On the Use of Instrumental Variables to Identify the Effect of a Mis-measured, Binary Regressor

Francis J. DiTraglia Camilo Garcia-Jimeno
University of Pennsylvania

October 16, 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 from econometricians, and the scant existing literature is contradictory. Frazis and Loewenstein (2003) provide a brief discussion of the prospects for identification in the parametric linear case, suggesting that, in general, identification should be expected to fail. In contrast, an important paper by Mahajan (2006) includes a proof of identification in the more general non-parametric model that relies on the availability of 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 consider 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. While allowing the instrument to take on additional values does increase the number of available moment conditions, we show that under the standard instrumental variables assumptions these additional moment conditions provide no additional identifying information about 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 Notation and Related Literature

Give the general model, the full notation, and the shorthand notation. Also give the "m" notation since this will make it easier to talk about the papers.

Why is this model interesting? What is known about it?

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, Bollinger (1996) has shown that the IV estimator does not recover the true treatment effect: while the IV estimator removes the effect of regressor endogeneity it is inconsistent and the asymptotic bias depends on the extent of measurement error. When the regressor of interest is in fact *exogenous*, however, and a valid instrument is available, it is possible to recover the treatment effect using a GMM approach. 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]

Unfortunately, IV is inconsistent under non-differential measurement error: removes only the effect of endogeneity, not measurement error. Papers that explain how to construct a method of moments estimator, not IV, that uses an instrument, or second measure, to eliminate non-differential measurement error when the treatment of interest is *exogenous*: Kane et al. (1999), Black et al. (2000), Frazis and Loewenstein (2003). Later generalizations of this idea by Lewbel (2007) and Mahajan (2006).

Only two papers discuss case in which the treatment is endogenous as well as mis-classified and they reach apparently contradictory conclusions. (Should also mention the Hausman et al. (1998) and Tanguay Brachet stuff, although identification here is really "at infinity" and depends crucially on the parametric specification.) While most of their paper is devoted to case of

[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 measure. 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.

exogenous treatment, Frazis and Loewenstein (2003) spend a few paragraphs discussing extension to endogenous treatment. Although do not provide a formal proof, argue the model won't be identified except under very strong parametric restrictions since relaxing exogeneity introduces two new parameters. Frazis and Loewenstein (2003) consider traditional linear IV model making no distinction between continuous and discrete instrument. In contrast Mahajan (2006), apparently unaware of Frazis and Loewenstein (2003), concludes that the treatment effect is identified under endogeneity and mis-classification. Clearly there is some confusion in the literature about these kinds of models, main goal of this paper is to clarify what can and cannot be learned about a causal effect in the presence of mis-classification and endogeneity. Present a general analysis of non-parametric identification in this setting. Key ingredient: put existing papers into a common framework, including Lewbel (2007) who works with an "instrument" that actually has a direct effect on the outcome. Explain the model we will work with here and how we hold covariates fixed as in the proofs of other papers, how this allows us to simplify notation. Introduce the "m" notation.

Summary of our findings goes here. First, the analysis in both Frazis and Loewenstein (2003) and Mahajan (2006) is flawed. Frazis and Loewenstein (2003) get the exogeneity assumption wrong while Mahajan (2006) assumes a contradiction: no first stage. We Consider ways to get extra moment conditions to try to achieve identification: homoskedasticity restriction and additional values of instrument, a la Lewbel (2007). The homoskedasticity condition yields a simple and informative partial identification result regardless of the number of values the instrument takes on. Presumably we will prove that no matter how many values the instrument takes on, we can't get identification, with or without the homoskedasticity condition. Possibly consider some additional restrictions on the m^*_{jk} that would yield identification: some kind of symmetry condition on selection or something. Probably these assumptions aren't very plausible in practice. Summary of paper. Example from development experiment? Proofs in appendix.

Should probably start off, possible in previous section, by writing out a general encompassing framework that will allow us to talk about all the papers in this section. Present our model before this section. Should also mention the assumption $1 - \alpha_0 - \alpha_1 > 0$.

Many examples want to estimate effect of binary treatment. Often treatment is mis-measured and possibly endogenous. Measurement error in binary regressor cannot be classical. This has been know for a while, see Aigner (1973) and Bollinger (1996). Intuition: can only mis-code true zero and one and true one as zero. Describe non-differential measurement error idea. Under this kind of measurement error, IV estimator is inconsistent for the causal effect: can remove the effect of endogeneity but not of measurement error. See for example Black et al. (2000); Frazis and Loewenstein (2003); Kane et al. (1999).

Now talk about Kane et al. (1999) and Black et al. (2000). Two measures of exogenous binary treatment with non-differential measurement error allow one to identify treatment effect. Method of moments estimator *not* IV. Relies on discreteness of the problem: construct "cells" for $E[y|z, T]$. Talk about how the two papers differ in their contribution. Black et al. (2000) consider not only the binary case but a continuous version that isn't identified.

Need to figure out how Card (1996) relates to these as well. It looks like he does not in fact use two measures to estimate the effect of union status on wages. Instead he uses a two-period panel dataset and examines external information comparing employer and employee reporting of union status. This leads him to propose the assumption that the "up" and "down" mis-classification probabilities are equal, since it fits this external dataset well. This is the "quasi-classical" measurement error case that we talked about previously. There is only one measurement error parameter and presumably the panel dataset allows him to identify it.

Frazis and Loewenstein (2003) point out that an instrument can be used in place of a second measure of T^* provided that T^* is still exogenous. Essentially the same estimator as in Black et al. (2000) and Kane et al. (1999) but more general since the instrument need not be binary: can in fact be continuous. But they make a mistake. They assume $E[zu] = 0$, $E[T^*u] = 0$ and non-differential measurement error and claim that this is sufficient to consistently estimate β. However, this is incorrect: we need the additional assumption that $E[zT^*u] = 0$ which is stronger. (They seem to think that this term only appears when you have an endogenous T^*.) While Frazis and Loewenstein (2003) are aware that there are some differences between two measures of T^*, as in Black et al. (2000), and an arbitrary instrument z, they seem to have missed one subtle point. The assumptions in Black et al. (2000) in fact imply that $E[u|T^*, z] = 0$.[2] From this it follows that $E[zT^*u] = E[zu] = E[T^*u] = 0$. However, if one takes the non-differential measurement assumption literally it is in fact sufficient in the case of two measures to assume only that $E[zu] = E[T^*u] = 0$:

$$\begin{aligned} E[zT^*u] &= E[(T^* + w)T^*u] = E[(T^*)^2u] + E[wuT^*] && (1) \\ &= E\left[E\left(u|T^*\right)\left(T^*\right)^2\right] + E\left[E\left(wu|T^*\right)T^*\right] && (2) \\ &= 0 + E\left[E(w|T^*)E(u|T^*)T^*\right] = 0 && (3) \end{aligned}$$

using the fact that $E[u|T^*] = 0$ and w is independent of u conditional on T^*. This argument does *not* necessarily apply to an arbitrary instrument z: $E[zu] = E[T^*u] = 0$ does not imply that $E[zT^*u] = 0$.

Put in our simple binary example.

While it might seem strange to assume in practice that $E[zu] = E[T^*u] = 0$ are exogenous but not that $E[zT^*u] = 0$ the point is merely that this is an additional assumption beyond the usual assumptions of lack of correlation.

Frazis and Loewenstein (2003) also briefly discuss case in which T^* is endogenous, basically conclude that all you can get in this case is bounds for

[2] This follows from Assumptions A1 and A2 combined with Equation 3.

the treatment effect.

Presumably we're going to show that this isn't the case!

Mahajan (2006) considers the case of a binary regressor subject to non-differential measurement error when one has available a binary instrument (he calls it an ILV). Doesn't seem to be aware of Frazis and Loewenstein (2003), similar to Black et al. (2000) and Kane et al. (1999) although he allows for non-parametric effects of covariates and allows the mis-classification error rates to depend on covariates. (Although since covariates are held fixed in the proofs, this isn't a big deal.) The crucial assumption for the instrument-like variable is that it is unrelated to the mis-classification probabilities. Uses the assumption $E[u|T^*, z]$ to get identification using same basic estimator as Black et al. (2000). Also says that he can get identification when T^* is endogenous, but this is wrong. He correctly shows that it would be sufficient to learn the mis-classification error rates (as we know, the IV estimator converges to $\beta/(1-\alpha_0-\alpha_1)$ even when T^* is endogenous). However, he then states that his earlier theorem has proven that these rates are identified. That theorem, however, relies crucially on the assumption that T^* is exogenous!

Another closely related paper is Lewbel (2007). Whereas Mahajan makes sufficient assumptions to identify β with a two-valued z, provided that T^* is exogenous, Lewbel works with a three-valued z. While Lewbel also assumes that $E[T^*u] = 0$, his "instrument" is really more like a covariate. He assumes that z is unrelated to the mis-classification probabilities but allows it to have a direct effect on y, as long as there is no interaction between T^* and z. Since this involves imposing fewer restrictions on the m_{ij}, Lewbel requires that z take on more values. There is also some kind of determinant condition that we don't fully understand yet, but will figure out soon!

3 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

$$E[u^2|z] = E[u^2]. \tag{4}$$

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 4 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 4 follows.

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

$$Var(y|z) = \beta^2 Var\left(T^*|z\right) + Var(u|z) + 2\beta Cov(T^*, u|z) \tag{5}$$

Under 4, $Var(u|z)$ 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

$$\Delta Var(y|z_a, z_b) = \beta^2 \Delta Var(T^*|z_a, z_b) + 2\beta \Delta Cov(T^*, u|z_a, z_b)] \tag{6}$$

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,

$$P(T^* = 1|z) = \frac{P(T = 1|z) - \alpha_0}{1 - \alpha_0 - \alpha_1} \tag{7}$$

and accordingly,

$$Var(T^*|z) = \frac{\left[P\left(T = 1|z\right) - \alpha_0\right]\left[1 - P\left(T = 1|z\right) - \alpha_1\right]}{\left(1 - \alpha_0 - \alpha_1\right)^2} \tag{8}$$

Thus, evaluating Equation 8 at z_a and z_b and simplifying,

$$\Delta Var(T^*|z_a, z_b) = \frac{\Delta Var(T|z_a, z_b) + \left(\alpha_0 - \alpha_1\right)\Delta E(T|z_a, z_b)}{\left(1 - \alpha_0 - \alpha_1\right)^2} \tag{9}$$

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

$$Cov(T^*, u|z) = E_{T^*|z}\left[E\left(T^* u|z, T^*\right)\right] = P(T^* = 1|z) E\left(u|T^* = 1, z\right) \tag{10}$$

since $E[z|u] = 0$. Combining this with Equation 7 and evaluating at z_a and z_b gives

$$\Delta Cov(T^*, u|z_a, z_b) = \frac{\left[E\left(T|z_a\right) - \alpha_0\right] m_{1a} - \left[E(T|z_b) - \alpha_0\right] m_{1b}}{1 - \alpha_0 - \alpha_1} \tag{11}$$

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

Both Equations 9 and 11 involve only observable quantities and the mis-classification rates α_0 and α_1. Equation 6, 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

$$\mathcal{W}(z_a, z_b) = \frac{E(y|z_a) - E(y|z_b)}{E(T|z_a) - E(T|z_b)} \tag{12}$$

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 7,

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

thus we find that

$$\beta = (1 - \alpha_0 - \alpha_1)\mathcal{W}(z_a, z_b). \tag{13}$$

Finally, combining Equations 6, 9, 11 and 13 we have

$$\begin{aligned} \Delta Var(y|z_a, z_b) = \mathcal{W}\left(z_a, z_b\right)^2 \left\{\Delta Var(T|z_a, z_b) + (\alpha_0 - \alpha_1)\Delta E(T|z_a, z_b)\right\} \\ + 2\mathcal{W}(z_a, z_b)\left\{\left[E\left(T|z_a\right) - \alpha_0\right] m_{1a} - \left[E(T|z_b) - \alpha_0\right] m_{1b}\right\} \end{aligned} \tag{14}$$

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

Equation 14 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 14 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 14. 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 $\tilde{z} \equiv T(z - E[z])$ is used as an instrument for T. We showed previously that

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

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

$$\widetilde{\beta}_{IV} = \beta\left[\frac{(1-p-\alpha_1)+\alpha_0}{(1-p)(1-\alpha_0-\alpha_1)}\right] + \left[\frac{P(z=1)\left[E(T|z=1)-\alpha_0\right]m_{11}}{(1-p)Cov(z,T)}\right] \quad (15)$$

Now we will show that the second term from Equation 14 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 7,

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

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 14 simplifies to

$$\Delta Var(y|z) = \mathcal{W}^2 \left\{ \Delta Var(T|z) + (\alpha_0 - \alpha_1) \Delta E(T|z) \right\} + 2\mathcal{W} \left\{ \frac{E(T|z=1) - \alpha_0}{P(z=0)} \right\} m_{11} \tag{16}$$

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.

Card, D., July 1996. The effects of unions on the structure of wages: a longitudinal analysis. Econometrica 64 (4), 957–979.

Frazis, H., Loewenstein, M. A., 2003. Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variabes. 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.

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.