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

Francis J. DiTraglia Camilo Garcia-Jimeno
University of Pennsylvania

October 17, 2015

Abstract

Abstract goes here.

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 valid instrument to identify the causal effect of an endogeneous, binary treatment that is subject to non-differential measurement error in a non-parametric regression 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 appears in an important paper by Mahajan (2006), who proves identification by relying on a discrete instrument that takes on at least two values. Here we show that this result is incorrect. To do so, we begin by providing a convenient notational framework within which to situate the problem. Using this framework we show that the proof in Appendix A.2 of Mahajan (2006) leads to a contradiction. Throughout the 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 in a model with additively separable errors, 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, borrowing an idea from Lewbel (2007), we explore whether expanding the support of the instrument, so that it takes on more than two values, yields identification. Allowing the instrument to take on additional values increases the number of available moment conditions. We show, however, that these additional moments cannot point identify the treatment effect. This holds true regardless of how many (finite) values the instrument takes on.

We then consider a new source of identifying information in the form of a conditional homoskedasticity assumption. In particular, we suppose that the conditional *variance* of the regression error term given the instrument is constant. While stronger than the usual mean independence assumption, this assumption holds automatically in a randomized controlled trial or a genuine natural experiment. To the best of our knowledge, this source of information has not been exploited in the extant literature on instrumental variables. We show that this assumption leads to a novel partial identification result that is easy to implement in practice and can be applied regardless of the number of values that the instrument takes on. Moreover, it can be used to obtain point identification in some special cases that nevertheless may be empirically relevant.

The remainder of this paper is organized as follows.

FILL IN!

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 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. 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 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 the other variables of 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 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

[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.

impose strong and somewhat unnatural conditions. The second paper that 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

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

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

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, 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]

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

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

and additionally that

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

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 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. While we do not rely on this condition below, it appears, for example in Mahajan (2006), and Lewbel (2007), among others.

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,

[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.

	$z=1$	$z=1$	$\ldots$	$z=K$
$T=0$	$\bar{y}_{01}$ \ 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 the moment, following the existing literature, we will restrict attention to the conditional mean 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,

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

	$z = 1$	$z = 1$	$\dots$	$z = K$
$T^* = 0$	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 $p^*_{1k} = q_k p^*_k$.

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 requires that the "instrument" take on at least three values.

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

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

which we can substitute into Equations 10 and 11 to yield

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

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

So what do we gain from the instrument validity assumption? 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. 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 $\hat{y}_{0k} + \hat{y}_{1k} = \mathbb{E}[y|z = z_k]$.

For any

4 Identification by Homoskedasticity

This section uses our notation rather than Mahajan's. We'll have to decide what notation we want to use in the paper itself but for the moment I'm trying to avoid confusion by talking about Mahajan's proofs using his own notation while keeping our derivations in the same notation we used on the whiteboard. I think that by assuming the instrument takes on three values (as in Lewbell) and imposing our homoskedasticity assumption we'll get identification in the case where T^* is endogenous so I've written out this derivation for arbitrary discrete z.

Now suppose that one is prepared to assume that

$$E[u^2|z] = E[u^2]. \tag{16}$$

When combined with the usual IV assumption, $E[u|z] = 0$, this implies $Var(u|z) = Var(u)$. Whether this assumption is reasonable, naturally, depends on the application. When z is the offer of treatment in a randomized controlled trial, for example, Equation 16 holds automatically as a consequence of the randomization. Similarly, in studies based on a "natural" rather than controlled experiment one typically argues that the instrument is not merely uncorrelated with u but *independent* of it, so that Equation 16 follows.

To see why homoskedasticity with respect to the instrument provides additional identifying information, first express the conditional variance of y as follows

$$Var(y|z) = \beta^2 Var\left(T^*|z\right) + Var(u|z) + 2\beta Cov(T^*, u|z) \tag{17}$$

Under 16, $Var(u|z)$ does not depend on z. Hence the *difference* of conditional variances evaluated at two values z_a and z_b in the support of z is simply

$$\Delta Var(y|z_a, z_b) = \beta^2 \Delta Var(T^*|z_a, z_b) + 2\beta \Delta Cov(T^*, u|z_a, z_b)] \tag{18}$$

Where $\Delta Var(y|z_a, z_b) = Var(y|z = z_a) - Var(y|z = z_b)$, and we define $\Delta Var(T^*|z_a, z_b)$ and $\Delta Cov(T^*, u|z_a, z_b)$ analogously.

First we simplify the $\Delta Var(T^*|z_a, z_b)$ term. Since T is conditionally independent of z given T^*,

$$\begin{aligned} P(T=1|z) &= E_{T^*|z}\left[E\left(T|z, T^*\right)\right] = E_{T^*|z}\left[E\left(T|T^*\right)\right] \\ &= P\left(T^*=1|z\right)\left(1-\alpha_1\right) + \left[1 - P\left(T^*=1|z\right)\right]\alpha_0 \\ &= \alpha_0 + \left(1-\alpha_0-\alpha_1\right) P(T^*=1|z) \end{aligned}$$

Rearranging,

$$P(T^* = 1|z) = \frac{P(T = 1|z) - \alpha_0}{1 - \alpha_0 - \alpha_1} \tag{19}$$

and accordingly,

$$Var(T^*|z) = \frac{[P(T = 1|z) - \alpha_0]\,[1 - P(T = 1|z) - \alpha_1]}{(1 - \alpha_0 - \alpha_1)^2} \tag{20}$$

Thus, evaluating Equation 20 at z_a and z_b and simplifying,

$$\Delta Var(T^*|z_a, z_b) = \frac{\Delta Var(T|z_a, z_b) + (\alpha_0 - \alpha_1)\,\Delta E(T|z_a, z_b)}{(1 - \alpha_0 - \alpha_1)^2} \tag{21}$$

Turning our attention to $\Delta Cov(T^*, u|z_a, z_b)$ first note that

$$Cov(T^*, u|z) = E_{T^*|z}\left[E\left(T^* u|z, T^*\right)\right] = P(T^* = 1|z)E\left(u|T^* = 1, z\right) \tag{22}$$

since $E[z|u] = 0$. Combining this with Equation 19 and evaluating at z_a and z_b gives

$$\Delta Cov(T^*, u|z_a, z_b) = \frac{[E(T|z_a) - \alpha_0]\, m_{1a} - [E(T|z_b) - \alpha_0]\, m_{1b}}{1 - \alpha_0 - \alpha_1} \tag{23}$$

where $m_{1a} = E\left[u|T^* = 1, z_a\right]$ and $m_{1b} = E[u|T^* = 1, z_b]$.

Both Equations 21 and 23 involve only observable quantities and the mis-classification rates α_0 and α_1. Equation 18, however, also involves β. Fortunately we can eliminiate this quantity as follows. First, let $\mathcal{W}(z_a, z_b)$ denote the Wald Estimator of β given by

$$\mathcal{W}(z_a, z_b) = \frac{E(y|z_a) - E(y|z_b)}{E(T|z_a) - E(T|z_b)} \tag{24}$$

Since $E(u|z) = 0$,

$$E(y|z_a) - E(y|z_b) = \beta\left[E\left(T^*|z_a\right) - E\left(T^*|z_b\right)\right]$$

and by Equation 19,

$$E(T|z_a) - E(T|z_b) = (1 - \alpha_0 - \alpha_1)\left[E(T^*|z_a) - E(T^*|z_b)\right]$$

thus we find that

$$\beta = (1 - \alpha_0 - \alpha_1)\mathcal{W}(z_a, z_b). \tag{25}$$

Finally, combining Equations 18, 21, 23 and 25 we have

$$\begin{aligned}\Delta Var(y|z_a, z_b) = \mathcal{W}\left(z_a, z_b\right)^2 \left\{\Delta Var(T|z_a, z_b) + (\alpha_0 - \alpha_1)\Delta E(T|z_a, z_b)\right\} \\ + 2\mathcal{W}(z_a, z_b)\left\{\left[E\left(T|z_a\right) - \alpha_0\right] m_{1a} - \left[E(T|z_b) - \alpha_0\right] m_{1b}\right\}\end{aligned} \tag{26}$$

an equation relating $\alpha_0, \alpha_1, m_{1a}$ and m_{1b} to various observable quantities.

Equation 26 provides an additional identifying restriction for each unique *pair* of values (z_a, z_b) in the support of z. If z takes on two values it provides one restriction, whereas if z takes on three values if provides two restrictions, and so on. To take a particularly simple example, suppose that z is binary and Mahajan's (2006) assumption that $E[u|z, T^*] = 0$ holds. Then Equation 26 reduces to

$$\Delta Var(y|1,0) = \left[\frac{Cov(z,y)}{Cov(z,T)}\right]^2 \left\{\Delta Var(T|1,0) + (\alpha_0 - \alpha_1)\left[\frac{Cov(z,T)}{Var(z)}\right]\right\}$$

Rearranging, we see that

$$\alpha_0 - \alpha_1 = \Delta Var(y|1,0)\left[\frac{Cov(z,T)Var(z)}{Cov(z,y)^2}\right] - \Delta Var(T|1,0)\left[\frac{Var(z)}{Cov(z,T)}\right]$$

In other words, the homoskedasticity restriction identifies the *difference* between the mis-classification rates. This makes intuitive sense. Provided that the variance of u is unrelated to z the only way that the variance of y can differ across values of z is if some values of z provide *more* information about the distribution of T^* than others. This is only possible if the mis-classification rates differ.

Of course, one need not impose the restriction that $E[u|z,T^*]=0$ to use the identifying information provided by Equation 26. Indeed, by exploiting homoskedasticity with respect to the instrument we can identify β using weaker conditions than Mahajan (2006) without requiring that z take on three or more values, as in Lewbel (2007). Moreover, when z does take on three or more values we can identify β even when T^* is endogenous.

I'm pretty sure this is true, but we do still need to prove it!

In the general case where we do not impose Mahajan's assumption that $E[u|z,T^*]=0$ the purpose of the homoskedasticity restrictions is to eliminate a quantity that appears in the moment condition that arises from the "modified IV estimator" in which $\widetilde{z} \equiv T\,(z-E[z])$ is used as an instrument for T. We showed previously that

$$\widetilde{\beta}_{IV} = \beta\left[\frac{(1-p-\alpha_1)+\alpha_0}{(1-p)(1-\alpha_0-\alpha_1)}\right] + \left[\frac{(1-\alpha_0-\alpha_1)\left\{E[zT^*u]-E[z]E[T^*u]\right\}}{(1-p)Cov(z,T)}\right]$$

First consider the case in which T^* is exogenous, so that $E(T^*u)=0$, and z is binary. Then the preceding reduces to

$$\widetilde{\beta}_{IV} = \beta\left[\frac{(1-p-\alpha_1)+\alpha_0}{(1-p)(1-\alpha_0-\alpha_1)}\right] + \left[\frac{(1-\alpha_0-\alpha_1)\,E[zT^*u]}{(1-p)Cov(z,T)}\right]$$

where

$$E[zT^*u] = E_{T^*,z}\left[E\left(zT^*u|z,T^*\right)\right] = p_{11}m_{11}$$

where $p_{jk}=P(T^*=j,z=k)$ and $m_{jk}=E[u|T^*=j,z=k]$. Note that by the definition of conditional probability we can equivalently express this as

$$E[zT^*u] = E(T^*|z=1)P(z=1)m_{11}$$

Thus we can rewrite the numerator of the second term in the expression for

$\widetilde{\beta}$ from above as

$$\begin{aligned} C &= (1-\alpha_0-\alpha_1)E(zT^*u) \\ &= (1-\alpha_0-\alpha_1)E(T^*|z=1)P(z=1)m_{11} \\ &= \left[E(T|z=1)-\alpha_0\right]P(z=1)m_{11} \end{aligned}$$

using Equation 19. Thus, when T^* is exogenous and z is binary, the expression for $\widetilde{\beta}_{IV}$ can be written as

$$\widetilde{\beta}_{IV} = \beta\left[\frac{(1-p-\alpha_1)+\alpha_0}{(1-p)(1-\alpha_0-\alpha_1)}\right] + \left[\frac{P(z=1)\left[E(T|z=1)-\alpha_0\right]m_{11}}{(1-p)Cov(z,T)}\right] \quad (27)$$

Now we will show that the second term from Equation 26 can be expressed in a similar fashion. The term in question is:

$$D = \left[E\left(T|z=1\right)-\alpha_0\right]m_{11} - \left[E\left(T|z=0\right)-\alpha_0\right]m_{10}$$

Imposing $Cov(T^*,u)=0$ gives $p_{10}m_{10}+p_{11}m_{11}=0$. Thus $m_{10}=-p_{11}m_{11}/p_{10}$. Now, by Equation 19,

$$-\frac{p_{11}}{p_{10}} = -\frac{P(T^*=1|z=1)P(z=1)}{P(T^*=1|z=0)P(z=0)} = -\frac{\left[E(T|z=1)-\alpha_0\right]P(z=1)}{\left[E(T|z=0)-\alpha_0\right]P(z=0)}$$

Substituting this into the expression for D, we have

$$D = \left[\frac{E(T|z=1)-\alpha_0}{P(z=0)}\right]m_{11}$$

and therefore, in the case where z is binary and T^* is exogenous Equation 26 simplifies to

$$\Delta Var(y|z) = \mathcal{W}^2\left\{\Delta Var(T|z)+(\alpha_0-\alpha_1)\Delta E(T|z)\right\}+2\mathcal{W}\left\{\frac{E(T|z=1)-\alpha_0}{P(z=0)}\right\}m_{11} \quad (28)$$

I think this will make things easier to solve because we could treat the quantity $[E(T|z=1)-\alpha_0]\, m_{11}$ as a unit and eliminate it from the system. But I could be wrong…

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., September 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 variabes. 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., May 2006. Identification and estimation of regression models with misclassification. Econometrica 74 (3), 631–665.