3

Identifiability in Measurement Error Models

Liqun Wang

Department of Statistics, University of Manitoba

CONTENTS

	Abstract	1
3.1	Introduction	1
3.2	Linear Measurement Error Model	3
3.3	Nonlinear Model with Berkson Error	5
3.4	Nonlinear Model with Classical Error	7
3.5	Conclusions and Discussion	12
	Acknowledgements	13

“nobreak

Abstract

Identifiability is a fundamental issue in measurement error models, because consistent estimation of all unknown parameters in the model is impossible if it is unidentifiable. In this Chapter we study this issue in linear and nonlinear regression models with either classical or Berkson type measurement error. We use a simple linear model to demonstrate how the measurement error in covariates causes its non-identifiability. We show that the nonlinear models with Berkson measurement error can be identified without prior restrictions on the parameters or extra data besides the main sample. We also show that the classical measurement error models can be identified using the instrumental variable approach and provide a sufficient rank condition for the identifiability. Some examples are provided for illustration.

Keywords:

Berkson measurement error, classical measurement error, errors in variables, identifiability, identification, consistent estimation, data uncertainty, regression models.

3.1 Introduction

Identifiability is a fundamental issue in the statistical estimation and inference in measurement error (ME) models. The problem is caused by the fact that measurement error in covariates introduces so much uncertainty that the main sample data, no matter how large the sample size, cannot provide sufficient amount of information to identify all unknown parameters.

Generally speaking, a parameter in a given model is identifiable, if it is uniquely determined by the sampling distribution of the observed random variables in the model. Specifically, suppose we observe a random vector Z and the sampling distribution $F(z;\theta)$, $z \in I\!R^k$, is known up to an unknown parameter vector $\theta \in \Theta$, where $\Theta \subseteq I\!R^p$ is the parameter space. Then θ is said to be identifiable, if for any possible values $\theta_1, \theta_2 \in \Theta$, $F(z;\theta_1) \equiv F(z;\theta_2)$ implies $\theta_1 = \theta_2$ (Hsiao, 1983; Fuller, 1987). The model is said to be identifiable, if all unknown parameters in it are identifiable. In other words, θ is identifiable if it is uniquely determined by the sampling distribution $F(z;\theta)$. Therefore θ cannot be consistently estimated if it is not identifiable.

While the problem of identifiability in linear ME models has been well studied, e.g., Hsiao (1983) and Fuller (1987), it has been a long-standing and challenging problem in nonlinear models. Usually direct and explicit answers are difficult to obtain because the analytic forms of the moments of the observed variables are not available. In such cases the problem has to be studied through the existence of consistent estimators for some or all of the unknown parameters in the model. In general, there are two different types of measurement error. One is the classical error, also called errors-in-variables, which is independent of the unobserved true covariates. The other is the Berkson error which is independent of the observed surrogate covariates. Examples of variables typically observed with classical ME include long-term average blood pressure or cholesterol level of a person measured during a clinic visit, or the viral load and CD4+ T cell counts measured in laboratories in an HIV trial. Examples of variables typically observed with Berkson ME include an individual's exposure to certain air pollutants that are measured at some monitoring stations in a city, or the actual absorption of a medical substance in a patient's blood stream measured by the predetermined doses.

To deal with classical ME models, some researchers used an instrumental variable (IV) approach. For example, Hausman et al. (1991) showed that the univariate polynomial model is identifiable, while Wang and Hsiao (1995, 1996, 2011) showed that identifiability holds in general nonlinear models. Furthermore, Wang and Hsiao (2007) developed consistent estimators for censored linear models, and Guan et al. (2019) used a similar approach for the ordinal probit models. A recent survey of the instrumental variable methods for the estimation in ME models is given by Lewbel (2021), while in longitudinal data models is given by Wang (2021). Some other researchers assumed that the ME variance is known or can be estimated by replicate measurements of the surrogate covariates. For example, Wang (1994, 1998) studied the censored linear (Tobit) model, Hausman et al. (1995) showed that the polynomial model is identifiable, whereas Li (2002), Li and Hsiao (2004), and Schennach (2004) proposed consistent estimators in generalized linear or nonlinear models. However, Huang and Huwang (2001) showed that a polynomial model of order greater than two where all unobserved variables and errors have normal distribution is identifiable without extra information.

Interestingly, nonlinear models with Berkson ME are generally identifiable without extra information or data besides the main sample. This was shown for logistic models by Rudemo et al. (1989), and for polynomial models by Huwang and Huang (2000). Later, Wang (2003, 2004) showed that a nonlinear model is generally identifiable using the first two conditional moments of the response variable given the observed surrogate covariates.

In this chapter, we provide an account of identifiability results for general nonlinear regression models with either Berkson or classical ME in the covariates. In particular, we show that the Berkson ME models are generally identifiable without extra data. We study the IV approach to identifiability in the classical ME models, where a semiparametric approach is also presented. The rest of this chapter is organized as follows. In Section 3.2, we demonstrate the problem of identifiability in a simple linear model. In Section 3.3, we study the identifiability in nonlinear models with Berkson measurement error. In Section 3.4, we use the instrumental variable method to study the identifiability in nonlinear models with classical measurement error. Finally, conclusions an discussion are given in Section 3.5.

3.2 Linear Measurement Error Model

In this section we demonstrate how measurement error causes the problem of identifiability using a simple linear model with classical ME. Specifically, consider

$$Y = \beta_0 + \beta_x X + \varepsilon \tag{3.1}$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and is uncorrelated with X. Suppose X is unobservable and instead the observed covariate is

$$X^* = X + U, \tag{3.2}$$

where U is the random ME independent of X and ε. Further, suppose $X \sim N(\mu_x, \sigma_x^2)$ and $U \sim N(0, \sigma_u^2)$. Then the sampling distribution of the observed variables (Y, X^*) is jointly normal which is completely characterized by its first two moments. Further, these two moments are related to the unknown parameters in model (3.1)–(3.2) as

$$\mu_y = \beta_0 + \beta_x \mu_x, \tag{3.3}$$

$$\mu_{x^*} = \mu_x, \tag{3.4}$$

$$\sigma_y^2 = \beta_x \sigma_{yx^*} + \sigma_\varepsilon^2, \tag{3.5}$$

$$\sigma_{yx^*} = \beta_x \sigma_x^2, \tag{3.6}$$

$$\sigma_{x^*}^2 = \sigma_x^2 + \sigma_u^2. \tag{3.7}$$

It is easy to see that the above five moment equations contain six unknown parameters $(\beta_0, \beta_x, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)$ on the right-hand sides. Therefore, except for μ_x, all other parameters cannot be uniquely determined by the observed moments on the left-hand sides. Hence the model is not identifiable.

In fact, by simple algebra we can find the range constraints of the unknown parameters imposed by (3.3)–(3.7). Suppose $\beta_x \neq 0$, then the constraints are

$$\frac{\sigma_{yx^*}}{\sigma_{x^*}^2} \leq \beta_x \leq \frac{\sigma_y^2}{\sigma_{yx^*}},$$

$$\mu_y - \frac{\mu_{x^*}\sigma_y^2}{\sigma_{yx^*}} \leq \beta_0 \leq \mu_y - \frac{\mu_{x^*}\sigma_{yx^*}}{\sigma_{x^*}^2},$$

$$\frac{\sigma_{yx^*}^2}{\sigma_y^2} \leq \sigma_x^2 \leq \sigma_{x^*}^2,$$

$$0 \leq \sigma_u^2 \leq \sigma_{x^*}^2 - \frac{\sigma_{yx^*}^2}{\sigma_y^2},$$

$$0 \leq \sigma_\varepsilon^2 \leq \sigma_y^2 - \frac{\sigma_{yx^*}^2}{\sigma_{x^*}^2},$$

if $\sigma_{yx^*} > 0$, and the inequalities for β_x, β_0 are reversed if $\sigma_{yx^*} < 0$. In other words, any set of parameter values falling in these intervals is compatible with the sampling distribution of (Y, X^*).

The lack of identifiability of the normal linear model remains true in the censored linear models. Wang (1994, 1998) studied the so-called Tobit model (3.1)–(3.2) where the response variable Y is censored so that one observes $Y^* = \max\{Y, 0\}$. Due to the censoring, now the parameters $\mu_y, \sigma_{yx^*}, \sigma_y^2$ on the left-hand sides of (3.3)–(3.7) are not directly observable. However, these parameters can be identified through the following observed (conditional) moments

$$\mu_{y^*} = \Phi(\eta)\mu_{y^*}^+, \tag{3.8}$$

$$\mu_{y^*}^+ = \mu_y + \sigma_y\phi(\eta)/\Phi(\eta), \tag{3.9}$$

$$\mu_{y^*x^*}^+ = \sigma_{yx*} + \mu_{x^*}\mu_{y^*}^+, \tag{3.10}$$

where $\mu_{y*}^+ = E(Y^* \mid Y^* > 0)$, $\eta = \mu_y/\sigma_y$, $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions. Therefore, as in the case of the linear model, prior restrictions on the parameters $(\beta_0, \beta_x, \sigma_x^2, \sigma_\varepsilon^2, \sigma_u^2)$ are needed to achieve model identifiability.

The situation in Berkson ME models is slightly different than in classical ME models. To illustrate, instead of (3.2) we assume that we observe X^*, where $X = X^* + U$ with U independent of X^* and ε. Then by substituting $X^* + U$ for X in equation (3.1), we obtain

$$Y = \beta_0 + \beta_x X^* + \delta,$$

where $\delta = \varepsilon + \beta_x U$ is uncorrelated with X^*. Hence β_0, β_x and $\sigma_\delta^2 = \sigma_\varepsilon^2 + \beta_x^2\sigma_u^2$ can be identified by the least squares regression of Y on X^*. However, without any extra information, only the regression parameters β_0, β_x are identifiable and the variance parameters $\sigma_\varepsilon^2, \sigma_u^2$ are not identifiable.

In practice, usually certain restrictions on unknown parameters are imposed in order to ensure identifiability. The commonly used restrictions are (1) ME variance σ_u^2 is known; (2) error variance ratio $\sigma_u^2/\sigma_\varepsilon^2$ is known; or (3) signal-to-noise (reliability, heritability) ratio $\sigma_x^2/\sigma_{x^*}^2$ is known. If there is no strong theoretical justification for these assumptions, these quantities can be estimated if additional data besides the main sample (Y, X^*) is available. For example, (1) validation data, i.e., an observed subsample on (Y, X); (2) replicate data, i.e., two or more independent and unbiased measurements X^* for the same X; (3) instrumental variables that are correlated with X but uncorrelated with errors U and ε. For more discussion see, e.g., Fuller (1987) and Carroll et al. (2006).

3.3 Nonlinear Model with Berkson Error

In this section we study the identifiability in nonlinear models with Berkson ME. Specifically, consider the model

$$Y = g(X, \beta) + \varepsilon, \tag{3.11}$$

where $Y \in I\!R$ is the response variable, $X \in I\!R^{p_x}$ is the vector of covariates, ε is the random error and $\beta \in I\!R^p$ is the vector of unknown parameters. In general, $g(x; \beta)$ is nonlinear in x. The unobserved true covariate X is related to its observed surrogate X^* as

$$X = X^* + U, \tag{3.12}$$

where U is the measurement error that has a density $f_u(u; \theta)$ with unknown parameters θ. Throughout this section we make the following assumptions.

Assumption 3.1. *(1) U is independent of X^* and $E(U) = 0$; (2) ε satisfies $E(\varepsilon \mid X, X^*) = 0$ and $E(\varepsilon^2 \mid X, X^*) = \sigma_\varepsilon^2$.*

Under these model assumptions the first two conditional moments of the response Y given the observed surrogate X^* are given by

$$\begin{aligned} E\left(Y \mid X^*\right) &= E[g(X, \beta) \mid X^*] \\ &= \int g\left(X^* + u, \beta\right) f_u(u; \theta)\, du \end{aligned} \tag{3.13}$$

and, similarly

$$\begin{aligned} E\left(Y^2 \mid X^*\right) &= E[g^2(X, \beta) \mid X^*] + E(\varepsilon^2 \mid X^*) \\ &= \int g^2\left(X^* + u, \beta\right) f_u(u; \theta)\, du + \sigma_\varepsilon^2. \end{aligned} \tag{3.14}$$

It is apparent that, when the conditional moments in (3.13) and (3.14) admit closed forms, the parameters $\beta, \theta, \sigma_\varepsilon^2$ can be identified by the nonlinear least squares method because both Y and X^* are observable. This is illustrated by the following examples.

Example 3.1. First, consider the quadratic model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon,$$

where $\beta_2 \neq 0$ and $Var(U) = \theta$. For this model the first conditional moment of Y given X^* is given by:

$$\begin{aligned} E(Y|X^*) &= \beta_0 + \beta_1 E[(X^* + U)|X^*] + \beta_2 E[(X^* + U)^2|X^*] \\ &= \beta_0 + \beta_1 X^* + \beta_2 X^{*2} + \beta_2\theta \\ &= \varphi_1 + \varphi_2 X^* + \varphi_3 X^{*2}, \end{aligned} \tag{3.15}$$

where $\varphi_1 = \beta_0 + \beta_2\theta$, $\varphi_2 = \beta_1$ and $\varphi_3 = \beta_2$. Similarly, the second conditional moment of Y is

$$\begin{aligned} E(Y^2|X^*) &= E[(\beta_0 + \beta_1 X + \beta_2 X^2)^2|X^*] + E(\varepsilon^2|X^*) \\ &= \varphi_4 + \varphi_5 X^* + \varphi_6 X^{*2} + 2\varphi_2\varphi_3 X^{*3} + \varphi_3^2 X^{*4}, \end{aligned} \tag{3.16}$$

where

$$\begin{aligned} \varphi_4 &= \beta_0^2 + (2\beta_0\beta_2 + \beta_1^2)\theta + 3\beta_2^2\theta^2 + \sigma_\varepsilon^2, \\ \varphi_5 &= 2\beta_1(\beta_0 + 3\beta_2\theta), \\ \varphi_6 &= 2\beta_0\beta_2 + \beta_1^2 + 6\beta_2^2\theta. \end{aligned}$$

It is easy to see that (φ_i) are identified by the nonlinear least squares regression using (3.15) and (3.16). Further, we can show that the mapping $(\beta_i, \theta, \sigma_\varepsilon^2) \mapsto (\varphi_i)$ is bijective. In fact, by using simple algebra we can find the inverse mapping

$$\begin{aligned} \beta_0 &= \varphi_1 - \varphi_3\theta, \\ \beta_1 &= \varphi_2, \\ \beta_2 &= \varphi_3, \\ \theta &= (\varphi_6 - 2\varphi_1\varphi_3 - \varphi_2^2)/(4\varphi_3^2), \\ \sigma_\varepsilon^2 &= \varphi_4 - \varphi_1^2 - \varphi_2^2\theta - 2\varphi_3^2\theta^2. \end{aligned}$$

Therefore, the original parameters $(\beta_i, \theta, \sigma_\varepsilon^2)$ are identifiable.

Example 3.2. Next, consider model $g(X;\beta) = \beta_1 X_1 + \beta_3 \exp(\beta_2 X_2)$, where $\beta_2\beta_3 \neq 0$. In addition, we assume that $U = (U_1, U_2)^\top \sim N(0, \theta I_2)$, where $\theta > 0$ and I_2 is the 2-dimensional identity matrix. Then using the expression of the normal moment generating function, we obtain the conditional moment of Y given X^* as

$$\begin{aligned} E(Y|X^*) &= \beta_1 X_1^* + \beta_1 E(U_1) + \beta_3 \exp(\beta_2 X_2^*) E\left(\exp(\beta_2 U_2)\right) \\ &= \varphi_1 X_1^* + \varphi_3 \exp(\varphi_2 X_2^*), \end{aligned} \tag{3.17}$$

where $\varphi_1 = \beta_1$, $\varphi_2 = \beta_2$ and $\varphi_3 = \beta_3 \exp(\beta_2^2\theta/2)$. Similarly, the second conditional moment of Y given X^* is

$$\begin{aligned}
E(Y^2|X^*) &= \beta_1^2 E[(X_1^* + U_1)^2|X^*] + \beta_3^2 E[\exp(2\beta_2(X_2^* + U_2))|X^*] \\
&\quad + 2\beta_1\beta_3 E[(X_1^* + U_1)\exp(\beta_2(X_2^* + U_2))|X^*] + E(\varepsilon^2|X^*) \\
&= \beta_1^2(X_1^{*2} + \theta) + \beta_3^2 \exp(2\beta_2 X_2^*) E\left[\exp(2\beta_2 U_2)\right] \\
&\quad + 2\beta_1\beta_3 X_1^* \exp(\beta_2 X_2^*) E[\exp(\beta_2 U_2)] + \sigma_\varepsilon^2 \\
&= \varphi_4 + \varphi_1^2 X_1^{*2} + \varphi_5 \exp(2\varphi_2 X_2^*) + 2\varphi_1\varphi_3 X_1^* \exp(\varphi_2 X_2^*), \qquad (3.18)
\end{aligned}$$

where $\varphi_4 = \beta_1^2\theta + \sigma_\varepsilon^2$ and $\varphi_5 = \beta_3^2 \exp(2\beta_2^2\theta)$. Again, (φ_i) are clearly identified by (3.17) and (3.18) and the nonlinear least squares method. Furthermore, it is straightforward to obtain the inverse relationship

$$\begin{aligned}
\beta_1 &= \varphi_1, \\
\beta_2 &= \varphi_2, \\
\beta_3 &= \varphi_3^2/\sqrt{\varphi_5}, \\
\theta &= \log(\varphi_5/\varphi_3^2)/\varphi_2^2, \\
\sigma_\varepsilon^2 &= \varphi_4 - \varphi_1^2 \log(\varphi_5/\varphi_3^2)/\varphi_2^2.
\end{aligned}$$

Therefore the mapping $(\beta_i, \theta, \sigma_\varepsilon^2) \mapsto (\varphi_i)$ is bijective and hence the original parameters $(\beta_i, \theta, \sigma_\varepsilon^2)$ are identifiable.

The above examples suggest that in many situations, the parameters in nonlinear models can be identified using the first two conditional moments of Y given X^* and the least squares method. This fact is shown for polynomial models by Huwang and Huang (2000), and for general nonlinear models by Wang (2003, 2004). When the closed forms of the conditional moments in (3.13) and (3.14) are not available, a simulation-based method can be used to identify and consistently estimate parameters $\beta, \theta, \sigma_\varepsilon^2$ (Wang, 2004).

3.4 Nonlinear Model with Classical Error

In this section, we consider the general nonlinear model (3.11) with classical ME, where the unobserved true covariate X is related to its observed surrogate X^* as

$$X^* = X + U \qquad (3.19)$$

and the ME U is independent of X. Unlike in the Berkson ME model (3.11)–(3.12), model (3.11) and (3.19) is generally unidentifiable without extra information. In this section, we show how the model can be identified by using the instrumental variable approach. In general, an IV is an observed variable that is correlated with the unobserved true covariate X but independent of measurement and model errors U and ε. Here, we assume that a vector of instrumental variables $W \in I\!R^{p_w}$ is available and is related to X through

$$X = \Gamma W + \delta, \qquad (3.20)$$

where Γ is a $p_x \times p_w$ matrix of unknown parameters and δ is the random error. In addition, we assume that the random errors in (3.11), (3.19) and (3.20) satisfy the following conditions.

Assumption 3.2. *(1) U satisfies $E(U \mid X, W) = 0$ and $E(UU^\top \mid X, W) = \Sigma_u$; (2) ε satisfies $E(\varepsilon \mid X, X^*, W) = 0$ and $E(\varepsilon^2 \mid X, X^*, W) = \sigma_\varepsilon^2$; (3) δ is independent of W and has $E(\delta) = 0$; (4) Γ has full row rank $p_x \leq p_w$.*

Note that the assumptions of zero conditional mean for the measurement and model errors in Assumption 3.2 (1) and (2) are weaker than the usual requirement for the IV that it is independent of the errors. Moreover, the independence assumption between W and δ in Assumption 3.2 (3) is made to simplify notation in the following derivations, where the conditional distribution of δ given W can be used instead of the marginal distribution of δ. Further, the full rank condition of Γ ensures that (3.20) does not contain redundant IVs. It also implies that at least as many IVs as the number of mismeasured covariates are needed in order to ensure model identification.

There is no distributional assumptions for X, U and ε, so model (3.11), (3.19) and (3.20) represent a semiparametric model. In the following, we show that the model is identifiable based on the observed data (Y, X^*, W). Our approach is based on the first two conditional moments of (Y, X^*) given the instrumental variables W.

First, substituting (3.20) in (3.19) results in a usual regression equation

$$E(X^* \mid W) = \Gamma W. \tag{3.21}$$

Therefore Γ can be identified by the least squares regression of X^* on W. Hence, subsequently we assume that Γ is identified. Further, the second moment of X^* given W is

$$\begin{aligned} E(X^* X^{*\top} \mid W) &= E(XX^\top \mid W) + E(UU^\top \mid W) \\ &= \Gamma W W^\top \Gamma^\top + E(\delta\delta^\top \mid W) + E(UU^\top \mid W) \\ &= \Gamma W W^\top \Gamma^\top + \Sigma_\delta + \Sigma_u. \end{aligned} \tag{3.22}$$

Therefore Σ_u can be identified by the above equation if Σ_δ is identified. In order to identify the other parameters, we need the conditional moments of the response variable Y given W. Under the model assumptions the conditional moment of Y given W is

$$\begin{aligned} E(Y \mid W) &= E[g(X, \beta) \mid W] \\ &= E[g(\Gamma W + \delta, \beta) \mid W] \\ &= \int g(x + \Gamma W, \beta)\, dF_\delta(x), \end{aligned} \tag{3.23}$$

where $F_\delta(\cdot)$ is the distribution function of δ. Similarly, the second order conditional moments are

$$\begin{aligned} E(YX^* \mid W) &= E[Xg(X, \beta) \mid W] + E[Ug(X, \beta) \mid W] + E(X\varepsilon \mid W) + E(U\varepsilon \mid W) \\ &= E[(\Gamma W + \delta) g(\Gamma W + \delta, \beta) \mid W] \\ &= \int (x + \Gamma W) g(x + \Gamma W, \beta)\, dF_\delta(x) \end{aligned} \tag{3.24}$$

and

$$
\begin{aligned}
E(Y^2 \mid W) &= E[g^2(X,\beta) \mid W] + E(\varepsilon^2 \mid W) \\
&= \int g^2(x + \Gamma W, \beta)\, dF_\delta(x) + \sigma_\varepsilon^2. \qquad (3.25)
\end{aligned}
$$

From the above moment equations we can see that, if β and the distribution F_δ (hence Σ_δ) can be identified by (3.23) and (3.24), then σ_ε^2 will be identified by (3.25). The following examples demonstrate how β and F_δ are identified by (3.23) and (3.24).

Example 3.1. Consider a polynomial model $g(x;\beta) = \beta_1 + \beta_2 x^2$, where $p_x = 1$, $\Gamma = 1$ and $Var(\delta) = \theta$. Then moments (3.23) and (3.24) are respectively

$$
\begin{aligned}
E(Y \mid W) &= (\beta_1 + \beta_2\theta) + \beta_2 W^2, \\
E(YX^* \mid W) &= (\beta_1 + 3\beta_2\theta) W + \beta_2 W^3.
\end{aligned}
$$

We can see that $\beta_1 + \beta_2\theta$ and β_2 can be identified by the least squares method from the first equation, while $\beta_1 + 3\beta_2\theta$ is identified by the second equation. It follows that β_1, β_2 and θ are identifiable.

Example 3.2. Another example is an exponential model $g(x;\beta) = \beta_1 \exp(\beta_2 x)$, where $\beta_1\beta_2 \neq 0$ and $\delta \sim N(0,\theta)$. In this model moments (3.23) and (3.24) are respectively

$$
\begin{aligned}
E(Y \mid W) &= \beta_1 \exp\left(\beta_2 W + \beta_2^2\theta/2\right), \\
E(YX^* \mid W) &= \beta_1(\beta_2\theta + W) \exp\left(\beta_2 W + \beta_2^2\theta/2\right).
\end{aligned}
$$

Again, β_2, $\beta_1 \exp\left(\beta_2^2\theta/2\right)$ and $\beta_1\beta_2\theta \exp\left(\beta_2^2\theta/2\right)$ can be identified by nonlinear least squares method, implying that β_1, β_2 and θ are identifiable.

In the rest of this section we derive sufficient conditions for β and F_δ to be identifiable based on moment equations (3.23)–(3.24). The basic idea is to apply Fourier deconvolution to these equations to separate β and F_δ. This approach is based on the following assumptions.

Assumption 3.3. *The distribution of W is absolutely continuous with respect to Lebesgue measure and has support $\mathbb{R}^{p_w}$.*

Assumption 3.4. *$g(x;\beta)(\|x\|+1) \in L^1(\mathbb{R}^{p_x})$, the space of absolutely integrable functions on $\mathbb{R}^{p_x}$. Further, the set $\mathcal{T} = \{t \in \mathbb{R}^{p_x} : \varphi_g(t;\beta) \neq 0\}$ is dense in $\mathbb{R}^{p_x}$, where $\varphi_g(t;\beta) = \int \exp(it^\top x) g(x;\beta)\, dx$ is the Fourier transform of $g(x;\beta)$ and $i = \sqrt{-1}$.*

The integrability of $g(x;\beta)$ in Assumption 3.4 implies the existence of the Fourier transform $\varphi_g(t;\beta)$ (Lukacs 1970). The second part of the assumption means that the zeros of $\varphi_g(t;\beta)$ are isolated points in $\mathbb{R}^{p_x}$. The examples given at the end of this section show that this assumption is fairly general. Define the functions

$$
\begin{aligned}
h_1(\beta) &= \int g(x;\beta)\, dx, \qquad (3.26) \\
h_2(\beta) &= \int x g(x;\beta)\, dx. \qquad (3.27)
\end{aligned}
$$

Then we have the following result.

Theorem 3.1. *Under Assumption 3.2–3.4, suppose $h_1(\beta)$ and $h_2(\beta)$ are differentiable in an open neighborhood of the true parameter value, and the Jacobian*

$$J(\beta) = \left(\frac{\partial h_1(\beta)}{\partial \beta}, \frac{\partial h_2^\top(\beta)}{\partial \beta} \right) \tag{3.28}$$

has full rank $p = \dim(\beta)$. Then β and F_δ are identifiable.

Proof. For every $v \in \mathbb{R}^{p_x}$, define

$$m_1(v) = \int g(x + v; \beta) dF_\delta(x). \tag{3.29}$$

Then by (3.23) $m_1(\Gamma W) = E(Y \mid W)$ which is fully observable on $\mathbb{R}^{p_x}$ because of Assumption 3.3 and that Γ has full rank. Moreover, by Assumption 3.4 we have $m_1(v) \in L^1(\mathbb{R}^{p_x})$ and, further, taking Fourier transformation on both sides of (3.29) yields

$$\begin{aligned} \varphi_{m_1}(t) &= \int \exp(it^\top v) m_1(v)\, dv \\ &= \int \exp(it^\top x) g(x; \beta)\, dx \cdot \int \exp(-it^\top v)\, dF_\delta(v) \\ &= \varphi_g(t; \beta) \varphi_\delta(-t), \end{aligned} \tag{3.30}$$

where $\varphi_\delta(t)$ is the characteristic function of δ. It follows from (3.30) that, for any $t \in \mathcal{T}$, $\varphi_\delta(t)$ is uniquely determined by

$$\varphi_\delta(t) = \frac{\varphi_{m_1}(-t)}{\varphi_g(-t; \beta)}. \tag{3.31}$$

Moreover, by Assumption 3.4 the value of $\varphi_\delta(t)$ at any zero of $\varphi_g(-t, \beta)$ is also uniquely determined, because any characteristic function is uniformly continuous in $\mathbb{R}^{p_x}$ (Lukacs, 1970). Therefore F_δ is identifiable as long as β is.

In order to obtain the identifiability condition for β, we define, for every $v \in \mathbb{R}^{p_x}$,

$$m_2(v) = \int (x + v) g(x + v; \beta)\, dF_\delta(x). \tag{3.32}$$

Then by (3.24) $m_2(\Gamma W) = E(YZ \mid W)$ which is observable. Further, by Assumption 3.4 $m_2(v) \in L^1(\mathbb{R}^{p_x})$ and, therefore, by integrating both sides of (3.29) and (3.32) and applying Fubini's Theorem we obtain

$$\int m_1(v) dv = \int g(x; \beta)\, dx = h_1(\beta), \tag{3.33}$$

$$\int m_2(v) dv = \int x g(x; \beta)\, dx = h_2(\beta). \tag{3.34}$$

Since the Jacobian $J(\beta)$ of the above equations has full rank p, by the Rank Theorem (Zeidler, 1986, p. 178), β is the unique solution of (3.33) and (3.34) in its neighborhood. It follows that β is identified by (3.23)–(3.24). □

Note that since $J(\beta)$ has dimensions $p \times (p_x + 1)$, for $J(\beta)$ to have full rank it is necessary that $p \le p_x + 1$. In addition, if $\varphi_\delta(t) \in L^1(I\!R^{p_x})$, then the density of δ exists and is given by

$$f_\delta(x) = \frac{1}{(2\pi)^{p_x}} \int \exp(it^\top x) \frac{\varphi_{m_1}(t)}{\varphi_g(t;\beta)}\, dt.$$

From a practical point of view, the integrability of $g(x;\beta)$ in Assumption 3.4 is not as restrictive as it appears, if X is bounded with probability one. In the case where X is unbounded, the truncated Fourier transform $\varphi_g^c(t;\beta) = \int_{-c}^{c} \exp(it'x) g(x;\beta)\, dx, c \to \infty$, can be used so that this assumption may be weaken to the condition that $E[|g(X;\beta)|(\|X\|+1)] < \infty$ (Wang and Hsiao, 2011).

In the rest of this section, we use some examples to illustrate how to apply Theorem 3.1 to check model identifiability. To simplify notation, we consider cases where all variables are scalars and $\Gamma = 1$.

Example 3.3. Consider an exponential model $g(x;\beta) = \exp(-\beta x^2)$ where $\beta > 0$ and hence $g(x;\beta)$ is integrable. Further, the Fourier transform of $g(x;\beta)$ is

$$\begin{aligned} \varphi_g(t;\beta) &= \int \exp(itx)\exp(-\beta x^2)\, dx \\ &= \sqrt{\frac{\pi}{\beta}} \exp\left(-\frac{t^2}{4\beta}\right), \end{aligned}$$

which clearly satisfies Assumption 3.4. To check the rank condition, we calculate

$$h_1(\beta) = \int \exp(-\beta x^2)\, dx = \sqrt{\frac{\pi}{\beta}}$$

and

$$h_2(\beta) = \int x \exp(-\beta x^2)\, dx = 0.$$

It follows that $J(\beta) = \left(-\sqrt{\pi}/(2\beta\sqrt{\beta}), 0\right)$ which has rank one. Therefore, by Theorem 3.1 the model is identifiable.

Example 3.4. Now consider the quadratic model $g(x;\beta) = \beta_1 x + \beta_2 x^2$ where $E\|X\|^3 < \infty$. It is clear that condition $E[|g(X;\beta)|(\|X\|+1)] < \infty$ holds. Further, for any $c > 0$, the truncated Fourier transform is

$$\begin{aligned} \varphi_g^c(t;\beta) &= \int_{-c}^{c} \exp(itx)(\beta_1 x + \beta_2 x^2)\, dx \\ &= 2c\cos(ct)\left(\frac{\beta_2}{t^2} - \frac{i\beta_1}{t}\right) + 2\sin(ct)\left(\frac{\beta_2 c^2}{t} + \frac{i\beta_1}{t^2} - \frac{2\beta_2}{t^3}\right), \end{aligned}$$

which satisfies Assumption 3.4. To check the rank condition, we calculate $h_1(\beta) = 2\beta_2 c^3/3$ and $h_2(\beta) = 2\beta_1 c^3/3$. Hence the Jacobian is

$$J(\beta) = \frac{2c^3}{3}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

which has rank two. Therefore, by Theorem 3.1 the model is identifiable.

Example 3.5. Finally consider another exponential $g(x;\beta) = \exp(\beta x), \beta \neq 0$ and X has a normal distribution. In this case condition $E[|g(X;\beta)|(\|X\|+1)] < \infty$ is satisfied. Further, for any $c > 0$ and $t \neq i\beta$, the truncated Fourier transform is

$$\varphi_g^c(t;\beta) = \frac{\exp[c(\beta+it)] - \exp[-c(\theta+it)]}{\beta+it},$$

which satisfies Assumption 3.4. Moreover, since $h_1(\beta) = [\exp(c\beta) - \exp(-c\beta)]/\beta$, and

$$h_2(\beta) = \frac{c(\exp(c\beta)+\exp(-c\beta))}{\beta} - \frac{\exp(c\beta)-\exp(-c\beta)}{\beta^2},$$

the Jacobian $J(\beta)$ is of rank one, implying that the model is identifiable.

3.5 Conclusions and Discussion

We have studied the problem of identifiability in regression models with covariate measurement error. We have shown that in general a nonlinear model with Berkson measurement error can be identified by using the first two conditional moments of the response variable given the surrogate predictors. However, for a nonlinear model with classical measurement error, or a linear model with either classical or Berkson error, to be identifiable, certain restrictions or extra data are required. We have shown how instrumental variables can help to achieve model identifiability. We derived a sufficient rank condition for the identifiability of the regression coefficients, which is easy to check in practice. As in most theoretical works on nonlinear systems, we only obtained the results for local identifiability. More restrictive assumptions and theoretical techniques would be needed to achieve global identifiability, if it is possible at all.

In practice, any variables that are correlated with the error-prone covariates but independent of the measurement and model errors can serve as instrumental variables, such as a second independent measurement taken with a different method, or repeated measurements at different time points in longitudinal studies. For example, a second measurement of the systolic blood pressure can be taken by a different physician at a different time, or the cholesterol level, which is measured by laboratory blood test, can be measured again by using the home test kit. A further example is the CD4+ T cell counts in a HIV patient tested by two different laboratories using two different methods. It is worthwhile to note that the assumption of instrumental variables is weaker than that of replicate data because they can be biased measurements of the unobserved true covariates (Carroll and Stefanski, 1994; Carroll et al., 2006). The instrumental variable approach is also used in the literature of causal inference to deal with the problem of unobserved confounders, where usually additional conditions are imposed to achieve model identification (Angrist et al., 1996; Baiocchi et al., 2014).

In this Chapter, we are mainly concerned with the case where the error-prone covariates and measurement errors are continuous random variables. For the misclassification problem and Bayesian approach, see Gustafson (2003, 2021) and Yi (2017). A survey of semiparametric methods for both measurement error and misclassification problems is given in Ma (2021). Moreover, we focused on the parametric identifiability in regression models. There is a large literature on nonparametric identification, which is usually called deconvolution problems, e.g., Schennach (2007); Zinde-Walsh (2014) and Delaigle and Van Keilegom (2021). For an overview of nonparametric methods of estimation with classical measurement errors, see Delaigle (2021), and Apanasovich and Liang (2021), who introduce kernel deconvolution methods of density and regression, respectively, or Kang and Qiu (2021) who survey Fourier transformation and other related approaches. For a survey of nonparametric methods for Berkson measurement error models, see Song (2021). See also Delaigle (2007) for nonparametric density estimation with a mixture of Berkson and classical errors.

Acknowledgements

The author is grateful to co-editor Aurore Delaigle for the kind invitation to write this chapter and for her helpful comments and suggestions on the previous version. This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Bibliography

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Apanasovich, T. and H. Liang (2021). Nonparametric measurement errors models for regression. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 14. CRC.

Baiocchi, M., J. Cheng, and D. S. Small (2014). Instrumental variable methods for causal inference. *Statistics in Medicine 33*(13), 2297–2340.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective.* CRC press.

Carroll, R. J. and L. A. Stefanski (1994). Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. *Statistics in Medicine 13*(12), 1265–1282.

Delaigle, A. (2007). Nonparametric density estimation from data with a mixture of berkson and classical errors. *Canadian Journal of Statistics 35*(1), 89–104.

Delaigle, A. (2021). Deconvolution kernel density estimator. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 10. CRC.

Delaigle, A. and I. Van Keilegom (2021). Deconvolution with unknown error distribution. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 12. CRC.

Fuller, W. A. (1987). *Measurement Error Models.* John Wiley & Sons.

Guan, J., H. Cheng, K. A. Bollen, D. R. Thomas, and L. Wang (2019). Instrumental variable estimation in ordinal probit models with mismeasured predictors. *Canadian Journal of Statistics 47*(4), 653–667.

Gustafson, P. (2003). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments.* CRC Press.

Gustafson, P. (2021). Partial learning of misclassification parameters. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 4. CRC.

Hausman, J. A., W. K. Newey, H. Ichimura, and J. L. Powell (1991). Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics 50*(3), 273–295.

Hausman, J. A., W. K. Newey, and J. L. Powell (1995). Nonlinear errors in variables estimation of some engel curves. *Journal of Econometrics 65*(1), 205–233.

Hsiao, C. (1983). Identification. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 1, pp. 223–283. North-Holland.

Huang, Y. S. and L. Huwang (2001). On the polynomial structural relationship. *Canadian Journal of Statistics 29*(3), 495–512.

Huwang, L. and Y. S. Huang (2000). On errors-in-variables in polynomial regression-berkson case. *Statistica Sinica 10*, 923–936.

Kang, Y. and P. Qiu (2021). Nonparametric deconvolution by fourier transformation and other related approaches. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 11. CRC.

Lewbel, A. (2021). Using instrumental variables to estimate models with mismeasured regressors. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 5. CRC.

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

Li, T. and C. Hsiao (2004). Robust estimation of generalized linear models with measurement errors. *Journal of Econometrics 118*(1-2), 51–65.

Lukacs, E. (1970). *Characteristic Functions.* Griffin.

Ma, Y. (2021). Semiparametric methods for measurement error and misclassification. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 9. CRC.

Rudemo, M., D. Ruppert, and J. Streibig (1989). Random-effect models in nonlinear regression with applications to bioassay. *Biometrics 45*, 349–362.

Schennach, S. M. (2004). Estimation of nonlinear models with measurement error. *Econometrica 72*(1), 33–75.

Schennach, S. M. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica 75*(1), 201–239.

Song, W. (2021). Nonparametric inference methods for berkson errors. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 13. CRC.

Wang, L. (1994). Identifiability and estimation of censored errors-in-variables models. In *Proceedings of the Business and Economic Statistics Section*, pp. 351–354. American Statistical Association (Alexandria, VA).

Wang, L. (1998). Estimation of censored linear errors-in-variables models. *Journal of Econometrics 84*(2), 383–400.

Wang, L. (2003). Estimation of nonlinear berkson-type measurement error models. *Statistica Sinica 13*, 1201–1210.

Wang, L. (2004). Estimation of nonlinear models with berkson measurement errors. *Annals of Statistics 32*(6), 2559–2579.

Wang, L. (2021). Estimation in mixed-effects models with measurement error. In G. Y. Yi, A. Delaigle, and P. Gustafson (Eds.), *Handbook of Measurement Error Models*, Chapter 17. CRC.

Wang, L. and C. Hsiao (1995). Simulation-based semiparametric estimation of nonlinear errors-in-variables models. *Working Paper, University of Southern California*.

Wang, L. and C. Hsiao (1996). A semiparametric estimation of nonlinear errors-in-variables models. In *Proceedings of the Business and Economic Statistics Section*, pp. 231–236. American Statistical Association (Alexandria, VA).

Wang, L. and C. Hsiao (2007). Two-stage estimation of limited dependent variable models with errors-in-variables. *The Econometrics Journal 10*(2), 426–438.

Wang, L. and C. Hsiao (2011). Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. *Journal of Econometrics 165*(1), 30–44.

Yi, G. Y. (2017). *Statistical Analysis with Measurement Error or Misclassification*. Springer.

Zeidler, E. (1986). *Nonlinear Functional Analysis and Its Applications I*. Springer.

Zinde-Walsh, V. (2014). Measurement error and deconvolution in spaces of generalized functions. *Econometric Theory 30*, 1207–1246.