

Returns to Lying? Identifying the Effects of Misreporting When the Truth is Unobserved

Yingyao Hu* Arthur Lewbel†
Johns Hopkins University Boston College

September 30, 2011

Abstract

Consider an observed binary regressor D and an unobserved binary variable D^* , both of which affect some other variable Y . This paper considers nonparametric identification and estimation of the effect of D on Y , conditioning on $D^* = 0$. For example, suppose Y is a person's wage, the unobserved D^* indicates if the person has been to college, and the observed D indicates whether the individual claims to have been to college. This paper then identifies and estimates the difference in average wages between those who falsely claim college experience versus those who tell the truth about not having college. We estimate this average effect of lying to be about 6% to 20%. Nonparametric identification without observing D^* is obtained either by observing a variable V that is roughly analogous to an instrument for ordinary measurement error, or by imposing restrictions on model error moments.

JEL Codes: C14, C13, C20, I2. Keywords: Binary regressor, misclassification, measurement error, unobserved factor, discrete factor, program evaluation, treatment effects, returns to schooling, wage model.

*Department of Economics, Johns Hopkins University, 440 Mergenthaler Hall, 3400 N. Charles Street, Baltimore, MD 21218, USA Tel: 410-516-7610. Email: yhu@jhu.edu, <http://www.econ.jhu.edu/people/hu/>

†Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467 USA. Tel: 617-552-3678. email: lewbel@bc.edu <http://www2.bc.edu/~lewbel>

1 Introduction

Consider an observed binary regressor D and an unobserved binary variable D^* , both of which affect some other variable Y . This paper considers nonparametric identification and estimation of the effect of D on Y , conditioning on a value of the unobserved D^* (and possibly on a set of other observed covariates X). Formally, what is identified is the function $R(D, X)$ defined by

$$R(D, X) = E(Y \mid D^* = 0, D, X).$$

This can then be used to evaluate

$$r(X) = R(1, X) - R(0, X)$$

and $r = E[r(X)]$, which are respectively, the conditional and unconditional effects of D on Y , holding D^* fixed. When D^* is observed, identification and estimation of R is trivial. Here we obtain identification and provide estimators when D^* is unobserved.

Assuming $E(Y \mid D^*, D, X)$ exists, define a model H and an error η by

$$Y = E(Y \mid D^*, D, X) + \eta = H(D^*, D, X) + \eta \tag{1}$$

where the function H is unknown and the error η is mean zero and uncorrelated with D , D^* , and X . Then, since D and D^* are binary, we may without loss of generality rewrite this model in terms of the unknown R , r , and an unknown function s as

$$Y = R(D, X) + s(D, X)D^* + \eta \tag{2}$$

or equivalently

$$Y = R(0, X) + r(X)D + s(D, X)D^* + \eta. \tag{3}$$

This paper provides conditions that are sufficient to point identify the unknown functions R and r , even though D^* is unobserved. We also show set (interval) identification under weaker assumptions.

For a specific example, suppose for a sample of individuals the observed D is one if an individual claims or is reported to have some college education (and zero otherwise), and the unobserved D^* is one if the individual actually has some college experience. Let Y be the individual's wage rate. Then r is the difference in average wages Y between those who claim to have a degree when they

actually do not, versus those who honestly report not having a college degree. This paper provides nonparametric identification and associated estimators of the function r . We empirically apply these methods to estimate this average difference in outcomes between truth tellers and liars, when the truth D^* is not observed. Notice that we are not focusing on the effects of misreporting on estimates of returns to schooling, as in, e.g., Ashenfelter and Krueger (1994), but rather on the direct effects of misreporting on wages.

Only responses and not intent can be observed, so we cannot distinguish between intentional lying and false beliefs about D^* . For example, suppose D^* as an actual treatment and D is a perceived treatment (i.e., D is the treatment an individual thinks he received, and so is a false belief rather than an intentional lie). Then r is the average placebo effect, that is, the average difference in outcomes between those who were untreated but believe they received treatment versus those who correctly perceive that they were untreated. This paper then provides identification and an estimator for this placebo effect when the econometrician does not observe who actually received treatment.

Given a Rubin (1974) type unconfoundedness assumption, r will equal the average placebo effect, or the average returns to lying (which could be positive or negative). Unconfoundedness may be a reasonable assumption in the placebo example, but is less likely to hold when lying is intentional. Without unconfoundedness, the difference r in outcomes Y that this paper identifies could be due in part to unobserved differences between truth tellers and liars. For example, r could be positive even if lying itself has no direct effect on wages, if those who misreport their education level are on average more aggressive in pursuing their goals than others, or if some of them have spent enough time and effort studying (more on average than other nongraduates) to rationalize claiming that they have college experience. Alternatively r could be negative even if the returns to lying itself is zero, if the liars are more likely to arouse suspicion, or if there exist other negative character flaws that correlate with misreporting. Even with unconfoundedness, r might not equal the true returns to lying if Y is self reported data and the propensity to misreport D^* is correlated with misreporting Y , e.g., individuals who lie about their education level may also lie about their income.

Given that unconfoundedness may often be implausible in this context, we will call r the "effects of lying," and use the phrase "returns to lying" only when unconfoundedness is assumed.

The interpretation of r as a placebo effect or as effects or returns to lying also assumes that D^* and D are respectively the true and reported values of the same variable. This paper's identification and associated estimator does not require D and D^* to be related in this way (they can be completely

different binary variables), however, for the purposes of interpreting the required assumptions and associated results, we will throughout this paper refer to D as the reported value of a true D^* .

Discreteness of D and D^* is also not essential for this paper's identification method, but it does simplify the associated estimators and limiting distribution theory. In particular, if we more generally have a reported Z and an unobserved Z^* , we could apply this paper's identification method for any particular values z and z^* of interest by letting $D^* = I(Z^* \neq z)$ and $D = I(Z \neq z)$, where I is the indicator function. Then $D = 1$ when $D^* = 0$ means lying by claiming a value z when the truth is not z . Although our identification theory still holds in that case, having D or D^* be zero could then be zero probability events, resulting in estimation problems analogous to weak instruments which we do not address here.

When D is a possibly mismeasured or misclassified observation of D^* , then $D - D^*$ is the measurement or misclassification error. Virtually all of the literature on mismeasured binary regressors (which goes back at least as far as Aigner 1973) that attempts to estimate or bound the effect of D^* on Y (a treatment effect) assumes $r(X) = 0$, or equivalently, that any misclassification or measurement errors have no effect on the outcome Y after conditioning on the true D^* . Examples include Ashenfelter and Krueger (1994), Kane and Rouse (1995), Card (1996), Bollinger (1996), Hotz, Mullin, and Sanders (1997), Klepper, (1988), Manski (1990), Hu (2008), Mahajan (2006), Lewbel (2007a), Chen, Hu, and Lewbel (2008a, 2008b), and Molinari (2008). The same is true for general endogenous binary regressor estimators when they are interpreted as arising from mismeasurement. See, e.g., Das (2004), Blundell and Powell (2004), Newey and Powell (2003), and Florens and Malavolti (2003). The assumption that $r(X) = 0$ may be reasonable if the reporting errors $D - D^*$ are due to data collection errors such as accidentally checking the wrong box on a survey form. Having $r(X) = 0$ would also hold if the outcome Y could not be affected by the individual's beliefs or reports regarding D , e.g., if D^* were an indicator of whether the individual owns stock and Y is the return on his investment, then that return will only depend on the assets he actually owns and not on his beliefs or self reports about what he owns. Still, there are many applications where it is not reasonable to assume a priori that $r(X)$ is zero, so even when $r(X)$ is not of direct interest, it may be useful to apply this paper's methods to test if it is zero, which would then permit the application of the existing mismeasured or misclassified binary regressor estimators that require that $r(X) = 0$.

We propose two different methods of obtaining nonparametric identification without observing D^* . One is by observing a variable V that has some special properties, analogous to an instrument.

The second way we obtain identification is through restrictions on the first three moments of the model error η . Identification using an instrument V requires V to have some of the properties of a repeated measurement. In particular, Kane and Rouse (1995) and Kane, Rouse, and Staiger (1999) obtain data on both self reports of educational attainment D , and on transcript reports. They provide evidence that this transcript data (like the self reports D) may contain considerable reporting errors on questions like, "Do you have some years of college?" These transcript reports therefore cannot be taken to equal D^* , but we show these transcripts may satisfy the conditions we require for use as an instrument V .

The alternative method we propose for identification does not require an instrument V , but is instead based primarily on assuming that the first three moments of the model error η are independent of the covariates. For example, if η is normal, as might hold by Gibrat's (1931) law for Y being log wages, and homoskedastic, then η will satisfy this assumption. This second method of identification is similar to Chen, Hu, and Lewbel (2008a, 2008b), though (as we will show later) those papers could not be used to identify the effects of lying in our context without additional information.

The next two sections describe identification with and without an instrument. We then propose estimators based on each of these methods of identification, and provide an empirical application estimating the effects on wages of lying about educational attainment.

2 Identification Using an Instrument

ASSUMPTION A1: *The variable Y , the binary variable D , and a (possibly empty) vector of other covariates X are all observable. The binary variable D^* is unobserved. $E(Y \mid D^*, D, X)$ exists. The functions H , R , r , s and the variable η are defined by equations (1), (2) and (3).*

ASSUMPTION A2: *A variable V is observed with*

$$E(\eta V \mid D, X) = 0, \tag{4}$$

$$E(V \mid D, D^* = 1, X) = E(V \mid D^* = 1, X), \tag{5}$$

$$E(V \mid D = 1, X) \neq E(V \mid X). \tag{6}$$

Equation (4) says that the instrument V is uncorrelated with the model error η for any value of the observable regressors D and X . A sufficient condition for equation (4) to hold is if $E(Y | D^*, D, X, V) = E(Y | D^*, D, X)$. This is a standard property for an instrument. The following very simple Lemmas are useful for interpreting and applying the other equations that comprise Assumption A2:

LEMMA 1: *Assume $E(D | D^* = 1, X) \neq 0$. Equation (5) holds if and only if*

$$\text{Cov}(D, V | D^* = 1, X) = 0 \tag{7}$$

LEMMA 2: *Assume $E(D | X) \neq 0$. Equation (6) holds if and only if*

$$\text{Cov}(D, V | X) \neq 0. \tag{8}$$

Proofs of Lemmas and Theorems are in the Appendix. As shown by Lemmas 1 and 2, equations (5) and (6) say that D and V are correlated, but at least for $D^* = 1$, this relationship only occurs through D^* . Equation (5) means that when $D^* = 1$, the variable D has no additional power to explain V given X . If V is a second mismeasurement of D^* , then (5) or its equivalent (7) is implied by a standard assumption of repeated measurements, namely, that the error in the measurement D be unrelated to the error in the measurement V , while equation (6) can be expected to hold because both measurements are correlated with the true D^* . Equation (6) is close to a standard instrument assumption, if we are thinking of V as an instrument for D (since we are trying to identify the effect of D on Y). Note that equation (6) or Lemma 2 can be easily tested, since they only depend on observables.

To facilitate interpretation of the identifying assumptions, we discuss them in the context of the example in which Y is a wage, D^* is the true indicator of whether an individual has some college experience, D is the individual's self report of college experience, and V is transcript reports of educational attainment, which are an alternative mismeasure of D^* . Let X denote a vector of other observable covariates we may be interested in that can affect either wages, schooling, and/or lying, so X could include observed attributes of the individual and of her job.

In the college and wages example, equation (4) will hold if wages depend on both actual and self reported education, i.e., D^* and D , but not on the transcript reports V . This should hold if employers rely on resumes and worker's actual knowledge and abilities, but don't see college

transcripts. Equation (5) or equivalently (7) makes sense, in that errors in college transcripts depend on the actual D^* , but not on what individuals later self report. However, this assumption could be violated if individuals see their own transcripts and base their decision to lie in part on what the transcripts say. Finally, (6) is likely to hold assuming transcripts and self reports are accurate enough on average to both be positively correlated with the truth.

Define the function $g(X)$ by

$$g(X) = E(V | D^* = 1, X).$$

THEOREM 1: *If Assumptions A1 and A2 hold then $R(D, X)$ satisfies*

$$R(D, X) = \frac{E(YV | D, X) - E(Y | D, X)g(X)}{E(V | D, X) - g(X)}. \quad (9)$$

and $r(X) = R(1, X) - R(0, X)$ satisfies

$$r(X) = E(Y | D = 1, X) - E(Y | D = 0, X) + \frac{\text{cov}(Y, V | D = 0, X)}{g(X) - E(V | D = 0, X)} - \frac{\text{cov}(Y, V | D = 1, X)}{g(X) - E(V | D = 1, X)}. \quad (10)$$

We now consider set identification of $r(X)$ based on equation (10), and then follow that with additional assumptions that suffice for point identification of $R(D, X)$, and hence of $r(X)$, based on equation (9).

2.1 Set Identification Bounds Using an Instrument

ASSUMPTION A3: *Assume that $0 \leq E(V | D = 0, X) < E(V | D = 1, X) \leq g(X)$*

Assumption A3 is a very mild set of inequalities. Having the support of V be nonnegative suffices to make the expectations in Assumption A3 nonnegative. $E(V | D = 0, X) < E(V | D = 1, X)$ essentially means that self reports are positively correlated with the instrument, which should hold since both would typically be positively correlated with the truth. In the college example, this inequality

is equivalent to $\Pr(V = 1 | D = 0, X) < \Pr(V = 1 | D = 1, X)$, meaning that people reporting going to college are more likely to have a transcript that says they went to college than people who report not going to college. Given equation (7), violation of this inequality would require a relatively large fraction of people to reverse lie, that is, claim to not have college when they have in fact gone to college.

Define $\delta^*(X)$ by

$$\delta^*(X) = g(X) - E(V | D = 1, X)$$

So the last inequality in Assumption A3 is $\delta^*(X) \geq 0$. When V is a mismeasure of D^* , having $\delta^*(X) \geq 0$ is equivalent to $\Pr(V = 1 | D = 1, X) \leq \Pr(V = 1 | D^* = 1, X)$, which basically says that the instrument is closer to the truth than to the self report. This holds if a transcript is more likely to say you went to college when you are in the set of people that actually did go to college than when you are in the set of people that claimed to have been to college. It can also be readily shown that this last equality holds if $\Pr(V = 1 | D = 1, D^* = 1, X) \geq \Pr(V = 1 | D = 1, D^* = 0, X)$, which means that among people who claim college, those who actually went to college have a higher chance of their transcript saying they went to college than those that who's claims to college are misreports. As with some earlier assumptions, this assumption in any of its forms will hold if people's decision to lie, or accidental misreporting, is unrelated to transcript errors.

COROLLARY 1.: *Let Assumptions A1, A2, and A3 hold. Then $r(X)$ lies in an identified interval that is bounded from below if $\text{cov}(Y, V | D = 0, X) > 0$ and bounded from above if $\text{cov}(Y, V | D = 0, X) < 0$. If there exists an identified positive $\delta(X)$ such that $\delta(X) \leq \delta^*(X)$ then $r(X)$ lies in an identified bounded interval.*

Corollary 1 provides bounds on $r(X)$ whether an identified $\delta(X)$ exists or not, but the bounds are improved given a $\delta(X)$. For an example of a $\delta(X)$, suppose that $E(V | D^* = 1, X) = E(V | D^* = 1)$, that is, the probability that a school produces the transcript error $V = 0$ when $D^* = 1$ is unrelated to an individual's observed attributes X , e.g., this would hold if all college graduates are equally likely to have the school lose their file or otherwise mistakenly report that they are not graduates. Then $g(X)$ is independent of X , and $\delta(X) = \sup_x E(V | D = 1, X = x) - E(V | D = 1, X)$ which may be strictly positive for many values of X .

Corollary 1 follows immediately from inspection of equation (10), as does the construction of bounds for $r(X)$. All of the terms on the right of equation (10) are moments of observable

data, and hence are identified, except for $g(X)$. By Assumption A3, a lower bound on $g(X)$ is $E(V | D = 1, X)$. An upper bound of $g(X)$ is $\sup[\text{supp}(V)]$, since $g(X)$ is an expectation of V and so cannot exceed the largest value V can take on. Note that when V is a mismeasure of D^* as in the college example, this upper bound of $g(X)$ is one. From Assumptions A1 and A2, all of the expectations and covariances on the right of equation (10) exist. The function $g(x)$ appears only in the denominators of the last two terms in equation (10). By Assumption A3, the third term in equation (10) lies in the interval bounded by the two points

$$\frac{\text{cov}(Y, V | D = 0, X)}{E(V | D = 1, X) - E(V | D = 0, X)} \quad \text{and} \quad \frac{\text{cov}(Y, V | D = 0, X)}{\sup[\text{supp}(V)] - E(V | D = 0, X)}$$

Both of which are finite. Similarly, the last term in equation (10) lies in the interval bounded by the two points

$$\frac{\text{cov}(Y, V | D = 1, X)}{\delta^*(X)} \quad \text{and} \quad \frac{\text{cov}(Y, V | D = 1, X)}{\sup[\text{supp}(V)] - E(V | D = 1, X)}$$

The second of these points is finite. Given only assumptions A1, A2, and A3, $\delta^*(X) \leq 0$ so the first of the above points can be infinite. Whether it is plus or minus infinity, and hence whether we only have a lower or upper bound for $r(X)$, depends on the sign of $\text{cov}(Y, V | D = 1, X)$. If we have a $\delta(X)$ with $0 < \delta(X) \leq \delta^*(X)$, then we instead obtain the finite bound $\text{cov}(Y, V | D = 1, X) / \delta(X)$.

To construct the identified interval that contains $r(X)$, we must consider four cases corresponding to the four possible pairs of signs that $\text{cov}(Y, V | D = 0, X)$ and $\text{cov}(Y, V | D = 1, X)$ can take on. Note that the denominators of the last two terms in equation (10) are positive. If $\text{cov}(Y, V | D = 0, X)$ and $\text{cov}(Y, V | D = 1, X)$ have opposite signs, then $r(X)$ is strictly increasing or decreasing in $g(X)$, so the interval that $r(X)$ can lie in is bounded by equation (10) evaluated at the lowest and highest values $g(X)$ can take on, the highest being $\sup[\text{supp}(V)]$ and lowest either $E(V | D = 1, X)$ or $E(V | D = 1, X) + \delta(X)$ if a $\delta(X)$ is known. If $\text{cov}(Y, V | D = 0, X)$ and $\text{cov}(Y, V | D = 1, X)$ have the same signs, then these could still be bounds on $r(X)$, but it is also possible in that case that $r(X)$ either first increases and then decreases in $g(X)$ or vice versa, in which case the point where the derivative of $r(X)$ with respect to $g(X)$ equals zero may also be a bound.

Although Assumption A3 is already rather weak, one could similarly obtain a looser bound by replacing it with the weaker assumption that $0 \leq E(V | D^* = 0, X) \leq E(V | D^* = 1, X)$. This is little more than the assumption that transcripts be right more often than they are wrong, that is,

people with college education will have a higher probability of transcripts reporting college education than those without college education.

2.2 Point Identification Using an Instrument

We now consider additional assumptions that permit point identification of $r(X)$.

COROLLARY 2: *Let Assumptions A1 and A2 hold. Assume the function $g(X)$ is known and $E(V | D, X) \neq g(X)$. Then $R(D, X)$ is identified by*

$$R(D, X) = \frac{E(Y(V - g(X)) | D, X)}{E((V - g(X)) | D, X)} \quad (11)$$

Identification of $r(X)$ is then given by $r(X) = R(1, X) - R(0, X)$. Corollary 2 follows immediately from Theorem 1 by substituting $g(X)$ for $E(V | D^* = 1, X)$ in equation (9), and observing that all the other terms in equation (9) are expectations of observables, conditioned on other observables, and hence are themselves identified. One way Corollary 2 might hold is if a form of validation data exists. For example if D and D^* refer to graduating from college, then $g(X)$ could be obtained from a survey of transcripts just of people known to have graduated college. A special case of this assumption holding is if V is a mismeasure of D^* , as when V is the transcript report, and $g(X) = 1$, that is, if transcript errors of the form $V = 0$ when $D^* = 1$ are ruled out.

Another example or variant of Corollary 2 is the following.

ASSUMPTION A4: *There exists an x_1 such that*

$$E(V | D^* = 1, X) = E(V | X = x_1) \quad (12)$$

and

$$E(V | D, X) \neq E(V | X = x_1) \quad (13)$$

Equation (12) assumes that V has the same mean for people who have $X = x_1$ as for people that have $D^* = 1$ and any value of X . One set of sufficient (but stronger than necessary) conditions for equation (12) to hold is if $E(V | D^* = 1, X = x_1) = E(V | D^* = 1)$, so for people having college

($D^* = 1$), the probability of a transcript error is unrelated to one's personal attribute information X , and if

$$\Pr(D^* = 1 \mid X = x_1) = 1, \tag{14}$$

so people who have $X = x_1$ are an observable subpopulation that definitely have some college. In our application, we use Corollary 3 below for identification and we take this subpopulation x_1 to be individuals with very high test scores and self reported advanced degrees. Note that if equation (14) holds then equation (12) would only be violated if colleges systematically made more or fewer errors when producing transcripts for individuals with attributes $X = x_1$ than for students with other attribute values.

Equation (13) is a technicality that, analogous to the assumption that $E(V \mid D, X) \neq g(X)$ in Corollary 2, will avoid division by zero in Corollary 3 below. It is difficult to see why it should not hold in general, and it is empirically testable since it depends only on observables. However, if both equations (12) and (14) hold then equation (13) will not hold for $X = x_1$. This means that $R(D, x_1)$ cannot be identified in this case, though we still identify $R(D, X)$ for $X \neq x_1$. This is logical because if all individuals having $X = x_1$ have $D^* = 1$ by equation (14), then none of them can be lying when reporting $D = 1$.

COROLLARY 3: *If Assumptions A1, A2, and A4 hold then $R(D, X)$ is identified by*

$$R(D, X) = \frac{E(YV \mid D, X) - E(Y \mid D, X) E(V \mid X = x_1)}{E(V \mid D, X) - E(V \mid X = x_1)}. \tag{15}$$

Corollary 3 follows Theorem 1, by substituting equation (12) into equation (9) to obtain equation (15), and equation (13) makes the denominator in equation (15) be nonzero.

Given identification of $R(D, X)$ by Corollary 2 or 3, the effects of lying $r(X)$ is also identified by $r(X) = R(1, X) - R(0, X)$.

Although rather more difficult to interpret and satisfy than the assumptions in Corollaries 2 and 3, yet another alternative set of identifying assumptions is equations (4), (6) and $Cov(D^*, V \mid D, X) = 0$, which by equation (3) implies $Cov(Y, V \mid X) = r(X)Cov(D, V \mid D, X)$ which can then be solved for, and hence identifies, $r(X)$.

3 Identification Without an Instrument

We now consider identification based on restrictions on moments of η rather than on the presence of an instrument. In particular, we will assume that the second and third moments of η do not depend on D^* , D , and X . The method of identification here is similar to that of Chen, Hu, and Lewbel (2008b), though that paper imposes the usual measurement error assumption that the outcome Y is conditionally independent of the mismeasure D , conditioning on the true D^* , or equivalently, it assumes that $r(X) = 0$. One could modify Chen, Hu, and Lewbel (2008b) to identify our effects of lying model in part by including D in the list of regressors and treating our V from the previous section as the observed mismeasure of D^* . However, in that case one would need both an instrument V with certain properties and restrictions on higher moments of η , while in the present paper these are alternative methods of identification.

ASSUMPTION B1:

$$E(\eta \mid D^*, D, X) = 0, \tag{16}$$

$$E(\eta^k \mid D^*, D, X) = E(\eta^k) \quad \text{for } k = 2, 3, \tag{17}$$

there exists an x_0 such that

$$\Pr(D = 0 \mid D^* = 1, X = x_0) = 0 \quad \text{and} \quad \Pr(D = 0 \mid X = x_0) > 0, \tag{18}$$

and

$$E(Y \mid D^* = 1, D, X) \geq E(Y \mid D^* = 0, D, X) \tag{19}$$

Equation (16) can be assumed to hold without loss of generality by definition of the model error η . Equation (17) says that the second and third moments of the model error η do not depend on D^* , D , X , and so would hold under the common modeling assumption that the error η in a wage equation is independent of the regressors.

Equation (18) implies that people, or at least those in some subpopulation $\{X = x_0\}$, will not underreport and claim to not have been to college if they in fact have been to college. At least in terms of wages, this is plausible in that it is hard to see why someone would lie to an employer by claiming to have less education or training than he or she really possesses.

Finally, equation (19) implies that the impact of D^* on Y conditional on D and X is known to be positive. This makes sense when Y is wages and D^* is the true education level, since ceteris paribus, higher education on average should result in higher wages on average.

Define

$$\begin{aligned}\sigma_{Y|D,X}^2(D, X) &= E(Y^2|D, X) - [E(Y|D, X)]^2, \\ v_{Y|D,X}^3(D, X) &= E([Y - E(Y|D, X)]^3 |D, X),\end{aligned}$$

$$\begin{aligned}\alpha(D, X) &= \sigma_{Y|D,X}^2(D, X) - \sigma_{Y|D,X}^2(0, x_0), \\ \beta(D, X) &= v_{Y|D,X}^3(D, X) - v_{Y|D,X}^3(0, x_0) + 2E(Y|D, X)\alpha(D, X), \\ \gamma(D, X) &= \alpha(D, X)^2 + [E(Y|D, X)]^2\alpha(D, X) - E(Y|D, X)\beta(D, X).\end{aligned}$$

THEOREM 2: *Suppose that Assumptions A1 and B1 hold and that $\alpha(D, X) \neq 0$ for $(D, X) \neq (0, x_0)$. Then, $R(D, X)$ and $s(D, X)$ are identified as follows:*

- i) if $(D, X) = (0, x_0)$, then $R(D, X) = E(Y|D, X)$;*
- ii) if $(D, X) \neq (0, x_0)$, then*

$$R(D, X) = \frac{\beta(D, X) - \sqrt{\beta(D, X)^2 + 4\alpha(D, X)\gamma(D, X)}}{2\alpha(D, X)},$$

and

$$s(D, X) = \frac{\alpha(D, X)}{E(Y|D, X) - R(D, X)} + E(Y|D, X) - R(D, X).$$

As before given $R(D, X)$ we may identify the effects of lying $r(X)$ using $r(x) = R(1, X) - R(0, X)$. Identification of $s(D, X)$ in Theorem 2 means that the entire conditional mean function H in equation 1 is identified.

Some intuition for this identification comes from observing that, conditional on X , the number of equality constraints imposed by the assumptions equal the number of unknowns. One of these equations is a quadratic, and the inequality (19) is only needed to identify which of the two roots is correct. Based on this intuition, identification based on alternative equality restrictions should be possible, e.g., in place of equation (18) one could consider the constraint that the third moment $E(\eta^3)$ equal zero. Also, dropping inequality assumptions like (19) will result in set rather than point identification, where the sets are finite and consist of only two or three possible values.

4 Unconfoundedness

By construction the function $r(X)$ is the difference in the conditional mean of Y (conditioning on D , X , and on $D^* = 0$) when D changes from zero to one. Assuming D is the reported response and D^* is the truth, here we formally provide the unconfoundedness condition required to have this $r(X)$ equal the returns to lying. Consider the weak version of the Rubin (1974) or Rosenbaum and Rubin (1984) unconfoundedness assumption given by equation (20), interpreting D as a treatment. Letting $Y(d)$ denote what Y equals given the response $D = d$, if

$$E[Y(d) \mid D, D^* = 0, X] = E[Y(d) \mid D^* = 0, X] \quad (20)$$

then it follows immediately from applying, e.g., Heckman, Ichimura, and Todd (1998), that $E[Y(1) - Y(0) \mid D^* = 0, X] = r(X)$ is the conditional average effect of D , and so is the conditional on X average returns to lying.

5 Estimation Using an Instrument

We now provide estimators of $R(D, X)$ and hence of $r(X)$ following from Corollary 2 or 3 of Theorem 1. We first describe nonparametric estimators that are based on ordinary sample averages, which can be used if X is discrete. We then discuss kernel based nonparametric estimation, and finally we provide a simple least squares based semiparametric estimator that does not require any kernels, bandwidths, or other smoothers regardless of whether X contains continuous or discrete elements.

5.1 Nonparametric, Discrete X Estimation

Note that while identification only requires Assumption A4 to hold for a single value of X , that is, x_1 , it may be the case that this assumption is known to hold for a range of values of x_1 . We may then replace $E(V \mid X = x_1)$ with the expected value of V conditional on X equalling any value in this range. This may then improve the accuracy with which we can estimate this conditional expectation. In particular if X has any continuous components then $E(V \mid X = x_1)$ for a single value of x_1 is conditioning on a zero probability event, the estimate of which will converge at a slower rate than conditioning on a range of values X that has nonzero probability. Therefore, define U_i to be a dummy

variable such that

$$U_i = I(X_i \in \{x_1: \text{Assumption A4 is known to hold}\}), \quad (21)$$

where $I(\cdot)$ is the indicator function. In other words, let U_i equal one if equations (12) and (13) are assumed to hold when replacing x_1 in those equations with X_i , otherwise let U_i equal zero. It then follows immediately from Corollary 3 that equation (15) holds replacing $E(V | X = x_1)$ with $E(V | U = 1)$, so

$$R(D, X) = \frac{E(YV | D, X) - E(Y | D, X) E(V | U = 1)}{E(V | D, X) - E(V | U = 1)}. \quad (22)$$

We first consider estimation in the simple case where X is discrete. Replacing the expectations in equation (22) with sample averages in this case gives the estimators

$$\widehat{R}(d, x) = \frac{\widehat{\mu}_{Y,V,X,d} - \widehat{\mu}_{Y,X,d}\widehat{\mu}}{\widehat{\mu}_{V,X,d} - \widehat{\mu}_{X,d}\widehat{\mu}}, \quad \widehat{r}(x) = \widehat{R}(1, x) - \widehat{R}(0, x). \quad (23)$$

with

$$\begin{aligned} \widehat{\mu}_{Y,V,X,d} &= \frac{1}{n} \sum_{i=1}^n Y_i V_i I(X_i = x, D_i = d), & \widehat{\mu}_{Y,X,d} &= \frac{1}{n} \sum_{i=1}^n Y_i I(X_i = x, D_i = d), \\ \widehat{\mu}_{V,X,d} &= \frac{1}{n} \sum_{i=1}^n V_i I(X_i = x, D_i = d), & \widehat{\mu}_{X,d} &= \frac{1}{n} \sum_{i=1}^n I(X_i = x, D_i = d), \\ \widehat{\mu}_{V,U} &= \frac{1}{n} \sum_{i=1}^n V_i U_i, & \widehat{\mu}_U &= \frac{1}{n} \sum_{i=1}^n U_i, & \widehat{\mu} &= \widehat{\mu}_{V,U} / \widehat{\mu}_U \end{aligned}$$

Estimation based on equation (11) is the same replacing $\widehat{\mu}$ with $g(X)$ in equation (23)

One may also consider the unconditional mean wages $R_d = E[R(d, X)]$ and unconditional average effects of lying $r = E[r(X)]$, which may be estimated by

$$\widehat{R}_d = \frac{1}{n} \sum_{i=1}^n \widehat{R}(d, X_i), \quad \widehat{r} = \frac{1}{n} \sum_{i=1}^n \widehat{r}(X_i). \quad (24)$$

Assuming independent, identically distributed draws of $\{Y_i, V_i, X_i, D_i, U_i\}$, and existence of relevant variances, it follows immediately from the Lindeberg-Levy central limit theorem and the delta method that $\widehat{R}(d, x)$, $\widehat{r}(x)$, \widehat{R}_d , and \widehat{r} are root n consistent and asymptotically normal, with variance formulas as provided in the appendix, or that can be obtained by an ordinary bootstrap. Analogous limiting distribution results will hold with heteroskedastic or dependent data generating processes, as long as a central limit theorem still applies.

5.2 General Nonparametric Estimation

Letting $\mu = E(V | U = 1)$, equation (22) can be rewritten as

$$R(D, X) = \frac{E[Y(V - \mu) | D, X]}{E[(V - \mu) | D, X]}. \quad (25)$$

Equation (11) can also be written in the form of equation (25) by replacing μ with $g(X)$.

Assume n independent, identically distributed draws of $\{Y_i, V_i, X_i, D_i, U_i\}$. Let $X_i = (Z_i, C_i)$ where Z and C are, respectively, the vectors of discretely and continuously distributed elements of X . Similarly let $x = (z, c)$. Let $\hat{\mu} = \hat{\mu}_{V,U}/\hat{\mu}_U$ if estimation is based on equation (22), otherwise replace $\hat{\mu}$ with $g(x)$. Using equation (25), a kernel based estimator for $R(D, X)$ is

$$\hat{R}(d, x) = \frac{\sum_{i=1}^n Y_i (V_i - \hat{\mu}) K[(C_i = c)/b] I(Z_i = z) I(D_i = d)}{\sum_{i=1}^n (V_i - \hat{\mu}) K[(C_i = c)/b] I(Z_i = z) I(D_i = d)} \quad (26)$$

where K is a kernel function and b is a bandwidth that goes to zero as n goes to infinity. Equation (26) is numerically identical to the ratio of two ordinary nonparametric Nadaraya-Watson kernel regressions of $Y(V - \hat{\mu})$ and $V - \hat{\mu}$ on X, D , which under standard conditions are consistent and asymptotically normal. These will have the same slower than root n rate of convergence as regressions that use a known μ in place of the estimator $\hat{\mu}$, because an estimated $\hat{\mu}$ converges at the rate root n by the law of large numbers. Alternatively, equation (25) can be rewritten as the conditional moment

$$E[(Y - R(D, X))(V - \mu) | D, X] = 0 \quad (27)$$

which may be estimated using, e.g., the functional GMM estimator of Ai and Chen (2003), or by Lewbel's (2007b) local GMM estimator, with limiting distributions as provided by those references.

Given $\hat{R}(d, x)$ from equation (26) we may as before construct $\hat{r}(x) = \hat{R}(1, x) - \hat{R}(0, x)$, and unconditional estimates \hat{R}_d and \hat{r} by equation (24). We also construct trimmed unconditional effects $\hat{r}_t = \frac{1}{n} \sum_{i=1}^n \hat{r}(X_i) I_{ti}$ and similarly for \hat{R}_{dt} , where I_{ti} is a trimming parameter that equals one for most observations i , but equals zero for tail observations. Assuming regularity conditions such as Newey (1994) these trimmed unconditional effects are root n consistent and asymptotically normal estimates of the trimmed means r_t and R_{dt} .

5.3 Simple Semiparametric Estimation

Assume we have a parameterization $R(D, X, \theta)$ for the function $R(D, X)$ with a vector of parameters θ . The function $s(D, X)$ and the distribution of the model error η are not parameterized. Then

based on the definition of μ and equation (27), θ and μ could be jointly estimated based on Corollary 3 by applying GMM to the moments

$$E[(V - \mu)U] = 0 \tag{28}$$

$$E[\psi(D, X)(Y - R(D, X, \theta))(V - \mu)] = 0 \tag{29}$$

for a chosen vector of functions $\psi(D, X)$. For estimation based on Corollary 2, the estimator would just use the moments given by equation (29) replacing μ with $g(X)$.

Let $W = (1, D, X)'$. If R has the linear specification $R(D, X, \theta) = W'\theta$ then let $\psi(D, X) = W$ to yield moments $E[W(Y - W'\theta)(V - \mu)] = 0$, so $\theta = E[(V - \mu)WW']^{-1}E[(V - \mu)WY]$. This then yields a weighted linear least squares regression estimator

$$\hat{\theta} = \left[\sum_{i=1}^n (V_i - \hat{\mu}) W_i W_i' \right]^{-1} \left[\sum_{i=1}^n (V_i - \hat{\mu}) W_i Y_i \right] \tag{30}$$

based on Corollary 3, or the same expression replacing $\hat{\mu}$ with $g(X_i)$ based on Corollary 2. Given $\hat{\theta}$ we then have $\hat{R}(D, X) = W'\hat{\theta}$. In this semiparametric specification $r(x)$ is a constant with $\hat{r}(x) = \hat{r} = \hat{\theta}_1$, the first element of $\hat{\theta}$. Note that both GMM based on equation (29) and the special case of weighted linear regression based on equation (30) do not require any kernels, bandwidths, or other smoothers for their implementation.

6 Estimation Without an Instrument

We now consider estimation based on Theorem 2. As in the previous section, let K be a kernel function, b be a bandwidth, and $X_i = (Z_i, C_i)$ where Z and C are, respectively, the vectors of discretely and continuously distributed elements of X . Also let $x = (z, c)$. For $k = 1, 2, 3$, define

$$\hat{E}(Y^k | D = d, X = x) = \frac{\sum_{i=1}^n Y_i^k K[(C_i = c)/b] I(Z_i = z) I(D_i = d)}{\sum_{i=1}^n K[(C_i = c)/b] I(Z_i = z) I(D_i = d)} \tag{31}$$

This is a standard Nadayara-Watson Kernel regression combining discrete and continuous data, which provides a uniformly consistent estimator of $E(Y^k | D = d, X = x)$ under standard conditions. Define

$$\begin{aligned} \hat{\sigma}_{Y|D,X}^2(d, x) &= \hat{E}(Y^2 | D = d, X = x) - \left[\hat{E}(Y | D = d, X = x) \right]^2, \\ \hat{v}_{Y|D,X}^3(d, x) &= \hat{E} \left(\left[Y - \hat{E}(Y | D = d, X = x) \right]^3 | D = d, X = x \right), \end{aligned}$$

$$\begin{aligned}
\hat{\alpha}(d, x) &= \hat{\sigma}_{Y|D,X}^2(d, x) - \hat{\sigma}_{Y|D,X}^2(0, x_0), \\
\hat{\beta}(d, x) &= \hat{v}_{Y|D,X}^3(d, x) - v_{Y|D,X}^3(0, x_0) + 2\hat{E}(Y|D = d, X = x)\hat{\alpha}(d, x), \\
\hat{\gamma}(d, x) &= \hat{\alpha}(d, x)^2 + \left[\hat{E}(Y|D = d, X = x)\right]^2 \hat{\alpha}(d, x) - \hat{E}(Y|D = d, X = x)\hat{\beta}(d, x).
\end{aligned}$$

Based on Theorem 2 and uniform consistency of the kernel regressions, a consistent estimator of $R(d, x)$ is then

$$\begin{aligned}
\hat{R}(0, x_0) &= \hat{E}(Y|D = 0, X = x_0), \\
\hat{R}(d, x) &= \frac{\hat{\beta}(d, x) - \sqrt{\hat{\beta}(d, x)^2 + 4\hat{\alpha}(d, x)\hat{\gamma}(d, x)}}{2\hat{\alpha}(d, x)} \quad \text{for } (d, x) \neq (0, x_0).
\end{aligned}$$

As before, every conditional expectation above that conditions on $X = x_0$ can be replaced by an expectation conditional on X equalling any value x having the property that the assumptions of Theorem 2 hold replacing x_0 with that value x .

If X does not contain any continuously distributed elements, then these estimators are smooth functions of cell means, and so are root n consistent and asymptotically normal by the Lindeberg Levy central limit theorem and the delta method. Given $\hat{R}(d, x)$ from equation (26) we may as before construct $\hat{r}(x) = \hat{R}(1, x) - \hat{R}(0, x)$, and unconditional effects \hat{R}_d and \hat{r} by equation (24). Also as before, root n consistent, asymptotically normal convergence of trimmed means of \hat{R}_d and \hat{r} is possible using regularity conditions as in Newey (1994) for two step plug in estimators.

7 Effects of Misreporting College Attainment

Here we report results of empirically implementing our estimators of $r(x)$ where D is self reports of schooling and Y is log wages. In this context, our effects of lying estimates should be interpreted only as the difference in means between accurate reporters and misreporters of college for a limited sample, rather than as actual returns to lying about schooling, for many reasons. First, our conditional mean estimates cannot control for the selection effects that are at the heart of the modern literature on wages and schooling going back at least to Heckman (1979). Similarly, unconfoundedness with respect to lying based on equation (20) may not hold. Also, people who misreport college may similarly misreport their wages. Our results may also differ from actual returns to lying by the fact that both the risks and the returns to misreporting on a survey are lower than for lying on a job

application, though presumably the cost of potentially being caught in a lie in any context provides some incentive to report the same education level on a survey as was reported to one's employer. Finally, our sample may not be representative of the general population.

7.1 Preliminary Data Analysis

Kane, Rouse, and Staiger (1999) estimate a model of wages as a function of having either some college, an associate degree or higher, or a bachelors degree or higher. Their model also includes other covariates, and they use data on both self reports and transcript reports of education level. Their data is from the National Longitudinal Study of High School Class of 1972 (NLS-72) and a Post-secondary Education Transcript Survey (PETS). We use their data set of $n = 5912$ observations to estimate the effects of lying, defining Y to be log wage in 1986, D to be one if an individual self reports having "some college" and zero otherwise, while V is one for a transcript report of having "some college" and zero otherwise (both before 1979). We also provide estimates where D and V are self and transcript reports of having an associate degree or more, and reports of having a bachelor's degree or more. We take X to be the same set of other regressors Kane, Rouse, and Staiger (1999) used, which are a 1972 standardized test score and zero-one dummy variables for female, black nonhispanic, hispanic, and other nonhispanic.

The means of D and V (which equal the fractions of our sample that report having that level of college or higher) are 0.6739 and 0.6539 for "some college," 0.4322 and 0.3884 respectively for "Associate degree," and 0.3557 and 0.3383 for "Bachelors degree." The average log wage Y is 2.228.

Table 1: Effects of Lying and Schooling Treating Transcripts as True

	Some college	Associate degree	Bachelor's degree
r if $V=D^*$	0.1266 (0.03129)	0.2322 (0.02748)	0.1948 (0.04451)
r if $V=D^*$, linear	0.07868 (0.02864)	0.1681 (0.02777)	0.1269 (0.04082)
s if $V=D^*$	0.2831 (0.01366)	0.2958 (0.01288)	0.3181 (0.01280)
$E(DV)$	0.6204	0.3794	0.3325
$E[D(1-V)]$	0.05345	0.05277	0.02317
$E[(1-D)V]$	0.03349	0.008965	0.005751
$E[(1-D)(1-V)]$	0.2926	0.5589	0.6385

Standard Errors are in Parentheses

If D^* were observed along with Y and D , then the functions $r(x)$ and $s(d, x)$ could be immediately estimated from equation (3). Table 1 provides preliminary estimates of r and s based on this equation, under the assumption that transcripts have no errors. The row "r if V=D*" in Table 1 is the sample estimates of $E(Y|V = 0, D = 1) - E(Y|V = 0, D = 0)$, which would equal an estimate of $r = E[r(X)]$ if $V = D^*$, that is, if the transcripts V were always correct. The row, "r if V=D*, linear" is the coefficient of D in a linear regression of Y on D , V , and X , and so is another estimate of r that would be valid if $V = D^*$ and given a linear model for log wages.

The third row of Table 1 is the sample analog of $E(Y|V = 1) - E(Y|V = 0)$, which if $V = D^*$ would be an estimate of the effects of schooling $s = E[s(D, X)]$ (that is, the difference in conditional means of log wages between those with $D^* = 1$, versus those with $D^* = 0$, which equals returns to schooling if the effects of schooling satisfy an unconfoundedness condition). In this and all other tables, standard errors are obtained by 400 bootstrap replications, and are given in parentheses.

Table 1 also shows the fraction of truth tellers and liars, if the transcripts V were always correct. The rows labeled $E(DV)$ and $E[(1-D)(1-V)]$ give the fraction of observations where self and transcript reports agree that the individual respectively either has or does not have the given level of college. The row labeled $E[D(1-V)]$ gives the fraction of relevant liars if the transcripts are correct, that is, it is the fraction who claim to have the given level of college, $D = 1$, while their transcripts say they do not, $V = 0$. This fraction is a little over 5% of the sample for some college or Associate degree, but only about half that amount appear to misreport having a Bachelor's degree.

If V has no errors, then Table 1 indicates a small amount of lying in the opposite direction, given by the row labeled $E[(1-D)V]$. These are people who self report having less education than is indicated by their transcripts, ranging from a little over half a percent of the sample regarding college degrees to almost 3% for "some college." It is difficult to see a motive for lying in this direction, which suggests ordinary reporting errors in self reports, transcript reports, or both.

Prior to estimating $r(x)$, we examined equation (6) of Assumption A2, which is testable. A sufficient condition for equation (6) to hold is that $E(V|D = 1) - E(V) \neq 0$. In our data the t-statistic for the null hypothesis $E(V|D = 1) = E(V)$ is over 40 for each of the three levels of schooling considered, which strongly supports this assumption.

7.2 Instrumental Variable Based Estimates

We now report instrumental variable based estimates, specifically, Table 2 summarizes estimates of $r(x)$ based on Corollary 3. We define U in equation (21) to equal one for individual's that both self report having a masters degree or a PhD and are in the top decile of the standardized test scores. We are therefore assuming that Assumption A4 holds for x_0 equal to any X that includes these attributes of a self reported advanced degree and a high test score.

In our data the mean of U is 0.03468, so about 3.5% of our sample have both very high test scores and self report an advance degree. We could have based U on transcript reports of a graduate degree instead, but then by construction we would have $\hat{\mu}_{V|U} = 1$. In our data, $\hat{\mu}_{V|U}$ is .971 for a Bachelor's degree, .981 for an Associate degree, and 1.000 for some college. Nonparametric estimates of $\hat{r}(x) = \hat{R}(1, x) - \hat{R}(0, x)$ are obtained with $\hat{R}(d, x)$ given by equation (26) with these estimates of $\hat{\mu}_{V|U}$, and where the variable C in X is the test score, while Z is the vector of other elements of X . The first row of Table 2 contains r , the sample average of $\hat{r}(X)$, while the second row has the estimated trimmed mean r_t , which is the sample average of $\hat{r}(X)$ after removing the highest 5% and lowest 5% of $\hat{r}(X)$ in the sample. Next are the lower quartile, middle quartile (median) and upper quartile r_{q1} , r_{med} , and r_{q3} , of $\hat{r}(X)$ in the sample. The final row, "r semi, linear" is a semiparametric estimate of r using equation (30). Standard errors, reported in parentheses, are based on 400 bootstrap replications. One set of sufficient regularity conditions for bootstrapping here is Theorem B in Chen, Linton, and Van Keilegom (2003).

Table 2: Effects of Lying, Nonparametric and Semiparametric Corollary 2 IV Estimates

	Some college	Associate degree	Bachelor's degree
r nonparametric	0.07052 (0.03420)	0.1696 (0.3335)	0.1250 (1.918)
r_t nonparametric	0.07355 (0.03166)	0.1796 (0.04158)	0.07109 (0.1217)
r_{q1} nonparametric	-0.05768 (0.04930)	0.09099 (0.06185)	-0.1654 (0.1841)
r_{med} nonparametric	0.06447 (0.03663)	0.1287 (0.04903)	0.06696 (0.1003)
r_{q3} nonparametric	0.1421 (0.03903)	0.3214 (0.05156)	0.3002 (0.1596)
r semi, linear	0.08008 (0.02940)	0.1610 (0.03362)	0.05613 (1.138)

For the nonparametric estimates, the kernel function K is a standard normal density function, with bandwidth $b = 0.1836$ given by Silverman's rule. Doubling or halving this bandwidth changed

most estimates by less than 10%, indicating that the results were generally not sensitive to bandwidth choice. An exception is that mean and trimmed mean estimates for the Bachelor's degree, which are small in Table 2, become larger (closer to the median r estimate) when the bandwidth is doubled. The results for the bachelor's degree are also much less precisely estimated than for some college or associate degree, with generally more than twice as large standard errors. Based on Table 1, we might expect that far fewer individuals misreport having a bachelor's degree, so the resulting imprecision in the Bachelor's degree estimates could be due to a much smaller fraction of data points that are informative about misreporting.

The nonparametric mean and median estimates of r are significant in Table 2, except for the Bachelor's degree. Overall, these results indicate that those who misreport by claiming to have some college have about 6% to 8% higher wages than those who tell the truth about not having any college on average, and those who misreport by claiming to have an associate degree have about 13% to 18% higher wages. The point estimates for lying about having a Bachelor's degree are lower, but they also have much larger standard errors. The variability in these estimated effects is quite large, ranging from a zero or negative effect at the first quartile to effects of 14% for some college to 32% for a degree at the third quartile. The semiparametric estimates of r are similar to the mean of the nonparametric estimates, though the variation in the quantiles of the nonparametric estimates suggests that the semiparametric specification, which assumes r is constant, is not likely to hold.

If transcripts V are very accurate, then V should be close to D^* , and the estimates of r in Table 1 should be close to those in Table 2. The linear model estimates in Table 1 are close to the semiparametric linear model estimates in Table 2 (for some college and associate degrees), however, the nonparametric estimates of r in Table 1 are much larger than the mean and median nonparametric estimates in Table 2. In linear models measurement error generally causes attenuation bias, but in contrast here the potentially mismeasured data estimates appear too large rather than too small. This could be due to nonlinearity, or because the potentially mismeasured variable V is highly correlated with another regressor, D .

We should expect that the effects of lying would be smaller than the returns to actually having some college or a degree. These effects of actual schooling are not identified from the assumptions in Corollary 2 or 3. Table 1 gives estimates of the effects of schooling s ranging from 28% for some college to 32% for a bachelor's degree, though these estimates are only reliable if transcripts V are accurate. These are indeed higher than the effects of lying, as one would expect. Also, while we

would expect the effects of schooling to increase monotonically with the level of schooling, we do not necessarily expect the effects of lying to increase in the same way, because those effects depend on other factors like the plausibility of the misreport.

Kane, Rouse, and Staiger (1999) report some substantial error rates in transcripts, however, those findings are based on model estimates that could be faulty, rather than any type of direct verification. It is possible that transcripts are generally accurate, and in that case the ability of our estimator to produce reasonable estimates of r would not be impressive, since one could then just as easily generate good estimates of r using regressions or cell means as in Table 1. Therefore, to check the robustness of our methodology, we reestimated the model after randomly changing 20% of the observations of V to $1 - V$, thereby artificially making V a much weaker instrument. The resulting estimates of the mean and trimmed mean of r were generally higher than those reported in Tables 1 and 2 (consistent with our earlier result that, in our application, measurement error in V seems to raise rather than lower estimates of the effects of lying). As with the other estimates, the numbers for bachelor’s degrees are unstable with very large standard errors. However, the estimates of the median of r with this noisy V data are very close to the median estimates in table 2 (though of course with larger standard errors) for some college and associate degree. Specifically, the r_{med} estimates with substantial measurement error added to V were 0.070, 0.133, and 0.190, compared to the r_{med} estimates in Table 2 of 0.064, 0.129, and 0.067.

Table 3: Nonparametric Corollary 3 IV Effects of Lying Linearized Coefficient Estimates

X	Some college	Associate degree	Bachelor’s degree
blacknh	-0.09208 (0.1246)	-0.2429 (2.674)	-0.3735 (2.288)
hispanic	0.01220 (0.1289)	-0.1529 (1.627)	-0.1541 (1.588)
othernh	0.2176 (0.1304)	0.1444 (1.265)	0.5398 (4.763)
female	0.09291 (0.06570)	0.2306 (0.5377)	0.2876 (3.370)
mscore	-0.009755 (0.03807)	0.03345 (0.3496)	-0.09489 (2.471)
constant	0.02449 (0.04635)	0.07127 (0.2840)	0.01803 (2.900)

To summarize how $\hat{r}(x)$ varies with regressors x , Table 3 reports the estimated coefficients from linearly regressing the nonparametric estimates $\hat{r}(x)$ on x and on a constant. The results show a few interesting patterns, including that women appear to have a larger effect of (possibly indicating higher

returns to) lying than men, and that individuals with above average high school test scores also have above average effects of misreporting a higher degree of education. These results are consistent with the notion that the effects of lying should be highest for those who can lie most plausibly (e.g., those with high ability) or for those who may be perceived as less likely to lie (such as women). However, these results should not be over interpreted, since they are mostly not statistically significant.

8 Alternative Estimates Without IV

To check the robustness of our results to alternative identifying assumptions, in Table 4 we report the effects of lying using the estimator based on Theorem 2, which does not use data on the instrument V . These estimates are based only on self reports, and so do not use the transcript data in any way. For these estimates we assume equation (18) holds for x_0 equal to any value of X , which implies the assumption that that no one understates their education level by reporting $D = 0$ when $D^* = 1$ (and hence that transcripts are wrong for the few observations in the data that have $D = 0$ and $V = 1$).

Table 4: Effects of Lying, Nonparametric and Semiparametric Theorem 2 Estimates Without IV

	Some college	Associate degree	Bachelor's degree
r nonparametric	-0.4127 (28.66)	0.1917 (2.915)	0.1247 (18.27)
r_t nonparametric	0.05064 (0.1402)	0.1684 (0.1738)	0.09186 (0.2489)
r_{q1} nonparametric	-0.05096 (0.1446)	-0.1065 (0.2406)	-0.5425 (0.3659)
r_{med} nonparametric	0.1179 (0.06115)	0.1495 (0.06191)	0.1958 (0.05549)
r_{q3} nonparametric	0.2570 (0.1019)	0.2813 (0.1428)	0.3308 (0.2038)

As should be expected, the estimates in Table 4 are mostly less precise than those in Table 2, in part because they do not exploit any transcript information, and they assume no heteroskedasticity in the model error η , which may not hold in this application. They are also more variable in part because they depend on higher moments of the data, and so will be more sensitive to outliers in the first stage nonparametric estimates. Still, the estimates in Table 4 are generally consistent with those in Table 2, and in particular almost all of the differences between Tables 2 and 4 are not statistically significant. Given the substantial differences in estimators and identifying assumptions

between Corollary 3 and Theorem 2, it is reassuring that the resulting estimates are robust across the two methodologies.

In the Appendix we report the estimates of $E[R(d, X)]$ corresponding to Tables 2 and 4. As one would expect, these are generally more stable than the estimates of $E[r(X)]$ reported in Tables 2 and 4, since $r(X)$ is a difference $R(1, X) - R(0, X)$ rather than a level $R(d, X)$.

9 Conclusions

We provide identification and associated estimators for the conditional mean of an outcome Y , conditioned upon an observed discrete variable D and an unobserved discrete variable D^* . In particular, we identify the effects of lying, that is, the average difference in the mean level of Y between individuals having the unobserved $D^* = 0$ and those having $D^* = 1$ when the observed $D = 0$. Given an unconfoundedness assumption this difference in conditional means equals either the returns to lying (if misreports of D are intentional) or a placebo effect.

In our empirical application, Y is log wages, while D and D^* are self reports and actual levels of educational attainment. We find that wages are on average about 6% to 12% higher for those who lie about having some college, and from 8% to 20% higher on average for those who lie about having a college degree, relative to those who tell the truth about not having college or a diploma. Median and trimmed mean estimates appear to be more reliable and robust than estimates of raw mean returns and returns at other quantiles. Our results are about the same based on either semiparametric or nonparametric estimation, and are roughly comparable whether identification and associated estimation is based on using transcript reports as an instrument, or is based on higher moment error independence assumptions without exploiting transcript data. Our results are also robust to artificially adding a great deal of noise to the instrument.

The plausibility of our particular identifying assumptions may be debated, but we believe much of the value of this paper is in demonstrating that these effects of misreporting can be identified at all, and we expect future research will yield alternative assumptions that may be better suited to this and other applications. It would be particularly useful in the future to investigate how these results may be extended to handle confounding correlations with the unobserved treatment D^* , to obtain returns to lying without unconfoundedness assumptions.

In this application D and D^* refer to the same binary event (educational attainment), with D

a self report of D^* . However, our theorems do require having D and D^* refer to the same binary event. More generally, one could estimate the average effect of any binary treatment or choice D (e.g., exposure to a law, a tax, or an advertisement) on any outcome Y (e.g., compliance with a law, income, expenditures on a product) where the effect is averaged only over some subpopulation of interest indexed by D^* (e.g., potential criminals, the poor, or a target audience of potential buyers), and where we do not observe exactly who is in the subpopulation of interest. Our identification strategy may thereby be relevant to a wide variety of applications, not just effects of lying.

10 Appendix

Proof of Lemmas 1 and 2: Consider Lemma 2 first:

$$\begin{aligned}
Cov(D, V | X) &= E(DV | X) - E(D | X) E(V | X) \\
&= E[DE(V | D, X) | X] - E(D | X) E(V | X) \\
&= \Pr(D = 1 | X) E(V | D = 1, X) - E(D | X) E(V | X) \\
&= E(D | X) [E(V | D = 1, X) - E(V | X)]
\end{aligned}$$

so $Cov(D, V | X) \neq 0$ if and only if the right side of the above expression is nonzero. The proof of Lemma 1 works exactly the same way.

Proof of Theorem 1:

First observe that

$$\begin{aligned}
E(D^*V | D, X) &= \sum_{d^*=0}^1 \Pr(D^* = d^* | D, X) E(D^*V | D^* = d^*, D, X) \\
&= \Pr(D^* = 1 | D, X) E(V | D^* = 1, D, X) \\
&= E(D^* | D, X) E(V | D^* = 1, X)
\end{aligned}$$

and using this result we have

$$\begin{aligned}
E(YV | D, X) &= R(D, X)E(V | D, X) + s(D, X)E[D^*V | D, X] + E(\eta V | D, X) \\
&= R(D, X)E(V | D, X) + s(D, X)E(D^* | D, X) E(V | D^* = 1, X).
\end{aligned}$$

Also

$$E(Y | D, X) = R(D, X) + s(D, X)E[D^* | D, X]$$

Use the latter equation to substitute $s(D, X)E[D^* | D, X]$ out of the former equation, and solve what remains for $R(D, X)$ to obtain equation (9). Equation (10) then follows immediately from equation (9) using $r(X) = R(1, X) - R(0, X)$ and the properties of a covariance.

Proof of Theorem 2: Begin with equation (2), $Y = R(D, X) + s(D, X)D^* + \eta$ with $R(D, X) = R(X) + r(X)D$. Assumption B1 implies that

$$\begin{aligned} \mu_{Y|D,X} &\equiv E(Y|D, X) & (32) \\ &= E((R(D, X) + s(D, X)D^*) | D, X) \\ &= R(D, X) + s(D, X)E(D^* | D, X), \end{aligned}$$

$$\begin{aligned} \mu_{Y^2|D,X} &\equiv E(Y^2 | D, X) & (33) \\ &= E((R(D, X) + s(D, X)D^* + \eta)^2 | D, X) \\ &= E((R(D, X) + s(D, X)D^*)^2 | D, X) + E\eta^2 \\ &= R(D, X)^2 + 2R(D, X)s(D, X)E(D^* | D, X) + s(D, X)^2E(D^* | D, X) + E\eta^2 \\ &= R(D, X)^2 + 2R(D, X)(\mu_{Y|D,X} - R(D, X)) + s(D, X)(\mu_{Y|D,X} - R(D, X)) + E\eta^2 \\ &= \mu_{Y|D,X}R(D, X) + (R(D, X) + s(D, X))(\mu_{Y|D,X} - R(D, X)) + E\eta^2, \end{aligned}$$

and

$$\begin{aligned} \mu_{Y^3|D,X} &\equiv E(Y^3 | D, X) & (34) \\ &= E((R(D, X) + s(D, X)D^* + \eta)^3 | D, X) \\ &= E[(R(D, X) + s(D, X)D^*)^3 | D, X] + 3E[(R(D, X) + s(D, X)D^*) | D, X] E\eta^2 + E\eta^3 \\ &= R(D, X)^3 + 3R(D, X)^2s(D, X)E(D^* | D, X) \\ &\quad + 3R(D, X)s(D, X)^2E(D^* | D, X) + s(D, X)^3E(D^* | D, X) \\ &\quad + 3\mu_{Y|D,X}E\eta^2 + E\eta^3. \end{aligned}$$

We now show that assumption B1 implies the identification of $E(\eta^k)$ for $k = 2, 3$. This assumption implies that

$$\begin{aligned}
& E(D^*|D = 0, X = x_0) \\
&= \Pr(D^* = 1|D = 0, X = x_0) \\
&= \Pr(D = 0|D^* = 1, X = x_0) \frac{\Pr(D^* = 1|X = x_0)}{\Pr(D = 0|X = x_0)} \\
&= 0,
\end{aligned}$$

and therefore,

$$\begin{aligned}
\mu_{Y|0,x_0} &\equiv E(Y|D = 0, X = x_0) \\
&= R(0, x_0) + s(0, x_0)E(D^*|D = 0, X = x_0) \\
&= R(0, x_0),
\end{aligned}$$

$$\begin{aligned}
\mu_{Y^2|0,x_0} &\equiv E(Y^2|D = 0, X = x_0) \\
&= R(0, x_0)^2 + 2R(D, X)s(D, X)E(D^*|D = 0, X = x_0) \\
&\quad + s(D, X)^2E(D^*|D = 0, X = x_0) + E\eta^2 \\
&= R(0, x_0)^2 + E\eta^2 \\
&= \mu_{Y|0,x_0}^2 + E\eta^2,
\end{aligned}$$

and

$$\begin{aligned}
\mu_{Y^3|0,x_0} &= E(Y^3|D = 0, X = x_0) \\
&= R(0, x_0)^3 + 3\mu_{Y|0,x_0}E\eta^2 + E\eta^3 \\
&= \mu_{Y|0,x_0}^3 + 3\mu_{Y|0,x_0}(E\eta^2 - \mu_{Y|0,x_0}^2) + E\eta^3.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
E\eta^2 &= \mu_{Y^2|0,x_0} - \mu_{Y|0,x_0}^2 \\
&\equiv \sigma_{Y|0,x_0}^2,
\end{aligned}$$

and

$$\begin{aligned}
E\eta^3 &= \mu_{Y^3|0,x_0} + 2\mu_{Y|0,x_0}^3 - 3\mu_{Y|0,x_0}\mu_{Y^2|0,x_0} \\
&= E\left((Y - \mu_{Y|0,x_0})^3 \mid D = 0, X = x_0\right) \\
&\equiv v_{Y|0,x_0}^3.
\end{aligned}$$

In the next step, we eliminate $s(D, X)$ and $E(D^*|D, X)$ in equations 32-34 to obtain a restriction only containing $R(D, X)$ and known variables. We will use the following two equations repeatedly.

$$(R(D, X) + s(D, X))(\mu_{Y|D,X} - R(D, X)) = \mu_{Y^2|D,X} - E\eta^2 - \mu_{Y|D,X}R(D, X) \quad (35)$$

$$s(D, X)E(D^*|D, X) = \mu_{Y|D,X} - R(D, X) \quad (36)$$

Notice that

$$s(D, X) = \frac{\mu_{Y^2|D,X} - \mu_{Y|D,X}^2 - \sigma_{Y|0,x_0}^2}{\mu_{Y|D,X} - R(D, X)} + \mu_{Y|D,X} - R(D, X)$$

which also implies that we can't identify $s(0, x_0)$ because $\mu_{Y|D=0,x_0} = R(0, x_0)$.

From here on we will for clarity drop the term (D, X) when it is obvious from context. Consider

$$\begin{aligned}
\mu_{Y^3|D,X} &\equiv E(Y^3|D, X) \\
&= E((R(D, X) + s(D, X)D^* + \eta)^3 | D, X) \\
&= E((R + sD^*)^3 | D, X) + 3E((R + sD^*) | D, X) E\eta^2 + E(\eta^3) \\
&= R(D, X)^3 + 3R(D, X)^2 s(D, X)E(D^*|D, X) \\
&\quad + 3R(D, X)s(D, X)^2 E(D^*|D, X) + s(D, X)^3 E(D^*|D, X) \\
&\quad + 3[R(D, X) + s(D, X)E(D^*|D, X)] E\eta^2 + E\eta^3 \\
&= R^3 + 3R^2(\mu_{Y|D,X} - R) + 3Rs(\mu_{Y|D,X} - R) + s^2(\mu_{Y|D,X} - R) + 3\mu_{Y|D,X}E\eta^2 + E\eta^3 \\
&= R^3 + 3R^2(\mu_{Y|D,X} - R) + 2Rs(\mu_{Y|D,X} - R) + s(R + s)(\mu_{Y|D,X} - R) + 3\mu_{Y|D,X}E\eta^2 + E\eta^3 \\
&= R^3 + 3R^2(\mu_{Y|D,X} - R) + 2Rs(\mu_{Y|D,X} - R) + s(\mu_{Y^2|D,X} - E\eta^2 - \mu_{Y|D,X}R) \\
&\quad + 3\mu_{Y|D,X}E\eta^2 + E\eta^3
\end{aligned}$$

Which, with a little algebra can be written as

$$\begin{aligned}
\mu_{Y^3|D,X} &= R(\mu_{Y^2|D,X} - E\eta^2) + (R + s)(\mu_{Y^2|D,X} - E\eta^2 - \mu_{Y|D,X}R) + 3\mu_{Y|D,X}E\eta^2 + E\eta^3 \\
&= R(\mu_{Y^2|D,X} - E\eta^2) + \frac{\mu_{Y^2|D,X} - E\eta^2 - \mu_{Y|D,X}R}{(\mu_{Y|D,X} - R)}(\mu_{Y^2|D,X} - E\eta^2 - \mu_{Y|D,X}R) \\
&\quad + 3\mu_{Y|D,X}E\eta^2 + E\eta^3.
\end{aligned}$$

That is

$$\begin{aligned}
0 &= (\mu_{Y^2|D,X} - E\eta^2 - \mu_{Y|D,X}R)^2 + (\mu_{Y^2|D,X} - E\eta^2)(\mu_{Y|D,X} - R)R \\
&\quad - (\mu_{Y^3|D,X} - (3\mu_{Y|D,X}E\eta^2 + E\eta^3))(\mu_{Y|D,X} - R).
\end{aligned}$$

The restrictions on R simplify to the quadratic equation

$$-\alpha R^2 + \beta R + \gamma = 0,$$

where

$$\begin{aligned}
\alpha &= -(\mu_{Y^2|D,X}^2 - (\mu_{Y^2|D,X} - E\eta^2)), \\
\beta &= (- (\mu_{Y^2|D,X} - E\eta^2) \mu_{Y|D,X} + \mu_{Y^3|D,X} - (3\mu_{Y|D,X}E\eta^2 + E\eta^3)), \\
\gamma &= (\mu_{Y^2|D,X} - E\eta^2)^2 - (\mu_{Y^3|D,X} - (3\mu_{Y|D,X}E\eta^2 + E\eta^3)) \mu_{Y|D,X}.
\end{aligned}$$

Notice that

$$\begin{aligned}
\sigma_{Y|D,X}^2 &= \mu_{Y^2|D,X} - \mu_{Y|D,X}^2, \\
v_{Y|D,X}^3 &\equiv E\left((Y - \mu_{Y|D,X})^3 | D, X\right) \\
&= \mu_{Y^3|D,X} + 2\mu_{Y|D,X}^3 - 3\mu_{Y|D,X}\mu_{Y^2|D,X}.
\end{aligned}$$

We then simplify the expressions of α , β , and γ as follows:

$$\begin{aligned}
\alpha &= -(\mu_{Y^2|D,X}^2 - (\mu_{Y^2|D,X} - E\eta^2)) \\
&= (\sigma_{Y|D,X}^2 - \sigma_{Y|0,x_0}^2),
\end{aligned}$$

$$\begin{aligned}
\beta &= \left(-(\mu_{Y^2|D,X} - E\eta^2) \mu_{Y|D,X} + \mu_{Y^3|D,X} - (3\mu_{Y|D,X}E\eta^2 + E\eta^3) \right) \\
&= (\mu_{Y^3|D,X} - 2\mu_{Y|D,X}E\eta^2 - E\eta^3 - \mu_{Y|D,X}\mu_{Y^2|D,X}) \\
&= v_{Y|D,X}^3 - 2\mu_{Y|D,X}^3 + 3\mu_{Y|D,X}\mu_{Y^2|D,X} - 2\mu_{Y|D,X}E\eta^2 - E\eta^3 - \mu_{Y|D,X}\mu_{Y^2|D,X} \\
&= v_{Y|D,X}^3 - E\eta^3 - 2\mu_{Y|D,X}^3 - 2\mu_{Y|D,X}E\eta^2 + 2\mu_{Y|D,X}\mu_{Y^2|D,X} \\
&= v_{Y|D,X}^3 - E\eta^3 - 2\mu_{Y|D,X}^3 - 2\mu_{Y|D,X}E\eta^2 + 2\mu_{Y|D,X}(\sigma_{Y|D,X}^2 + \mu_{Y|D,X}^2) \\
&= v_{Y|D,X}^3 - E\eta^3 + 2\mu_{Y|D,X}(\sigma_{Y|D,X}^2 - E\eta^2) \\
&= v_{Y|D,X}^3 - v_{Y|0,x_0}^3 + 2\mu_{Y|D,X}(\sigma_{Y|D,X}^2 - \sigma_{Y|0,x_0}^2) \\
&= v_{Y|D,X}^3 - v_{Y|0,x_0}^3 + 2\mu_{Y|D,X}\alpha,
\end{aligned}$$

$$\begin{aligned}
\gamma &= (\mu_{Y^2|D,X} - E\eta^2)^2 - (\mu_{Y^3|D,X} - (3\mu_{Y|D,X}E\eta^2 + E\eta^3)) \mu_{Y|D,X} \\
&= (\sigma_{Y|D,X}^2 + \mu_{Y|D,X}^2 - E\eta^2)^2 - (\mu_{Y^3|D,X} - (3\mu_{Y|D,X}E\eta^2 + E\eta^3)) \mu_{Y|D,X} \\
&= \mu_{Y|D,X}^4 + 2\mu_{Y|D,X}^2(\sigma_{Y|D,X}^2 - E\eta^2) + (\sigma_{Y|D,X}^2 - E\eta^2)^2 \\
&\quad - \mu_{Y^3|D,X}\mu_{Y|D,X} + 3\mu_{Y|D,X}^2E\eta^2 + \mu_{Y|D,X}E\eta^3 \\
&= \mu_{Y|D,X}^4 + 2\mu_{Y|D,X}^2\sigma_{Y|D,X}^2 + (\sigma_{Y|D,X}^2 - E\eta^2)^2 - \mu_{Y^3|D,X}\mu_{Y|D,X} + \mu_{Y|D,X}^2E\eta^2 + \mu_{Y|D,X}E\eta^3 \\
&= \mu_{Y|D,X}^4 + 2\mu_{Y|D,X}^2\sigma_{Y|D,X}^2 + (\sigma_{Y|D,X}^2 - E\eta^2)^2 \\
&\quad - (v_{Y|D,X}^3 - 2\mu_{Y|D,X}^3 + 3\mu_{Y|D,X}\mu_{Y^2|D,X}) \mu_{Y|D,X} + \mu_{Y|D,X}^2E\eta^2 + \mu_{Y|D,X}E\eta^3 \\
&= \mu_{Y|D,X}^4 + 2\mu_{Y|D,X}^2\sigma_{Y|D,X}^2 + (\sigma_{Y|D,X}^2 - E\eta^2)^2 \\
&\quad + 2\mu_{Y|D,X}^4 - 3\mu_{Y|D,X}^2\mu_{Y^2|D,X} + \mu_{Y|D,X}^2E\eta^2 + \mu_{Y|D,X}(E\eta^3 - v_{Y|D,X}^3) \\
&= \mu_{Y|D,X}^4 + 2\mu_{Y|D,X}^2\sigma_{Y|D,X}^2 + (\sigma_{Y|D,X}^2 - E\eta^2)^2 \\
&\quad + 2\mu_{Y|D,X}^4 - 3\mu_{Y|D,X}^2(\sigma_{Y|D,X}^2 + \mu_{Y|D,X}^2) + \mu_{Y|D,X}^2E\eta^2 + \mu_{Y|D,X}(E\eta^3 - v_{Y|D,X}^3) \\
&= (\sigma_{Y|D,X}^2 - E\eta^2)^2 - \mu_{Y|D,X}^2(\sigma_{Y|D,X}^2 - E\eta^2) - \mu_{Y|D,X}(v_{Y|D,X}^3 - E\eta^3) \\
&= (\sigma_{Y|D,X}^2 - \sigma_{Y|0,x_0}^2)^2 - \mu_{Y|D,X}^2(\sigma_{Y|D,X}^2 - \sigma_{Y|0,x_0}^2) - \mu_{Y|D,X}(v_{Y|D,X}^3 - v_{Y|0,x_0}^3) \\
&= \alpha^2 - \mu_{Y|D,X}^2\alpha - \mu_{Y|D,X}(\beta - 2\mu_{Y|D,X}\alpha) \\
&= \alpha^2 + \mu_{Y|D,X}^2\alpha - \mu_{Y|D,X}\beta.
\end{aligned}$$

In summary, we have

$$-\alpha R^2 + \beta R + \gamma = 0$$

$$\begin{aligned}
\alpha &= \sigma_{Y|D,X}^2 - \sigma_{Y|0,x_0}^2 \\
\beta &= v_{Y|D,X}^3 - v_{Y|0,x_0}^3 + 2\mu_{Y|D,X}\alpha \\
\gamma &= \alpha^2 + \mu_{Y|D,X}^2\alpha - \mu_{Y|D,X}\beta
\end{aligned}$$

That means

$$R = \frac{\beta + \sqrt{\beta^2 + 4\alpha\gamma}}{2\alpha} \text{ or } \frac{\beta - \sqrt{\beta^2 + 4\alpha\gamma}}{2\alpha}.$$

In fact, we may show that equations 36 and 35 implies

$$\alpha \geq 0$$

Consider

$$\begin{aligned}
s &= \frac{\mu_{Y^2|D,X} - \mu_{Y|D,X}^2 - E\eta^2}{\mu_{Y|D,X} - R} + \mu_{Y|D,X} - R \\
&= \frac{\alpha}{\mu_{Y|D,X} - R} + \mu_{Y|D,X} - R
\end{aligned}$$

and

$$\begin{aligned}
E(D^*|D, X) &= \frac{\mu_{Y|D,X} - R}{s} \\
&= \frac{(\mu_{Y|D,X} - R)^2}{(\mu_{Y|D,X} - R)^2 + \alpha}.
\end{aligned}$$

Therefore, $0 \leq E(D^*|D, X) \leq 1$ implies that $\alpha \geq 0$.

The last step is to eliminate one of the two roots to achieve point identification. Notice that

$$E(Y|D^*, D, X) = R(D, X) + s(D, X)D^*.$$

Assumption B1 implies that

$$s(D, X) \geq 0.$$

Consider

$$\begin{aligned}
\mu_{Y|D,X} &= R + sE(D^*|D, X) \\
&= R[1 - E(D^*|D, X)] + (R + s)E(D^*|D, X).
\end{aligned}$$

Therefore, $0 \leq E(D^*|D, X) \leq 1$ and $s(D, X) \geq 0$ imply

$$R \leq \mu_{Y|D,X} \leq s + R,$$

Thus, we may identify R as the smaller root if $\mu_{Y|D,X}$ is between the two roots. , i.e.,

$$-\alpha\mu_{Y|D,X}^2 + \beta\mu_{Y|D,X} + \gamma \geq 0,$$

which holds because

$$\begin{aligned} & -\alpha\mu_{Y|D,X}^2 + \beta\mu_{Y|D,X} + \gamma \\ = & -\alpha\mu_{Y|D,X}^2 + \beta\mu_{Y|D,X} + \alpha^2 + \mu_{Y|D,X}^2\alpha - \mu_{Y|D,X}\beta \\ = & \alpha^2 \geq 0. \end{aligned}$$

Therefore, we have

$$R(D, X) = \frac{\beta - \sqrt{\beta^2 + 4\alpha\gamma}}{2\alpha}.$$

Notice that R equals the larger root if $s(D, X) \leq 0$. The function $s(D, X)$ then follows.

Discrete Limiting Distributions for equation (15). Let

$$\begin{aligned} \hat{\alpha}(x) &= \left(\hat{\mu}_{Y,V,X,1}, \hat{\mu}_{Y,V,X,0}, \hat{\mu}_{Y,X,1}, \hat{\mu}_{Y,X,0}, \hat{\mu}_{V,X,1}, \hat{\mu}_{V,X,0}, \hat{\mu}_{X,1}, \hat{\mu}_{X,0}, \hat{\mu}_{VU}, \hat{\mu}_U \right)^T, \\ \alpha_0 &= E[\hat{\alpha}(x)], \\ \hat{R}(d, \hat{\alpha}(x)) &\equiv \frac{\left(\hat{\mu}_{Y,V,X,1}^d \hat{\mu}_{Y,V,X,0}^{1-d} \right) \hat{\mu}_U - \left(\hat{\mu}_{Y,X,1}^d \hat{\mu}_{Y,X,0}^{1-d} \right) \hat{\mu}_{VU}}{\left(\hat{\mu}_{V,X,1}^d \hat{\mu}_{V,X,0}^{1-d} \right) \hat{\mu}_U - \left(\hat{\mu}_{X,1}^d \hat{\mu}_{X,0}^{1-d} \right) \hat{\mu}_{VU}}, \\ \hat{r}(x) &= \hat{R}(1, \hat{\alpha}(x)) - \hat{R}(0, \hat{\alpha}(x)), \end{aligned}$$

$$\begin{aligned} \gamma &= \left. \frac{\partial}{\partial t} R(d, \alpha_0 + t(\hat{\alpha} - \alpha_0)) \right|_{t=0} \\ &\equiv G(d, \alpha_0)^T (\hat{\alpha} - \alpha_0), \end{aligned}$$

$$V(\hat{\alpha}(x)) = n \times E \left[(\hat{\alpha} - \alpha_0) (\hat{\alpha} - \alpha_0)^T \right].$$

Assuming independent, identically distributed draws and existence of $V(\hat{\alpha}(x))$, by the Lindeberg-Levy central limit theorem and the delta method

$$\begin{aligned}\sqrt{n} \left[\widehat{R}(d, x) - R(d, x) \right] &\rightarrow {}^d N(0, \Omega_R) \\ \Omega_R &= G(d, \alpha_0(x))^T V(\hat{\alpha}(x)) G(d, \alpha_0(x))\end{aligned}$$

and

$$\begin{aligned}\sqrt{n} [\widehat{r}(x) - r(x)] &\rightarrow {}^d N(0, \Omega_r) \\ \Omega_r &= [G(1, \alpha_0(x)) - G(0, \alpha_0(x))]^T V(\hat{\alpha}(x)) [G(1, \alpha_0(x)) - G(0, \alpha_0(x))].\end{aligned}$$

Table 5: R(0,X), Nonparametric and Semiparametric Corollary 3 IV Estimates

	Some college	Associate degree	Bachelor's degree
R0 nonparametric	2.072 (0.01514)	2.125 (0.009665)	2.143 (0.007997)
R0 _t nonparametric	2.065 (0.01536)	2.125 (0.01016)	2.144 (0.008579)
R0 _{q1} nonparametric	1.863 (0.02520)	1.939 (0.01940)	1.975 (0.01834)
R0 _{med} nonparametric	2.003 (0.03859)	2.089 (0.03225)	2.143 (0.02788)
R0 _{q3} nonparametric	2.309 (0.02681)	2.326 (0.01763)	2.319 (0.01665)
R0 semi, linear	2.025 (0.01174)	2.094 (0.008754)	2.114 (0.007451)

Table 6: R(1,X), Nonparametric and Semiparametric Corollary 3 IV Estimates

	Some college	Associate degree	Bachelor's degree
R1 nonparametric	2.142 (0.03011)	2.295 (0.3326)	2.268 (1.918)
R1 _t nonparametric	2.152 (0.02986)	2.319 (0.04103)	2.223 (0.1219)
R1 _{q1} nonparametric	1.997 (0.04430)	2.181 (0.06094)	2.092 (0.1694)
R1 _{med} nonparametric	2.173 (0.04633)	2.380 (0.04084)	2.189 (0.1026)
R1 _{q3} nonparametric	2.340 (0.04635)	2.449 (0.04508)	2.397 (0.1731)
R1 semi, linear	2.188 (0.02898)	2.341 (0.03397)	2.267 (1.149)

Table 7: $R(0,X)$, Nonparametric and Semiparametric Theorem 2 Estimates Without IV

	Some college	Associate degree	Bachelor's degree
R0 nonparametric	2.078 (0.01383)	2.126 (0.009571)	2.144 (0.007897)
$R0_t$ nonparametric	2.074 (0.01447)	2.123 (0.01025)	2.148 (0.008459)
$R0_{q1}$ nonparametric	1.891 (0.02270)	1.942 (0.01916)	1.974 (0.01820)
$R0_{med}$ nonparametric	2.022 (0.03761)	2.095 (0.03227)	2.146 (0.02767)
$R0_{q3}$ nonparametric	2.288 (0.02247)	2.321 (0.01708)	2.324 (0.01620)

Table 8: $R(1,X)$, Nonparametric and Semiparametric Theorem 2 Estimates Without IV

	Some college	Associate degree	Bachelor's degree
R1 nonparametric	1.666 (28.66)	2.318 (2.915)	2.269 (18.27)
$R1_t$ nonparametric	2.141 (0.1418)	2.310 (0.1719)	2.227 (0.2483)
$R1_{q1}$ nonparametric	1.832 (0.1459)	2.069 (0.2132)	1.525 (0.3052)
$R1_{med}$ nonparametric	2.223 (0.07273)	2.247 (0.08727)	2.222 (0.07330)
$R1_{q3}$ nonparametric	2.419 (0.09065)	2.501 (0.1239)	2.552 (0.2033)

References

- [1] AI, C. AND X. CHEN (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1844.
- [2] AIGNER, D. J. (1973), "Regression With a Binary Independent Variable Subject to Errors of Observation," *Journal of Econometrics*, 1, 249-60.
- [3] ASHENFELTER, O. AND A. KRUEGER (1994), "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review*, 84, 1157-1173
- [4] BLUNDELL, R. AND J. L. POWELL, (2004), "Endogeneity in Semiparametric Binary Response Models" *Review of Economic Studies*, 71, 655-679.

- [5] BOLLINGER, C. R. (1996), "Bounding Mean Regressions When a Binary Regressor is Mismeasured," *Journal of Econometrics*, 73, 387-399.
- [6] CARD, D. (1996), "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64, 957-979.
- [7] CHEN, X., Y. HU AND A. LEWBEL, (2008a), "Nonparametric Identification of Regression Models Containing a Misclassified Dichotomous Regressor Without Instruments," *Economics Letters*, 100, 381-384.
- [8] CHEN, X., Y. HU AND A. LEWBEL, (2008b), "A Note on the Closed-form Identification of Regression Models with a Mismeasured Binary Regressor," *Statistics and Probability Letters*, 78, 1473-1479.
- [9] CHEN, X., O. LINTON, AND I. VAN KEILEGOM, (2003) "Estimation of Semiparametric Models when the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591-1608,
- [10] DAS, M., (2004), "Instrumental Variables Estimators of Nonparametric Models With Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335-361.
- [11] FLORENS, J.-P. AND L. MALAVOLTI, (2003), "Instrumental Regression with Discrete Variables," unpublished manuscript.
- [12] GIBRAT, R. (1931), *Les Inegalites Economiques*, Librairie du Recueil Sirey, Paris
- [13] HECKMAN, J. J. (1979), "Sample selection bias as a specification error," *Econometrica*, 47, 153-161.
- [14] HECKMAN, J. J., H. ICHIMURA AND P. TODD, (1998), "Matching as an Econometric Evaluations Estimator," *Review of Economic Studies*, 65, 261-294.
- [15] HOTZ, V. J., C. MULLIN, AND S. SANDERS, (1997), "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing," *Review of Economic Studies*, 64, 575-603.

- [16] HU, Y. (2008), "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, vol. 144, issue 1, pages 27-61.
- [17] KANE, T. J., AND C. E. ROUSE, (1995), "Labor market returns to two- and four- year college," *American Economic Review*, 85, 600-614
- [18] KANE, T. J., C. E. ROUSE, AND D. STAIGER, (1999), "Estimating Returns to Schooling When Schooling is Misreported," NBER working paper #7235.
- [19] KLEPPER, S., (1988), "Bounding the Effects of Measurement Error in Regressions Involving Dichotomous Variables," *Journal of Econometrics*, 37, 343-359.
- [20] LEWBEL, A., (2007a), "Estimation of Average Treatment Effects With Misclassification," *Econometrica*, 75, 537-551.
- [21] LEWBEL, A., (2007b), "A Local Generalized Method of Moments Estimator," *Economics Letters*, 94, 124-128.
- [22] MAHAJAN, A. (2006) "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74, 631-665.
- [23] MANSKI, C. F. (1990) "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings*, 80, 319-323.
- [24] NEWEY, W. K., (1994), "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233-253.
- [25] NEWEY, W. K. AND J. L. POWELL, (2003), "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565-1578.
- [26] MOLINARI, F. (2008), "Partial Identification of Probability Distributions with Misclassified Data," *Journal of Econometrics*, 144, 81-117.
- [27] ROSENBAUM, P. AND D. RUBIN, (1984), "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, 516-524.

- [28] RUBIN, D. B. (1974), “Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies, ” *Journal of Educational Psychology*, 76, 688-701.