Misclassification Error and Instrumental Variables

Yingyao Hu[1]

The University of Texas at Austin

(this version: Sept, 2004, the first version: April 2004)

Abstract

This paper provides a new estimator in the estimation of nonlinear models with measurement error using instrumental variables . The latent true variable considered here is a 0-1 dichotomous variable; therefore, the measurement error is misclassification error. The results show that instrumental variables are actually very powerful under weak assumptions to estimate nonlinear errors-in-variables models. It is shown that a latent model can be nonparametrically identifiable and directly estimable using instrumental variables. The new estimator is shown to be $\sqrt{n}$ consistent. The simulations suggest good finite sample properties of the estimator. An empirical illustration is also provided.

JEL classification: C14, C41.

Keywords: nonlinear errors-in-variables model, instrumental variable, misclassification error.

[1]I am grateful to Stephen Donald, Jerry Hausman, Cheng Hsiao, Arthur Lewbel, Susanne Schennach, and Zhong Zhao for their suggestions and comments. I especially thank Robert Moffitt and Geert Ridder for their advice and support. Contact Information: Department of Economics, University of Texas at Austin, 1 University Station C3100, BRB 1.116, Austin, TX 78712, hu@eco.utexas.edu, http://www.eco.utexas.edu/~hu/.

1 Introduction

Many models used in empirical research in microeconomics are nonlinear in the explanatory variables. Examples are nonlinear regression models, models for limited-dependent variables (logit, probit, tobit, etc.), and duration models. Often the parameters of such nonlinear models are estimated using data in which one or more independent variables are measured with error. Measurement error is a pervasive problem in economic data (Bound, Brown, and Mathiowetz, 2001). The measurement error associated with discrete variables is called misclassification error. Consistent estimation of a nonlinear model with measurement error is considered to be a difficult problem and usually requires parametric assumptions or additional sample information, such as instrumental variables (IV), repeated measurements, or validation samples (see Carroll, Ruppert, and Stefanski, 1995, for a survey). This paper provides a new method for using the instrumental variable in the estimation of nonlinear models with misclassification error without imposing strong restrictions on the instrumental variable beyond its definition. The latent model is shown to be nonparametrically identifiable and directly estimable using instrumental variable. The misclassification error can be correlated with not only the latent true value but also with other explanatory variables. The estimator is very easy to derive and highly applicable.

In a measurement error model, a valid instrument is a variable that (a) can be excluded from the model, (b) is correlated with the latent true value, and (c) is independent of the measurement error. The IV method was developed for models that are linear in the mismeasured variables. In general, IV estimators are biased in nonlinear models (Amemiya, 1985). However, Amemiya and Fuller (1988) and Carroll and Stefanski (1990) obtain a consistent IV estimator in nonlinear models under the assumption that the measurement error vanishes if the sample size increases. Buzas (1997) derives an instrumental variable estimator that is approximately consistent for general nonlinear models. Hausman, Ichimura, Newey, and Powell (1991) and Hausman, Newey, and Powell (1995) extend IV estimation to a polynomial regression model. Lewbel (1998) shows a consistent estimator for a specially specified latent variable model with instrumental variables and a strong exclusion restriction. Newey (2001) and Schennach (2004) consider the nonlinear regression model using a prediction equation

with instrumental variables independent of the prediction error. The method outlined in this paper shows that the instrumental variable is actually very powerful in the estimation of nonlinear models with measurement error under weak assumptions. The latent model is shown to be nonparametrically identifiable and directly estimable.

The misclassification error has been analyzed in a few studies. Aigner (1973) and Bollinger (1996) consider the issue of misclassifed binary regressors. Freeman (1984) investigates the misclassification error in the union status in the longitudinal sample. Hausman, Abrevaya, and Scott-Morton (1998) and Lewbel (2000) discuss the effect of misclassification error on the dependent variable. Horowitz and Manski (1995) and Molinari (2004) derive bounds on parameters when the misclassified variable is a response variable. Ramalho (2002) deals with the presence of misclassification in the response variable in choice-based samples. Black, Berger, and Scott (2000) estimate the slope coefficient in a regression model when a secondary measurement is available. Kane, Rouse, and Staiger, (1999) and Lewbel (2003) also use instruments to solve misclassification in treatment effect models. A recent paper by Mahajan (2003) provides point estimators for binary choice models with misclassified regressors. In this paper, we consider general nonlinear models with misclassification errors. The misclassification error can be correlated with not only the true latent variable but also with other explanatory variables.

This paper is organized as follows. Section 2 introduces the model and assumptions on instrumental variable. Section 3 shows the nonparametric identification of the estimator. Section 4 develops a $\sqrt{n}-$ consistent semiparametric maximum likelihood estimator. Section 5 presents Monte Carlo evidence of the finite sample performance of the estimator. Section 6 applies the estimator to a probit model of labor supply. Section 7 concludes the paper. The proofs are in the appendix.

2 The model

The model considered in this paper contains three variables, y, x^* and w. The variable y is a dependent variable, w is the accurately measured variable, and x^* is the latent true discrete variable which is subject to misclassification error. Suppose the conditional density of the

dependent variable y on x^*, w is

$$f_{y|x^*w}(y|x^*, w). \tag{1}$$

In an i.i.d. sample, we observe y, x, w and z, where x is a proxy of x^* and z is an instrumental variable satisfying:

Assumption 1: $f_{y|x^*wxz}(y|x^*w, x, z) = f_{y|x^*w}(y|x^*, w)$.
Assumption 2: $f_{x|x^*wz}(x|x^*, w, z) = f_{x|x^*w}(x|x^*, w)$.

Assumption 1 means the misclassified variable x and the instrumental variable z do not contain any useful information on the dependent variable y beyond x^* and w. It also implies that the misclassification error is independent of the dependent variable y conditional on explanatory variables.The detailed discussion on differential measurement error can be found in Bound, Brown, and Mathiowetz (2001) and Carroll, Ruppert, and Stefanski, (1995). Assumption 2 implies the misclassification error between x^* and x is independent of the instrumental variable z conditional on the latent variable and other explanatory variables. We will discuss the correlation between the instrumental variable z and the true value x in the following sections. These assumptions are widely used in most of the relevant studies. Note that the instrumental variable z can contain measurement error. As long as the measurement error in z is independent of y and x conditional on x^* and w, the mismeasured instrumental is still an instrument, and the method in this paper still applies.

The key difference between Mahajan (2003) and this paper lies in Assumption 2. The comparable part, i.e., section 4, of the former paper relies on the assumption as follows:

Assumption 3: $f_{x|x^*wz}(x|x^*, w, z) = f_{x|x^*}(x|x^*)$.

Assumption 3 means the misclassification error is independent of the IV and other explanatory variables conditional on the latent variable. Obviously, Assumption 2 considered in this paper is much weaker than Assumption 3. Besides this difference in assumptions there are important differences in the identification strategy. Under Assumption 3, the misclassification probability $f_{x|x^*}$ is treated as additional unknown parameters (or constants) in Mahajan (2003). The identification of the parameter of interest together with $f_{x|x^*}$ is proved by contradiction in his Lemma 4 and hence is not constructive in the sense that it leads directly to

an estimator. The parameters are estimated together as a "plug-in" semiparametric MLE (Newey and McFadden, 1994). [2] In this paper, the identification approach is to solve a nonlinear inversion problem. The model $f_{y|x^*w}$ is expressed as a known function of densities of observed distributions, and, therefore, is nonparametrically identified. The likelihood function is much more complicated than that in Mahajan (2003) but the advantage is that it can allow for a more general form of the misclassification error model. Moreover, the estimator is still a simple "plug-in" semiparametric MLE.

Under Assumption 2, Mahajan (2003) considers the case where repeated measurements are available. Although there is some similarity between IV and repeated measurements, the likelihood based approach in his paper exploits the specific structure imposed by repeated measurements and makes it hard to compare his section 5 with this paper. However, there are still two points worthy mentioning. First, his identification in Lemma 9 still relies on Assumption 3, which is stronger than the assumption used in this paper. Second, if one treats the secondary measurement as an IV, this paper suggests that the likelihood can be written in such a way that there exists a simple "plug-in" semiparametric MLE. Therefore, a sieve estimator of nuisance functions in his section 5 is redundant.[3]

In this paper, we consider general nonlinear models with misclassification errors. As shown in Assumption 2, the misclassification error can be correlated with not only the true latent variable but also with all the other explanatory variables. This paper shows that the distribution function of the dependent variable conditional on explanatory variables $f_{y|x^*w}$ can be expressed as a known function of densities of observed distributions, and, therefore, is nonparametrically identified. With this expression of $f_{y|x^*w}$, we can easily extend the method in this paper to a GMM framework, such as $E_{y|x^*w}\left[m(y, x^*, w; \theta_0)\right] = 0$. The powerful identification in this paper implies that Assumptions 3 is unnecessarily strong and that a

[2] Assumption 3 can be relaxed to $f_{x|x^*wz}(x|x^*, w_1, w_2, z) = f_{x|x^*w_1}(x|x^*, w_1)$ with $w = (w_1, w_2)$. This assumption means that the misclassification probability has to be independent of part of the explanatory variable w_2 and the IV z conditional on the rest of the explanatory variables x^* and w_2. If one wants to generalize the estimator under Assumption 3 to this case, the misclassification probability $f_{x|x^*w_1}$ has to be estimated as a nuisance unknown function, or an infinitely dimensional unknown parameter. The estimator can not be as easy as a "plug-in" semiparametric MLE anymore. A sieve estimator of $f_{x|x^*w_1}$ has to be used in the estimation of the parameter of interest. The identification condition still needs to be found in this case.

[3] I am thankful to Geert Ridder for his suggestions on comparison between Mahajan (2003) and this paper.

secondary measurement is not needed. The estimator in this paper is a simple "plug-in" semiparametric MLE. The parameter of interest is estimated through the maximization of the likelihood whose unknown parts only contain the parameter of interest and unknown density functions, i.e. the "plug-in" part, whose corresponding sample distributions are directly available in the data. This feature makes the estimator highly applicable.

3 The Identification

We first illustrate the key idea of this paper using the 0-1 dichotomous case with misclassification error.

Assumption 2.1: x,x^* and z are 0-1 dichotomous variables.

The key is to express the conditional density of y on x, w and z, $f_{yx|wz}(y, x|w, z)$, as a function of $f_{y|x^*w}(y|x^*, w)$ and directly estimable densities. By law of total probability, we have

$$f_{yx|wz}(y, x|w, z) = \sum_{x^*=0,1} f_{y|x^*w}(y|x^*, w) f_{x|x^*w}(x|x^*, w) f_{x^*|wz}(x^*|w, z). \tag{2}$$

First, we can solve for $f(x^*|w, z)$ through

$$f_{y|wz}(y|w, z) = \sum_{x^*=0,1} f_{y|x^*w}(y|x^*, w) f_{x^*|wz}(x^*|w, z). \tag{3}$$

Define $f_{x^*|wz}(1|w, z) := f_{x^*|wz}(x^* = 1|w, z)$. We make the following assumption:

Assumption 2.2: $f_{y|x^*w}(y|1, w) \neq f_{y|x^*w}(y|0, w)$.

This assumption means that x^* is a valid explanatory variable so that it is correlated with the dependent variable y. It also implies that one cannot test for a zero effect of x^* on y conditional on w. We then have

$$f_{x^*|wz}(1|w, z) = \frac{f_{y|wz}(y|w, z) - f_{y|x^*w}(y|0, w)}{f_{y|x^*w}(y|1, w) - f_{y|x^*w}(y|0, w)}, \tag{4}$$

$$f_{x^*|wz}(0|w, z) = \frac{f_{y|x^*w}(y|1, w) - f_{y|wz}(y|w, z)}{f_{y|x^*w}(y|1, w) - f_{y|x^*w}(y|0, w)}. \tag{5}$$

Second, we solve for the misclassification probability $f_{x|x^*w}(x|x^*, w)$ using the instrumental variable z through

$$f_{x|wz}(x|w,z) = \sum_{x^*=0,1} f_{x|x^*w}(x|x^*,w) f_{x^*|wz}(x^*|w,z). \tag{6}$$

The assumption that z is an instrument is equivalent to

Assumption 2.3: $f_{x^*|wz}(x^*|w,0) \neq f_{x^*|wz}(x^*|w,1)$.

As shown below in equations (7), and (8), this assumption is necessary for identification. Assumption 2.3 does not specify whether the instrument is weak or not. If the instrument is weak, one would expect the difference between $f_{x^*|wz}(x^*|w,0)$ and $f_{x^*|wz}(x^*|w,1)$ is small. Since the difference in the denominator in equations (7), and (8), the weakness of the instrument may still be a problem in the estimation. Since the major purpose is to introduce a new estimator, we assume the instrument z is valid. We then have

$$f_{x|x^*w}(x|1,w) = \frac{f_{x|wz}(x|w,1) f_{x^*|wz}(0|w,0) - f_{x|wz}(x|w,0) f_{x^*|wz}(0|w,1)}{f_{x^*|wz}(0|w,0) - f_{x^*|wz}(0|w,1)}, \tag{7}$$

$$f_{x|x^*w}(x|0,w) = \frac{f_{x|wz}(x|w,0) f_{x^*|wz}(1|w,1) - f_{x|wz}(x|w,1) f_{x^*|wz}(1|w,0)}{f_{x^*|wz}(0|w,0) - f_{x^*|wz}(0|w,1)}. \tag{8}$$

Combining equations (4), (5), (7), and (8), we obtain

$$f_{x|x^*w}(x|1,w) = \frac{1}{A}[f_{y|x^*w}(y|1,w) - B], \tag{9}$$

$$f_{x|x^*w}(x|0,w) = \frac{1}{A}[f_{y|x^*w}(y|0,w) - B], \tag{10}$$

where

$$A = \frac{f_{y|wz}(y|w,1) - f_{y|wz}(y|w,0)}{f_{x|wz}(x|w,1) - f_{x|wz}(x|w,0)},$$

$$B = \frac{f_{y|wz}(y|w,0) f_{x|wz}(x|w,1) - f_{y|wz}(y|w,1) f_{x|wz}(x|w,0)}{f_{x|wz}(x|w,1) - f_{x|wz}(x|w,0)}.$$

Again, the denominator should not be equal to zero if z is a valid instrumental variable for

x^*. Finally, we obtain the expression of $f_{yx|wz}(y, x|w, z)$:

$$f_{yx|wz}(y, x|w, z) = \frac{f_{y|wz}(y|w, z)}{A}\{f_{y|x^*w}(y|1, w) + f_{y|x^*w}(y|0, w) \tag{11}$$

$$-\frac{1}{f_{y|wz}(y|w, z)} f_{y|x^*w}(y|1, w) f_{y|x^*w}(y|0, w) - B\}.$$

This equation holds for $z = 1$ and $z = 0$ so that we can solve for $f_{y|x^*w}$ as follows:

$$f_{y|x^*w}(y|1, w) + f_{y|x^*w}(y|0, w) = C + B, \tag{12}$$

$$f_{y|x^*w}(y|1, w) f_{y|x^*w}(y|0, w) = D, \tag{13}$$

where

$$C = \frac{f_{yx|wz}(y, x|w, 1) - f_{yx|wz}(y, x|w, 0)}{f_{x|wz}(x|w, 1) - f_{x|wz}(x|w, 0)},$$

$$D = \frac{f_{y|wz}(y|w, 0) f_{yx|wz}(y, x|w, 1) - f_{y|wz}(y|w, 1) f_{yx|wz}(y, x|w, 0)}{f_{x|wz}(x|w, 1) - f_{x|wz}(x|w, 0)}.$$

Therefore, the density $f_{y|x^*w}$ can be solved as follows:

$$f_{y|x^*w}(y|x^*, w)|_{x^*=0,1} = \frac{1}{2}[(C + B) \pm \sqrt{(C + B)^2 - 4D}]. \tag{14}$$

Obviously, $f_{y|x^*w}(y|1, w)$ and $f_{y|x^*w}(y|0, w)$ are symmetric because the values 0 or 1 are just symbols for two different statuses. The exact solution of $f_{y|x^*w}(y|1, w)$ and $f_{y|x^*w}(y|0, w)$ can be obtained if we know the sign of $f_{y|x^*w}(y|1, w) - f_{y|x^*w}(y|0, w)$. First, if we know x^* and z are positively correlated conditional on w, we have $f_{x^*|wz}(1|w, 1) > f_{x^*|wz}(1|w, 0)$.

Assumption 2.4: $f_{x^*|wz}(1|w, 1) > f_{x^*|wz}(1|w, 0)$.

Therefore, the sign of $f_{y|x^*w}(y|1, w) - f_{y|x^*w}(y|0, w)$ can be determined through equation (4) as follows:

$$f_{x^*|wz}(1|w, 1) - f_{x^*|wz}(1|w, 0) = \frac{f_{y|wz}(y|w, 1) - f_{y|wz}(y|w, 0)}{f_{y|x^*w}(y|1, w) - f_{y|x^*w}(y|0, w)}. \tag{15}$$

Second, if we assume the misclassification error is not very severe so that x^* and x are still

positively correlated conditional on w, then we have $f_{x|x^*w}(1|1,w) > f_{x|x^*w}(1|0,w)$.

Assumption 2.5: $f_{x|x^*w}(1|1,w) > f_{x|x^*w}(1|0,w)$.

The solution of $f_{y|x^*w}(y|1,w)$ and $f_{y|x^*w}(y|0,w)$ can be determined from (9-10) as follows:

$$f_{x|x^*w}(1|1,w) - f_{x|x^*w}(1|0,w) = \frac{1}{A}\Big|_{x=1}[f_{y|x^*w}(y|1,w) - f_{y|x^*w}(y|0,w)]. \tag{16}$$

Therefore, we have the following theorem:

Theorem 1 *(Identification) Suppose that Assumptions 1, 2, 2.1, 2.2, 2.3 and 2.4 (or 2.5) are satisfied. Then the model $f_{y|x^*w}$ is nonparametrically identifiable and directly estimable.*

Obviously, if the instrumental variable has more values in its support, the model is over-identified.

4 The semi-parametric MLE

This section considers the estimation of a parametric model in the form of a conditional density function of y as follows:

$$f_{y|x^*w}(y|x^*, w; \theta_0), \tag{17}$$

where $f_{y|x^*w}$ is known up to the unknown parameter θ_0, and x^* is a 0-1 dichotomous variable subject to misclassification error. We observe x as a proxy of x^* with y and w. Define the nuisance parameters as

$$\gamma_0 = [f_{ywz}(y,w,z), f_{xwz}(x,w,z), f_{wz}(w,z)]^T$$

From equation (11), we have

$$f_{y|xwz}(y|x,w,z;\theta_0,\gamma_0) = \frac{f_{y|wz}(y|w,z)}{f_{x|wz}(x|w,z)}\frac{1}{A}\{f_{y|x^*w}(y|1,w;\theta_0) + f_{y|x^*w}(y|0,w;\theta_0) \tag{18}$$

$$-\frac{1}{f_{y|wz}(y|w,z)}f_{y|x^*w}(y|1,w;\theta_0)f_{y|x^*w}(y|0,w;\theta_0) - B\}.$$

The nuisance parameter γ_0 can be estimated nonparametrically as follows:

$$\widehat{\gamma} = \left(\widehat{f}_{ywz}(y,w,z), \widehat{f}_{xwz}(x,w,z), \widehat{f}_{wz}(w,z)\right)^T$$

where

$$\widehat{f}_{ywz}(y,w,z) = \frac{1}{n}\sum_{i=1}^{n} I(z_i = z)\left[\frac{1}{h^{r+1}}K\left((\frac{y-y_i}{h}, \frac{w-w_i}{h})^T\right)\right],$$

$$\widehat{f}_{xwz}(x,w,z) = \frac{1}{n}\sum_{i=1}^{n} I(x_i = x)I(z_i = z)\left[\frac{1}{h^{r}}K(\frac{w-w_i}{h})\right].$$

$$\widehat{f}_{wz}(w,z) = \frac{1}{n}\sum_{i=1}^{n} I(z_i = z)\left[\frac{1}{h^{r}}K(\frac{w-w_i}{h})\right].$$

The constant r is the dimension of w. The function $I(.)$ is an indicator function and the function $K(.)$ is a known kernel function with bandwidth h.

We then can semiparametrically estimate the unknown parameter of interest, $\theta_0 \in \Theta$, through the density function $f_{y|xwz}$. The semi-parametric MLE is defined as

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \ln f_{y|xwz}(y_i|x_i, w_i, z_i; \theta, \widehat{\gamma}) \tag{19}$$

with $f_{y|xwz}(y_i|, x_i, w_i, z_i; \theta)$ the conditional density in which we replace $f_{y|wz}$ and $f_{x|wz}$ with their nonparametric estimators.

The semiparametic estimator in this paper falls into the framework introduced in section 8.3 of Newey and McFadden (1994). The estimator in this paper can be considered as an application of the general semiparametric estimator in their chapter. We will therefore make similar assumptions and will just give a brief discussion if it has been covered in the chapter. Let $\omega := (y, x, w, z)$. Define the score function as

$$g(\omega, \theta, \gamma) = \nabla_\theta \ln f_{y|xwz}(\omega, \theta, \gamma)$$

To guarantee the consistency of the estimator, we make the following assumptions:

Assumption 4.1: θ_0 is identifiable in $f_{y|x^*w}(y|x^*w;\theta)$, $\theta_0 \in \Theta$, and Θ is compact.

Assumption 4.2: $\omega \in \mathcal{W}$ *and* $\mathcal{W}$ *is a compact set.*

Assumption 4.3: $f_{y|x^*w}(y|x^*w;\theta)$ *is continuously differentiable in* θ*, and* $E\left[\left\|f_{y|x^*w}(y|x^*w;\theta)\right\|^2\right] < \infty$.

Assumption 4.4: There are constants $0 < m_0 < m_1 < \infty$ *such that for all* $x^* \in \{0,1\}$, $\omega \in \mathcal{W}$, *and* $\theta \in \Theta$

$$m_0 \leq f_{y|x^*w}(y|x^*,w;\theta) \leq m_1,$$

$$\left|\nabla_\theta f_{y|x^*w}(y|x^*,w;\theta)\right| \leq m_1,$$

$$m_0 \leq f_{wz}(w,z).$$

Assumption 4.5: $K(u)$ *is differentiable of order* d*, the derivatives of order* d *are bounded.* $K(u)$ *is zero outside a bounded set.* $\int_{-\infty}^{\infty} K(x)dx = 1$ *, and there is a positive integer* m *such that for all* $j < m$, $\int_{-\infty}^{\infty} K(u)\left[\otimes_{l=1}^{j} u\right] du = 0$.

Assumption 4.6: There is a version of $\gamma_0(\omega)$ *that is continuously differentiable to order* d *with bounded derivatives on an open set containing* $\mathcal{W}$.

Assumption 4.7: $nh^{r+1}/\ln n \to \infty$, $\sqrt{n}h^{2m} \to 0$, *and* $\sqrt{n}\ln n/(nh^{r+1+2d}) \to 0$ *as* $n \to \infty$.

Theorem 2 *(Consistency) Supposed that Assumptions 4.1-4.7 and the assumptions of Theorem 1 are satisfied. Then*

$$\hat{\theta} \xrightarrow{p} \theta_0.$$

Proof. See the Appendix. ■

To obtain the asymptotic normality of $\hat{\theta}$, we need the following assumptions:

Assumption 4.8: $\hat{\theta} \xrightarrow{p} \theta_0 \in interior(\Theta)$.

Assumption 4.9: There are function $b(\omega)$, $\varepsilon > 0$ with $E\left|b(\omega)\right| < \infty$ and

$$\left\|\nabla_{\theta\theta} f_{y|x^*w}(y|x^*,w;\theta) - \nabla_{\theta\theta} f_{y|x^*w}(y|x^*,w;\theta_0)\right\| \leq b(\omega)\left\|\theta - \theta_0\right\|^{\varepsilon}.$$

Assumption 4.10: $\nabla_{\theta\theta} f_{y|x^*w}(y|x^*w;\theta_0)$ *exists and is nonsingular.*

Theorem 3 *(Asymptotic Normality) Supposed that Assumptions 4.8-4.10 and the assumptions of Theorem 2 are satisfied. Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, G_\theta^{-1} \Omega G_\theta^{-1\prime}), \tag{20}$$

where

$$G_\theta = E\left[\nabla_\theta g(\omega, \theta_0, \gamma_0)\right],$$

$$\Omega = Var\left[g(\omega, \theta_0, \gamma_0) + \delta(\omega)\right],$$

$$\delta(\omega) = v(\omega) - E\left[v(\omega)\right],$$

$$v(\widetilde{\omega}) = E\left[\left. \nabla_\gamma g(\omega, \theta_0, \gamma)|_{\gamma=\gamma_0(\widetilde{\omega})} \right| \widetilde{\omega}\right].$$

Proof. See the Appendix. ■

5 Simulation

This section applies the method developed above to a probit model with a mismeasured 0-1 dichotomous explanatory variable. The conditional density function of the probit model is

$$\begin{aligned} f^*(y|x^*, w; \theta) &= P(y, x^*, w; \theta)^y (1 - P(y, x^*, w; \theta))^{1-y} \\ P(y, x^*, w; \theta) &= \Phi(\beta_0 + \beta_1 x^* + \beta_2 w) \end{aligned} \tag{21}$$

where $\theta = (\beta_0, \beta_1, \beta_2)'$ and Φ is the standard normal cdf. Three estimators are considered: The first is the ML probit estimator that uses mismeasured covariate x in the primary sample as if it were accurate, i.e., it ignores the measurement error. The MLE is not consistent. The second estimator is the infeasible ML probit estimator that uses the latent true x^* as covariate. This estimator is consistent and has the smallest asymptotic variance of all estimators that we consider. The conditional density function is $f^*(y|x^*, w; \theta)$. The third estimator is the semi-parametric MLE developed above that uses the instrumental variable. For each estimator, we report Root Mean Squared Error (RMSE), the average bias of estimates, and the standard deviation of the estimates over the replications. Table 1 shows that the MLE

which ignores the misclassification error is significantly biased as expected. The bias of the coefficient of the mismeasured independent variable is larger than the bias of the coefficient of the other covariate or the constant. The small-sample biases in the new semi-parametric MLE are similar to those of the other consistent estimator..

In all cases the MSE of the infeasible MLE is much smaller than that of the other consistent estimators. The loss of precision is associated with the fact that x^* is not observed, but that we must integrate with respect to its distribution given x, w and z. As the misclassification probability changes, the semi-parametric MLE performs well in each case.

6 Empirical Illustration

The section applies the estimator to a probit model of labor supply, which investigate the impact of education on women's labor supply. The population considered are all the women at the age from 18 to 65 who still lives with their parents and have left the school. The parents' education level is used as the instrumental variable in the model. The education level is treated as a dichotomous variable, which equals zero if an individual finished high school education or less. If a sample contains the information on parents' education or other instrumental variables, we can consider a lager population. The dependent variable is a dichotomous indicator of employment status. The independent variable contains education, age, work experience, and race. The marital status is not contained in the model because only less than 2 percent of 1457 women are married in the sample. The data are from the March supplement of the 2002 Current Population Survey (CPS). The joint distribution of women and their parents' education level is shown in Table 2. Table 3 contains the descriptive statistics of other variables.

The developed estimator uses parents' education level as an instrument, and allows the misclassification error to be correlated with other explanatory variables, such as age, work experience, and race. The estimation results in Table 4 contain two estimates. The second and the third columns contain the estimate ignoring misclassification error. And the estimates of the developed method is shown in the last two columns. The asymptotic standard error of the new estimator is estimated through equation 8.18 in Newey and McFadden

(1994). The results show that the impact of education on women's labor supply is much larger than in the case that the misclassification error is ignored.

7 Conclusion

This paper provides a new method for using instrumental variables in the estimation of nonlinear models with measurement error. The results show that instrumental variables are actually very powerful under weak assumptions to estimate nonlinear errors-in-variables models. It is shown that a latent model can be nonparametrically identifiable and directly estimable using instrumental variables. The misclassification error can be correlated with not only the latent true value but also with other explanatory variables. The new estimator is shown to be $\sqrt{n}$ consistent, very easy to derive, and highly applicable.

8 Appendix

Proof. Theorem 2 (consistency)

We use Theorem 2.1 in Newey and McFadden (1994) to show the consistency of the estimator. Define

$$Q_n(\theta,\gamma) = \frac{1}{n}\sum_{i=1}^{n} g(\omega_i,\theta,\gamma)$$

$$Q_0(\theta,\gamma) = Eg(\omega,\theta,\gamma)$$

Then

$$\sup_{\theta\in\Theta}|Q_n(\theta,\widehat{\gamma}) - Q_0(\theta,\gamma_0)| \leq \sup_{\theta\in\Theta}|Q_n(\theta,\widehat{\gamma}) - Q_n(\theta,\gamma_0)| + \sup_{\theta\in\Theta}|Q_n(\theta,\gamma_0) - Q_0(\theta,\gamma_0)|$$

The first term on the right hand side can be bounded as follows

$$\sup_{\theta\in\Theta}|Q_n(\theta,\widehat{\gamma}) - Q_n(\theta,\gamma_0)| \leq \sup_{\omega\in\mathcal{W}}|\widehat{\gamma}(\omega) - \gamma_0(\omega)| \times \sup_{\theta\in\Theta}\sup_{\|\gamma-\gamma_0\|_\infty\leq\varepsilon}|\nabla_\gamma Q_n(\theta,\gamma)|$$

for some $\varepsilon \to 0$ as $n \to \infty$. By assumption 4.5-4.7, $\sup_{\omega\in\mathcal{W}}|\widehat{\gamma}(\omega) - \gamma_0(\omega)| = o_p(1)$. $\nabla_\gamma g(\theta,\gamma)$ can be found through equation 18. By assumptions 2.3 and 4.4, $\nabla_\gamma g(\theta,\gamma)$ is bounded, and

therefore, the second term on the right hand side is bounded. Thus we have

$$\sup_{\theta\in\Theta}|Q_n(\theta,\widehat{\gamma})-Q_n(\theta,\gamma_0)|=o_p(1)$$

By assumption 4.4, $\sup_{\theta\in\Theta}|Q_n(\theta,\gamma_0)-Q_0(\theta,\gamma_0)|=o_p(1)$. By Lemma 2.4 in Newey and McFadden (1994), we have

$$\sup_{\theta\in\Theta}|Q_n(\theta,\widehat{\gamma})-Q_0(\theta,\gamma_0)|=o_p(1).$$

The estimator $\widehat{\theta}$ is then consistent by Theorem 2.1 in Newey and McFadden (1994). ■

Proof. Theorem 3 (asymptotic normality)

We use Theorems 8.11 and 8.12 in Newey and McFadden (1994) to show the asymptotic normality of the estimator. First, we have

$$\sqrt{n}(\widehat{\theta}-\theta_0)=-\left[\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta g(\omega_i,\widetilde{\theta},\widehat{\gamma})\right]^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g(\omega_i,\theta_0,\widehat{\gamma})\right)$$

where $\widetilde{\theta}$ is between $\widehat{\theta}$ and θ_0. The key is to find the distribution of the second term on the right hand side. Note that $\nabla_\gamma g(\omega_i,\theta_0,\gamma)$ and $\nabla_{\gamma\gamma}g(\omega_i,\theta_0,\gamma)$ can be found easily from 18, and is a specific algebraic function of γ and $f_{y|x^*w}$. $\nabla_\gamma g(\omega_i,\theta_0,\gamma)$ and $\nabla_{\gamma\gamma}g(\omega_i,\theta_0,\gamma)$ are, therefore, bounded by assumptions 2.3 and 4.4. Expanding $g(\omega_i,\theta_0,\gamma)$ with respect to γ gives

$$\begin{aligned}g(\omega,\theta_0,\gamma) &= g(\omega,\theta_0,\gamma_0)+\nabla_\gamma g(\omega,\theta_0,\gamma_0)^T[\gamma(\omega)-\gamma_0(\omega)]\\ &\quad+[\gamma(\omega)-\gamma_0(\omega)]^T\nabla_{\gamma\gamma}g(\omega,\theta_0,\gamma_0)[\gamma(\omega)-\gamma_0(\omega)]+o(\|\gamma(\omega)-\gamma_0(\omega)\|^2)\end{aligned}$$

Therefore, for all γ with $\|\gamma(\omega)-\gamma_0(\omega)\|$ small enough, there exists a function $b(\omega)$ with $E[b(\omega)]<\infty$ and

$$\left|g(\omega,\theta_0,\gamma)-g(\omega,\theta_0,\gamma_0)-\nabla_\gamma g(\omega,\theta_0,\gamma_0)^T[\gamma(\omega)-\gamma_0(\omega)]\right|\le b(\omega)\|\gamma(\omega)-\gamma_0(\omega)\|^2$$

Define

$$G(\omega, \gamma - \gamma_0) = \nabla_\gamma g(\omega, \theta_0, \gamma_0)^T \left[\gamma(\omega) - \gamma_0(\omega)\right].$$

We have shown that $E\left|\nabla_\gamma g(\omega, \theta_0, \gamma_0)\right| < \infty$. Let

$$v(\widetilde{\omega}) = E\left[\left.\nabla_\gamma g(\omega, \theta_0, \gamma)|_{\gamma=\gamma_0(\widetilde{\omega})}\right| \widetilde{\omega}\right].$$

We then have

$$\int G(\omega, \gamma) dF_0(\omega) = \int v(\omega)\gamma(\omega) d\omega.$$

Moreover, $v(\widetilde{\omega})$ is bounded and continuos almost everywhere by assumptions 4.3 and 4.6. By theorem 8.11 in Newey and McFadden (1994), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(\omega_i, \theta_0, \widehat{\gamma}) \xrightarrow{d} N(0, Var\left\{g(\omega, \theta_0, \gamma_0) + \delta(\omega)\right\})$$

with $\delta(\omega) = v(\omega) - E\left[v(\omega)\right]$. By assumption 4.8-4.10 with the same argument as before, we have

$$\frac{1}{n} \sum_{i=1}^{n} \nabla_\theta g(\omega_i, \overline{\theta}, \widehat{\gamma}) \xrightarrow{p} E\left[\nabla_\theta g(\omega, \theta_0, \gamma_0)\right]$$

Theorem 8.12 in Newey and McFadden (1994) then gives

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, G_\theta^{-1} \Omega G_\theta^{-1\prime})$$

where

$$G_\theta = E\left[\nabla_\theta g(\omega, \theta_0, \gamma_0)\right]$$

$$\Omega = Var\left[g(\omega, \theta_0, \gamma_0) + \delta(\omega)\right].$$

■

References

[1] Aigner, D., 1973, "Regression with a binary independent variable subject to errors of observations," Journal of Econometrics, 1, pp. 49-60.

[2] Amemiya, Y., 1985, "Instrumental variable estimator for the nonlinear errors-in-variables model," Journal of Econometrics, 28, pp 273-289.

[3] Amemiya, Y. and Fuller, W.A., 1988, "Estimation for the nonlinear functional relationship," the Annals of Statistics, 16, pp. 147-160.

[4] Black, D., M. C. Berger, and F. A. Scott, 2000, "Bounding parameter estimates with nonclassical measurement error," Journal of the American Statistical Association, 95, pp. 739-748.

[5] Bollinger, C., 1996, "Bounding mean regressions when a binary regressor is mismeasured," Journal of Econometrics, 73 (1996), pp 387-399.

[6] Bound, J. C. Brown, and N. Mathiowetz, 2001, "Measurement error in survey data," in J.J. Heckman and E. Leamer eds., Handbook of Econometrics Vol 5.

[7] Buzas, J. S., 1997, "Instrumental variable estimation in nonlinear measurement error models," Communications in Statistics, Part A - Theory and Methods, 26, pp 2861-2877.

[8] Carroll, R.J., D. Ruppert, and L.A. Stefanski, 1995, Measurement Error in Nonlinear Models. Chapman & Hall, New York.

[9] Carroll, .J. and L.A. Stefanski, 1990, "Approximate quasi-likelihood estimation in models with surrogate predictors," Journal of the American Statistical Association 85, pp. 652-663.

[10] Freeman, R., 1984, "Longitudinal analyses of the effects of trade unions," Journal of Labor Economics 2, pp1-26.

[11] Hausman, J., J. Abreveya and F. Scott-Morton, 1998, "Misclassification of the dependent variable in a discrete response setting," Journal of Econometrics, 87, pp239-269.

[12] Hausman, J., Ichimura, H., Newey, W., and Powell, J., 1991, "Identification and estimation of polynomial errors-in-variables models," Journal of Econometrics, 50, pp. 273-295.

[13] Hausman, J.A., W.K. Newey, and J.L. Powell, 1995, "Nonlinear errors in variables: estimation of some Engel curves," Journal of Econometrics 65, pp. 205-233.

[14] Horowitz, J., and C. Manski, 1995, "Identification and robustness with contaminated and corrupt data," Econometrica, 63, pp. 281-302.

[15] Kane, T. J., C. E. Rouse, and D. Staiger, (1999), "Estimating returns to schooling when schooling is misreported," NBER working paper #7235.

[16] Lewbel, A., 1998, "Semiparametric latent variable model estimation with endogenous or mismeasured regressors," Econometrica, vol. 66, pp. 105-121

[17] Lewbel, A., 2000, "Identification of the binary choice model with misclassification, Econometric Theory, 16 (2000), pp. 603-60.

[18] Lewbel, A., 2003, "Estimation of average treatment effects with misclassification," memo.

[19] Mahajan, A., 2003, "Misclassified regressors in binary choice models," memo.

[20] Molinari, F., 2004, "Partial identification of probability distributions with misclassified data," memo.

[21] Newey, W.K., 2001, "Flexible simulated moment estimation of nonlinear errors-in-variables models," Review of Economics and Statistics, 83(4), pp. 616-627.

[22] Newey, W. and D. McFadden, 1994, Large sample esitmation and hypothesis testing, in Handbook of Econometrics, R. Engle and D. McFadden, eds, vol. 4, pp2111-2245.

[23] Ramalho, E., 2002, "Regression models for choice-based samples with misclassification in the response variable," Journal of Econometrics 106, pp171-201

[24] Schennach, S., 2004, "Instrumental variable estimation of nonlinear errors-in-variables models," memo.

Table 1: Simulation results of Probit model: sample size 500; number of repetitions 200.

	β_1			β_2			β_0		
$p = .3$ $q = .2$	Root MSE	Mean bias	Std. dev.	Root MSE	Mean bias	Std. dev.	Root MSE	Mean bias	Std. dev.
Ignoring meas. error	.541	-.523	.139	.181	-.103	.149	.293	.277	.095
True x^*	.160	.015	.159	.166	-.013	.165	.104	.000	.104
I.V.	.421	-.095	.410	.307	-.087	.295	.263	.045	.259
$p = .3 - .1w$ $q = .2 + .1w$	Root MSE	Mean bias	Std. dev.	Root MSE	Mean bias	Std. dev.	Root MSE	Mean bias	Std. dev.
Ignoring meas. error	.538	-.520	.136	.210	-.150	.147	.290	.275	.094
True x^*	.157	.012	.157	.165	-.011	.164	.104	.000	.104
I.V.	.409	-.124	.389	.332	-.138	.302	.238	.061	.230
$p = .3 + .1w$ $q = .2 + .1w$	Root MSE	Mean bias	Std. dev.	Root MSE	Mean bias	Std. dev.	Root MSE	Mean bias	Std. dev.
Ignoring meas. error	.509	-.491	.137	.176	-.094	.149	.279	.263	.093
True x^*	.160	.015	.159	.165	-.014	.165	.104	-.001	.104
I.V.	.318	-.108	.299	.307	-.071	.298	.205	.052	.198

Note:
1) $\beta_1 = 1$,$\beta_2 = 1$,$\beta_0 = .5$, $x^* = I(\epsilon < .6)$; $z = I(\epsilon + \delta < .6)$, $\epsilon \sim Uniform(0, 1)$, $\delta \sim N(0, .04)$, $(\rho_{x^*z} \approx .67)$, $w \sim N(0, .25)$.
2) $\Pr(x = 0|x^* = 1, w) = \min(1, \max(0, p))$, $\Pr(x = 1|x^* = 0, w) = \min(1, \max(0, q))$.
3) $K(x) = .5(3 - x^2)\phi(x)$ and $h = .2$, where $\phi(x)$ is the standard normal density.

Table 2: Joint distribution of education (1457 observations)

	parents' education		
education	high school or lower	college or higher	total
high school or lower	.337	.163	.500
college or higher	.204	.296	.500
total	.541	.459	1

Table 3: Summary statistics of variables (1457 observations)

	mean	std.dev	min	max
employment	.822	.382	0	1
number of children	.577	.906	0	8
weeks worked in last year	41.704	12.720	26	52
age	24.173	6.133	18	56
race (white=1)	.781	.414	0	1

Table 4: Estimation results

	Ignoring meas. error		I.V.	
	estimate	std.dev	estimate	std.dev
education	.3342	.0859	.9578	.1992
work experience	.0296	.0032	.0235	.0057
age	.0155	.0080	.0100	.0146
number of kids	-.0311	.0447	-.0222	.0996
race	.2599	.0955	.1839	.1838
constant	-1.2664	.2388	-1.1367	.6700