displayed Equation 26: the system no longer has a unique solution so identification fails.[8]

Proposition 1 (Lack of a First Stage). *Suppose that Assumptions 1(i) and 3 hold and* $\mathbb{E}[\varepsilon|T^*] \neq 0$. *Then* $\mathbb{P}(T^* = 1|z_1) = \mathbb{P}(T^* = 1|z_2)$, *violating Assumption 2(ii).*

To understand the economic intuition behind Proposition 1, consider a simple example in which we randomize the offer of a job training program to a sample of workers to study the impact on future earnings. In this context z indicates whether a particular individual is *offered* job training by the experimenter while T^* indicates whether she actually *obtains* job training from any source, inside or outside of the experiment. We observe not T^* but a self-report T that is measured with error. In this example u contains all of the unobservable factors that determine an individual's wage.

Assumption 3 allows for endogenous treatment receipt: $\mathbb{E}[u|T^* = 1]$ may be different from $\mathbb{E}[u|T^* = 0]$. We might expect, for example, that individuals who obtain job training are more motivated than those who do not, and hence earn higher wages on average. However, Assumption 3 imposes that $\mathbb{E}\left[u|T^* = t, z_1\right] = \mathbb{E}\left[u|T^* = t, z_2\right]$ for $t = 0, 1$. This has two implications. First, it means that, among those who do not obtain job training, the average value of u is the same for those who were offered training and those who were not. Second, it means that, among those who *did* obtain job training, the average value of u is the same for those who were offered training and those who were not. In other words, Assumption 3 requires that there is *no selection on unobservables*. This is exactly the opposite of what we would expect in the job training

[7]Our Lemma 3 does not in fact appear in Mahajan (2006), but it is an implicit step in his proof in Appendix A2.

[8]Notice that the root of the problem is the attempt to use *one* instrument to solve both the measurement error and endogeneity problems. In a setting where one had a second mis-measured surrogate for T^* *in addition* to an instrument that is conditionally mean independent of ε one could use the second surrogate as an instrument for the first to estimate α_0 and α_1 via Lemma 1 and then use the additional instrumental variable to estimate $\beta/(1 - \alpha_0 - \alpha_1)$ via the familiar Wald IV estimator. This is effectively the approach used by Battistin et al. (2014) to evaluate the returns to schooling in a setting with multiple misreported measures of educational qualifications.

setting. For example, individuals who are offered job training but refuse it, are likely to be very different from those who are not offered training and fail to obtain it from an outside source. And herein lies the problem: Assumption 3 simultaneously allows endogeneity and rules out selection. Given that the offer of job training is randomly assigned, and hence a valid instrument, the only way to avoid a contradiction is if there is no first stage: the fraction of individuals who take up job training cannot depend on the offer of training.

3.3 Lack of Identification From Conditional Means

We have seen that Mahajan (2006)'s approach based on a binary instrument cannot identify β when the treatment is endogenous: Assumption 3 in fact implies that the instrument is *irrelevant*. We now show that, regardless of how many values the instrument takes on, conditional mean information is insufficient for identification.

To begin, consider the model in Equation 3.1 without any restrictions on the m^*_{tk}, that is *without* imposing the IV restriction given in Equation 3.3. In this fully general case, the $2K+3$ unknown parameters are $\beta, \alpha_0, \alpha_1$ and the conditional means of u, namely m^*_{tk}. In contrast, there are only $2K$ available moment conditions.

Lemma 4. *Suppose that* $\mathbb{E}\left[\varepsilon|T^*, T, z\right] = \mathbb{E}\left[\varepsilon|T^*, z\right]$. *Then, under Assumption 1 (ii)–(iv),*

$$\hat{y}_{0k} = \frac{\alpha_1(p_k - \alpha_0)(\beta + m^*_{1k}) + (1-\alpha_0)(1-p_k-\alpha_1)m^*_{0k}}{1-\alpha_0-\alpha_1} \tag{3.4}$$

$$\hat{y}_{1k} = \frac{(1-\alpha_1)(p_k - \alpha_0)(\beta + m^*_{1k}) + \alpha_0(1-p_k-\alpha_1)m^*_{0k}}{1-\alpha_0-\alpha_1} \tag{3.5}$$

where $\hat{y}_{0k} = (1-p_k)\bar{y}_{0k}$ *and* $\hat{y}_{1k} = p_k\bar{y}_{1k}$.

Notice that the observable "weighted" cell mean $\hat{y}_{tk}$ defined in the preceding lemma depends on both m^*_{tk} *and* $m^*_{1-t,k}$ since the cell in which $T=t$ from Table 1 is in fact a mixture of both the cells $T^*=0$ and $T^*=1$ from Table 2, for a particular column k. Clearly we have fewer equations than unknowns. What additional restrictions could we consider imposing on the system? In a very interesting paper, Lewbel (2007) proposes using a three-valued "instrument" that does *not* satisfy the exclusion restriction. By assuming instead that there is no *interaction* between the instrument and the treatment, he is able to prove identification of the treatment effect. Using our notation it is easy to see why Lewbel (2007) requires a three-valued instrument. His moment conditions are equivalent to Equations 3.4 and 3.5 with the additional restriction that $m^*_{0k} = m^*_{1k}$ for all $k = 1, \ldots, K$. This leaves the number of equations unchanged at $2K$, but reduces the number of unknowns to $K + 3$. The smallest K for which $K + 3$ is at least as large as $2K$ is 3.[9]

Unlike Lewbel (2007) we, along with Mahajan (2006) and others, assume that $\mathbb{E}[\varepsilon|z] = 0$ so that Equation 3.3 holds.

Corollary 1. *Suppose that $\mathbb{E}[\varepsilon|T^*, T, z] = \mathbb{E}[\varepsilon|T^*, z]$. Then, under Assumption 1,*

$$\hat{y}_{0k} = \alpha_1(p_k - \alpha_0)\left(\frac{\beta}{1-\alpha_0-\alpha_1}\right) + (1-\alpha_0)c - (p_k - \alpha_0)m^*_{1k} \tag{3.6}$$

$$\hat{y}_{1k} = (1-\alpha_1)(p_k - \alpha_0)\left(\frac{\beta}{1-\alpha_0-\alpha_1}\right) + \alpha_0 c + (p_k - \alpha_0)m^*_{1k} \tag{3.7}$$

where $\hat{y}_{0k} = (1 - p_k)\bar{y}_{0k}$ and $\hat{y}_{1k} = p_k\bar{y}_{1k}$.

Equations 3.6 and 3.7 also make it clear why the IV estimator is inconsistent in the face of non-differential measurement error, and that this inconsistency does not depend on the endogeneity of the treatment, as noted by Frazis and Loewenstein (2003). Adding together Equations 3.6 and 3.7 yields

$$\hat{y}_{0k} + \hat{y}_{1k} = c + (p_k - \alpha_0)\left(\frac{\beta}{1-\alpha_0-\alpha_1}\right)$$

completely eliminating the m^*_{1k} from the system. Taking the difference of the preceding expression evaluated at two different values of the instrument, z_k and z_ℓ, and rearranging

$$\mathcal{W} = \frac{(\hat{y}_{0k} + \hat{y}_{1k}) - (\hat{y}_{0\ell} + \hat{y}_{1\ell})}{p_k - p_\ell} = \frac{\beta}{1-\alpha_0-\alpha_1} \tag{3.8}$$

which is the well-known Wald IV estimator, since $\hat{y}_{0k} + \hat{y}_{1k} = \mathbb{E}[y|z = z_k]$.

Imposing $\mathbb{E}[\varepsilon|z] = 0$ replaces the K unknown parameters $\{m^*_{0k}\}_{k=1}^K$ in Equations 3.4–3.5 with a single parameter c, leaving us with the same $2K$ equations but only $K + 4$ unknowns. When $K = 2$ (a binary instrument) we have 4 equations and

[9] The context considered by Lewbel (2007) is slightly different from the one we pursue here, in that his "instrument" is more like a covariate: it is allowed to have a direct effect on the outcome of interest. For this reason, Lewbel (2007) cannot use the exogeneity of the treatment to obtain identification based on a two-valued instrument.

6 unknowns. So how can one identify β in this case? The literature has imposed additional assumptions which, using our notation, can once again be mapped into restrictions on the m^*_{tk}. Black et al. (2000), Kane et al. (1999), and Mahajan (2006) make a *joint* exogeneity assumption on (T^*, z), namely $\mathbb{E}[\varepsilon|T^*, z] = 0$. Notice that this is strictly stronger than assuming that the instrument is valid and the treatment is exogenous. In our notation, this joint exogeneity assumption is equivalent to imposing $m^*_{tk} = c$ for all t, k. This reduces the parameter count to 4 regardless of the value of K. Thus, when the instrument is binary, we have exactly as many equations as unknowns. The arguments in Black et al. (2000), Kane et al. (1999), and Mahajan (2006) are all equivalent to solving Equations 3.6 and 3.7 for β under the added restriction that $m^*_{1k} = c$, establishing identification for this case. Frazis and Loewenstein (2003) use the same argument in a linear model with a potentially continuous instrument, but impose only the weaker conditions that the treatment is exogenous and the instrument is valid. Nevertheless, a crucial step in their derivation implicitly assumes the stronger joint exogeneity assumption used by Black et al. (2000), Kane et al. (1999) and Mahajan (2006). Without this assumption, their proof does not in fact go through.

If one wishes to allow for an endogenous treatment, the joint exogeneity assumption $m^*_{tk} = c$ is unusable and we have $2K$ equations in $K + 4$ unknowns. Based on the identification arguments described above, there would seem to be two possible avenues for identification of the treatment effect when a valid instrument is available. One idea would be to impose alternative conditions on the m^*_{tk} that are compatible with an endogenous treatment. If z is binary, two additional restrictions would suffice to equate the counts of moments and unknowns. As we showed in Proposition 1, however, this approach fails. Another idea, inspired by Lewbel (2007), would be to rely on an instrument that takes on more than two values. Following this approach would suggest a 4-valued instrument, the smallest value of K for which $2K = K + 4$. Unfortunately this approach fails as well, as we now show.

Theorem 1 (Lack of Identification)**.** *Suppose that Assumption 1 holds and additionally that $\mathbb{E}[\varepsilon|T^*, T, z] = \mathbb{E}[\varepsilon|T^*, z]$ (non-differential measurement error). Then regardless of how many values z takes on, generically β is unidentified based on the observables contained in Table 1.*

The preceding argument establishes lack of identification by deriving a parametric relationship between β and $\alpha_0, \alpha_1, \{m^*_{1k}\}_{k=1}^K$. So long as we adjust the other parameters according to this relationship, we are free to vary β while leaving all observable moments unchanged. This holds regardless of the number of values, K, that the instrument takes on.

3.4 Identification Based on Higher Moments

Having shown that the moment conditions from Table 1 do not identify β regardless of the value of K, we now consider exploiting the information contained in higher moments of y. When z is not merely mean-independent but in fact *statistically* independent of ε, as in a randomized controlled trial or a true natural experiment, the following assumptions hold automatically.

Assumption 4 (Second Moment Independence). $\mathbb{E}[\varepsilon^2|z] = \mathbb{E}[\varepsilon^2]$

Assumption 5 (Third Moment Independence). $\mathbb{E}[\varepsilon^3|z] = \mathbb{E}[\varepsilon^3]$

Theorem 2. *Under Assumption 4 and the conditions of Theorem 1 the difference of mis-classification rates, $(\alpha_1 - \alpha_0)$ is identified provided that z takes on at least two values.*

The preceding result can be used in several ways. One possibility is to test for the presence of mis-classification error. If the treatment is measured without error, then α_0 must equal α_1. By examining the identified quantities $\mathcal{R}$ and $\mathcal{W}$, one could possibly discover that this requirement it violated.[10] Moreover, in some settings mis-classification may be one-sided. In a smoking and birthweight example, it seems unlikely that mothers who did *not* smoke during pregnancy would falsely claim to have smoked. If either of α_0, α_1 is known, Theorem 2 point identifies the unknown error rate and hence β, using the fact that $\beta = \mathcal{W}(1 - \alpha_0 - \alpha_1)$. When neither of the error rates is known *a priori*, the same basic idea can be used to construct *bounds* for β. We now show that by augmenting Theorem 2 with information on conditional *third* moments, we can point identify β.

Theorem 3. *Under Assumptions 4-5 and the conditions of Theorem 1, the mis-classification rates α_0 and α_1 and the treatment effect β are identified provided that $\alpha_0 + \alpha_1 < 1$ and z takes on at least two values.*

Note that, unlike Theorem 2, Theorem 3 requires the assumption that $\alpha_0 + \alpha_1 < 1$. Without this assumption, we identify β only up to sign.

[10] Note that Theorem 2 requires $\alpha_0 + \alpha_1 \neq 1$, since the Wald estimator would otherwise be undefined. Accordingly, we cannot test for measurement error in the non-generic case $\alpha_0 = \alpha_1 = 1/2$.

3.5 Estimation

We now briefly describe how the identification results from above can be applied in practice. In our arguments above we suppressed dependence on the covariates $\mathbf{x}$. We now make this dependence explicit to illustrate how one can estimate the ATE $\tau(\mathbf{x})$ and the mis-classification rates $\alpha_0(\mathbf{x})$ and $\alpha_1(\mathbf{x})$ non-parametrically as a function of $\mathbf{x}$. Our proofs, which appear below in the appendix, provide closed-form expressions for each of these quantities in terms of three objects: $\mathcal{W}(\mathbf{x})$, defined in Equation 3.8, $\mathcal{R}(\mathbf{x})$, defined in Equation A.11 and $\mathcal{S}(\mathbf{x})$, defined in Equation A.20. Provided that $p_k(\mathbf{x}) \neq p_\ell(\mathbf{x})$ for all $k \neq \ell$, which follows from Assumption 2(ii) by Equation 3.2, $\alpha_0(\mathbf{x}), \alpha_1(\mathbf{x})$ and $\tau(\mathbf{x})$ are smooth functions of $(\mathcal{W}(\mathbf{x}), \mathcal{R}(\mathbf{x}), \mathcal{S}(\mathbf{x}))$. Under the same condition, these in turn are smooth functions of the underlying conditional expectation functions $\mathbb{E}\left[T|z,\mathbf{x}\right], \mathbb{E}\left[y|z,\mathbf{x}\right]$, $\mathbb{E}\left[y^2|z,\mathbf{x}\right]$, $\mathbb{E}\left[y^3|z,\mathbf{x}\right]$, $\mathbb{E}\left[yT|z,\mathbf{x}\right]$, and $\mathbb{E}\left[y^2T|z,\mathbf{x}\right]$. One could employ any number of existing nonparametric techniques to estimate these conditional expectation functions, at which point one is left with a standard, two-step method of moments estimation problem.[11]

4 Conclusion

This paper has presented the first point identification result for the effect of an endogenous, binary, mis-measured treatment using a discrete instrument. While our results require us to impose stronger conditions on the instrument, these conditions are satisfied in a number of empirically relevant examples, for example randomized controlled trials and true natural experiments. We obtain identification by augmenting conditional first moments with additional information contained in second and third moments and further derive a partial identification result based on first and second moments alone. In addition, and contrary to an incorrect previous result in Mahajan (2006), we showed that appealing to higher moments is necessary if one wishes to obtain identification: first moment information alone cannot identify the causal effect of an endogenous, mis-classified binary treatment regardless of the number of values the instrument may take. We have focused here on establishing identification and partial identification results using a particular set of moment conditions. More generally one could consider the use of additional moment conditions based on an independence

[11]For more details of the implementation of such a procedure see, e.g., Mahajan (2006) Section 4 or Lewbel (2007) Section 4.

assumption for the instrument. From the standpoint of estimation, rather than identification, the use of such moments could provide efficiency gains. Another possible avenue for future research would be to extend our framework to the fuzzy regression discontinuity setting.

Appendix A Proofs

Proof of Lemma 1. Since z is a valid instrument that does not influence the misclassification probabilities

$$\mathbb{E}[y|z_k] = c + \beta\mathbb{E}[T^*|z_k] + \mathbb{E}[\varepsilon|z_k] = c + \beta p_k^* = c + \beta\left(\frac{p_k - \alpha_0}{1 - \alpha_0 - \alpha_1}\right)$$

by Equation 3.2. Since p_k is observed, and z takes on two values, this is a system of two linear equations in c, β provided that α_0, α_1 are known. A unique solution exists if and only if $p_1 \neq p_2$. □

Proof of Lemma 2. See Mahajan (2006) Appendix A.1. □

Proof of Lemma 3. Taking conditional expectations of the causal model,

$$\mathbb{E}[y|T^*] = c + \beta T^* + \mathbb{E}[\varepsilon|T^*]$$

which implies that

$$\nu = y - c - \beta T^* - \mathbb{E}[\varepsilon|T^*] = \varepsilon - \mathbb{E}[\varepsilon|T^*].$$

Now, taking conditional expectations of both sides given T^*, T, z, we see that

$$\begin{aligned}\mathbb{E}[\nu|T^*, T, z] &= \mathbb{E}[\varepsilon|T^*, T, z] - \mathbb{E}\left[\mathbb{E}\left(\varepsilon|\, T^*\right) | \, T, T^*, z\right] \\ &= \mathbb{E}[\varepsilon|T^*, T, z] - \mathbb{E}\left[\varepsilon|\, T^*\right] = 0\end{aligned}$$

by Assumption 3, since $\mathbb{E}[\varepsilon|T^*]$ is (T^*, T, z)–measurable. □

Proof of Proposition 1. By the Law of Iterated Expectations,

$$\mathbb{E}[\varepsilon|T^*, z] = \mathbb{E}_{T|T^*,z}\left[\mathbb{E}\left(\varepsilon|T^*, T, z\right)\right] = \mathbb{E}_{T|T^*,z}\left[\mathbb{E}\left(\varepsilon|T^*\right)\right] = \mathbb{E}\left[\varepsilon|T^*\right] \tag{A.1}$$

where the second equality follows from Assumption 3 and the final equality comes from the fact that $\mathbb{E}[\varepsilon|T^*]$ is (T^*, z)–measurable. Using our notation from above let $u = c+\varepsilon$ and define $m^*_{tk} = \mathbb{E}[u|T^* = t, z = z_k]$. Since c is a constant, by Equation A.1 we see that $m^*_{01} = m^*_{02}$ and $m^*_{11} = m^*_{12}$. Now, by Assumption 1(i) we have $\mathbb{E}[\varepsilon|z] = 0$ so that $\mathbb{E}[u|z_1] = \mathbb{E}[u|z_2] = c$. Again using iterated expectations,

$$\mathbb{E}\left[u|z_1\right] = \mathbb{E}_{T^*|z_1}\left[\mathbb{E}\left(u|T^*, z_1\right)\right] = (1-p^*_1)m^*_{01} + p^*_1 m^*_{11} = c$$
$$\mathbb{E}\left[u|z_2\right] = \mathbb{E}_{T^*|z_2}\left[\mathbb{E}\left(u|T^*, z_2\right)\right] = (1-p^*_2)m^*_{02} + p^*_2 m^*_{12} = c$$

The preceding two equations, combined with $m^*_{01} = m^*_{02}$ and $m^*_{11} = m^*_{12}$ imply that $p^*_1 = p^*_2$ unless $m^*_{01} = m^*_{11} = m^*_{02} = m^*_{12} = c$. But this four-way equality is ruled out by the assumption that $\mathbb{E}[\varepsilon|T^*] \neq 0$. □

Proof of Lemma 4. The result follows by combining Equation 3.2 with Bayes' rule and the Law of Iterated Expectations applied to Equation 3.1. □

Proof of Corollary 1. Using Equation 3.2 and rearranging,

$$\frac{(1-p_k-\alpha_1)m^*_{0k}}{1-\alpha_0-\alpha_1} = c - \frac{(p_k-\alpha_0)m_{1k^*}}{1-\alpha_0-\alpha_1}.$$

The result follows by substituting into Equations 3.4–3.5 from Lemma 4. □

Proof of Theorem 1. Recall from the discussion preceding Equation 3.8 that the Wald estimator $\mathcal{W} = \beta/(1-\alpha_0-\alpha_1)$ is identified in this model so long as K is at least 2. Rearranging, we find that:

$$\alpha_0 = (1-\alpha_1) - \beta/\mathcal{W}$$
$$(p_k - \alpha_0) = p_k - (1-\alpha_1) + \beta/\mathcal{W}$$
$$1-\alpha_0 = \alpha_1 + \beta/\mathcal{W}$$

Substituting these into Equations 3.6 and 3.7 and summing the two, we find, after some algebra, that

$$\hat{y}_{0k} + \hat{y}_{1k} + \mathcal{W}(1-p_k) = c + \beta + \mathcal{W}\alpha_1.$$

Since the left-hand side of this expression depends only on observables and the identified quantity $\mathcal{W}$, this shows that the right-hand side is itself identified in this model. For simplicity, we define $\mathcal{Q} = c + \beta + \mathcal{W}\alpha_1$. Since $\mathcal{W}$ and $\mathcal{Q}$ are both identified, varying either *necessarily* changes the observables, so we must hold both of them constant. We now show that Equations 3.6 and 3.7 can be expressed in terms of $\mathcal{W}$ and $\mathcal{Q}$. Conveniently, this eliminates α_0 from the system. After some algebra,

$$\hat{y}_{0k} = \alpha_1(\mathcal{Q} - m^*_{1k}) + \beta(c - m^*_{1k})/\mathcal{W} + (1 - p_k)\left[m^*_{1k} - \mathcal{W}\alpha_1\right] \quad \text{(A.2)}$$

$$\hat{y}_{1k} = (1 - \alpha_1)\mathcal{Q} + \beta(m^*_{1k} - c)/\mathcal{W} - (1 - p_k)\left[m^*_{1k} + \mathcal{W}(1 - \alpha_1)\right] \quad \text{(A.3)}$$

Now, rearranging Equation A.3 we see that

$$\mathcal{Q} - \hat{y}_{1k} - \mathcal{W}(1 - p_k) = \alpha_1(\mathcal{Q} - m^*_{1k}) + \beta(c - m^*_{1k})/\mathcal{W} + (1 - p_k)\left[m^*_{1k} - \mathcal{W}\alpha_1\right] \quad \text{(A.4)}$$

Notice that the right-hand side of Equation A.4 is the *same* as that of Equation A.2 and that $\mathcal{Q} - \hat{y}_{1k} - \mathcal{W}(1 - p_k)$ is precisely $\hat{y}_{0k}$. In other words, given the constraint that $\mathcal{W}$ and $\mathcal{Q}$ must be held fixed, we only have *one* equation for each value that the instrument takes on. Finally, we can solve this equation for m^*_{1k} as

$$m^*_{1k} = \frac{\mathcal{W}(\hat{y}_{0k} - \alpha_1\mathcal{Q}) - \beta(\mathcal{Q} - \beta - \mathcal{W}\alpha_1) + \mathcal{W}^2(1 - p_k)\alpha_1}{\mathcal{W}(1 - p_k - \alpha_1) - \beta} \quad \text{(A.5)}$$

using the fact that $c = \mathcal{Q} - \beta - \mathcal{W}\alpha_1$. Equation A.5 is a manifold parameterized by (β, α_1) that is *unique* to each value that the instrument takes on. Thus, by adjusting $\{m^*_{1k}\}_{k=1}^K$ according to Equation A.5 we are free to vary β while holding all observable moments fixed. □

Proof of Theorem 2. First define

$$\mu^*_{k\ell} = (p_k - \alpha_0)m^*_{1k} - (p_\ell - \alpha_0)m^*_{k\ell} \quad \text{(A.6)}$$

$$\Delta\overline{y^2} = \mathbb{E}(y^2|z_k) - \mathbb{E}(y^2|z_\ell) \quad \text{(A.7)}$$

$$\Delta\overline{yT} = \mathbb{E}(yT|z_k) - \mathbb{E}(yT|z_\ell) \quad \text{(A.8)}$$

By iterated expectations it follows, after some algebra, that

$$\Delta\overline{y^2} = \beta\mathcal{W}(p_k - p_\ell) + 2\mathcal{W}\mu^*_{k\ell} \quad \text{(A.9)}$$

$$\Delta\overline{yT} = (1 - \alpha_1)\mathcal{W}(p_k - p_\ell) + \mu^*_{k\ell} \quad \text{(A.10)}$$

Now, solving Equation A.10 for $\mu^*_{k\ell}$, substituting the result into Equation A.9 and rearranging,

$$\mathcal{R} \equiv \beta - 2(1-\alpha_1)\mathcal{W} = \frac{\Delta\overline{y^2} - 2\mathcal{W}\Delta\overline{yT}}{\mathcal{W}(p_k - p_\ell)}. \tag{A.11}$$

Since $\mathcal{W}$ is identified it follows that $\mathcal{R}$ is identified. Rearranging the preceding equality and substituting $\beta = \mathcal{W}(1-\alpha_0-\alpha_1)$ to eliminate β, we find that

$$\alpha_1 - \alpha_0 = 1 + \mathcal{R}/\mathcal{W}. \tag{A.12}$$

Because both $\mathcal{R}$ and $\mathcal{W}$ are identified, it follows that the difference of error rates is also identified. □

Proof of Theorem 3. First define

$$v^*_{tk} = \mathbb{E}(u^2|T^* = t, z = z_k) \tag{A.13}$$
$$\lambda^*_{k\ell} = (p_k - \alpha_0)v^*_{1k} - (p_\ell - \alpha_0)v^*_{1\ell} \tag{A.14}$$
$$\Delta\overline{y^3} = \mathbb{E}(y^3|z_k) - \mathbb{E}(y^3|z_\ell) \tag{A.15}$$
$$\Delta\overline{y^2T} = \mathbb{E}(y^2T|z_k) - \mathbb{E}(y^2T|z_\ell) \tag{A.16}$$

where u, as above, is defined as $\varepsilon + c$. By iterated expectations it follows, after some algebra, that

$$\Delta\overline{y^3} = \beta^2\mathcal{W}(p_k - p_\ell) + 3\beta\mathcal{W}\mu^*_{k\ell} + 3\mathcal{W}\lambda^*_{k\ell} \tag{A.17}$$
$$\Delta\overline{y^2T} = \beta(1-\alpha_1)\mathcal{W}(p_k - p_\ell) + 2(1-\alpha_1)\mathcal{W}\mu^*_{k\ell} + \lambda^*_{k\ell} \tag{A.18}$$

where, as above, the identified quantity $\mathcal{W}$ equals $\beta/(1-\alpha_0-\alpha_1)$ and $\mu^*_{k\ell}$ is as defined in Equation A.6. Now, substituting for $\lambda^*_{k\ell}$ in Equation A.17 using Equation A.18 and rearranging, we find that

$$\Delta\overline{y^3} - 3\mathcal{W}\Delta\overline{y^2T} = \beta\mathcal{W}(p_k - p_\ell)\left[\beta - 3\mathcal{W}(1-\alpha_1)\right] + 3\mathcal{W}\mathcal{R}\mu^*_{k\ell} \tag{A.19}$$

where $\mathcal{R}$ is as defined in Equation A.11. Now, using Equation A.10 to eliminate $\mu^*_{k\ell}$ from the preceding equation, we find after some algebra that

$$\mathcal{S} \equiv \beta^2 - 3\mathcal{W}(1-\alpha_1)(\beta + \mathcal{R}) = \frac{\Delta\overline{y^3} - 3\mathcal{W}\left[\Delta\overline{y^2T} + \mathcal{R}\Delta\overline{yT}\right]}{\mathcal{W}(p_k - p_\ell)}. \tag{A.20}$$

Notice that $\mathcal{S}$ is identified. Finally, by eliminating β from the preceding expression using Equation A.11, we obtain a quadratic equation in $(1-\alpha_1)$, namely

$$2\mathcal{W}^2(1-\alpha_1)^2 + 2\mathcal{R}\mathcal{W}(1-\alpha_1) + (\mathcal{S}-\mathcal{R}^2) = 0. \tag{A.21}$$

Note that, since, $\mathcal{W}, \mathcal{R}$ and $\mathcal{S}$ are all identified, we can solve Equation A.21 for $(1-\alpha_1)$. The solutions are as follows

$$(1-\alpha_1) = \frac{1}{2}\left(-\frac{\mathcal{R}}{\mathcal{W}} \pm \frac{1}{\mathcal{W}}\sqrt{3\mathcal{R}^2 - 2\mathcal{S}}\right) \tag{A.22}$$

It can be shown that $3\mathcal{R}^2 - 2\mathcal{S} = [\mathcal{R} + 2\mathcal{W}(1-\alpha_1)]^2$ so the quantity under the radical is guaranteed to be positive, yielding two real solutions. One of these is $(1-\alpha_1)$. Using Equation A.12 we can re-express Equation A.21 as a quadratic in α_0. After simplifying, we obtain a quadratic with *identical* coefficients, implying that the other root of Equation A.21 identifies α_0. Now, let r_{max} denote the larger of the two roots of Equation A.21 and r_{min} the smaller. (By assumption $\alpha_0 + \alpha_1 \neq 1$ which implies $r_{max} \neq r_{min}$.) We claim that $r_{max} = 1 - \alpha_1$ and hence that $r_{min} = \alpha_0$. Suppose that this were not the case. Then $r_{max} = \alpha_0$ and $r_{min} = 1 - \alpha_1$ and accordingly

$$1 - \alpha_0 - \alpha_1 = r_{min} - r_{max} < 0$$

which violates the assumption $\alpha_0 + \alpha_1 < 1$. Therefore α_0 and α_1 are identified and multiplying the Wald estimator $\mathcal{W}$ by $(1-\alpha_0-\alpha_1)$ identifies β. □

References

Aigner, D. J., 1973. Regression with a binary independent variable subject to errors of observation. Journal of Econometrics 1, 49–60.

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

Black, D. A., Berger, M. C., Scott, F. A., 2000. Bounding parameter estimates with nonclassical measurement error. Journal of the American Statistical Association 95 (451), 739–748.

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

Bollinger, C. R., van Hasselt, M., 2015. Bayesian moment-based inference in a regression models with misclassification error, working Paper.

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

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

Chen, X., Hu, Y., Lewbel, A., 2008a. Nonparametric identification of regression models containing a misclassified dichotomous regressor with instruments. Economics Letters 100, 381–384.

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

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

Hausman, J., Abrevaya, J., Scott-Morton, F., 1998. Misclassification of the dependent variable in a discrete-response setting. Journal of Econometrics 87, 239–269.

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

Hu, Y., Shennach, S. M., January 2008. Instrumental variable treatment of nonclassical measurement error models. Econometrica 76 (1), 195–216.

Hu, Y., Shiu, J.-L., Woutersen, T., 2015. Identification and estimation of single index models with measurement error and endogeneity. The Econometrics Journal (Forthcoming).

Kane, T. J., Rouse, C. E., Staiger, D., July 1999. Estimating returns to schooling when schooling is misreported. Tech. rep., National Bureau of Economic Research, NBER Working Paper 7235.

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

Lewbel, A., 1997. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. Econometrica, 1201–1213.

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

Lewbel, A., 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. Journal of Business & Economic Statistics 30 (1), 67–80.

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

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

Schennach, S. M., 2004. Estimation of nonlinear models with measurement error. Econometrica 72 (1), 33–75.

Schennach, S. M., 2007. Instrumental variable estimation of nonlinear errors-in-variables models. Econometrica 75 (1), 201–239.

Schennach, S. M., 2013. Measurement error in nonlinear models – a review. In: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), Advances in Economics and Econometrics. Vol. 3. Cambridge University Press, pp. 296–337.

Shiu, J.-L., 2015. Identification and estimation of endogenous selection models in the presence of misclassification errors. Economic Modelling (Forthcoming).

Song, S., 2015. Semiparametric estimation of models with conditional moment restrictions in the presence of nonclassical measurement errors. Journal of Econometrics 185 (1), 95–109.

Song, S., Schennach, S. M., White, H., 2015. Semiparametric estimation of models with conditional moment restrictions in the presence of nonclassical measurement errors. Quantitative Economics (Forthcoming).

Ura, T., November 2015. Heterogeneous treatment effects with mismeasured endogenous treatment. Tech. rep., Duke University Department of Economics.

van Hasselt, M., Bollinger, C. R., 2012. Binary misclassification and identification in regression models. Economics Letters 115, 81–84.