Economics Letters 115 (2012) 81–84

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Binary misclassification and identification in regression models

Martijn van Hasselt [a,*], Christopher R. Bollinger [b]

[a] Behavioral Health Economics Program, RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709, United States

[b] Department of Economics, Gatton College of Business and Economics, The University of Kentucky, Lexington, KY 40506, United States

ARTICLE INFO

Article history:
Received 10 October 2011
Accepted 30 November 2011
Available online 13 December 2011

JEL classification:
C18
C21
C25

Keywords:
Misclassification error
Binary regressors
Partial identification
Homoscedasticity

ABSTRACT

We study a regression model with a binary explanatory variable that is subject to misclassification errors. The regression coefficient is then only partially identified. We derive several results that relate different assumptions about the misclassification probabilities and the conditional variances to the size of the identified set.

1. Introduction

In this note, we study a regression model with a binary explanatory variable that may be misclassified. That is, there is some nonzero probability of observing a "false positive" or a "false negative". For example, such errors can be simple data coding errors, or result from the underreporting of certain behaviors (e.g., illicit drug use) in surveys. In general, the regression parameter is not identified in the presence of misclassification, and different sets of identifying assumptions have been investigated. Chen et al. (2008a,b), Mahajan (2006) and Lewbel (2007) present sufficient conditions for identification. Alternatively, Klepper (1988), Manski (1990), Bollinger (1996), and Deng and Hu (2009) take a partial identification approach and derive bounds of the identified set that are nonparametrically identified under weak assumptions. This note adds to this literature by presenting several new partial identification results. Specifically, we demonstrate to what extent homoscedasticity and restrictions on the misclassification probabilities shrink the size of the identified set. We present the model, assumptions and main results in Section 2. We provide a brief discussion in Section 3. All proofs are collected in the Appendix.

2. Main results

The regression model for the outcome variable Y_i is given by

$$Y_i = \alpha + \beta Z_i + U_i, \qquad E(U_i|Z_i) = 0, \qquad \sigma_U^2 = E(U_i^2) < \infty, \quad (1)$$

where U_i is an unobserved error term. The regressor $Z_i \in \{0, 1\}$ is binary, with $\Pr\{Z_i = 1\} = \pi$ and $\pi \in (0, 1)$. Note that linearity of the model is not restrictive since Z_i is binary. The econometrician does not observe Z_i, but a binary variable X_i, which is a potentially misclassified version of Z_i. Specifically, we assume the following.

Assumption 2.1. $\Pr\{X_i = 1|Z_i, Y_i\} = (1 - q)Z_i + p(1 - Z_i)$.

Assumption 2.2. $p + q < 1$.

Assumption 2.1 introduces the misclassification probabilities p and q, and states that X_i is conditionally independent of the outcome Y_i; hence, the misclassification error contains no information about Y_i, or vice versa.[1] In some applications, however, this assumption may be untenable (e.g., Kreider and Pepper, 2007). Assumption 2.2 ensures that the covariance between Z_i and X_i is positive, so that X_i is a better predictor of Z_i than a purely random guess.

The mean, variance and covariance of (X_i, Y_i) are given by[2]

$$\begin{aligned} \mu_X &= (1 - \pi)p + \pi(1 - q), \\ \mu_Y &= \alpha + \beta\pi, \\ \sigma_{XY} &= \beta\pi(1 - \pi)(1 - p - q), \\ \sigma_Y^2 &= \beta^2\pi(1 - \pi) + \sigma_U^2. \end{aligned} \quad (2)$$

[1] The misclassification error is then called "nondifferential" (Carroll et al., 1995).

[2] We omit the variance of X_i, since it is a function of μ_X.

* Corresponding author. Tel.: +1 919 541 6925; fax: +1 919 485 5555.
E-mail address: mvhasselt@rti.org (M. van Hasselt).

0165-1765/$ – see front matter
doi:10.1016/j.econlet.2011.11.031

These moments of (X_i, Y_i) are nonparametrically identified. If $\sigma_{XY} = 0$, then it follows from (2) that $(\alpha, \beta, \sigma_U^2)$ is identified. To avoid this trivial case, we impose the following:

Assumption 2.3. $\sigma_{XY} > 0$.

Bollinger (1996, Theorem 1) shows that under Assumptions 2.1–2.3:

$$\frac{\sigma_{XY}}{\sigma_X^2} \le \beta \le \max\left\{\mu_X \frac{\sigma_{XY}}{\sigma_X^2} + (1-\mu_X)\frac{\sigma_Y^2}{\sigma_{XY}},\right. \left.(1-\mu_X)\frac{\sigma_{XY}}{\sigma_X^2} + \mu_X \frac{\sigma_Y^2}{\sigma_{XY}}\right\}. \quad (3)$$

The first and second bounds on the right-hand side apply for $\mu_X \le \frac{1}{2}$ and $\mu_X > \frac{1}{2}$, respectively. Let $\sigma_j^2 \equiv \mathrm{Var}(Y_i|X_i = j)$ for $j = 0, 1$ be the conditional variance of Y_i given X_i. Our first result shows that the bounds can be tightened if the regression error U_i is homoscedastic:

Lemma 1. *If Assumptions 2.1–2.3 hold and $E(U_i^2|Z_i) = E(U_i^2) = \sigma_U^2$, then*

$$\frac{\sigma_{XY}}{\sigma_X^2} \le \beta \le \max\left\{(1-\mu_X)\frac{\sigma_{XY}}{\sigma_X^2} + \mu_X \frac{\sigma_1^2(1-\mu_X)^2}{\sigma_{XY}},\right. \left.\mu_X \frac{\sigma_{XY}}{\sigma_X^2} + (1-\mu_X)\frac{\sigma_0^2\mu_X^2}{\sigma_{XY}}\right\}.$$

The first and second bound on the right-hand side apply for $\mu_X \le \frac{1}{2}$ and $\mu_X > \frac{1}{2}$, respectively. The upper bound is sharper than the one in (3).

By comparing the upper bounds of Lemma 1 and Eq. (3) it is easy to see when homoscedasticity is effective in reducing the size of the identified set. For example, suppose that $\mu_X < \frac{1}{2}$ and consider increasing values of σ_0^2; this leads to a higher value of σ_Y^2, which in turn increases the bound in (3). The bound in Lemma 1, however, is unaffected. Intuitively, when $\mu_X < \frac{1}{2}$ it is more likely that $X_i = 0$. At the same time a large value of σ_0^2 implies substantial variation in the outcome when $X_i = 0$, making it 'harder' to identify β. The homoscedasticity assumption is then more valuable in terms of reducing the (absolute and relative) size of the identified set.

The following results consider the effect of various assumptions about p and q on the bounds for β. Lemma 2 complements Theorem 2 in Bollinger (1996):

Lemma 2. *Suppose that Assumptions 2.1–2.3 hold. If $q = 0$ and $p > 0$, then*

$$\frac{\sigma_{XY}}{\sigma_X^2} \le \beta \le \mu_X \frac{\sigma_{XY}}{\sigma_X^2} + (1-\mu_X)\frac{\sigma_Y^2}{\sigma_{XY}}.$$

If $p = 0$ and $q > 0$, then

$$\frac{\sigma_{XY}}{\sigma_X^2} \le \beta \le (1-\mu_X)\frac{\sigma_{XY}}{\sigma_X^2} + \mu_X \frac{\sigma_Y^2}{\sigma_{XY}}.$$

In some cases the restriction that either p or q is equal to zero, may be reasonable. For example, individuals are likely to underreport the use of illegal drugs ($q > 0$), but it is highly implausible that a non-user reports actual use ($p = 0$). The upper bounds in Lemma 2 are again sharper than the one in (3). For example, when $\mu_X > \frac{1}{2}$ the additional assumption that $q = 0$ reduces the upper bound by an amount

$$(1-\mu_X)\frac{\sigma_{XY}}{\sigma_X^2} + \mu_X \frac{\sigma_Y^2}{\sigma_{XY}} - \left(\mu_X \frac{\sigma_{XY}}{\sigma_X^2} + (1-\mu_X)\frac{\sigma_Y^2}{\sigma_{XY}}\right) = (2\mu_X - 1)\left(\frac{\sigma_Y^2}{\sigma_{XY}} - \frac{\sigma_{XY}}{\sigma_X^2}\right) > 0.$$

Next, suppose that false positives are at least as likely as false negatives ($q \le p$), or vice versa ($p \le q$). Let ρ_{XY} denote the correlation coefficient. The identified set for β then takes the following form:

Lemma 3. *Suppose that Assumptions 2.1–2.3 hold. If $q \le p$, then*

$$\frac{\sigma_{XY}}{\sigma_X^2} \le \beta \le \max\left\{\mu_X \frac{\sigma_{XY}}{\sigma_X^2} + (1-\mu_X)\frac{\sigma_Y^2}{\sigma_{XY}},\right. \left.\frac{\sigma_Y^2}{\sigma_{XY}}\sqrt{1 - 4\sigma_X^2(1-\rho_{XY}^2)}\right\}.$$

Conversely, if $p \le q$, then

$$\frac{\sigma_{XY}}{\sigma_X^2} \le \beta \le \max\left\{\frac{\sigma_Y^2}{\sigma_{XY}}\sqrt{1 - 4\sigma_X^2(1-\rho_{XY}^2)},\right. \left.(1-\mu_X)\frac{\sigma_{XY}}{\sigma_X^2} + \mu_X \frac{\sigma_Y^2}{\sigma_{XY}}\right\}$$

the first and second bounds on the right-hand side apply for $\mu_X \le \frac{1}{2}$ and $\mu_X > \frac{1}{2}$, respectively.

When $\mu_X \le \frac{1}{2}$, a comparison with Eq. (3) shows that the additional information $q \le p$ does not sharpen the upper bound on β; however, for $\mu_X > \frac{1}{2}$ the upper bound is reduced to $(\sigma_Y^2/\sigma_{XY})\sqrt{1 - 4\sigma_X^2(1-\rho_{XY}^2)}$.

Finally, suppose that $p = q$. Misclassification is now symmetric: conditional on (Z_i, Y_i) false positives and false negatives are equally likely. Under Assumptions 2.2 and 2.3 the bounds for β follow immediately from Lemma 3:

$$\frac{\sigma_{XY}}{\sigma_X^2} \le \beta \le \frac{\sigma_Y^2}{\sigma_{XY}}\sqrt{1 - 4\sigma_X^2(1-\rho_{XY}^2)}.$$

If, in addition, U_i is homoscedastic, then β is typically identified:

Lemma 4. *Suppose that Assumptions 2.1–2.3 hold with $p = q$ and $E(U_i^2|Z_i) = E(U_i^2) = \sigma_U^2$. Then β is identified, except when $\mu_X = \frac{1}{2}$.*

The identification result can be understood in terms of solving the system of equations in (2). Symmetry eliminates one of the unknowns, and homoscedasticity leads to two equations for σ_0^2 and σ_1^2, instead of a single one for σ_Y^2. This yields a system of 5 equations in 5 unknowns, which has a unique solution when $\mu_X \ne \frac{1}{2}$.

3. Discussion

In this paper we have analyzed partial identification of the regression coefficient of a binary misclassified variable. In particular, we have shown how various assumptions about the misclassification probabilities and the regression error variance affect the bounds of the identified set. Interestingly, these assumptions only affect the upper bound. In the case of Lemma 4 the regression parameter is identified. For most applications, however, the assumptions of homoscedasticity and symmetric misclassification are too strong. The identification strategies of Mahajan (2006) and Lewbel (2007), based on the availability of an instrumental variable, may then be more appropriate.

The use of conditional moments is common in identification analyses. Chen et al. (2008b) show that with homoscedasticity and the additional assumption that $E(U_i^3|Z_i) = E(U_i^3)$, the regression parameter is identified. We demonstrate here (Lemma 1) that dropping the restriction on the third moment results in partial identification. Our bounds are sharper, however, than those for the general heteroscedastic case in Bollinger (1996). On the other hand, Deng and Hu (2009) show that the identified set can be unbounded if the misclassification error is related to the outcome Y_i. The assumption of nondifferential measurement error therefore carries important identifying information, because it bounds the identified set. Additional information about the misclassification rates, if deemed plausible, can further sharpen the bounds (Lemmas 2 and 3).

Appendix

Proof of Lemma 1. From Assumption 2.1 and Bayes' rule it follows that

$$\Pr\{Z_i = 1|X_i = 0\} = \frac{\pi q}{1-\mu_X},$$

$$\Pr\{Z_i = 1|X_i = 1\} = \frac{\pi(1-q)}{\mu_X}.$$

The conditional variance of Z_i given X_i is

$$V(Z_i|X_i = 0) = \frac{\pi(1-\pi)q(1-p)}{(1-\mu_X)^2},$$

$$V(Z_i|X_i = 1) = \frac{\pi(1-\pi)p(1-q)}{\mu_X^2}.$$

Assumption 2.1 and homoscedasticity imply that

$$\begin{aligned} E(U_i|X_i) &= E[E(U_i|Z_i, X_i)|X_i] \\ &= E[E(U_i|Z_i)|X_i] \\ &= 0, \end{aligned}$$

$$\begin{aligned} E(U_i^2|X_i) &= E[E(U_i^2|Z_i)|X_i] \\ &= \sigma_U^2. \end{aligned}$$

The conditional variance of Y_i given X_i can now be calculated as

$$\sigma_0^2 = \beta^2 \frac{\pi(1-\pi)q(1-p)}{(1-\mu_X)^2} + \sigma_U^2, \tag{A.1}$$

$$\sigma_1^2 = \beta^2 \frac{\pi(1-\pi)(1-q)p}{\mu_X^2} + \sigma_U^2. \tag{A.2}$$

From the first and third equations in (2) we can solve for β as

$$\begin{aligned} \beta &= \frac{\sigma_{XY}(1-p-q)}{(\mu_X - p)(1-\mu_X - q)} \\ &= b(p, q). \end{aligned} \tag{A.3}$$

Substituting β and $\pi = (\mu_X - p)/(1-p-q)$ in (A.1) and (A.2), and using $\sigma_U^2 \geq 0$, we obtain:

$$0 \leq p \leq \mu_X \left[\frac{\mu_X^2 \sigma_1^2 (1-\mu_X - q)}{\mu_X^2 \sigma_1^2 (1-\mu_X - q) + \sigma_{XY}^2 (1-q)} \right], \tag{A.4}$$

$$0 \leq q \leq (1-\mu_X) \left[\frac{(1-\mu_X)^2 \sigma_0^2 (\mu_X - p)}{(1-\mu_X)^2 \sigma_0^2 (\mu_X - p) + \sigma_{XY}^2 (1-p)} \right]. \tag{A.5}$$

The upper (lower) bound of the identified set is the maximum (minimum) of $b(p, q)$, subject to the restrictions in (A.4) and (A.5). Since $\partial b/\partial p, \partial b/\partial q > 0$ the lower bound for β is attained at $p = q = 0$. For the upper bound, suppose first that

$$\left[\frac{\partial b(p, q)}{\partial p}\right]_{p=q=0} = \frac{\sigma_{XY}}{\mu_X^2} \geq \frac{\sigma_{XY}}{(1-\mu_X)^2} = \left[\frac{\partial b(p, q)}{\partial q}\right]_{p=q=0},$$

or $\mu_X \leq \frac{1}{2}$. Starting at $(0, 0)$ it is optimal to increase p. Moreover, since $d^2b/dp^2 > 0$ for $q = 0$ and all feasible values of p, it is optimal to increase p until the upper bound of (A.4) is binding. The first upper bound is then obtained by setting $q = 0$ and

$$p = \mu_X \left[\frac{\mu_X^2 \sigma_1^2 (1-\mu_X)}{\mu_X^2 \sigma_1^2 (1-\mu_X) + \sigma_{XY}^2} \right],$$

in (A.3). A similar argument shows that for $\mu_X > \frac{1}{2}$ the second upper bound follows by substituting $p = 0$ and

$$q = (1-\mu_X) \left[\frac{(1-\mu_X)^2 \sigma_0^2 \mu_X}{(1-\mu_X)^2 \sigma_0^2 \mu_X + \sigma_{XY}^2} \right],$$

into (A.3).

To show the second statement of Lemma 1, we first derive an expression for σ_Y^2. From Bayes' rule:

$$\begin{aligned} E(Z_i|X_i) &= \Pr\{Z_i = 1|X_i = 1\} X_i + \Pr\{Z_i = 1|X_i = 0\}(1 - X_i) \\ &= \frac{\pi q}{1-\mu_X} + \frac{X_i \pi (1-\mu_X - q)}{\mu_X(1-\mu_X)}, \end{aligned}$$

$$V[E(Z_i|X_i)] = \frac{\pi^2 (1-\mu_X - q)^2}{\mu_X(1-\mu_X)}.$$

Using $\pi = (\mu_X - p)/(1-p-q)$ and Eq. (A.3), we get

$$\begin{aligned} \sigma_Y^2 &= E[V(Y_i|X_i)] + V[E(Y_i|X_i)] \\ &= \mu_X \sigma_1^2 + (1-\mu_X)\sigma_0^2 + V[\alpha + \beta E(Z_i|X_i) + E(U_i|X_i)] \\ &= \mu_X \sigma_1^2 + (1-\mu_X)\sigma_0^2 + \frac{\beta^2 \pi^2 (1-\mu_X - q)^2}{\mu_X(1-\mu_X)} \\ &= \mu_X \sigma_1^2 + (1-\mu_X)\sigma_0^2 + \frac{\sigma_{XY}^2}{\sigma_X^2}. \end{aligned}$$

Suppose $\mu_X \leq \frac{1}{2}$, and consider the difference between the corresponding bounds in (3) and Lemma 1. Using the expression for σ_Y^2 given above:

$$\begin{aligned} \Delta_1 &= \left(\mu_X \frac{\sigma_{XY}}{\sigma_X^2} + (1-\mu_X) \frac{\sigma_Y^2}{\sigma_{XY}} \right) \\ &\quad - \left((1-\mu_X) \frac{\sigma_{XY}}{\sigma_X^2} + \mu_X \frac{\sigma_1^2 (1-\mu_X)^2}{\sigma_{XY}} \right) \\ &= \mu_X \frac{\sigma_{XY}}{\sigma_X^2} + (1-\mu_X) \left(\mu_X \frac{\sigma_1^2}{\sigma_{XY}} + (1-\mu_X) \frac{\sigma_0^2}{\sigma_{XY}} + \frac{\sigma_{XY}}{\sigma_X^2} \right) \\ &\quad - (1-\mu_X) \frac{\sigma_{XY}}{\sigma_X^2} - \mu_X \frac{\sigma_1^2 (1-\mu_X)^2}{\sigma_{XY}} \\ &= \mu_X \frac{\sigma_{XY}}{\sigma_X^2} + \frac{1}{\sigma_{XY}} \left[\sigma_1^2 \mu_X \sigma_X^2 + \sigma_0^2 (1-\mu_X)^2 \right] \\ &> 0. \end{aligned}$$

Similarly, for $\mu_X > \frac{1}{2}$ the difference is

$$\begin{aligned} \Delta_2 &= (1-\mu_X) \frac{\sigma_{XY}}{\sigma_X^2} + \frac{1}{\sigma_{XY}} \left[\sigma_1^2 \mu_X^2 + \sigma_0^2 (1-\mu_X) \sigma_X^2 \right] \\ &> 0. \quad \square \end{aligned}$$

Proof of Lemma 2. Suppose first that $q = 0$, so that $\pi = (\mu_X - p)/(1-p)$, and

$$\beta = \frac{\sigma_{XY}(1-p)}{(\mu_X - p)(1-\mu_X)} = b(p).$$

We now need to minimize and maximize $b(p)$, subject to constraints. From $\sigma_U^2 \geq 0$ and the system (2), it follows that

$$0 \leq p \leq \mu_X (1 - \rho_{XY}^2).$$

Since $db(p)/dp > 0$, the maximum and minimum of $b(p)$ are attained at $\mu_X(1-\rho_{XY}^2)$ and 0, respectively. Substituting these values in $b(p)$ yields the result. The argument for the case $p = 0$ is completely analogous. □

Proof of Lemma 3. Consider $b(p, q)$ in (A.3). Since $b(p, q)$ is increasing in both p and q, the lower bound is obtained at $p = q = 0$, and $b(0, 0) = \sigma_{XY}/\sigma_X^2$. For the upper bound, first consider the case $\mu_X \le \frac{1}{2}$. Then $\partial b(p, q)/\partial p \ge \partial b(p, q)/\partial q > 0$ at (0, 0), and it is optimal to increase p until the nonnegativity of σ_U^2 is binding. This occurs at $p = \mu_X(1-\rho_{XY}^2)$, and substitution of this into $b(p, q)$ yields the first bound in Lemma 3. Now suppose that $\mu_X > \frac{1}{2}$, so that $0 < \partial b(p, q)/\partial p < \partial b(p, q)/\partial q$ at the point (0, 0). It would be optimal to increase q, but now the constraint $q \le p$ is binding. At the optimum we therefore must have $p > 0$. It is easy to show that σ_U^2 is zero when

$$q = q(p) = \frac{\sigma_X^2(1-\rho_{XY}^2) - p(1-\mu_X)}{\mu_X - p}. \tag{A.6}$$

This curve intersects with the line $q = p$ at the point

$$p^* = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4\sigma_X^2(1-\rho_{XY}^2)}. \tag{A.7}$$

For maximizing $b(p, q)$ the constraint $q = p$ is binding when $p < p^*$, whereas (A.6) is binding for $p > p^*$. It remains to determine at which point the maximum can be found. Suppose that (A.6) holds at the solution. Substituting this into $b(p, q)$ and taking the derivative it follows that

$$\frac{\partial b(p, q(p))}{\partial p} = -\frac{\sigma_Y^2}{\sigma_{XY}} - \frac{\sigma_{XY}\sigma_X^2(1-\rho_{XY}^2)}{\rho_{XY}^2(\mu_X - p)^2} < 0.$$

Therefore, the value of $b(p, q)$ can be increased by decreasing p to the point p^*, where also $q = p^*$ (note that a point (q, p^*) with $q < p^*$ is not optimal, since $b(p, q)$ can be increased by increasing q). Substituting $q = p^*$ and (A.7) into the expression for $b(p, q)$ yields the second bound. The proof for the case $p \le q$ is analogous and omitted here. □

Proof of Lemma 4. Assume first that $\mu_X \neq \frac{1}{2}$. Following the proof of Lemma 1 it can be shown that

$$\sigma_0^2 = \frac{\beta^2\pi(1-\pi)p(1-p)}{(1-\mu_X)^2} + \sigma_U^2,$$

$$\sigma_1^2 = \frac{\beta^2\pi(1-\pi)p(1-p)}{\mu_X^2} + \sigma_U^2.$$

Taking the difference of these two equations eliminates the structural parameter σ_U^2. Substituting $\pi = (\mu_X - p)/(1-2p)$ and $b(p, q)$ with $p = q$ into the difference $\sigma_1^2 - \sigma_0^2$, we find

$$p = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4\frac{c}{(1+c)}\sigma_X^2},$$

$$c \equiv \frac{(\sigma_1^2 - \sigma_0^2)\sigma_X^4}{\sigma_{XY}^2(1-2\mu_X)}.$$

Substituting the solution for p back into (A.3) we can solve for the coefficient as

$$\beta = \frac{\sigma_{XY}}{\sigma_X^2}(1+c)\sqrt{1 - 4\frac{c}{(1+c)}\sigma_X^2}.$$

When $\mu_X = \pi = \frac{1}{2}$ the conditional variance of Y_i does not depend on X_i and we can no longer eliminate σ_U^2. The remaining parameters $(\alpha, \beta, \sigma_U^2, p)$ are now no longer identified. □

References

Bollinger, C.R., 1996. Bounding mean regressions when a binary regressor is mismeasured. Journal of Econometrics 73, 387–399.

Carroll, R., Ruppert, D., Stefanski, L., 1995. Measurement Error in Nonlinear Models. Chapman & Hall, London.

Chen, X., Hu, Y., Lewbel, A., 2008a. Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments. Economics Letters 100, 381–384.

Chen, X., Hu, Y., Lewbel, A., 2008b. A note on the closed-form identification of regression models with a mismeasured binary regressor. Statistics and Probability Letters 78, 1473–1479.

Deng, P., Hu, Y., 2009. Bounding the effect of a dichotomous regressor with arbitrary measurement errors. Economics Letters 105, 256–260.

Klepper, S., 1988. Bounding the effects of measurement error in regressions involving dichotomous variables. Journal of Econometrics 37, 343–359.

Kreider, B., Pepper, J.V., 2007. Disability and employment: reevaluating the evidence in light of reporting errors. Journal of the American Statistical Association 102, 432–441.

Lewbel, A., 2007. Estimation of average treatment effects with misclassification. Econometrica 75, 537–551.

Mahajan, A., 2006. Identification and estimation of regression models with misclassification. Econometrica 74, 631–666.

Manski, C.F., 1990. Nonparametric bounds on treatment effects. American Economic Review 80, 319–323.