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

Francis J. DiTraglia Camilo Garcia-Jimeno

University of Pennsylvania

This Version: November 2, 2015, First Version: October 31, 2015

Abstract

This paper studies the use of a discrete instrumental variable to identify the causal effect of a endogenous, mis-measured, binary treatment in a homogeneous effects model with additively separable errors. 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 provide a convenient notational framework to address this question and use it to derive a number of results. First, we prove that the treatment effect is unidentified based on conditional first-moment information, regardless of the number of values that the instrument may take. Second, we 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 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 completely characterizes the measurement error process.

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

JEL Codes: C10, C18, C25, C26

1 Introduction

This paper studies the use of a discrete instrumental variable to identify the causal effect of an endogenous, mis-measured, binary treatment in a homogeneous effects model with additively separable errors. Although a relevant case for applied work, this setting 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 providing a convenient notational framework within which to situate the problem. Using this framework we then show 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. 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 imply the lack of a first-stage, violating 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. Mahajan (2006) and related papers, discussed below, use only conditional means of the outcome to identify the treatment effect. However, 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. To the best of our knowledge, this source of information has not been exploited in the extant literature on instrumental variables. Under this stronger condition on the instrument, we first show that conditional second moments of the outcome variable identify the *difference* of mis-classification rates in the mis-measured regressor: the probability that a true one is classified as a zero minus the probability that a true zero is classified as a one. 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 conditional third moments point identify both mis-classification rates. Thus, combining conditional first, second and third moment information point identifies the treatment effect.

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 then lays out the econometric model, its assumptions, and our notational framework. Section 4 considers identification based on conditional means, showing that Mahajan's proof is incorrect and that increasing the support of the instrument cannot be used to obtain identification. Section 5 presents our results under stronger conditions on the instrument, based on conditional second and third moments of the outcome variable. Section 6 concludes.

2 Related Literature

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. As is well known, instrumental variables (IV) based on a single valid instrument suffices to recover the treatment effect in a linear model with a single endogenous regressor subject to classical measurement errors. As is less well known, classical measurement error is in fact impossible when the regressor of interest is binary: 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).

Measurement error in a binary regressor is usually called *mis-classification*. 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 uncorrelated with all other variables in the system. Even under this comparatively mild departure from classical measurement error, the IV estimator is inconsistent (Black et al., 2000; Kane et al., 1999). Moreover, the probability limit of the IV estimator does not depend on whether the treatment is endogenous or not (Frazis and Loewenstein, 2003).

When the treatment is in fact *exogenous*, however, a valid instrument suffices to recover the treatment effect using a non-linear GMM estimator. Black et al. (2000) and Kane et al. (1999) more-or-less simultaneously pointed this out in a setting in which *two* alternative measures of treatment are available, both subject to non-differential measurement error. In essence, one measure serves as an instrument for the other although the estimator is quite different

from IV.[1] 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 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. Lewbel (2007) provides a related identification result in 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 the treatment, in addition to suffering from non-differential measurement error, is also endogenous. Only two papers consider this case. 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. The second paper that

[1] 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.

considers 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. Although their discussion does not apply to the non-parametric case, Frazis and Loewenstein's intuition turns out to be right: Mahajan's proof is incorrect, as we prove below using a convenient notational framework introduced in the following section.

3 The Model and Notation

Let T^* be a binary indicator of true treatment status, possibly endogenous, $\mathbf{x}$ be a vector of exogenous covariates, and y be an outcome of interest where

$$y = h(T^*, \mathbf{x}) + \varepsilon \tag{1}$$

and ε is mean zero. Since T^* is potentially endogenous, $\mathbb{E}[\varepsilon|T^*, \mathbf{x}]$ may not be zero. Now let z be a discrete instrumental variable with support set $\{z_k\}_{k=1}^K$ satisfying the usual instrumental variables assumption, namely $\mathbb{E}[\varepsilon|z, \mathbf{x}] = 0$. We assume throughout that z is a relevant instrument for T^*, in other words

$$\mathbb{P}(T^* = 1|z_j, \mathbf{x}) \neq \mathbb{P}(T^* = 1|z_k, \mathbf{x}), \quad \forall k \neq j. \tag{2}$$

Our goal is to estimate the average treatment effect (ATE) function

$$\tau(\mathbf{x}) = h(1, \mathbf{x}) - h(0, \mathbf{x}). \tag{3}$$

We maintain throughout that $\tau(\mathbf{x}) \neq 0$. If it were zero, this would imply that T^* is irrelevant for y which can be directly tested regardless of whether any mis-classification is present and regardless of whether T^* is endogenous.[2]

[2]This is because, as we will see below, the Wald Estimator is identified and is proportional to the treatment effect. This estimator exists provided that we have a valid and relevant instrument that takes on at least two values.

Now, suppose we observe not T^* but a noisy measure T polluted by non-differential measurement error. In particular, we assume that

$$\mathbb{P}(T = 1|T^* = 0, z, \mathbf{x}) = \alpha_0(\mathbf{x}) \tag{4}$$
$$\mathbb{P}(T = 0|T^* = 1, z, \mathbf{x}) = \alpha_1(\mathbf{x}) \tag{5}$$

and additionally that

$$\mathbb{E}[\varepsilon|T^*, T, z, \mathbf{x}] = \mathbb{E}[\varepsilon|T^*, z, \mathbf{x}] \tag{6}$$

Equations 4–5 amount to the assumption that z and T are conditionally independent given $(T^*, \mathbf{x})$. In other words, z provides no additional information about the process that causes T to be mis-classified above that already contained in T^* and $\mathbf{x}$. In contrast, we allow for the possibility that the measurement error process *does* depend on the exogenous covariates $\mathbf{x}$. Equation 6 states that, given knowledge of true treatment status, the instrument and the exogenous covariates, the *observed* treatment status contains no information about the mean of the regression error term. The assumptions on the measurement error process contained in Equations 4–6 are standard in the literature. Another standard assumption is the condition

$$\alpha_0(\mathbf{x}) + \alpha_1(\mathbf{x}) < 1 \tag{7}$$

which rules out the possibility that $1 - T$ is a better measure of T^* than T is, and vice-versa. This condition is imposed in the literature (Black et al., 2000; Frazis and Loewenstein, 2003; Kane et al., 1999; Lewbel, 2007; Mahajan, 2006) because in its absence the treatment effect would only be identified up to sign. Our results will not in fact require the condition in Equation 7 to hold.

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,

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

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

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 can be re-expressed as a simple linear model, namely

$$y = \beta T^* + u \tag{8}$$

where we define $u = c + \varepsilon$, an error term that need not be mean zero. In the context of Equation 8 the only observable information consists of the moments of y, conditional on 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 in section 5 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 $q_k = \mathbb{P}(z=z_k)$. Table 1 depicts the observable 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,

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

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

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

since the misclassification probabilities do not depend on z by Equations 4–5. Rearranging,

$$p^*_k = \frac{p_k - \alpha_0}{1-\alpha_0-\alpha_1}, \quad 1-p^*_k = \frac{1-p_k-\alpha_1}{1-\alpha_0-\alpha_1}. \tag{9}$$

Equation 9 implies that p^*_k is observable up to knowledge of the mis-classification rates α_0, α_1 since p_k is observable. Thus, the full set of parameters needed to characterize the model in Equation 8 consists of $\beta, \alpha_0, \alpha_1$ and the conditional means of u, namely m^*_{tk} for a total of $2K+3$ parameters. In contrast, there are only $2K$ available moment conditions, namely:

$$\hat{y}_{0k} = \frac{\alpha_1(p_k-\alpha_0)(\beta+m^*_{1k}) + (1-\alpha_0)(1-p_k-\alpha_1)m^*_{0k}}{1-\alpha_0-\alpha_1} \tag{10}$$

$$\hat{y}_{1k} = \frac{(1-\alpha_1)(p_k-\alpha_0)(\beta+m^*_{1k}) + \alpha_0(1-p_k-\alpha_1)m^*_{0k}}{1-\alpha_0-\alpha_1} \tag{11}$$

by the Law of Iterated Expectations, where the observables on the left hand side are defined according to $\hat{y}_{0k} = (1-p_k)\bar{y}_{0k}$ and $\hat{y}_{1k} = p_k\bar{y}_{1k}$. Notice that the observable "weighted" cell mean $\hat{y}_{tk}$ 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 very easy to see why and how Lewbel's argument works. His moment conditions are equivalent to Equations 10 and 11 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, which makes it clear why Lewbel's proof must require that the "instrument" take on at least three values.[3]

Unlike Lewbel (2007), we, along with Mahajan (2006) and others, assume that z satisfies the exclusion restriction. This implies a different constraint on the m^*_{tk} from Table 2. Since $u = c + \varepsilon$, $\mathbb{E}[\varepsilon|z] = 0$ implies that

$$\mathbb{E}[u|z] = E[u] = c. \tag{12}$$

By the Law of Iterated Expectations, this can be expressed as

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

for all $k = 1, \ldots, K$. This restriction imposes that a particular weighted sum over the rows of a given column of Table 2 takes the same value *across* columns. Using Equation 9 and rearranging gives

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

[3]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.

which we can substitute into Equations 10 and 11 to yield

$$\hat{y}_{0k} = \alpha_1(p_k - \alpha_0)\left(\frac{\beta}{1-\alpha_0-\alpha_1}\right) + (1-\alpha_0)c - (p_k - \alpha_0)m^*_{1k} \tag{13}$$

$$\hat{y}_{1k} = (1-\alpha_1)(p_k - \alpha_0)\left(\frac{\beta}{1-\alpha_0-\alpha_1}\right) + \alpha_0 c + (p_k - \alpha_0)m^*_{1k}. \tag{14}$$

Equations 13 and 14 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 13 and 14 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 expression evaluated at two different values of the instrument, z_k and z_ℓ, and rearranging

$$\mathcal{W} = \frac{(\hat{y}_{0k} + \hat{y}_{1k}) - (\hat{y}_{0\ell} + \hat{y}_{1\ell})}{p_k - p_\ell} = \frac{\beta}{1-\alpha_0-\alpha_1} \tag{15}$$

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

Imposing Equation 12 replaces the K unknown parameters $\{m^*_{0k}\}_{k=1}^K$ 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 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 13 and 14 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, clearly the joint exogeneity assumption $m^*_{tk} = c$ is unusable: we are back to $2K$ 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. A first possibility 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. This is the route followed by Mahajan (2006) in his proof of identification with a binary instrument and endogenous treatment. His Equation (11), expressed in our notation, amounts to adding two cross-column restrictions in Table 2: $m^*_{11} = m^*_{12}$ and $m^*_{01} = m^*_{02}$. Another possibility, suggested by Lewbel's approach, 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 $2K = K + 4$. In the following section we present two of our main results: first Mahajan's approach leads to a contradiction, and second, regardless of the value of K, β is unidentified based on conditional mean information.

4 Non-identification Based on First Moments

4.1 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) $\alpha_0 + \alpha_1 \neq 1$

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

Proof of Lemma 1. Since z is a valid instrument that does not influence the mis-classification 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 9. 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 $p_1 \neq p_2$. □

In his Theorem 1, Mahajan (2006) proves that α_0, α_1 can in fact be identified under the following assumptions.[4]

[4]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.

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) $\mathbb{E}[\nu|T^*, T, z] = 0$.[5]

(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* $\mathbb{E}[y|T^*]$*.*

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

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 misclassification 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*

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

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

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).*

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. □

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. 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.

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).*

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

$$\mathbb{E}[\varepsilon|T^*, z] = \mathbb{E}_{T|T^*,z}\left[\mathbb{E}\left(\varepsilon|T^*, T, z\right)\right] = \mathbb{E}_{T|T^*,z}\left[\mathbb{E}\left(\varepsilon|T^*\right)\right] = \mathbb{E}\left[\varepsilon|T^*\right] \tag{16}$$

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 16 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,

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

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 $\mathbb{E}[\varepsilon|T^*] \neq 0$. □

4.2 Generic Lack of Identification

We have seen that Mahajan (2006)'s approach cannot identify β when the treatment is endogenous: Assumption 3 in fact implies that the instrument is *irrelevant*. But this alone does not establish that a valid instrument is insufficient to identify β when the treatment is endogenous. In particular, our equation counts from above appear to suggest that a valid instrument that takes on at least four values might suffice for identification. Unfortunately, this is not the case 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.*

Proof of Theorem 1. The assumptions of this Theorem are the same as those used to derive Equations 13 and 14. These expressions, for $k = 1, \ldots, K$ constitute the full set of available moment conditions. To establish lack of identification, we derive a parametric relationship between β and the other model parameters such that, varying β along this parametric relationship, the observables $(\hat{y}_{0k}, \hat{y}_{1k})$ are unchanged for all k.

Recall from the discussion preceding Equation 15 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:

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

Substituting these into Equations 13 and 14 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 13 and 14 can be expressed in terms of $\mathcal{W}$ and $\mathcal{Q}$. Conveniently, this eliminates α_0 from the system. After some algebra,

$$\hat{y}_{0k} = \alpha_1(\mathcal{Q} - m^*_{1k}) + \beta(c - m^*_{1k})/\mathcal{W} + (1 - p_k)\left[m^*_{1k} - \mathcal{W}\alpha_1\right] \tag{17}$$

$$\hat{y}_{1k} = (1 - \alpha_1)\mathcal{Q} + \beta(m^*_{1k} - c)/\mathcal{W} - (1 - p_k)\left[m^*_{1k} + \mathcal{W}(1 - \alpha_1)\right] \tag{18}$$

Now, rearranging Equation 18 we see that

$$\mathcal{Q} - \hat{y}_{1k} - \mathcal{W}(1 - p_k) = \alpha_1(\mathcal{Q} - m^*_{1k}) + \beta(c - m^*_{1k})/\mathcal{W} + (1 - p_k)\left[m^*_{1k} - \mathcal{W}\alpha_1\right] \tag{19}$$

Notice that the right-hand side of Equation 19 is the *same* as that of Equation 17 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

$$m^*_{1k} = \frac{\mathcal{W}(\hat{y}_{0k} - \alpha_1\mathcal{Q}) - \beta(\mathcal{Q} - \beta - \mathcal{W}\alpha_1) + \mathcal{W}^2(1 - p_k)\alpha_1}{\mathcal{W}(1 - p_k - \alpha_1) - \beta} \tag{20}$$

using the fact that $c = \mathcal{Q} - \beta - \mathcal{W}\alpha_1$. Equation 20 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 20 we are free to vary β while holding all observable moments fixed. □

5 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.*

Proof of Theorem 2. First define

$$\begin{aligned} \mu^*_{k\ell} &= (p_k - \alpha_0)m^*_{1k} - (p_\ell - \alpha_0)m^*_{k\ell} && (21) \\ \Delta\overline{y^2} &= \mathbb{E}(y^2|z_k) - \mathbb{E}(y^2|z_\ell) && (22) \\ \Delta\overline{yT} &= \mathbb{E}(yT|z_k) - \mathbb{E}(yT|z_\ell) && (23) \end{aligned}$$

By iterated expectations it follows, after some algebra, that

$$\begin{aligned} \Delta\overline{y^2} &= \beta\mathcal{W}(p_k - p_\ell) + 2\mathcal{W}\mu^*_{k\ell} && (24) \\ \Delta\overline{yT} &= (1 - \alpha_1)\mathcal{W}(p_k - p_\ell) + \mu^*_{k\ell} && (25) \end{aligned}$$

Now, solving Equation 25 for $\mu^*_{k\ell}$, substituting the result into Equation 24 and rearranging,

$$\mathcal{R} \equiv \beta - 2(1 - \alpha_1)\mathcal{W} = \frac{\Delta\overline{y^2} - 2\mathcal{W}\Delta\overline{yT}}{\mathcal{W}(p_k - p_\ell)}. \quad (26)$$

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

$$\alpha_1 - \alpha_0 = 1 + \mathcal{R}/\mathcal{W}. \quad (27)$$

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

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. 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 z takes on at least two values.*

Proof of Theorem 3. First define

$$\begin{aligned} v^*_{tk} &= \mathbb{E}(u^2|T^* = t, z = z_k) && (28)\\ \lambda^*_{k\ell} &= (p_k - \alpha_0)v^*_{1k} - (p_\ell - \alpha_0)v^*_{1\ell} && (29)\\ \Delta\overline{y^3} &= \mathbb{E}(y^3|z_k) - \mathbb{E}(y^3|z_\ell) && (30)\\ \Delta\overline{y^2T} &= \mathbb{E}(y^2T|z_k) - \mathbb{E}(y^2T|z_\ell) && (31) \end{aligned}$$

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

$$\begin{aligned} \Delta\overline{y^3} &= \beta^2\mathcal{W}(p_k - p_\ell) + 3\beta\mathcal{W}\mu^*_{k\ell} + 3\mathcal{W}\lambda^*_{k\ell} && (32)\\ \Delta\overline{y^2T} &= \beta(1 - \alpha_1)\mathcal{W}(p_k - p_\ell) + 2(1 - \alpha_1)\mathcal{W}\mu^*_{k\ell} + \lambda^*_{k\ell} && (33) \end{aligned}$$

where, as above, the identified quantity $\mathcal{W}$ equals $\beta/(1 - \alpha_0 - \alpha_1)$ and $\mu^*_{k\ell}$ is as defined in Equation 21. Now, substituting for $\lambda^*_{k\ell}$ in Equation 32 using Equation 33 and rearranging, we find that

$$\Delta\overline{y^3} - 3\mathcal{W}\Delta\overline{y^2T} = \beta\mathcal{W}(p_k - p_\ell)\left[\beta - 3\mathcal{W}(1 - \alpha_1)\right] + 3\mathcal{W}\mathcal{R}\mu^*_{k\ell} \qquad (34)$$

where $\mathcal{R}$ is as defined in Equation 26. Now, using Equation 25 to eliminate

$\mu^*_{k\ell}$ from the preceding equation, we find after some algebra that

$$\mathcal{S} \equiv \beta^2 - 3\mathcal{W}(1-\alpha_1)(\beta+\mathcal{R}) = \frac{\Delta\overline{y^3} - 3\mathcal{W}\left[\Delta\overline{y^2T} + \mathcal{R}\Delta\overline{yT}\right]}{\mathcal{W}(p_k - p_\ell)}. \tag{35}$$

Notice that $\mathcal{S}$ is identified. Finally, by eliminating β from the preceding expression using Equation 26, we obtain a quadratic equation in $(1-\alpha_1)$, namely

$$2\mathcal{W}^2(1-\alpha_1)^2 + 2\mathcal{R}\mathcal{W}(1-\alpha_1) + (\mathcal{S} - \mathcal{R}^2) = 0. \tag{36}$$

Note that, since, $\mathcal{W}, \mathcal{R}$ and $\mathcal{S}$ are all identified, we can solve Equation 36 for $(1-\alpha_1)$. The solutions are as follows

$$(1-\alpha_1) = \frac{1}{2}\left(-\frac{\mathcal{R}}{\mathcal{W}} \pm \frac{1}{\mathcal{W}}\sqrt{3\mathcal{R}^2 - 2\mathcal{S}}\right) \tag{37}$$

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 $(1-\alpha_1)$, but what about the other root? Using Equation 27 we can re-express Equation 36 as a quadratic in α_0. Surprisingly, after simplifying, we obtain a quadratic with *identical* coefficients. This implies that the second root of Equation 36 identifies α_0. Since we know the sign of the difference $\alpha_1 - \alpha_0$ from Theorem 2, we know which mis-classification rate is larger and hence can correctly label the two roots. Finally, substituting into $\beta = \mathcal{W}(1-\alpha_0-\alpha_1)$, we identify the treatment effect. □

Note that, in contrast to all other results in the literature (Black et al., 2000; Frazis and Loewenstein, 2003; Kane et al., 1999; Lewbel, 2007; Mahajan, 2006), our proof does *not* require the assumption that $\alpha_0 + \alpha_1 < 1$ to identify β.

6 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. By appealing to higher moments we can accommodate any amount of mis-classification, dispensing with a standard assumption from the literature that mis-classification is not "too severe." 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. While we have restricted our attention in this paper to the case of homogeneous treatment effects, a promising avenue for future research would be to consider the heterogenous case.

References

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

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.

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

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.

Lewbel, A., March 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.