Nonparametric Identification of Regression Models Containing a Misclassified Dichotomous Regressor Without Instruments

Authors: Xiaohong Chen, Yingyao Hu, Arthur Lewbel

This work is posted on eScholarship@BC,
Boston College University Libraries.

Boston College Working Papers in Economics, 2007

Originally posted on: http://ideas.repec.org/p/boc/bocoec/675.html

Nonparametric Identification of Regression Models Containing a Misclassified Dichotomous Regressor Without Instruments

Xiaohong Chen* Yale University

Yingyao Hu† Johns Hopkins University

Arthur Lewbel‡ Boston College

First version: November 2006; Revised July 2007.

Abstract

This note considers nonparametric identification of a general nonlinear 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.

Keywords: misclassification error; identification; nonparametric regression;

*Department of Economics, Yale University, Box 208281, New Haven, CT 06520-8281, USA. Tel: 203-432-5852. Email: xiaohong.chen@yale.edu.

†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.

‡Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. Tel: 617-522-3678. Email: lewbel@bc.edu.

1 Motivation

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, Brown, and Mathiowetz (2001) for a survey). Misclassification may also arise from ordinary coding or reporting errors, e.g., Kane, Rouse, and Staiger (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; see, e.g., Hsiao (1991), Hausman, Arevaya and Scott-Morton (1998), and Hong and Tamer (2003). Additional sample information often used to obtain identification includes an instrumental variable or a repeated measurement in the same sample (see, e.g., Hausman, Ichimura, Newey and Powell (1991), Li (2002), Schennach (2004), Carroll, Ruppert, Crainiceanu, Tosteson and Karagas (2004), Mahajan (2006), Lewbel (2007a), Hu (2006) and Hu and Schennach (2006)), or a secondary sample (see, e.g., Lee and Sepanski (1995), Chen, Hong, and Tamer (2005), Hu and Ridder (2006), and Chen and Hu (2006)). See, e.g., Carroll, Ruppert, Stefanski and Crainiceanu (2006), and Chen, Hong and Nekipelov (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. In particular we show that, given some mild regularity conditions, a nonparametric mean regression with a misclassified 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.

This note is organized as follows. Section 2 provides the identification results with a closed-form identification of the latent model. Section 3 describes possible estimators and concludes the note. Proofs are in the appendix.

2 Identification

We are interested in a regression model as follows:

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

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

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

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:

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

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 $q(w) = f_{X^*|X,W}(0|1,w)$. Equation (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

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

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 equation (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 equation (2.4) is invertible, which implies

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

If we then plug in the expressions for $p(w)$ and $q(w)$ in equation (2.6), we obtain for $j = 0, 1$

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

Equation (2.7) is our vehicle for identification. Given any information about the distribution of the regression error η, equation (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 equation (2.7) implies that for any real t

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

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 $j=0,1$

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

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 equation (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*

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

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

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. It

should not be surprising that our identification based on a third moment restriction exploits skewness. An analogous result is Lewbel (1997), who obtains identification in a classical measurement error context without auxiliary data by assuming no skewness of the measurement error and skewness of the underlying regressor distribution. In contrast, in the present context it is the regression model error that is assumed to have zero skewness.

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 $E[\left(Y - \mu_j(w)\right)^3 | X = j, W = w]$ 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 equation (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 equation (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

$$\begin{aligned} E(\left(Y-\mu_j(w)\right)I(X=j)|W &= w)=0, \\ E(\left(Y^2-\upsilon_j(w)\right)I(X=j)|W &= w)=0, \\ E(\left(Y^3-\kappa_j(w)\right)I(X=j)|W &= w)=0, \end{aligned}$$

$$E\left(\begin{array}{c} 2m_j(W)^3-3\frac{\upsilon_1(w)-\upsilon_0(w)}{\mu_1(w)-\mu_0(w)}m_j(W)^2 \\ -\frac{3\upsilon_0(w)\mu_1(w)-3\upsilon_1(w)\mu_0(w)+\kappa_0(w)-\kappa_1(w)}{\mu_1(w)-\mu_0(w)}m_j(W)+\frac{\mu_1(w)\kappa_0(w)-\mu_0(w)\kappa_1(w)}{\mu_1(w)-\mu_0(w)} \end{array}|W=w\right)=0.$$

See the Appendix, particularly equation (A.8). 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.

References

[1] Ai, C. and X. Chen (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," Econometrica, 71, 1795-1844.

[2] Bound, J., C. Brown, and N. Mathiowetz (2001): "Measurement Error in Survey Data," in *Handbook of Econometrics*, vol. 5, ed. by J. J.Heckman and E. Leamer, Elsevier Science.

[3] Carroll, R., D. Ruppert, L. Stefanski and C. Crainiceanu (2006): *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. CRI, New York.

[4] Carroll, R., D. Ruppert, C. Crainiceanu, T. Tostenson and M. Karagas (2004): "Nonlinear and Nonparametric Regression and Instrumental Variables," *Journal of the American Statistical Association*, 99 (467), pp. 736-750.

[5] Chen, X., H. Hong, and E. Tamer (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72, 343-366.

[6] Chen, X., H. Hong, and D. Nekipelov (2007): "Measurement Error Models," survey prepared for *Journal of Economic Literature*.

[7] Chen, X. and Y. Hu (2006): "Identification and inference of nonlinear models using two samples with arbitrary measurement errors," Cowles Foundation Discussion Paper no. 1590.

[8] Hausman, J., J. Arevaya and F. Scott-Morton (1998): "Misclassification of the Dependent Variable in a Discrete-response Setting," *Journal of Econometrics*, 87, pp. 239-269.

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

[10] Hsiao, C. (1991): "Identification and estimation of dichotomous latent variables models using panel data," *Review of Economic Studies* 58, pp. 717-731.

[11] Hong, H, and E. Tamer (2003): "A simple estimator for nonlinear error in variable models," *Journal of Econometrics*, Volume 117, Issue 1, Pages 1-19

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

[13] Hu, Y., and G. Ridder (2006): "Estimation of Nonlinear Models with Measurement Error Using Marginal Information," Working Paper, University of Southern California.

[14] Hu, Y., and S. Schennach (2006): "Identification and estimation of nonclassical nonlinear errors-in-variables models with continuous distributions using instruments," Cemmap Working Papers CWP17/06.

[15] Kane, T. J., C. E. Rouse, and D. Staiger, (1999), "Estimating Returns to Schooling When Schooling is Misreported," NBER working paper #7235.

[16] Lee, L.-F., and J.H. Sepanski (1995): "Estimation of linear and nonlinear errors-in-variables models using validation data," *Journal of the American Statistical Association*, 90 (429).

[17] 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.

[18] Lewbel, A. (2007a): "Estimation of average treatment effects with misclassification," *Econometrica*, 2007, 75, 537-551.

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

[20] Li, T. (2002): "Robust and consistent estimation of nonlinear errors-in-variables models," *Journal of Econometrics*, 110, pp. 1-26.

[21] Mahajan, A. (2006): "Identification and estimation of regression models with misclassification," *Econometrica*, vol. 74, pp. 631-665.

[22] Schennach, S. (2004): "Estimation of nonlinear models with measurement error," *Econometrica*, vol. 72, no. 1, pp. 33-76.

4 Appendix

Proof. (Theorem 2.1) Frist, 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 equation (2.3), which is equivalent to

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

Using the notations above, we have

$$\begin{pmatrix} f_{Y|X,W}(y|0,w) \\ f_{Y|X,W}(y|1,w) \end{pmatrix} = \begin{pmatrix} 1 - p(w) & p(w) \\ q(w) & 1 - q(w) \end{pmatrix} \begin{pmatrix} f_{\eta|X^*,W}(y - m_0(w)|0,w) \\ f_{\eta|X^*,W}(y - m_1(w)|1,w) \end{pmatrix}. \tag{A.2}$$

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:

$$p(w) = \frac{\mu_0(w) - m_0(w)}{m_1(w) - m_0(w)} \text{ and } q(w) = \frac{m_1(w) - \mu_1(w)}{m_1(w) - m_0(w)}. \tag{A.3}$$

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 equation (A.3), we have

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

and

$$\begin{aligned}
\begin{pmatrix} f_{\eta|X^*,W}(y-m_0(w)\,|0,w) \\ f_{\eta|X^*,W}(y-m_1(w)\,|1,w) \end{pmatrix} &= \begin{pmatrix} 1-\frac{m_0(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)} & \frac{m_0(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)} \\ \frac{\mu_1(w)-m_1(w)}{\mu_1(w)-\mu_0(w)} & 1-\frac{\mu_1(w)-m_1(w)}{\mu_1(w)-\mu_0(w)} \end{pmatrix} \begin{pmatrix} f_{Y|X,W}(y|0,w) \\ f_{Y|X,W}(y|1,w) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\mu_1(w)-m_0(w)}{\mu_1(w)-\mu_0(w)} & \frac{m_0(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)} \\ \frac{\mu_1(w)-m_1(w)}{\mu_1(w)-\mu_0(w)} & \frac{m_1(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)} \end{pmatrix} \begin{pmatrix} f_{Y|X,W}(y|0,w) \\ f_{Y|X,W}(y|1,w) \end{pmatrix}.
\end{aligned}$$

In other words, we have

$$f_{\eta|X^*,W}(y-m_j(w)\,|j,w) = \frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)} f_{Y|X,W}(y|0,w) + \frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)} f_{Y|X,W}(y|1,w). \tag{A.4}$$

Let ϕ denote the characteristic function and

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

Equation (A.4) 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

$$\ln\left(e^{itm_j(w)}\phi_{\eta|X^*=j,w}(t)\right) = \ln\left(\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=0,w}(t) + \frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=1,w}(t)\right). \tag{A.5}$$

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 implies that for $j = 0, 1$

$$0 = i\frac{\partial^3}{\partial t^3}\ln\left(\frac{\mu_1(w) - m_j(w)}{\mu_1(w) - \mu_0(w)}\phi_{Y|X=0,w}(t) + \frac{m_j(w) - \mu_0(w)}{\mu_1(w) - \mu_0(w)}\phi_{Y|X=1,w}(t)\right)\Bigg|_{t=0}. \tag{A.6}$$

Given that $\frac{\partial}{\partial t}\ln f(t) = \frac{f'}{f}$, $\frac{\partial^2}{\partial t^2}\ln f(t) = \frac{f''}{f} - \left(\frac{f'}{f}\right)^2$ and

$$\begin{aligned}\frac{\partial^3}{\partial t^3}\ln f(t) &= \frac{f'''}{f} - \frac{f''f'}{f^2} - 2\frac{f'}{f}\left[\frac{f''}{f} - \left(\frac{f'}{f}\right)^2\right] \\ &= \frac{f'''}{f} - 3\frac{f''f'}{f^2} + 2\left(\frac{f'}{f}\right)^3,\end{aligned}$$

we have

$$\begin{aligned}0 = {} & i\frac{\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\frac{\partial^3}{\partial t^3}\phi_{Y|X=0,w}(t) + \frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\frac{\partial^3}{\partial t^3}\phi_{Y|X=1,w}(t)}{\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=0,w}(t) + \frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=1,w}(t)} \\ & -3i\frac{\left(\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\frac{\partial^2}{\partial t^2}\phi_{Y|X=0,w}(t) + \frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\frac{\partial^2}{\partial t^2}\phi_{Y|X=1,w}(t)\right)}{\left(\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=0,w}(t) + \frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=1,w}(t)\right)^2} \times \\ & \times\left(\frac{\mu_1(w) - m_j(w)}{\mu_1(w) - \mu_0(w)}\frac{\partial}{\partial t}\phi_{Y|X=0,w}(t) + \frac{m_j(w) - \mu_0(w)}{\mu_1(w) - \mu_0(w)}\frac{\partial}{\partial t}\phi_{Y|X=1,w}(t)\right) \\ & +2i\left(\frac{\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\frac{\partial}{\partial t}\phi_{Y|X=0,w}(t) + \frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\frac{\partial}{\partial t}\phi_{Y|X=1,w}(t)}{\frac{\mu_1(w)-m_j(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=0,w}(t) + \frac{m_j(w)-\mu_0(w)}{\mu_1(w)-\mu_0(w)}\phi_{Y|X=1,w}(t)}\right)^3.\end{aligned} \tag{A.7}$$

When $t = 0$, we have $\phi_{Y|X,W}(0) = 1$,

$$\frac{\partial}{\partial t}\phi_{Y|X=j,w}(0) = iE(Y|X = j, W = w) = i\mu_j,$$

$$\frac{\partial^2}{\partial t^2}\phi_{Y|X=j,w}(0) = -E(Y^2|X = j, W = w) = -\upsilon_j,$$

and

$$\frac{\partial^3}{\partial t^3}\phi_{Y|X=j}(0) = -iE(Y^3|X = j) = -i\kappa_j.$$

Furthermore, equation (A.7) becomes with $t = 0$,

$$\begin{aligned}0 = {} & i\left(\frac{\mu_1(w) - m_j(w)}{\mu_1(w) - \mu_0(w)}(-i\kappa_0(w)) + \frac{m_j(w) - \mu_0(w)}{\mu_1(w) - \mu_0(w)}(-i\kappa_1(w))\right) \\ & -3i\left(\frac{\mu_1(w) - m_j(w)}{\mu_1(w) - \mu_0(w)}(-\upsilon_0(w)) + \frac{m_j(w) - \mu_0(w)}{\mu_1(w) - \mu_0(w)}(-\upsilon_1(w))\right) \times \\ & \times\left(\frac{\mu_1(w) - m_j(w)}{\mu_1(w) - \mu_0(w)}(i\mu_0(w)) + \frac{m_j(w) - \mu_0(w)}{\mu_1(w) - \mu_0(w)}(i\mu_1(w))\right) \\ & +2i\left(\frac{\mu_1(w) - m_j(w)}{\mu_1(w) - \mu_0(w)}(i\mu_0(w)) + \frac{m_j(w) - \mu_0(w)}{\mu_1(w) - \mu_0(w)}(i\mu_1(w))\right)^3.\end{aligned}$$

Furthermore, we have

$$G(m_j) = 0,$$

where

$$\begin{aligned} G(z) \equiv\ & 2z^3 - 3\frac{\upsilon_1(w) - \upsilon_0(w)}{\mu_1(w) - \mu_0(w)}z^2 \\ & -\frac{3\upsilon_0(w)\mu_1(w) - 3\upsilon_1(w)\mu_0(w) + \kappa_0(w) - \kappa_1(w)}{\mu_1(w) - \mu_0(w)}z + \frac{\mu_1(w)\kappa_0(w) - \mu_0(w)\kappa_1(w)}{\mu_1(w) - \mu_0(w)}. \end{aligned} \tag{A.8}$$

Obviously, this equation has two real roots, i.e., $m_0(w)$ and $m_1(w)$. This means this cubic equation has three real roots because all the coefficients are real. Suppose the three roots of this equation are $r_a(w) \leq r_b(w) \leq r_c(w)$ for each given w. Since $m_0(w) \neq m_1(w)$, we will never have $r_a(w) = r_b(w) = r_c(w)$. The unknowns $m_j(w)$ corresponds to two of the three roots. 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 $m_1(w) \geq \mu_1(w) > \mu_0(w) \geq m_0(w)$. Given the shape of the cubic function, we know

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

That means the second largest of the three roots r_b is between $\mu_0(w)$ and $\mu_1(w)$ if

$$G(\mu_1(w)) < 0 \text{ and } G(\mu_0(w)) > 0.$$

We next show that $G(\mu_1(w)) = E\left[(Y - \mu_1(w))^3 | X = 1, W = w\right]$ and $G(\mu_0(w)) = E\left[(Y - \mu_0(w))^3 | X =\right.$ First, we consider $G(\mu_1(w))$:

$$\begin{aligned} G(\mu_1(w)) &= 2\mu_1^3(w) - 3\frac{\upsilon_1(w) - \upsilon_0(w)}{\mu_1(w) - \mu_0(w)}\mu_1(w) \\ &\quad -\frac{3\upsilon_0(w)\mu_1(w) - 3\upsilon_1(w)\mu_0(w) + \kappa_0(w) - \kappa_1(w)}{\mu_1(w) - \mu_0(w)}\mu_1(w) \\ &\quad +\frac{\mu_1(w)\kappa_0(w) - \mu_0(w)\kappa_1(w)}{\mu_1(w) - \mu_0(w)} \\ &= \kappa_1(w) - 3\upsilon_1(w)\mu_1(w) + 2\mu_1^3 \\ &= E\left[(Y - \mu_1(w))^3 | X = 1, W = w\right] \end{aligned}$$

Second, we consider $G(\mu_0(w))$

$$\begin{aligned} G(\mu_0(w)) &= 2\mu_0^3(w) - 3\frac{\upsilon_1(w) - \upsilon_0(w)}{\mu_1(w) - \mu_0(w)}\mu_0(w) \\ &\quad - \frac{3\upsilon_0(w)\mu_1(w) - 3\upsilon_1(w)\mu_0(w) + \kappa_0(w) - \kappa_1(w)}{\mu_1(w) - \mu_0(w)}\mu_0(w) \\ &\quad + \frac{\mu_1(w)\kappa_0(w) - \mu_0(w)\kappa_1(w)}{\mu_1(w) - \mu_0(w)} \\ &= \kappa_0(w) - 3\upsilon_0(w)\mu_0(w) + 2\mu_0^3 \\ &= E\left[(Y - \mu_0(w))^3 | X = 0, W = w\right] \end{aligned}$$

Therefore, assumption 2.5 implies that $G(\mu_1(w)) \leq 0$ and $G(\mu_0(w)) \geq 0$. Given the graph of the cubic function $G(\cdot)$, 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 $E\left[(Y - \mu_1(w))^3 | X = 1, W = w\right] = 0$, i.e., $G(\mu_1(w)) = 0$, $\mu_1(w)$ is a root of $G(\cdot)$. Given the graph of the cubic function $G(\cdot)$, we know

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

That means the condition $\left.\frac{dG(z)}{dz}\right|_{z=\mu_1(w)} > 0$ guarantees that $\mu_1(w)$ is the largest root and equal to $m_1(w)$. If $E\left[(Y - \mu_0(w))^3 | X = 0, W = w\right] = 0$, i.e. $G(\mu_0(w)) = 0$, $\mu_0(w)$ is a root of $G(\cdot)$. The condition $\left.\frac{dG(z)}{dz}\right|_{z=\mu_0(w)} > 0$ 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 equation (A.3), and the density $f_{\eta|X^*,W}$(or $f_{Y|X^*,W}$) is also identified from equation (A.4). 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. ■