Identifying the Returns to Lying When the Truth is Unobserved*

Yingyao Hu — Johns Hopkins University

Arthur Lewbel — Boston College

original April, 2007, revised Nov. 2007

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 returns to lying to be about 7% 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.

*We would like to thank Xiaohong Chen for her help on this paper. We also thank participants of UCL, IFS, CEMMAP, BC, Montreal, and Brown seminars for helpful comments, and Douglas Staiger for providing data. All errors are our own.

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 identify the unknown functions R and r, even though D^* is unobserved.

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.

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 econometrican 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 willing to lie about 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 lying.

The interpretation of r as a placebo effect 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), and does not require unconfoundedness, 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^* and refer to r as the returns to lying.

Discreteness of D and D^* is also not essential for this paper's identification method, but will simplify the associated estimators. 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.

When D is a possibly mismeasured or misclassified observation of D^*, then $D - D^*$ is the measurement or misclassification error. Most of the literature on mismeasured binary regressors attempts to estimate the effect of D^* on Y (a treatment effect) and assumes $r(X) = 0$, or equivalently, that the measurement error has no effect on the outcome Y after conditioning on the true D^*. Recent examples include Hu (2006), Mahajan (2006), Lewbel (2007a), and Chen, Hu, and Lewbel (2007). 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 $r(X) = 0$ will be reasonable if the reporting errors $D - D^*$ are due to data collection errors such as accidently 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 many of the existing treatment or mismeasured or misclassified regressor estimators which all require that $r(X) = 0$.

We propose two different methods to obtain 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 be 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.

The next two sections describe identification with and without an instrument. We then propose estimators based on the 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*

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

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

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

The following Lemmas are useful for interpreting and applying Assumption A2:

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

$$Cov\left(D, V \mid D^* = 1, X\right) = 0 \tag{7}$$

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

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

Proofs of Lemmas and Theorems are in the Appendix. 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 $E(Y \mid D^*, D, X, V) = E(Y \mid D^*, D, X)$. This is a standard property for an instrument.

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.

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

$$R(D, X) = \frac{E\left(YV \mid D, X\right) - E\left(Y \mid D, X\right) E\left(V \mid D^* = 1, X\right)}{E\left(V \mid D, X\right) - E\left(V \mid D^* = 1, X\right)}. \tag{9}$$

It follows immediately from Theorem 1 that $R(D, X)$ is identified if $E\left(V \mid D, X\right) \neq E\left(V \mid D^* = 1, X\right)$ to avoid division by zero and if $E\left(V \mid D^* = 1, X\right)$ can be identified, because the other terms in equation (9) are expectations of observables, conditioned on other observables, and hence are themselves identified.

We now consider two alternative methods of satisfying these conditions needed to identify $R(D, X)$.

ASSUMPTION A3: *Assume*

$$E\left(V \mid D^{*}=1, X\right)=1, \tag{10}$$

and

$$E\left(V \mid D, X\right) \neq 1 \tag{11}$$

Note that if $V \in \{0, 1\}$ (as is the case when V is a mismeasure of D^{*}, like when V is the transcript report) then equation (10) is equivalent to $\Pr\left(V=1 \mid D^{*}=1, X\right) = 1$. This equation (10) rules out transcript errors of the form $V = 0$ when $D^{*} = 1$, and therefore requires that only one type of transcript error be possible, namely, $V = 1$ when $D^{*} = 0$. For example, if D and D^{*} refer to graduating from college then equation (10) says that anyone who has a diploma will have an accurate transcript, but people who did not graduate may have transcript errors.

Equation (11) requires that there not exist a value of D, X that always yields $V = 1$, or more precisely, that if such a D, X exists, then we cannot identify $R\left(D, X\right)$ for that D, X, since for those people we will not observe any variation in the instrument V. Equation (11) is empirically testable since it depends only on observables.

COROLLARY 1: *If Assumptions A1, A2, and A3 hold then* $R(D, X)$ *is identified by*

$$R(D, X)=\frac{E\left(YV \mid D, X\right)-E\left(Y \mid D, X\right)}{E\left(V \mid D, X\right)-1}=\frac{E\left(Y\left(V-1\right) \mid D, X\right)}{E\left(\left(V-1\right) \mid D, X\right)} \tag{12}$$

Corollary 1 follows from Theorem 1 by substituting equation (10) into equation (9). Equation (11) then ensures that the denominator of equation (12) is nonzero.

Equation (10), which in the wage application rules out one kind of transcript error, may be overly strong. We now consider an alternative assumption and associated identification that does not require this restriction.

ASSUMPTION A3': *There exists an observed binary* $U \in \{0, 1\}$ *(having* U *be an element or subset of* X *is permitted but not required) such that*

$$E\left(V \mid D^{*}=1, X\right)=E\left(V \mid U=1\right) \tag{13}$$

and

$$E(V \mid D, X) \neq E(V \mid U = 1) \tag{14}$$

Equation (13) assumes that V has the same mean for people who have $U = 1$ as for people that have $D^* = 1$ and any value of X. One set of sufficient conditions for equation (13) is if $E(V \mid D^* = 1, X, U) = E(V \mid D^* = 1)$, so for people having college ($D^* = 1$), the probability of a transcript error is unrelated to one's personal attribute information X and U, and if

$$\Pr\left(D^* = 1 \mid U = 1\right) = 1, \tag{15}$$

so people who have $U = 1$ are an observable subpopulation (e.g., medical doctors or PhD's) that definitely have some college. If equation (15) holds then equation (13) would only be violated if colleges systematically made more errors when producing transcripts for individuals with some value of attributes X, U than for students with other attribute values.

Equation (14) is a technicality that, like equation (11) in Corollary 1, will avoid division by zero in Corollary 2 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 (13) and (15) hold then equation (14) will not hold for values x_0 such that $\Pr(U = 1 \mid X = x_0) = 1$. This means that $R(D, X)$ cannot be identified for $X = x_0$, which is logical because all members of subgroup x_0 have $U = 1$ which then means they have $D^* = 1$ by equation (15), and therefore none of them can be lying when reporting $D = 1$.

COROLLARY 2: *If Assumptions A1, A2, and A3' hold then* $R(D, X)$ *is identified by*

$$R(D, X) = \frac{E(YV \mid D, X) - E(Y \mid D, X)\, E(V \mid U = 1)}{E(V \mid D, X) - E(V \mid U = 1)}. \tag{16}$$

Corollary 2 follows Theorem 1, by substituting equation (13) into equation (9) to obtain equation (16), and equation (14) makes the denominator in equation (16) be nonzero.

Given identification of $R(D, X)$ by Corollary 1 or 2, the returns to 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 1 and 2, 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. The method of identification here is similar to that of Chen, Hu, and Lewbel (2007), 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 $r(X) = 0$.

ASSUMPTION B2:

$$E\left(\eta \mid D^*, D, X\right) = 0, \tag{17}$$

$$E\left(\eta^k \mid D^*, D, X\right) = E\left(\eta^k\right) \quad \text{for } k = 2, 3, \tag{18}$$

there exists an x_0 *such that*

$$\Pr\left(D = 0 \mid D^* = 1, X = x_0\right) = 0 \quad \text{and} \quad \Pr\left(D = 0 \mid X = x_0\right) > 0, \tag{19}$$

and

$$E\left(Y \mid D^* = 1, D, X\right) \geq E\left(Y \mid D^* = 0, D, X\right) \tag{20}$$

Equation (17) can be assumed to hold without loss of generality by definition of the model error η. Equation (18) 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 (19) 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 (20) 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

$$\sigma^2_{Y|D,X}(D,X) = E\left(Y^2|D,X\right) - [E(Y|D,X)]^2,$$
$$\upsilon^3_{Y|D,X}(D,X) = E\left([Y - E(Y|D,X)]^3|D,X\right),$$

$$\alpha(D,X) = \sigma^2_{Y|D,X}(D,X) - \sigma^2_{Y|D,X}(0,x_0),$$
$$\beta(D,X) = \upsilon^3_{Y|D,X}(D,X) - \upsilon^3_{Y|D,X}(0,x_0) + 2E(Y|D,X)\,\alpha(D,X),$$
$$\gamma(D,X) = \alpha(D,X)^2 + [E(Y|D,X)]^2\,\alpha(D,X) - E(Y|D,X)\,\beta(D,X).$$

THEOREM 2: *Suppose that Assumptions A1 and B2 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 returns to 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.

The proof of Theorem 2 shows that $R(D,X)$ satisfies a quadratic equation, and equation (20) is only needed to identify which of the two roots is correct.

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 (21), interpreting D as a treatment. Letting $Y(d)$ denote what Y equals given the response $D = d$, if

$$E[Y(d) \mid D, D^* = 0, X] = E[Y(d) \mid D^* = 0, X] \tag{21}$$

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 $r(X)$ based on Corollaries 1 and 2 of Theorem 1. We first describe nonparametric estimation 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

When X is discrete, replacing the expectations in equation (16) with sample averages gives the estimators

$$\widehat{R}(d,x) = \frac{\widehat{\mu}_{Y,V,X,d} - \widehat{\mu}_{Y,X,d}\widehat{\mu}}{\widehat{\mu}_{V,X,d} - \widehat{\mu}_{X,d}\widehat{\mu}}, \quad \widehat{r}(x) = \widehat{R}(1,x) - \widehat{R}(0,x). \tag{22}$$

with

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

Estimation based on equation (12) is the same replacing $\widehat{\mu}$ with the number one in equation (22)

We also consider the unconditional mean returns $R_d = E[R(d, X)]$ and unconditional average returns to lying $r = E[r(X)]$, which may be estimated by

$$\widehat{R}_d = \frac{1}{n}\sum_{i=1}^{n}\widehat{R}(d, X_i), \qquad \widehat{r} = \frac{1}{n}\sum_{i=1}^{n}\widehat{r}(X_i). \tag{23}$$

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 nonindependent data generating processes, as long as a central limit theorem still applies.

5.2 General Nonparametric Estimation

Letting $\mu = E(V \mid U = 1)$, equation (16) can be rewritten as

$$R(D, X) = \frac{E[Y(V-\mu) \mid D, X]}{E[(V-\mu) \mid D, X]}. \tag{24}$$

Equation (12) can also be written in the form of equation (24) by letting $\mu = 1$.

Assume n independent, identically distributed draws of $\{Y_i, V_i, X_i, D_i, U_i\}$. Let $\widehat{\mu} = \widehat{\mu}_{V,U}/\widehat{\mu}_U$ if estimation is based on equation (16), otherwise let $\widehat{\mu} = 1$ if estimation is based on equation (12). 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 $x = (z, c)$. Based on equation (24), a kernel based estimator for $R(D, X)$ is

$$\widehat{R}(d, x) = \frac{\Sigma_{i=1}^{n} Y_i (V_i - \widehat{\mu}) K[(C_i = c)/b] I(Z_i = z) I(D_i = d)}{\Sigma_{i=1}^{n} (V_i - \widehat{\mu}) K[(C_i = c)/b] I(Z_i = z) I(D_i = d)} \tag{25}$$

where K is a kernel function and b is a bandwidth that goes to zero as n goes to infinity. Equation (25) is numerically identical to the ratio of two ordinary nonparametric Nadaraya-Watson kernel regressions of $Y(V - \widehat{\mu})$ and $V - \widehat{\mu}$ 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 used the constant μ in place of the estimator $\widehat{\mu}$, because $\widehat{\mu}$ either equals the constant one, or it converges at the rate root n by the law of large numbers. Alternatively, equation (24) can be rewritten as the conditional moment condition

$$E[(Y - R(D, X))(V - \mu) \mid D, X] = 0 \tag{26}$$

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 $\widehat{R}(d,x)$ from equation (25) we may as before construct $\widehat{r}(x) = \widehat{R}(1,x) - \widehat{R}(0,x)$, and unconditional returns $\widehat{R}_d$ and $\widehat{r}$ by equation (23). We also construct trimmed unconditional returns $\widehat{r}_t = \frac{1}{n}\sum_{i=1}^{n}\widehat{r}(X_i)I_{ti}$ and similarly for $\widehat{R}_{dt}$, 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 returns are root n consistent and asymptotically normal of 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 (26), θ and μ could be jointly estimated based on Corollary 2 by applying GMM to the moments

$$E\left[(V-\mu)\,U\right] = 0 \tag{27}$$

$$E\left[\psi\,(D,X)\,(Y - R(D,X,\theta))\,(V-\mu)\right] = 0 \tag{28}$$

for a chosen vector of functions $\psi\,(D,X)$. For estimation based on Corollary 1, the estimator would just use the moments given by equation (28) with $\mu = 1$.

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

$$\widehat{\theta} = \left[\sum_{i=1}^{n}(V_i - \widehat{\mu})\,W_iW_i'\right]^{-1}\left[\sum_{i=1}^{n}(V_i - \widehat{\mu})\,W_iY_i\right] \tag{29}$$

based on Corollary 2, or the same expression with $\widehat{\mu} = 1$ based on Corollary 1. Given $\widehat{\theta}$ we then have $\widehat{R}(D,X) = W'\widehat{\theta}$. In this semiparametric specification $r(x)$ is a constant with $\widehat{r}(x) = \widehat{r} = \widehat{\theta}_1$, the first element of $\widehat{\theta}$. Note that both GMM based on equation (28) and the special case of weighted linear regression based on equation (29) 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

$$\widehat{E}\left(Y^k | D = d, X = x\right) = \frac{\Sigma_{i=1}^n Y_i^k K[(C_i = c)/b] I(Z_i = z) I(D_i = d)}{\Sigma_{i=1}^n K[(C_i = c)/b] I(Z_i = z) I(D_i = d)} \tag{30}$$

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

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

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

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

$$\widehat{R}(0, x_0) = \widehat{E}\left(Y | D = 0, X = x_0\right),$$

$$\widehat{R}(d, x) = \frac{\widehat{\beta}(d, x) - \sqrt{\widehat{\beta}(d, x)^2 + 4\widehat{\alpha}(d, x)\widehat{\gamma}(d, x)}}{2\widehat{\alpha}(d, x)} \quad \text{for} \quad (d, x) \neq (0, x_0).$$

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 $\widehat{R}(d, x)$ from equation (25) we may as before construct $\widehat{r}(x) = \widehat{R}(1, x) - \widehat{R}(0, x)$, and unconditional returns $\widehat{R}_d$ and $\widehat{r}$ by equation (23). Also as before, Root n consistent, asymptotically normal convergence of trimmed means of $\widehat{R}_d$ and $\widehat{r}$ is possible using regularity conditions as in Newey (1994) for two step plug in estimators.

7 Returns to Lying about College

Here we report results of empirically implementing our estimators of $r(x)$ where D is self reports of schooling and Y is log wages. We will for convenience refer to these results as returns to lying, but strong caveats are required for that interpretation. First, we are only estimating conditional means, so our results fail to control for the selection effects that lie at the heart of the modern literature on wages and schooling going back at least to Heckman (1979). In addition, unconfoundedness with respect to lying based on equation (21) may not hold for the reasons listed in the introduction. Finally, our sample is likely to not be representative of the general population. It is therefore safest to interpret the estimates here as simply difference in means between truth tellers and liars for a limited sample, rather than as formal returns to lying.

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 returns to 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 mean of U is 0.03468, so about 3.5% of our sample have transcripts indicating graduate degrees, and the average log wage Y is 2.228.

If D^* were observed along with Y and D, then the functions $r(x)$ and $s(d, x)$ could be immediately estimated from equation 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, $D * V$, and X, and so is another estimate of r that would be valid if 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 returns to schooling $s = E[s(D, X)]$ (or more precisely, the difference in conditional means of log wages between those with $D^* = 1$, versus those with $D^* = 0$, which is 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)] gives 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 lie about having 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.

Table 1: Returns to 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

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 Estimates Based on Corollary 1

Table 2 summarizes estimates of $r(x)$ based on Corollary 1. Nonparametric estimates of $\widehat{r}(x) = \widehat{R}(1, x) - \widehat{R}(0, x)$ are obtained with $\widehat{R}(d, x)$ given by equation (25) with $\widehat{\mu}_{V|U} = 1$, where the variable C in X is the test score, and Z is the vector of other elements of X. The first row of Table 2 contains r, the sample average of $\widehat{r}(X)$, while the second row has the estimated trimmed mean r_t, which is the sample average of $\widehat{r}(X)$ after removing the highest 5% and lowest 5% of $\widehat{r}(X)$ in the sample. Next are the lower quartile, middle quartile (median) and upper quartile r_{q1}, r_{med}, and r_{q3}, of $\widehat{r}(X)$ in the sample. The final row, "r semi, linear" is a semiparametric estimate of r using equation (29). As before, standard errors 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).

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 bachelor's degree are also much less precisely estimated than for some college or associate degree, with generally twice as large standard errors. Based on Table 1, we might expect that far fewer individuals lie about 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 lying.

The nonparametric mean and median estimates of r are generally significant in Table 2, except for the mean estimates for the Bachelor's degree. Overall, these results indicate that those who lie by claiming to have have some college have about 7% higher wages than those who tell the truth about not having any college on average, and those who lie by claiming to have an associate or bachelor's degree have about 18% higher wages. However, the variability in these returns is large, ranging from zero or negative returns at the first quartile to returns of 14% for some college to 31% 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.

Recall that estimation based on Corollary 1 assumes no observations with $D^* = 1$ and $D = 0$. If transcripts V are very accurate, then V should be close to D^*, so $E[(1 - D)V]$ in Table 1 should be close to zero, and the estimates of r in Table 1 should be close to those in Table 2. The evidence on this is mixed. $E[(1 - D)V]$ is close to zero for the two types of degrees, but less so for "some college." The linear model estimates in Table 1 are close to the semiparametric linear model estimates in Table 1, however, the nonparametric estimates of r in Table 1 are rather 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 returns to lying would be smaller than the returns of actually having some college or a degree. These returns to actual schooling are not identified from the assumptions in Corollary 1 or 2. Table 1 gives estimates of returns to schooling s of 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 returns to lying, as one would expect. Also, while we would expect the returns to schooling to increase monotonically with the level of schooling, we do not necessarily expect the returns to lying to increase in the same way, because those returns depend on other factors like the plausibility of the lie.

Table 2: Returns to Lying, Nonparametric and Semiparametric Corollary 1 IV Estimates

	Some college	Associate degree	Bachelor's degree
r nonparametric	0.07051 (0.03420)	0.1759 (0.02998)	0.04140 (0.07665)
r_t nonparametric	0.07355 (0.03166)	0.1896 (0.02894)	0.09440 (0.07203)
r_{q1} nonparametric	-0.05768 (0.04930)	0.09691 (0.04219)	-0.04684 (0.09336)
r_{med} nonparametric	0.06447 (0.03663)	0.1992 (0.03995)	0.1748 (0.05683)
r_{q3} nonparametric	0.1421 (0.03903)	0.3111 (0.03920)	0.2478 (0.06094)
r semi, linear	0.08008 (0.02940)	0.1702 (0.02668)	0.1281 (0.04353)

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. Based on our empirical results comparing Tables 1 and 2, 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 returns to lying), but 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 much larger standard errors). Specifically, the r_{med} estimates with substantial measurement error added to V were 0.070, 0.190, and 0.170, compared to the r_{med} estimates in Table 2 of 0.064, 0.199, and 0.175.

To summarize how $\widehat{r}(x)$ varies with regressors x, Table 3 reports the estimated coefficients from linearly regressing the nonparametric estimates $\widehat{r}(x)$ on x and on a constant. The results show a few interesting patterns, including that women appear to have a higher return to lying than men, and that for individuals with above average high school test scores also have above average returns to lying about a higher degree of education. These results are consistent with the notion that returns to lying should be highest for those 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 not particularly stable and many are not statistically insignificant.

Table 3: Nonparametric Corollary 1 IV Returns to Lying Linearized Coefficient Estimates

X	Some college	Associate degree	Bachelor's degree
blacknh	-0.09208 (0.1246)	-0.1521 (0.1114)	0.04640 (0.2464)
hispanic	0.01220 (0.1289)	-0.1492 (0.1968)	0.05146 (0.5439)
othernh	0.2176 (0.1304)	0.03830 (0.1844)	-0.005045 (0.4755)
female	0.09291 (0.06570)	0.1791 (0.05585)	0.01029 (0.1449)
mscore	-0.009755 (0.03807)	0.05248 (0.03832)	0.1608 (0.09928)
constant	0.02449 (0.04635)	0.09574 (0.04338)	-0.001377 (0.1018)

8 Alternative Estimates

To check the robustness of our results to alternative identifying assumptions, we now provide estimates based on Corollary 2 and Theorem 2. First consider estimation based on Corollary 2, which replaces Assumption A3 with Assumption A3', and so requires an additional variable U. We define U 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 could have based U on transcript reports of a graduate degree instead, but then by construction we would have $\widehat{\mu}_{V|U} = 1$, which would then yield numerically identical estimates to those previously reported based on Corollary 1. In our data $\widehat{\mu}_{V|U}$ is .971 for a Bachelor's degree, .981 for an Associate degree, and 1.000 for some college (so the estimates for "some college" in Table 4 are the same as in Table 2). The estimates of returns to lying are somewhat lower in Table 4 than in Table 2. In a few cases they are much lower (e.g., the median returns to lying about a bachelor's degree are only 7% in Table 4 versus 17% in Table 2) but the standard errors are also larger in Table 4, so the differences between the tables are not statistically significantly.

Table 4: Returns to 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)

In Table 5 we report the returns to lying using the estimator based on Theorem 2, which does not use data on either the instrument V or U. These estimates are based only on self reports, and so do not use the transcript data in any way. For these estimates we let $X_0 = X$, which (as with the estimates based on Corollary 1) implies the assumption that that no one reports $D = 0$ when $D^* = 1$ (and hence that transcripts are wrong for the observations in the data that have $D = 0$ and $V = 1$).

As should be expected, the estimates in Table 5 are 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 5 are generally consistent with those in Table 2, in particular, as with Table 4, almost all of the differences between Tables 2 and 5 are not statistically significantly

Table 5: Returns to 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)

Given the substantial differences in estimators and identifying assumptions between Corollary 1 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, 4, and 5. As one would expect, these are generally more stable than the estimates of $E[r(X)]$ reported in Tables 2, 4, and 5, 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 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 7% higher for those who lie about having some college, and from 7% 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 those degrees. Estimates at the median appear to be more reliable and robust than estimates of the mean returns. Our median 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.

In this application D and D^* refer to the same binary event (educational attainment), with D a self report of D^*, so what is identified is the mean effects of having D either agree with or contradict D^*, which given unconfoundedness identifies either the returns to lying (if misreports of D are intentional) or a placebo effect.

To apply our methodology, it is not necessary for D and D^* to 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 returns to lying.

10 Appendix

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

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

so $Cov\left(D, V \mid X\right) \neq 0$ 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

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

and using this result we have

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

Also

$$E\left(Y \mid D,X\right) = R(D,X) + s(D,X)E\left[D^{*}|D,X\right]$$

Use the latter equation to substitute $s(D,X)E\left[D^{*}|D,X\right]$ out of the former equation, and solve what remains for $R(D,X)$ to obtain equation (9).

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 B2.1-2 implies that

$$\begin{aligned} \mu_{Y|D,X} &\equiv E\left(Y|D,X\right) \\ &= E\left(\left(R(D,X) + s(D,X)D^{*}\right)|D,X\right) \\ &= R(D,X) + s(D,X)E\left(D^{*}|D,X\right), \end{aligned} \tag{31}$$

$$\begin{aligned} \mu_{Y^{2}|D,X} &\equiv E\left(Y^{2}|D,X\right) \\ &= E\left(\left(R(D,X) + s(D,X)D^{*} + \eta\right)^{2}|D,X\right) \\ &= E\left(\left(R(D,X) + s(D,X)D^{*}\right)^{2}|D,X\right) + E\eta^{2} \\ &= R(D,X)^{2} + 2R(D,X)s(D,X)E\left(D^{*}|D,X\right) + s(D,X)^{2}E\left(D^{*}|D,X\right) + E\eta^{2} \\ &= R^{2} + 2R\left(\mu_{Y|D,X} - R\right) + s\left(\mu_{Y|D,X} - R\right) + E\eta^{2} \\ &= \mu_{Y|D,X}R + (R+s)\left(\mu_{Y|D,X} - R\right) + E\eta^{2}, \end{aligned} \tag{32}$$

and

$$\begin{aligned} \mu_{Y^{3}|D,X} &\equiv E\left(Y^{3}|D,X\right) \\ &= E\left(\left(R(D,X) + s(D,X)D^{*} + \eta\right)^{3}|D,X\right) \\ &= E\left[\left(R(D,X) + s(D,X)D^{*}\right)^{3}|D,X\right] + 3E\left[\left(R(D,X) + s(D,X)D^{*}\right)|D,X\right]E\eta^{2} + E\eta^{3} \\ &= R(D,X)^{3} + 3R(D,X)^{2}s(D,X)E\left(D^{*}|D,X\right) \\ &\quad +3R(D,X)s(D,X)^{2}E\left(D^{*}|D,X\right) + s(D,X)^{3}E\left(D^{*}|D,X\right) \\ &\quad +3\mu_{Y|D,X}E\eta^{2} + E\eta^{3}. \end{aligned} \tag{33}$$

We then show that assumption B2.3 implies the identification of $E\left(\eta^k\right)$ for $k = 2, 3.$ This assumption implies that

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

and therefore,

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

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

and

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

Therefore, we have

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

and

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

In the next step, we eliminate $s(D, X)$ and $E(D^*|D, X)$ in equations 31-33 to obtain a restriction only containing $R(D, X)$ and known variables. We will use the following two equations repeatedly.

$$(R + s)\left(\mu_{Y|D,X} - R\right) = \mu_{Y^2|D,X} - E\eta^2 - \mu_{Y|D,X}R \tag{34}$$

$$sE\left(D^*|D, X\right) = \mu_{Y|D,X} - R \tag{35}$$

Notice that

$$s = \frac{\mu_{Y^2|D,X} - \mu^2_{Y|D,X} - \sigma^2_{Y|0,x_0}}{\mu_{Y|D,X} - R} + \mu_{Y|D,X} - R$$

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

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

That is

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

The restrictions on R simplify to the quadratic equation

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

where

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

Notice that

$$
\sigma^2_{Y|D,X} = \mu_{Y^2|D,X} - \mu^2_{Y|D,X},
$$

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

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

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

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

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

In summary, we have

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

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

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 35 and 34 implies

$$\alpha \geq 0$$

Consider

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

and

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

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

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

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

Assumption B2.4 implies that

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

Consider

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

Therefore, $0 \leq E\left(D^*|D, X\right) \leq 1$ 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.,

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

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 (16). Let

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

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

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

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

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

and

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

Table 6: R(0,X), Nonparametric and Semiparametric Corollary 1 IV Estimates

	Some college	Associate degree	Bachelor's degree
R0 nonparametric	2.072 (0.01514)	2.125 (0.009659)	2.143 (0.007992)
$R0_t$ nonparametric	2.065 (0.01536)	2.125 (0.01016)	2.144 (0.008568)
$R0_{q1}$ nonparametric	1.863 (0.02520)	1.940 (0.01939)	1.975 (0.01834)
$R0_{med}$ nonparametric	2.003 (0.03859)	2.089 (0.03224)	2.143 (0.02788)
$R0_{q3}$ nonparametric	2.309 (0.02681)	2.326 (0.01762)	2.319 (0.01663)
R0 semi, linear	2.025 (0.01174)	2.094 (0.008751)	2.114 (0.007448)

Table 7: R(1,X), Nonparametric and Semiparametric Corollary 1 IV Estimates

	Some college	Associate degree	Bachelor's degree
R1 nonparametric	2.142 (0.03011)	2.301 (0.02844)	2.184 (0.07708)
$R1_t$ nonparametric	2.152 (0.02985)	2.324 (0.02751)	2.240 (0.07308)
$R1_{q1}$ nonparametric	1.997 (0.04430)	2.179 (0.05181)	2.131 (0.09785)
$R1_{med}$ nonparametric	2.173 (0.04633)	2.382 (0.02936)	2.312 (0.05680)
$R1_{q3}$ nonparametric	2.340 (0.04635)	2.455 (0.03637)	2.424 (0.05945)
R1 semi, linear	2.188 (0.02898)	2.351 (0.02604)	2.339 (0.04339)

Table 8: R(0,X), Nonparametric and Semiparametric Corollary 2 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 9: R(1,X), Nonparametric and Semiparametric Corollary 2 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 10: 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 11: 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] BLUNDELL, R. AND J. L. POWELL, (2004), "Endogeneity in Semiparametric Binary Response Models" *Review of Economic Studies*, 71, 655-679.

[3] CHEN, X., Y. HU AND A. LEWBEL, (2007), “Nonparametric Identification of Regression Models Containing a Misclassified Dichotomous Regressor Without Instruments," unpublished manuscript.

[4] CHEN, X., O. LINTON, AND I. VAN KEILEGOM, (2003) "Estimation of Semiparametric Models when the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591-1608,

[5] DAS, M., (2004), "Instrumental Variables Estimators of Nonparametric Models With Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335-361.

[6] FLORENS, J.-P. AND L. MALAVOLTI, (2003), "Instrumental Regression with Discrete Variables," unpublished manuscript.

[7] GIBRAT, R. (1931), Les Inegalites Economiques, Librairie du Recueil Sirey, Paris

[8] HECKMAN, J. J. (1979), "Sample selection bias as a specification error," *Econometrica*, 47, 153–161.

[9] HECKMAN, J. J., H. ICHIMURA AND P. TODD, (1998), "Matching as an Econometric Evaluations Estimator, Review of Economic Studies, 65, 261-294.

[10] HU, Y. (2006), "Identification and estimation of nonlinear models with misclassification error using instrumental variables," U. Texas at Austin unpublished manuscript.

[11] KANE, T. J., AND C. E. ROUSE, (1995), "Labor market returns to two- and four- year college," *American Economic Review*, 85, 600-614

[12] KANE, T. J., C. E. ROUSE, AND D. STAIGER, (1999), “Estimating Returns to Schooling When Schooling is Misreported,” NBER working paper #7235.

[13] LEWBEL, A., (2007a), "Estimation of Average Treatment Effects With Misclassification," *Econometrica*, 75, 537-551forthcoming.

[14] LEWBEL, A., (2007b), "A Local Generalized Method of Moments Estimator," *Economics Letters*, 94, 124-128.

[15] MAHAJAN, A. (2006) "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74, 631-665.

[16] NEWEY, W. K., (1994), "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233-253.

[17] NEWEY, W. K. AND J. L. POWELL, (2003), "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565-1578.

[18] 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.

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