Economics Letters 100 (2008) 381–384

Contents lists available at ScienceDirect

Economics Letters

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

Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments

Xiaohong Chen [a,1], Yingyao Hu [b,2], Arthur Lewbel [c,*]

[a] Department of Economics, Yale University, Box 208281, New Haven, CT 06520-8281, USA
[b] Department of Economics, Johns Hopkins University, 440 Mergenthaler Hall, 3400 N. Charles Street, Baltimore, MD 21218, USA
[c] Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA

ARTICLE INFO

Article history:
Received 6 December 2007
Received in revised form 28 February 2008
Accepted 6 March 2008
Available online 20 March 2008

Keywords:
Misclassification error
Identification
Nonparametric regression

JEL classification:
C14
C20

ABSTRACT

We observe a dependent variable and some regressors, including a mismeasured binary regressor. We provide identification of the nonparametric regression model containing this misclassified dichotomous regressor. We obtain identification without parameterizations or instruments, by assuming the model error isn't skewed.

1. Motivation

We provide identification of a nonparametric regression model with a dichotomous regressor subject to misclassification error. The available sample information consists of a dependent variable and a set of regressors, one of which is binary and error-ridden with misclassification error that has unknown distribution. Our identification strategy does not parameterize any regression or distribution functions, and does not require additional sample information such as instrumental variables, repeated measurements, or an auxiliary sample. Our main identifying assumption is that the regression model error has zero conditional third moment. The results include a closed-form solution for the unknown distributions and the regression function.

Dichotomous (binary) variables, such as union status, smoking behavior, and having a college degree or not, are involved in many economic models. Measurement errors in dichotomous variables take the form of misclassification errors, i.e., some observations where the variable is actually a one may be misclassified as a zero, and vice versa. A common source of misclassification errors is self-reporting, where people may have psychological or economic incentives to misreport dichotomous variables (see Bound et al. (2001) for a survey). Misclassification may also arise from ordinary coding or reporting errors, e.g., Kane et al. (1999) report substantial classification errors in both self-reports and transcript reports of educational attainment. Unlike ordinary mismeasured regressors, misclassified regressors cannot possess the properties of classically mismeasured variables, in particular, classification errors are not independent of the underlying true regressor, and are in general not mean zero.

As with ordinary mismeasured regressors, estimated regressions with a misclassified regressor are inconsistent, and the latent true regression model based just on conditionally mean zero model errors is generally not identified in the presence of a misclassified regressor. To identify the latent model, we must either impose additional assumptions or possess additional sample information. One popular additional assumption is to assume the measurement error distribution belong to some parametric family. Additional sample information often used to obtain identification includes an instrumental variable or a repeated measurement in the same sample, or a secondary sample. See, e.g., Carroll et al. (2006), and Chen et al. (2007) for detailed recent reviews on existing approaches to measurement error problems.

In this note we obtain identification without parameterizing errors and without auxiliary information like instrumental variables, repeated measurements, or a secondary sample. A related result is Chen et al. (2008). We show here that, given some mild regularity conditions, a nonparametric mean regression with a misclassified

* Corresponding author. Tel.: +1 617 522 3678.
E-mail addresses: xiaohong.chen@yale.edu (X. Chen), yhu@jhu.edu (Y. Hu), lewbel@bc.edu (A. Lewbel).
[1] Tel.: +1 203 432 5852.
[2] Tel.: +1 410 516 7610.

0165-1765/$ – see front matter
doi:10.1016/j.econlet.2008.03.015

binary regressor is identified (and can be solved in closed form) if the latent regression error has zero conditional third moment, as would be the case if the regression error were symmetric. We also briefly discuss how simple estimators might be constructed based on our identification method.

2. Identification

We are interested in a regression model as follows:

$$Y = m(X^*, W) + \eta, \quad E(\eta|X^*, W) = 0 \tag{2.1}$$

where Y is the dependent variable, $X^* \in \mathcal{X} = \{0, 1\}$ is the dichotomous regressor subject to misclassification error, and W is an error-free covariate vector. We are interested in the nonparametric identification of the regression function $m()$. The regression error η need not be independent of the regressors X^* and W, so we have conditional density functions

$$f_{Y|X^*,W}(y|x^*, w) = f_{\eta|X^*,W}(y - m(x^*, w)|x^*, w). \tag{2.2}$$

In a random sample, we observe $(X, Y, W) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{W}$, where X is a proxy or a mismeasured version of X^*. We assume

Assumption 2.1. $f_{Y|X^*,W,X}(y|x^*,w,x) = f_{Y|X^*,W}(y|x^*,w)$ for all $(x,x^*,y,w) \in \mathcal{X} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{W}$.

This assumption implies that the measurement error in X is independent of the dependent variable Y conditional on the true value X^* and the covariate W, and so X is independent of the regression error η conditional on X^* and W. This is analogous to the classical measurement error assumption of having the measurement error independent of the regression model error. This assumption may be problematic in applications where the same individual who provides the source of misclassification by supplying X also helps determine the outcome Y, however, this is a standard assumption in the literature of mismeasured and misclassified regressors. See, e.g., Li (2002), Schennach (2004), Mahajan (2006), Lewbel (2007a) and Hu (2006).

By construction, the relationship between the observed density and the latent ones are as follows:

$$\begin{aligned} f_{Y|X,W}(y|x, w) &= \sum_{x^*} f_{Y|X^*,W,X}(y|x^*, w, x) f_{X^*|X,W}(x^*|x, w) \\ &= \sum_{x^*} f_{\eta|X^*,W}\left(y - m(x^*, w)|x^*, w\right) f_{X^*|X,W}(x^*|x, w). \end{aligned} \tag{2.3}$$

Using the fact that X and X^* are 0–1 dichotomous, define the following simplifying notation: $m_0(w) = m(0, w)$, $m_1(w) = m(1, w)$, $\mu_0(w) = E(Y|X=0, w)$, $\mu_1(w) = E(Y|X=1, w)$, $p(w) = f_{X^*|X,W}(1|0, w)$, and $q(w) = f_{X^*|X,W}(0|1, w)$. Eq. (2.3) is then equivalent to

$$\begin{pmatrix} f_{Y|X,W}(y|0, w) \\ f_{Y|X,W}(y|1, w) \end{pmatrix} = \begin{pmatrix} 1 - p(w) & p(w) \\ q(w) & 1 - q(w) \end{pmatrix} \begin{pmatrix} f_{\eta|X^*,W}(y - m_0(w)|0, w) \\ f_{\eta|X^*,W}(y - m_1(w)|1, w) \end{pmatrix}. \tag{2.4}$$

Since $f_{\eta|X^*,W}$ has zero mean, we obtain

$$\begin{aligned} \mu_0(w) &= (1 - p(w))m_0(w) + p(w)m_1(w) \text{ and } \mu_1(w) \\ &= q(w)m_0(w) + (1 - q(w))m_1(w). \end{aligned} \tag{2.5}$$

Assume

Assumption 2.2. $m_1(w) \neq m_0(w)$ for all $w \in \mathcal{W}$.

This assumption means that X^* has a nonzero effect on the conditional mean of Y, and so is a relevant explanatory variable, given W. We may now solve Eq. (2.5) for $p(w)$ and $q(w)$, yielding

$$p(w) = \frac{\mu_0(w) - m_0(w)}{m_1(w) - m_0(w)} \text{ and } q(w) = \frac{m_1(w) - \mu_1(w)}{m_1(w) - m_0(w)} \tag{2.6}$$

Without loss of generality, we assume,

Assumption 2.3. for all $w \in \mathcal{W}$, (i) $\mu_1(w) > \mu_0(w)$; (ii) $p(w) + q(w) < 1$.

Assumption 2.3(i) is not restrictive because one can always redefine X as $1 - X$ if needed. Assumption 2.3(ii) implies that the ordering of $m_1(w)$ and $m_0(w)$ is the same as that of $\mu_1(w)$ and $\mu_0(w)$ because $1 - p(w) - q(w) = \frac{\mu_1(w) - \mu_0(w)}{m_1(w) - m_0(w)}$. The intuition of Assumption 2.3 (ii) is that the total misclassification probability is not too large so that $\mu_1(w) > \mu_0(w)$ implies $m_1(w) > m_0(w)$ (see, e.g., Lewbel, 2007a) for a further discussion of this assumption). In summary, we have

$$m_1(w) \geq \mu_1(w) > \mu_0(w) \geq m_0(w).$$

The condition $p(w) + q(w) \neq 1$ also guarantees that the matrix $\begin{pmatrix} 1 - p(w) & p(w) \\ q(w) & 1 - q(w) \end{pmatrix}$ in Eq. (2.4) is invertible. If we then plug into Eq. (2.4) the expressions for $p(w)$ and $q(w)$ in Eq. (2.6), we obtain for $j = 0,1$

$$\begin{aligned} f_{\eta|X^*,W}(y - m_j(w)|j, w) &= \frac{\mu_1(w) - m_j(w)}{\mu_1(w) - \mu_0(w)} f_{Y|X,W}(y|0, w) \\ &\quad + \frac{m_j(w) - \mu_0(w)}{\mu_1(w) - \mu_0(w)} f_{Y|X,W}(y|1, w). \end{aligned} \tag{2.7}$$

Eq. (2.7) is our vehicle for identification. Given any information about the distribution of the regression error η, Eq. (2.7) provides the link between that information and the unknowns $m_0(w)$ and $m_1(w)$, along with the observable density $f_{Y|X,W}$ and observable conditional means $\mu_0(w)$ and $\mu_1(w)$. The specific assumption about η that we use to obtain identification is this:

Assumption 2.4. $E(\eta^3|X^*, W) = 0$.

A sufficient though much stronger than necessary condition for this assumption to hold is that $f_{\eta|X^*,W}$ be symmetric for each $x^* \in \mathcal{X}$ and $w \in \mathcal{W}$. Notice that the regression model error η need not be independent of the regressors X^*, W, and in particular our assumptions permit η to have heteroskedasticity of completely unknown form.

Let ϕ denote the characteristic function and

$$\begin{aligned} \phi_{\eta|X^*=j,w}(t) &= \int e^{it\eta} f_{\eta|X^*,W}(\eta|j, w) d\eta \\ \phi_{Y|X=j,w}(t) &= \int e^{ity} f_{Y|X,W}(y|j, w) dy. \end{aligned}$$

Then Eq. (2.7) implies that for any real t

$$\begin{aligned} &\ln\left(e^{itm_j(w)} \phi_{\eta|X^*=j,w}(t)\right) \\ &= \ln\left(\frac{\mu_1(w) - m_j(w)}{\mu_1(w) - \mu_0(w)} \phi_{Y|X=0,w}(t) + \frac{m_j(w) - \mu_0(w)}{\mu_1(w) - \mu_0(w)} \phi_{Y|X=1,w}(t)\right). \end{aligned} \tag{2.8}$$

Notice that

$$\begin{aligned} \frac{\partial^3}{\partial t^3} \ln\left(e^{itm_j(w)} \phi_{\eta|X^*=j,w}(t)\right)\Big|_{t=0} &= \frac{\partial^3}{\partial t^3} \ln \phi_{\eta|X^*=j,w}(t)\Big|_{t=0} \\ &= -iE\left(\eta^3|X^* = j, W = w\right). \end{aligned}$$

Assumption 2.4 therefore implies that for $j = 0,1$

$$G(m_j(w)) = 0, \tag{2.9}$$

where

$$G(z) \equiv i \frac{\partial^3}{\partial t^3} \ln\left(\frac{\mu_1(w) - z}{\mu_1(w) - \mu_0(w)} \phi_{Y|X=0,w}(t) + \frac{z - \mu_0(w)}{\mu_1(w) - \mu_0(w)} \phi_{Y|X=1,w}(t)\right)\Big|_{t=0}.$$

This equation shows that the unknowns $m_0(w)$ and $m_1(w)$ are two roots of the cubic function $G(\cdot)$ in Eq. (2.9). Suppose the three roots of this equation are $r_a(w) \leq r_b(w) \leq r_c(w)$. In fact, we have

$$r_a(w) \leq m_0(w) \leq \mu_0(w) < \mu_1(w) \leq m_1(w) \leq r_c(w),$$

which implies bounds on $m_0(w)$ and $m_1(w)$. To obtain point identification of $m_j(w)$, we need to be able to uniquely define which roots of the cubic function $G(\cdot)$ correspond to $m_0(w)$ and $m_1(w)$. This is provided by the following assumption.

Assumption 2.5. Assume

$$E\left[(Y-\mu_0(w))^3|X=0,W=w\right]\geq 0\geq E\left[(Y-\mu_1(w))^3|X=1,W=w\right]$$

and, when an equality with $X=j$ holds, assume $\frac{dG(z)}{dz}|_{z=\mu_j(w)}>0$.

It follows from Assumption 2.5 that

$$r_a(w)\leq\mu_0(w)<r_b(w)<\mu_1(w)\leq r_c(w).$$

Since $m_0(w)\leq\mu_0(w)<\mu_1(w)\leq m_1(w)$, we then have point identification by $m_0(w)=r_a(w)$ and $m_1(w)=r_c(w)$. Note that Assumption 2.5 is directly testable from the data. Based on the definition of skewness of a distribution and $\mu_0(w)<E(Y|W=w)<\mu_1(w)$, Assumption 2.5 implies that the distributions $f_{Y|X,W}(y|1, w)$ and $f_{Y|X,W}(y|0, w)$ are skewed towards the unconditional mean $E(Y|W=w)$ compared with each conditional mean. An analogous result is Lewbel (1997), who obtains identification in a classical measurement error context without auxiliary data exploiting skewness in a different way.

Notice that Assumption 2.4 implies that $E[(Y-m_0(w))^3|X=0, W=w]=0$ and $E[(Y-m_1(w))^3|X=1, W=w]=0$. Assumption 2.5 then implies

$$E\left[(Y-\mu_0(w))^3|X=0,W=w\right]\geq E\left[(Y-m_0(w))^3|X=0,W=w\right]$$

and

$$E\left[(Y-\mu_1(w))^3|X=1,W=w\right]\leq E\left[(Y-m_1(w))^3|X=1,W=w\right].$$

The third moments on the left-hand sides are observed from the data and the right-hand sides contain the latent third moments. We may treat the third moments $E[(Y-\mu_j(w))^3|X=j, W=w]$ as a naive estimator of the true moments $E[(Y-m_j(w))^3|X=j, W=w]$. Assumption 2.4 implies that the latent third moments are known to be zero. Assumption 2.5 implies that the sign of the bias of the naive estimator is different in two subsamples corresponding to $X=0$ and $X=1$.

We leave the detailed proof to the Appendix and summarize the result as follows:

Theorem 2.1. *Suppose that* Assumptions 2.1–2.5 *hold in* Eq. (2.1). *Then, the density $f_{Y,X,W}$ uniquely determines $f_{Y|X*,W}$ and $f_{X*,X,W}$.*

Identification of the distributions $f_{Y|X*,W}$ and $f_{X*,X,W}$ by Theorem 2.1 immediately implies that the regression function $m(X^*, W)$, the conditional distribution of the regression error, $f_{\eta|X*,W}$, and the conditional distribution of the misclassification error (the difference between X and X^*) are all identified.

3. Conclusions and possible estimators

We have shown that a nonparametric regression model containing a dichotomous misclassified regressor can be identified without any auxiliary data like instruments, repeated measurements, or a secondary sample (such as validation data), and without any parametric restrictions. The only identifying assumptions are some regularity conditions and the assumption that the regression model error has zero conditional skewness.

We have focused on identification, so we conclude by briefly describing how estimators might be constructed based on our identification method. One possibility would be to substitute consistent estimators of the conditional means $\mu_j(w)$ and characteristic functions $\phi_{Y|X,w}(t)$ into Eq. (2.9), and solve the resulting cubic equation for estimates of $m_j(w)$. Another possibility is to observe that, based on the proof of our main theorem, the identifying equations can be written in terms of conditional mean zero expectations as

$$\begin{aligned}
&E\big((Y-\mu_j(w))I(X=j)|W=w\big)=0,\\
&E\big((Y^2-\upsilon_j(w))I(X=j)|W=w\big)=0,\\
&E\big((Y^3-\kappa_j(w))I(X=j)|W=w\big)=0,\\
&E\left(\begin{array}{c}2m_j(W)^3-3\dfrac{\upsilon_1(w)-\upsilon_0(w)}{\mu_1(w)-\mu_0(w)}m_j(W)^2\\ -\dfrac{3\upsilon_0(w)\mu_1(w)-3\upsilon_1(w)\mu_0(w)+\kappa_0(w)-\kappa_1(w)}{\mu_1(w)-\mu_0(w)}m_j(W)+\dfrac{\mu_1(w)\kappa_0(w)-\mu_0(w)\kappa_1(w)}{\mu_1(w)-\mu_0(w)}|W=w\end{array}\right)=0.
\end{aligned} \tag{2.10}$$

See the Appendix, particularly Eq. (A.6). We might then apply Ai and Chen (2003) to these conditional moments to obtain sieve estimates of $m_j(w)$, $\mu_j(w)$, $\upsilon_j(w)$, and $\kappa_j(w)$. Alternatively, the local GMM estimator of Lewbel (2007b) could be employed. If w is discrete or empty, or if these functions of w are finitely parameterized, then these estimators could be reduced to ordinary GMM.

Appendix A

Proof (Theorem 2.1). First, we introduce notations as follows: for $j=0,1$, $m_j(w)=m(j, w)$, $\mu_j(w)=E(Y|X=j, W=w)$, $p(w)=f_{X*|X,W}(1|0, w)$, $q(w)=f_{X*|X,W}(0|1, w)$, $\upsilon_j(w)=E(Y^2|X=j, W=w)$, and $\kappa_j(w)=E(Y^3|X=j, W=w)$.We start the proof with Eq. (2.3), which is equivalent to

$$\begin{pmatrix} f_{Y|X,W}(y|0,w)\\ f_{Y|X,W}(y|1,w)\end{pmatrix}=\begin{pmatrix}1-p(w) & p(w)\\ q(w) & 1-q(w)\end{pmatrix}\begin{pmatrix} f_{\eta|X*,W}(y-m_0(w)|0,w)\\ f_{\eta|X*,W}(y-m_1(w)|1,w)\end{pmatrix}. \tag{A.1}$$

Assumption 2.4 implies that $f_{\eta|X*,W}$ has zero mean. Therefore, we have

$$\begin{aligned}
\mu_0(w)&=(1-p(w))m_0(w)+p(w)m_1(w),\\
\mu_1(w)&=q(w)m_0(w)+(1-q(w))m_1(w).
\end{aligned}$$

By Assumption 2.2, we may solve for $p(w)$ and $q(w)$ as follows:

$$p(w)=\frac{\mu_0(w)-m_0(w)}{m_1(w)-m_0(w)}\ \text{and}\ q(w)=\frac{m_1(w)-\mu_1(w)}{m_1(w)-m_0(w)}. \tag{A.2}$$

We also have $1-p(w)-q(w)=\frac{\mu_1(w)-\mu_0(w)}{m_1(w)-m_0(w)}$. As discussed before, Assumption 2.3 implies that $m_1(w)\geq\mu_1(w)>\mu_0(w)\geq m_0(w)$ and

$$\begin{pmatrix} f_{\eta|X*,W}(y-m_0(w)|0,w)\\ f_{\eta|X*,W}(y-m_1(w)|1,w)\end{pmatrix}=\frac{1}{1-p(w)-q(w)}\begin{pmatrix}1-q(w) & -p(w)\\ -q(w) & 1-p(w)\end{pmatrix}\begin{pmatrix} f_{Y|X,W}(y|0,w)\\ f_{Y|X,W}(y|1,w)\end{pmatrix}.$$

Plug-in the expression of $p(w)$ and $q(w)$ in Eq. (A.2), we have

$$\begin{aligned}
f_{\eta|X*,W}(y-m_j(w)|j,w)=&\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}f_{Y|X,W}(y|0,w)\\
&+\frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}f_{Y|X,W}(y|1,w).
\end{aligned} \tag{A.3}$$

Let ϕ denote the characteristic function, $\phi_{\eta|X*=j,w}(t)=\int e^{it\eta}f_{\eta|X*,W}(\eta|j, w)d\eta$, and $\phi_{Y|X=j,w}(t)=\int e^{ity}f_{Y|X,W}(y|j, w)dy$. Eq. (A.3) implies that for any real t

$$e^{itm_j(w)}\phi_{\eta|X*=j,w}(t)=\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=0,w}(t)+\frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=1,w}(t).$$

We then consider the log transform

$$\ln\left(e^{itm_j(w)}\phi_{\eta|X*=j,w}(t)\right)=\ln\left(\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=0,w}(t)+\frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=1,w}(t)\right). \tag{A.4}$$

Assumption 2.4 implies that for $j=0, 1$

$$0=i\frac{\partial^3}{\partial t^3}\ln\left(\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=0,w}(t)+\frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=1,w}(t)\right)|_{t=0}. \tag{A.5}$$

When $t=0$, we have $\phi_{Y|X,W}(0)=1$, $\frac{\partial}{\partial t}\phi_{Y|X=j,W}(0)=i\mu_j$, $\frac{\partial^2}{\partial t^2}\phi_{Y|X=j,W}(0)=-\upsilon_j$, and $\frac{\partial^3}{\partial t^3}\phi_{Y|X=j}(0)=-i\kappa_j$. Furthermore, Eq. (A.5) implies $G(m_j)=0$ where

$$G(z)\equiv 2z^3-3\frac{\upsilon_1(w)-\upsilon_0(w)}{\mu_1(w)-\mu_0(w)}z^2 \quad (A.6)$$
$$-\frac{3\upsilon_0(w)\mu_1(w)-3\upsilon_1(w)\mu_0(w)+\kappa_0(w)-\kappa_1(w)}{\mu_1(w)-\mu_0(w)}z+\frac{\mu_1(w)\kappa_0(w)-\mu_0(w)\kappa_1(w)}{\mu_1(w)-\mu_0(w)}.$$

This cubic equation has two real roots $m_0(w)$ and $m_1(w)$, and has all real coefficients. Therefore its third root is also real. Suppose the three roots are $r_a(w)\le r_b(w)\le r_c(w)$ for each given w. Since $m_0(w)\ne m_1(w)$, we will never have $r_a(w)=r_b(w)=r_c(w)$. If the second largest of the three roots is between $\mu_0(w)$ and $\mu_1(w)$, i.e., $\mu_0(w)<r_b(w)<\mu_1(w)$, then we know the largest root $r_c(w)$ equals $m_1(w)$ and the smallest root $r_a(w)$ equals $m_0(w)$ because $m_1(w)\ge\mu_1(w)>\mu_0(w)\ge m_0(w)$. Given the shape of the cubic function, we know

$$G(z)\begin{cases} <0 & \text{if} \quad z<r_a(w) \\ >0 & \text{if} \quad r_a(w)<z<r_b(w) \\ <0 & \text{if} \quad r_b(w)<z<r_c(w) \\ >0 & \text{if} \quad r_c(w)<z \end{cases}.$$

That means the second largest of the three roots r_b is between $\mu_0(w)$ and $\mu_1(w)$ if $G(\mu_1(w))<0$ and $G(\mu_0(w))>0$. It is tedious but straightforward to show that $G(\mu_1(w))=E[(Y-\mu_1(w))^3|X=1, W=w]$ and $G(\mu_0(w))=E[(Y-\mu_0(w))^3|X=0, W=w]$. Therefore, Assumption 2.5 implies that $G(\mu_1(w))\le 0$ and $G(\mu_0(w))\ge 0$. Given the graph of the cubic function $G(\bullet)$, if $G(\mu_1(w))<0$, then $\mu_1(w)<m_1(w)$ implies that $m_1(w)$ equals the largest of the three roots. When $G(\mu_0(w))>0$, then $m_0(w)<\mu_0(w)$ implies that $m_0(w)$ equals the smallest of the three roots.

In case $E[(Y-\mu_1(w))^3|X=1, W=w]=0$, i.e., $G(\mu_1(w))=0$, $\mu_1(w)$ is a root of $G(\bullet)$. Given the graph of the cubic function $G(\bullet)$, we know

$$\frac{dG(z)}{dz}\begin{cases} \ge 0 & \text{at} \quad z=r_a(w) \\ \le 0 & \text{at} \quad z=r_b(w) \\ \ge 0 & \text{at} \quad z=r_c(w) \end{cases}.$$

That means the condition $\frac{dG(z)}{dz}|_{z=\mu_1(w)}>0$ guarantees that $\mu_1(w)$ is the largest root and equal to $m_1(w)$. If $E[(Y-\mu_0(w))^3|X=0, W=w]=0$, i.e. $G(\mu_0(w))=0$, $\mu_0(w)$ is a root of $G(\bullet)$. The condition $\frac{dG(z)}{dz}|_{z=\mu_0(w)}>0$ guarantees that $\mu_0(w)$ is the smallest root and equal to $m_0(w)$. In summary, Assumption 2.5 guarantees that $m_0(w)$ and $m_1(w)$ can be identified out of the three directly estimable roots.

After we have identified $m_0(w)$ and $m_1(w)$, $p(w)$ and $q(w)$ (or $f_{X^*|X,w}$) are identified from Eq. (A.2), and the density $f_{\eta|X^*,W}$ (or $f_{Y|X^*,W}$) is also identified from Eq. (A.3). Since X and W are observed in the data, identification of $f_{X^*|X,W}$ implies that of $f_{X^*,X,W}$. Thus, we have identified the latent densities $f_{Y|X^*,W}$ and $f_{X^*,X,W}$ from the observed density $f_{Y,X,W}$ under Assumptions 2.1–2.5. □

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71, 1795–1844.

Bound, J., Brown, C., Mathiowetz, N., 2001. In: Heckman, J.J., Leamer, E. (Eds.), Measurement Error in Survey Data. . Handbook of Econometrics, vol. 5. Elsevier Science.

Carroll, R., Ruppert, D., Stefanski, L., Crainiceanu, C., 2006. Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition. CRI, New York.

Chen, X., Hong, H., Nekipelov, D., 2007. Measurement error models, unpublished manuscript, Yale University.

Chen, X., Hu, Y., Lewbel, A., 2008. Nonparametric Identification and Estimation of Nonclassical Errors-in-Variables Models Without Additional Information, Statistica Sinica, forthcoming.

Hu, Y. (2006): "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables," Working Paper, University of Texas at Austin.

Kane, T.J., Rouse, C.E., Staiger, D., 1999. Estimating Returns to Schooling When Schooling is Misreported, NBER working paper #7235.

Lewbel, A., 1997. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. Econometrica 65, 1201–1213.

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

Lewbel, A., 2007b. A local generalized method of moments estimator. Economics Letters 94, 124–128.

Li, T., 2002. Robust and consistent estimation of nonlinear errors-in-variables models. Journal of Econometrics 110, 1–26.

Mahajan, A., 2006. Identification and estimation of regression models with misclassification. Econometrica 74, 631–665.

Schennach, S., 2004. Estimation of nonlinear models with measurement error. Econometrica 72, 33–75.