Identification of Semiparametric Measurement Error Models*

Yingyao Hu and Ji-Liang Shiu†

August 2011

Abstract

We consider the identification of semiparametric models in nonclassical errors-in-variables models with an unobserved regressor. The identification strategy does not require additional data information, such as instruments, double measurements, or validation data are available. Our main identifying assumptions only include the completeness of several families of observable conditional distributional functions or injectivity of integral operators constructed by observable density functions. The detailed discussions of the identification in continuous and discrete cases are provided. While estimators in discrete cases are directly by the identification steps, estimators in continuous cases are constructed using a sieve maximum likelihood estimator (MLE). The finite-sample properties of these estimators are investigated through Monte Carlo simulations. An empirical application of measurement errors in consumption is demonstrated.

Keywords: Nonclassical measurement error, Identification, Completeness, Semiparametric measurement error models , Conditional maximum likelihood estimation.

---

*We thank the participants at the 2011 Asian Meeting of Econometric Society, Seoul, Korea, for various suggestions regarding this paper. The authors are solely responsible for any remaining errors.

†The authors can be reached at jishiu@ccu.edu.tw and yhu@jhu.edu.

1. Introduction

In many econometric models, the parameter of interest θ is usually defined through a family of conditional densities of y given x^*,

$$f_{Y|X^*}(y|x^*;\theta), \tag{1}$$

where y is the endogenous variables, x^* is the unobserved explanatory, and $\theta \in \Theta$ is a parameter which specify the exact structure of the model. That is: there exists $\theta_{\text{true}} \in \Theta$ such that $f_{Y|X^*}(y|x^*;\theta_{\text{true}}) = f_{Y|X^*}(y|x^*)$. When x^* is observed in the sample, the identifiability of the parametric system can be approached via the nonsingularity of Fisher's information matrix evaluated at the true value of the parameter, θ_{true}. However, if x^* is not observed, only its error-contaminated counterpart, x, is observed, complications arise due to the presence of measurement error. In this paper, we consider parametric econometric models (1) with nonclassical measurement error when no additional data, such as validation data or double measurements, are available. The measurement error of the nonclassical type creates an identification problem that complicates consistent estimation of the parameter θ_{true}. To solve the identification problem, we impose several assumptions related to the completeness of families of conditional distributional functions. The intuition of these identification assumptions is to allows us to obtain a family of conditional distribution in which the standard conditional maximum likelihood estimate (CMLE) can be applied to. Under these assumptions, the parametric family $f_{Y|X^*}(y|x^*;\theta)$ can be transformed into the family of probability density functions, $f(x;\theta)$ without losing variation of the parameter θ. This family is then used by CMLE with the observed data to estimate the parameters of interest.

Our approach relies on completeness of families of distributions. This type of assumptions are quite weak and are commonly used in the literature on nonparametric instrumental variable models. Completeness assumptions are often phrased in terms of injectivity assumptions of an integral operator whose kernel function are the corresponding density functions. Completeness has been considered by several studies e.g. Newey and Powell (2003), Blundell, Chen, and Kristensen (2007), Hu and Shum (2008), Shiu and Hu (2010), and Carrasco, Florens, and Renault (2007), etc. The detailed discussions of the property can be found in D'Haultfoeuille (2011), Andrews (2011), and Hu and Shiu (2011).

The validity of the classical measurement error assumption has raised great concern in a number of practical applications.[1] Measurement error in a discrete variable, such as education, gender, or union status, is considered to be discrete, while the error in a continuous variable, such as wage or income, is believed to be continuous.[2] In the case of discretely distributed regressors, estimators using instrumental variables in nonclassical measurement error models have been developed. Mahajan (2006) studies a nonparametric regression model where one of the true regressors is a binary variable and shows that the regression function is nonparametrically identified in the presence of an instrumental variable that is correlated with the unobserved true underlying variable but unrelated to the measurement error. Hu (2008) provides a solution to nonlinear models with multi-value misclassification error using instrumental variables. The misclassification error is also allowed to be correlated with all the explanatory variables in the model. Lewbel (2007) considers a nonparametric or semiparametric regression model with mismeasured binary regressor. The model is similar to a model of average treatment effect where treatment may be misclassified. On the other hand, Chen, Hu, and Lewbel (2009) obtain nonparametric identification without additional sample information, such as instrumental variables or a secondary sample.

As for nonclassical nonlinear errors-in-variables models with continuously distributed variables (in both linear or nonlinear models), Chen, Hong, and Tamer (2005) and Chen, Hong, and Tarozzi (2008) make use of an auxiliary data set containing correctly measured observations to obtain consistent estimates of parameters in moment conditions. Hu and Schennach (2008) propose a approach in which the main identifying assumption is that some measures of location of the distribution of the measurement error (e.g. its mean, mode or median) equals zero. They show that the joint densities of observable variables are related to the joint densities of unobservable variables by an integral equation and the equation is shown to define the operator equivalent relationship. The identification of the model relies on the unique eigenvalue-eigenfunction decomposition of the integral operator, and the estimation is through sieve maximum likelihood estimation.

The rest of the paper is organized as follows. Section 2 introduces the completeness assumptions and provide the derivation of identification. Section 3 gives detailed discussion

[1] Studies in Bollinger (1998), Bound, Brown, Duncan, and Rodgers (1994), and Bound, Brown, and Mathiowetz (2001) provide provide evidences of nonclassical measurement errors in economics data sets.

[2] Discrete measurement error is also called misclassification error.

of the identification in continuous and discrete cases. Section 4 provides concluding remarks.

2. Identification

2.1. Assumptions and Results

Set the domains of the parametric density function $f_{Y|X^*}(y|x^*;\theta)$ and the true density function of interest $f_{Y|X^*}(y|x^*)$ be $\mathcal{Y}_\theta$ and $\mathcal{Y}$ respectively. Let $\mathcal{L}^2(\mathfrak{X},\omega)=\{h(\cdot):\int_{\mathfrak{X}}|h(x)|^2\omega(x)dx<\infty,\}$ be a weighted L^2 space such that $\int_{\mathfrak{X}}\omega(x)dx=1$. For a given parameter θ, define an operator as follows:

$$L_{f_{Y|X^*;\theta}}:\mathcal{L}^2(\mathcal{X}^*,\omega)\to\mathcal{L}^2(\mathcal{Y}_\theta)\text{ with}\tag{2}$$

$$(L_{f_{Y|X^*;\theta}}h)(y)=\int f_{Y|X^*}(y|x^*;\theta)h(x^*)\omega(x^*)dx^*,\quad\forall y\in\mathcal{Y}_\theta.\tag{3}$$

The following definition is introduced to define a generalization of the completeness.

Definition 2.1. *A family of functions $\{f(x,z):x\in\mathcal{X}\}$ satisfies a completeness condition if for $h(z)\in\mathcal{L}^2(\mathcal{Z})$ such that*

$$\int f(x,z)h(z)dz=0\quad\textit{for all } x.\tag{4}$$

then $h(z)=0$. That is: there does not exist a nonzero function in $\mathcal{L}^2(\mathcal{Z})$ orthogonal to the family of the functions $\{f(x,z):x\in\mathcal{X}\}$.

Given a conditional density function $f(x|z)$, we can define two families of functions, $\{f(x|z):x\in\mathcal{X}\}$ and $\{f(x|z):z\in\mathcal{Z}\}$. The completeness of the second family $\{f(x|z):z\in\mathcal{Z}\}$ is the same as the conditional expectation version of completeness in Andrews (2011), Newey and Powell (2003), and Hu and Shiu (2011), i.e., for $h(x)\in\mathcal{L}^2(\mathcal{X})$ if $E[h|z]=0$ for all z then $h=0$.

Assumption 2.1. *(Dependence between Y and X^*) Assume that for each $\theta\in\Theta$, the family of the latent density functions $\{f_{Y|X^*}(y|x^*;\theta):y\in\mathcal{Y}_\theta\}$ is complete over $\mathcal{L}^2(\mathcal{X}^*,\omega)$.*

This assumption guarantees that the operators $L_{f_{Y|X^*;\theta}}$ are invertible for all θ and also secures dependence between y and x^*. If y and x^* are independent then it violates completeness of $\{f_{Y|X^*}(y|x^*;\theta_{\text{true}}):y\in\mathcal{Y}\}$. Since the operator $L_{f_{Y|X^*;\theta}}$ is invertible for each θ, define

$f(x^*, x; \theta) \equiv \omega(x^*) \cdot L^{-1}_{f_{Y|X^*;\theta}}(f_{Y,X}(y,x))$. Although completeness of the family of functions, $\{f_{Y|X^*}(y|x^*;\theta) : y \in \mathcal{Y}_\theta\}$ makes $f(x^*, x; \theta)$ exist, we might lose variation of the parameter θ. That is: $f(x^*, x; \theta)$ is independent of the parameter θ or $f(x^*, x; \theta_1) = f(x^*, x; \theta_2)$ for θ_1 and θ_2 in Θ. Define

$$L_{\frac{f_{X^*,X;\theta}}{\omega(x^*)}} : \mathcal{L}^2(\mathcal{X}^*, \omega) \to \mathcal{L}^2(\mathcal{X}) \text{ with } (L_{\frac{f_{X^*,X;\theta}}{\omega(x^*)}} h)(x) = \int \frac{f(x^*, x; \theta)}{\omega(x^*)} h(x^*)\omega(x^*)dx^*.$$

The next assumption prevents this loss.

Assumption 2.2. *(Dependence between X and X^*) Assume the following conditions:*

(i) the family of the joint density functions $\{f_{Y,X}(y,x) : x \in \mathcal{X}\}$ is complete over $\mathcal{L}^2(\mathcal{Y})$;

(ii) the family of the parametric density functions $\{f_{Y|X^*}(y|x^*;\theta_{true}) : x^* \in \mathcal{X}^*\}$ is complete over $\mathcal{L}^2(\mathcal{Y})$, i.e., $\int_{\mathcal{Y}} f_{Y|X^*}(y|x^*;\theta_{true})h(y)dy = 0$ for any x^* implies $h = 0$.

The second part of the assumption is different from Assumption 2.1.[3] Consider some equivalent condition for Assumption 2.2. Define operators

$$L_{f_{Y,X}} : \mathcal{L}^2(\mathcal{Y}_\theta \cap \mathcal{Y}) \to \mathcal{L}^2(\mathcal{X}) \text{ with } (L_{f_{Y,X}} h)(x) = \int_{\mathcal{Y}_\theta \cap \mathcal{Y}} f_{Y,X}(y,x)h(y)dy,$$
$$\widetilde{L}_{f_{Y|X^*;\theta}} : \mathcal{L}^2(\mathcal{Y}_\theta \cap \mathcal{Y}) \to \mathcal{L}^2(\mathcal{X}^*, \omega) \text{ with } (\widetilde{L}_{f_{Y|X^*;\theta}} h)(x^*) = \int f_{Y|X^*}(y|x^*;\theta)h(y)dy.$$

For arbitrary $h \in \mathcal{L}^2(\mathcal{Y}_\theta \cap \mathcal{Y})$,

$$(L_{f_{Y,X}} h)(x) \tag{5}$$
$$= \int f_{Y,X}(y,x)h(y)dy \tag{6}$$
$$= \int \left(\int_{\mathcal{X}^*} f_{Y|X^*}(y|x^*;\theta) f(x^*, x; \theta)dx^* \right) h(y)dy \tag{7}$$
$$= \int_{\mathcal{X}^*} \left(\int f_{Y|X^*}(y|x^*;\theta)h(y)dy \right) f(x^*, x; \theta)dx^* \tag{8}$$
$$= \int_{\mathcal{X}^*} \frac{f(x^*, x; \theta)}{\omega(x^*)} \left(\widetilde{L}_{f_{Y|X^*;\theta}} h)(x^*) \right) \omega(x^*)dx^* \tag{9}$$
$$= (L_{\frac{f_{X^*,X;\theta}}{\omega(x^*)}} \widetilde{L}_{f_{Y|X^*;\theta}} h)(x). \tag{10}$$

[3]Consider $y = bx^* + \eta$ with $\mathcal{X}^* = \{0, 1\}$ and η is from standard normal distribution truncated by $[-\frac{1}{2}, \frac{1}{2}]$. The family $\{f_{Y|X^*}(y|x^*;\theta) : x^* \in \mathcal{X}^*\}$ is not complete in $\mathcal{L}^2(\mathcal{Y}_\theta)$ but the family $\{f_{Y|X^*}(y|x^*;\theta) : y \in \mathcal{Y}_\theta\}$ is complete in $\mathcal{L}^2(\mathcal{X}^*)$ by the results in Newey and Powell (2003) which are introduced as Theorem 2.2 & 2.2 in subsection 2.2.

It follows that $L_{f_{Y,X}} = L_{\frac{f_{X^*,X;\theta}}{\omega(x^*)}} \widetilde{L}_{f_{Y|X^*;\theta}}$. The equation suggests that the invertibility of the operators $L_{f_{Y,X}}$ and $\widetilde{L}_{f_{Y|X^*;\theta_{\text{true}}}}$ make the operator $L_{\frac{f_{X^*,X}}{\omega(x^*)}}$ invertible. Since the completeness of the families $\{f_{Y,X}(y,x) : x \in \mathcal{X}\}$ and $\{f_{Y|X^*}(y|x^*;\theta_{\text{true}}) : x^* \in \mathcal{X}^*\}$ in Assumption 2.2 ensures the invertibility of the operators $L_{f_{Y,X}}$ and $\widetilde{L}_{f_{Y|X^*;\theta_{\text{true}}}}$, the operator $L_{\frac{f_{X^*,X}}{\omega(x^*)}}$ is invertible under Assumption 2.2. This invertibility leads to the completeness of $\{\frac{f_{X^*,X}(x^*,x)}{\omega(x^*)} : x \in \mathcal{X}\}$ over $\mathcal{L}^2(\mathcal{X}^*, \omega)$. Since $\omega(x^*)$ is positive and bounded above, the family $\{f_{X^*,X}(x^*,x) : x \in \mathcal{X}\}$ is also complete over $\mathcal{L}^2(\mathcal{X}^*, \omega)$. Notice that (1) Assumption 2.2(i) imposes restrictions on the observables joint density functions $f_{Y,X}(y,x)$ so that it is testable; (2) the independence between X^* and X fails Assumption 2.2. By the definition, $f(x^*,x;\theta_{\text{true}}) = f(x^*,x)$. If $\theta \neq \theta_{\text{true}}$ then $f_{Y|X^*}(y|x^*;\theta) \neq f_{Y|X^*}(y|x^*;\theta_{\text{true}})$. Follow the definition of $f(x^*,x;\theta)$,

$$f_{Y,X}(y,x) = \int f_{Y|X^*}(y|x^*;\theta) f(x^*,x;\theta) dx^*, \qquad \forall y \in \mathcal{Y}_\theta \cap \mathcal{Y} \tag{11}$$

$$f_{Y,X}(y,x) = \int f_{Y|X^*}(y|x^*;\theta_{\text{true}}) f(x^*,x;\theta_{\text{true}}) dx^* \quad \forall y \in \mathcal{Y}. \tag{12}$$

Then,

$$0 = \int f_{Y|X^*}(y|x^*;\theta) f(x^*,x;\theta) - f_{Y|X^*}(y|x^*;\theta_{\text{true}}) f(x^*,x;\theta_{\text{true}}) dx^*, \tag{13}$$

$$\begin{aligned} &= \int (f_{Y|X^*}(y|x^*;\theta) f(x^*,x;\theta) - f_{Y|X^*}(y|x^*;\theta) f(x^*,x;\theta_{\text{true}}) dx^* \qquad (14) \\ &\quad + \int f_{Y|X^*}(y|x^*;\theta) f(x^*,x;\theta_{\text{true}}) - f_{Y|X^*}(y|x^*;\theta_{\text{true}}) f(x^*,x;\theta_{\text{true}}) dx^*, \\ &= \int f_{Y|X^*}(y|x^*;\theta) \left(f(x^*,x;\theta) - f(x^*,x;\theta_{\text{true}})\right) dx^* \qquad (15) \\ &\quad + \int \left(f_{Y|X^*}(y|x^*;\theta) - f_{Y|X^*}(y|x^*;\theta_{\text{true}})\right) f(x^*,x;\theta_{\text{true}}) dx^*. \end{aligned}$$

Suppose that $f(x^*,x;\theta) = f(x^*,x;\theta_{\text{true}})$ for $\theta \neq \theta_{\text{true}}$. Plugging the equation into Eq. (14) implies that for all $(y,x) \in (\mathcal{Y}_\theta \cap \mathcal{Y}) \times \mathcal{X}$,

$$0 = \int \left(f_{Y|X^*}(y|x^*;\theta) - f_{Y|X^*}(y|x^*;\theta_{\text{true}})\right) \frac{f(x^*,x;\theta_{\text{true}})}{\omega(x^*)} \omega(x^*) dx^*.$$

The completeness of the family $\{\frac{f_{X^*,X}(x^*,x)}{\omega(x^*)} : x \in \mathcal{X}\}$ over $\mathcal{L}^2(\mathcal{X}^*, \omega)$ under Assumption 2.2 results in $f_{Y|X^*}(y|x^*;\theta) = f_{Y|X^*}(y|x^*;\theta_{\text{true}})$ for $\theta \neq \theta_{\text{true}}$, a contradiction. Hence, $f(x^*,x;\theta)$ is well defined and $f(x^*,x;\theta)$ is still a function of θ, i.e., $f(x^*,x;\theta) \neq f(x^*,x;\theta_{\text{true}})$ if $\theta \neq \theta_{\text{true}}$.

To identify θ from observable distribution, it is necessary to integrate out the unobservable true covariate x^*. Denote $\bar{f}(x;\theta) \equiv \int f(x^*, x; \theta) dx^*$. However, the parameter θ might disappear or lose variation after integrating out x^* in $f(x^*, x; \theta)$. The following assumption related to the structure of the parameter θ is needed to make the parameter survive after the integration.

Assumption 2.3. *(Variation in Parameter) The family of derivatives of the parametric density functions over θ, $\{\frac{1}{\omega(x^*)} \frac{\partial}{\partial \theta} f_{Y|X^*}(y|x^*; \theta_{true}) : x^* \in \mathcal{X}^*\}$, is complete over $\mathcal{L}^2(\mathcal{Y})$.*

This assumption of derivation has an intuitive sense since it implies that completeness is well preserved when the family $\{f_{Y|X^*}(y|x^*; \theta_{\text{true}}) : x^* \in \mathcal{X}^*\}$ varies around the true parameter θ_{true}. Denote $K_{A;\theta}(x^*, x) = \frac{f(x^*,x;\theta) - f(x^*,x)}{\omega(x^*)}$ and $K_{B;\theta}(y, x^*) = \frac{f_{Y|X^*}(y|x^*;\theta) - f_{Y|X^*}(y|x^*)}{\omega(x^*)}$ for $\theta \neq \theta_{\text{true}}$. Define operators

$$(16) \qquad L_{K_{A;\theta}} : \mathcal{L}^2(\mathcal{X}^*, \omega) \to \mathcal{L}^2(\mathcal{X}) \text{ with } (L_{K_{A;\theta}} h)(x) = \int K_{A;\theta}(x^*, x) h(x^*) \omega(x^*) dx^*,$$

$$(17) \qquad L_{K_{B;\theta}} : \mathcal{L}^2(\mathcal{Y}_\theta \cap \mathcal{Y}) \to \mathcal{L}^2(\mathcal{X}^*, \omega) \text{ with } (L_{K_{B;\theta}} h)(x^*) = \int K_{B;\theta}(y, x^*) h(y) dy,$$

$$(18) \qquad L_{f_{X^*,X;\theta}} : \mathcal{L}^2(\mathcal{X}^*, \omega) \to \mathcal{L}^2(\mathcal{X}) \text{ with } (L_{f_{X^*,X;\theta}} h)(x) = \int f(x^*, x; \theta) h(x^*) \omega(x^*) dx^*.$$

From Eq. (13)-(15), for all $(y, x) \in (\mathcal{Y}_\theta \cap \mathcal{Y}) \times \mathcal{X}$

$$(19) \qquad 0 = \int f_{Y|X^*}(y|x^*; \theta) K_{A;\theta}(x^*, x) \omega(x^*) dx^* + \int K_{B;\theta}(y, x^*) f(x^*, x) \omega(x^*) dx^*.$$

For arbitrary $h \in \mathcal{L}^2(\mathcal{Y}_\theta \cap \mathcal{Y})$,

$$(L_{K_{A;\theta}} \widetilde{L}_{f_{Y|X^*;\theta}} h)(x) \tag{20}$$

$$= \int_{x^*} K_{A;\theta}(x^*, x)(\widetilde{L}_{f_{Y|X^*;\theta}} h)(x^*)\omega(x^*)dx^* \tag{21}$$

$$= \int_{x^*} K_{A;\theta}(x^*, x) \int_y f_{Y|X^*}(y|x^*;\theta)h(y)dy\omega(x^*)dx^* \tag{22}$$

$$= \int_y \left(\int_{x^*} f_{Y|X^*}(y|x^*;\theta) K_{A;\theta}(x^*, x)\omega(x^*)dx^* \right) h(y)dy \tag{23}$$

$$= -\int_y \left(\int_{x^*} K_{B;\theta}(y, x^*) f(x^*, x)\omega(x^*)dx^* \right) h(y)dy \tag{24}$$

$$= -\int_{x^*} f(x^*, x) \left(\int_y K_{B;\theta}(y, x^*)h(y)dy \right) \omega(x^*)dx^* \tag{25}$$

$$= -\int_{x^*} f(x^*, x)(L_{K_{B;\theta}} h)(x^*)\omega(x^*)dx^* \tag{26}$$

$$= -(L_{f_{X^*,X}} L_{K_{B;\theta}} h)(x), \tag{27}$$

where we have used (i) Eq. (19), (ii) an interchange of the order of integration (justified by Fubini's theorem), and (iii) the definitions of $L_{K_{A;\theta}}$, $\widetilde{L}_{f_{Y|X^*;\theta}}$, $L_{f_{X^*,X}}$, and and $L_{K_{B;\theta}}$. It follows that

$$0 = L_{K_{A;\theta}} \widetilde{L}_{f_{Y|X^*;\theta}} + L_{f_{X^*,X}} L_{K_{B;\theta}}. \tag{28}$$

Consider a parameter $\theta = \theta_{\text{true}} + h$ for some small $h \neq 0$. Plug this equation into Eq. (28) and divide the equation by h,

$$0 = \left(\frac{1}{h} L_{K_{A;\theta_{\text{true}}+h}} \right) \widetilde{L}_{f_{Y|X^*;\theta_{\text{true}}+h}} + L_{f_{X^*,X}} \left(\frac{1}{h} L_{K_{B;\theta_{\text{true}}+h}} \right). \tag{29}$$

Define

$$L_{dK_{A;\theta}} : \mathcal{L}^2(\mathcal{X}^*, \omega) \to \mathcal{L}^2(\mathcal{X}) \text{ with} \tag{30}$$

$$(L_{dK_{A;\theta}} h)(x) = \int \left(\frac{1}{\omega(x^*)} \frac{\partial}{\partial \theta} f(x^*, x; \theta) \right) h(x^*)\omega(x^*)dx^*, \tag{31}$$

$$L_{dK_{B;\theta}} : \mathcal{L}^2(\mathcal{Y}) \to \mathcal{L}^2(\mathcal{X}^*, \omega) \text{ with} \tag{32}$$

$$(L_{dK_{B;\theta}} h)(x^*) = \int \left(\frac{1}{\omega(x^*)} \frac{\partial}{\partial \theta} f_{Y|X^*}(y|x^*;\theta) \right) h(y)dy. \tag{33}$$

Let $h \mapsto 0$, Eq. (29) becomes

$$0 = L_{dK_{A;\theta_{\text{true}}}} \widetilde{L}_{f_{Y|X^*;\theta_{\text{true}}}} + L_{f_{X^*,X}} L_{dK_{B;\theta_{\text{true}}}}. \tag{34}$$

Assumption 2.2(ii) and 2.3 guarantee the operators $\widetilde{L}_{f_{Y|X^*;\theta_{\text{true}}}}$, and $L_{dK_{B;\theta_{\text{true}}}}$ are invertible respectively. Applying these results to Eq. (34) with the invertibility of $L_{f_{X^*,X}}$ from Assumption 2.2 leads to the invertibility of the operator $L_{dK_{A;\theta_{\text{true}}}}$. This implies that the family of the derivatives over θ, $\{\frac{1}{\omega(x^*)}\frac{\partial}{\partial\theta} f(x^*, x; \theta_{\text{true}}) : x \in \mathcal{X}\}$, is complete over $\mathcal{L}^2(\mathcal{X}^*, \omega)$.

Suppose $\bar{f}(x;\theta) = \bar{f}(x;\theta_{\text{true}})$ for $\theta \neq \theta_{\text{true}}$. This is $\int f(x^*, x; \theta)dx^* = \int f(x^*, x)dx^*$. It follows that

$$\int \left(\frac{1}{\omega(x^*)} \frac{\partial}{\partial\theta} f(x^*, x; \theta_{\text{true}}) \right) \omega(x^*) dx^* = 0 \quad \text{for all } x. \tag{35}$$

Hence, a constant function is orthogonal to the family $\{\frac{1}{\omega(x^*)}\frac{\partial}{\partial\theta} f(x^*, x; \theta_{\text{true}}) : x \in \mathcal{X}\}$. Since the constant function is in $\mathcal{L}^2(\mathcal{X}^*, \omega)$, Eq. (35) contravenes completeness of $\{\frac{1}{\omega(x^*)}\frac{\partial}{\partial\theta} f(x^*, x; \theta_{\text{true}}) : x \in \mathcal{X}\}$. Therefore, under Assumptions 2.1-2.3, $\bar{f}(x;\theta)$ is well defined and $\bar{f}(x;\theta) \neq \bar{f}(x;\theta_{\text{true}})$ if $\theta \neq \theta_{\text{true}}$.

Remark: one of necessary conditions from Assumptions 2.1-2.3 is that $\mathcal{Y}_\theta = \mathcal{Y}$ for all θ is ruled out. Suppose that $\mathcal{Y}_\theta = \mathcal{Y}$. Integrating out y over the domain $\mathcal{Y}$ in Eq. (11) and (12) and interchanging integrations, we obtain

$$f_X(x) = \int_{x^*} \left(\int_y f_{Y|X^*}(y|x^*;\theta) dy \right) f(x^*, x; \theta) dx^* = \int_{x^*} f(x^*, x; \theta) dx^* \quad \text{for all } \theta \in \Theta. \tag{36}$$

There is no variation of the parameter θ after the integration. This observation shows that Assumptions 2.1-2.3 restricts the domains, $\mathcal{Y}_\theta$ and $\mathcal{Y}$.

Since the functions $\bar{f}(x;\theta)$ have a valid parametric form, the parametric family $\{\bar{f}(x;\theta) : \theta \in \Theta\}$ might provide a parameterized family of probability density functions (pdf) for $f(x)$. There are two requirements for being a parametric pdf, (1) it is nonnegative everywhere, and (2) its integral over the entire space is equal to one. In order to satisfy these conditions, consider the normalized function $f(x;\theta) \equiv \frac{|\bar{f}(x;\theta)|}{\int_{\mathcal{X}} |\bar{f}(x;\theta)| dx}$ which is positive and its integral over the entire space is equal to one. Similar to the previous consideration, this normalization might not carry the variation of θ. The next assumption makes the values of $\bar{f}(x;\theta)$ positive for any parameter in Θ and then $f(x;\theta) = \frac{\bar{f}(x;\theta)}{\int_{\mathcal{X}} \bar{f}(x;\theta)dx}$ under the assumption.

Assumption 2.4. *(Positivity around θ_{true}) Assume Θ is an open neighbor of θ_{true} such that*

(i) the operator $L_{f_{Y|X^*;\theta}}$ is bounded below for $\theta \in \Theta$, i.e.,

$$c_1 \leq \sup_{f \in \mathcal{L}^2(\mathcal{X}^*,\omega)} \frac{\left\| L_{f_{Y|X^*;\theta}}(f) \right\|}{\|f\|} \equiv \left\| L_{f_{Y|X^*;\theta}} \right\| \quad \forall \theta \in \Theta;$$

(ii) $\int_y \int_{x^*} \left(f_{Y|X^*}(y|x^*;\theta) - f_{Y|X^*}(y|x^*;\theta_{true}) \right)^2 \omega(x^*) dx^* dy \leq c_2 |\theta - \theta_{true}| \quad \forall \theta \in \Theta$;

(iii) $\|f_{Y,X}(\cdot,x)\|^2 = \int_{\mathcal{Y}} f_{Y,X}(y,x)^2 dy \leq M$ for all $x \in \mathcal{X}$;

(iv) the density function $f(x)$ is bounded below, i.e., $f(x) > c_3 > 0$.

Lemma 2.1. *Suppose that Assumptions 2.1 and 2.4 hold. Then $\bar{f}(x;\theta) > 0$ for all x and $\theta \in \Theta$.*

Proof Given $h_1 \in \mathcal{L}^2(\mathcal{Y}_\theta \cap \mathcal{Y})$,

$$h_1 = L_{f_{Y|X^*;\theta}} L^{-1}_{f_{Y|X^*;\theta}} h_1 \tag{37}$$

$$h_1 = L_{f_{Y|X^*;\theta_{\text{true}}}} L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1. \tag{38}$$

Subtracting Eq. (38) from Eq. (37),

$$\begin{aligned} 0 &= L_{f_{Y|X^*;\theta}} L^{-1}_{f_{Y|X^*;\theta}} h_1 - L_{f_{Y|X^*;\theta}} L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1 \\ &\quad + L_{f_{Y|X^*;\theta}} L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1 - L_{f_{Y|X^*;\theta_{\text{true}}}} L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1 \\ &= L_{f_{Y|X^*;\theta}} \left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right) h_1 + \left(L_{f_{Y|X^*;\theta}} - L_{f_{Y|X^*;\theta_{\text{true}}}} \right) L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1. \end{aligned}$$

Applying the definition of the operator norm to the above equation leads to

$$\left\| L_{f_{Y|X^*;\theta}} \left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right) \right\| = \left\| \left(L_{f_{Y|X^*;\theta}} - L_{f_{Y|X^*;\theta_{\text{true}}}} \right) L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right\|.$$

Conditions (i) implies $\left\| L^{-1}_{f_{Y|X^*;\theta}} \right\| \leq \frac{1}{\left\| L_{f_{Y|X^*;\theta}} \right\|} \leq \frac{1}{c_2}$. In addition, for $h_2 \in \mathcal{L}^2(\mathcal{X}^*, \omega)$,

$$\begin{aligned}
&\left\| \left(L_{f_{Y|X^*;\theta}} - L_{f_{Y|X^*;\theta_{\text{true}}}} \right) h_2 \right\|^2 \\
&= \int_y \left(\int_{x^*} \left(f_{Y|X^*}(y|x^*;\theta) - f_{Y|X^*}(y|x^*;\theta_{\text{true}}) \right) h_2(x^*)\omega(x^*)dx^* \right)^2 dy \\
&\leq \int_y \left(\int_{x^*} \left(f_{Y|X^*}(y|x^*;\theta) - f_{Y|X^*}(y|x^*;\theta_{\text{true}}) \right)^2 \omega(x^*)dx^* \right) \left(\int_{x^*} h_2(x^*)^2\omega(x^*)dx^* \right) dy \\
&\equiv \left\| f_{Y|X^*;\theta} - f_{Y|X^*;\theta_{\text{true}}} \right\|^2 \left\| h_2 \right\|^2 .
\end{aligned}$$

It follows that $\left\| L_{f_{Y|X^*;\theta}} - L_{f_{Y|X^*;\theta_{\text{true}}}} \right\| \leq \left\| f_{Y|X^*;\theta} - f_{Y|X^*;\theta_{\text{true}}} \right\|$. Combining these results with Condition (ii) , we obtain

$$\begin{aligned}
c_2 \left\| L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right\| &\leq \left\| L_{f_{Y|X^*;\theta}} \left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right) \right\| \\
&= \left\| \left(L_{f_{Y|X^*;\theta}} - L_{f_{Y|X^*;\theta_{\text{true}}}} \right) L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right\| \\
&\leq \left\| L_{f_{Y|X^*;\theta}} - L_{f_{Y|X^*;\theta_{\text{true}}}} \right\| \left\| L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right\| \\
&\leq \frac{1}{c_1} \left\| f_{Y|X^*;\theta} - f_{Y|X^*;\theta_{\text{true}}} \right\| \\
&\leq \frac{c_2}{c_1} |\theta - \theta_{\text{true}}|.
\end{aligned}$$

Using these results to compute the difference between $\bar{f}(x;\theta)$ and $f(x)$,

$$\begin{aligned}
&|\bar{f}(x;\theta) - f(x)|^2 \\
&= \left| \int_{x^*} f(x^*, x;\theta) - f(x^*, x;\theta_{\text{true}})dx^* \right|^2 \\
&= \left| \int_{x^*} \omega(x^*) \left(L^{-1}_{f_{Y|X^*;\theta}}(f_{Y,X}(y,x)) - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}(f_{Y,X}(y,x)) \right) dx^* \right|^2 \\
&= \left| \int_{x^*} \omega(x^*) \left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right) (f_{Y,X}(y,x))dx^* \right|^2 \\
&\leq \int_{x^*} \left[\left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right) (f_{Y,X}(y,x)) \right]^2 \omega(x^*)dx^* \int_{x^*} \omega(x^*)dx^* \\
&\leq c_4 \left\| L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} \right\|^2 \left\| f_{Y,X}(\cdot, x) \right\|^2 \\
&\leq \frac{c_2^2 c_4 M}{c_1^4} |\theta - \theta_{\text{true}}|^2.
\end{aligned}$$

This suggests $f(x) - \bar{f}(x;\theta) \leq |\bar{f}(x;\theta) - f(x)| \leq c_5|\theta - \theta_{\text{true}}|$. Therefore, $\bar{f}(x;\theta)$ is bounded below by

$$0 < c_3 - c_5|\theta - \theta_{\text{true}}| < f(x) - c_5|\theta - \theta_{\text{true}}| \le \bar{f}(x;\theta) \quad \forall \theta \in \Theta,$$

where we have used (i) Θ is an open neighbor of θ_{true}, (ii) Condition (iv), $f(x) > c_3 > 0$.

Q.E.D.

Assumption 2.5. *(Normalization) Suppose that the family of the functions* $\{\frac{\partial}{\partial x} f_{Y,X}(y,x) - f'(x) f_{Y|X}(y|x) : x \in \mathcal{X}\}$ *is complete over* $\mathcal{L}^2(\mathcal{Y})$*.*

Lemma 2.2. *Under Assumption 2.2(ii), Assumption 2.3, and Assumption 2.5, the values of the function* $f(x;\theta)$ *still has a variation in* θ*, i.e., there exists* $\theta \neq \theta_{true}$ *such that* $f(x;\theta) \neq f(x;\theta_{true})$*.*

Proof Set $h_1(y,x) = \frac{\partial}{\partial x} f_{Y,X}(y,x) - f'(x) f_{Y|X}(y|x)$. Use Eq. (12) to rewrite $h_1(y,x)$ as,

$$\begin{aligned} h_1(y,x) &= \frac{\partial}{\partial x} f_{Y,X}(y,x) - f'(x)\frac{f_{Y,X}(y,x)}{f(x)} \\ &= \int f_{Y|X^*}(y|x^*;\theta_{\text{true}}) \left[\frac{\partial}{\partial x} f(x^*,x;\theta_{\text{true}}) - f'(x)\frac{f(x^*,x;\theta_{\text{true}})}{f(x)}\right] dx^*. \end{aligned} \tag{39}$$

This implies that

$$L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1(y,x) = \frac{\partial}{\partial x} f(x^*,x;\theta_{\text{true}}) - f'(x)\frac{f(x^*,x;\theta_{\text{true}})}{f(x)}.$$

Define operators

$$\begin{aligned} &L_{h_1(y,x)} : \mathcal{L}^2(\mathcal{Y}) \to \mathcal{L}^2(\mathcal{X}) \text{ with } (L_{h_1(y,x)}h)(x) = \int_{\mathcal{Y}} h_1(y,x)h(y)dy, \\ &L_{K_{D_1}} : \mathcal{L}^2(\mathcal{X}^*,\omega) \to \mathcal{L}^2(\mathcal{X}) \text{ with} \\ &(L_{K_{D_1}}h)(x) = \int \left(L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1(y,x)\right) h(x^*)\omega(x^*)dx^*. \end{aligned}$$

With these definitions of operators, we can transform Eq. (39) into an operator relationship $L_{h_1(y,x)} = L_{K_{D_1}} \widetilde{L}_{f_{Y|X^*;\theta}}$. Since $\widetilde{L}_{f_{Y|X^*;\theta}}$ is invertible by Assumption 2.2(ii) and the operator $L_{h_1(y,x)}$ is invertible by the completeness of the family $\{h_1(y,x) : x \in \mathcal{X}\}$ over $\mathcal{L}^2(\mathcal{Y})$, the operator $L_{K_{D_1}}$ is invertible. This amounts to the family $\{L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1(y,x) : x \in \mathcal{X}\}$ is complete over $\mathcal{L}^2(\mathcal{X}^*,\omega)$. This implies that the family $\{\omega(x^*)L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}h_1(y,x) : x \in \mathcal{X}\}$ is also complete over $\mathcal{L}^2(\mathcal{X}^*,\omega)$.

On the other hand,

$$\begin{aligned}
h_1(y,x) &= L_{f_{Y|X^*;\theta}} L^{-1}_{f_{Y|X^*;\theta}} h_1(y,x)\\
h_1(y,x) &= L_{f_{Y|X^*;\theta_{\text{true}}}} L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}} h_1(y,x).
\end{aligned}$$

It follows that

$$\begin{aligned}
0 &= L_{f_{Y|X^*;\theta}}\left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}\right)h_1(y,x) + \left(L_{f_{Y|X^*;\theta}} - L_{f_{Y|X^*;\theta_{\text{true}}}}\right)L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}h_1(y,x),\\
&= \int_{x^*} f_{Y|X^*}(y|x^*;\theta)\left[\left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}\right)h_1(y,x)\right]\omega(x^*)dx^*\\
&\quad + \int_{x^*}\left(f_{Y|X^*;\theta} - f_{Y|X^*;\theta_{\text{true}}}\right)\left[L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}h_1(y,x)\right]\omega(x^*)dx^*
\end{aligned}$$

Set $K_{C;\theta}(x^*,x) = \left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}\right)h_1(y,x)$ and $K_{D_2}(x^*,x) = \omega(x^*)L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}h_1(y,x)$. Then, the above equation becomes

$$0 = \int_{x^*} f_{Y|X^*}(y|x^*;\theta)K_{C;\theta}(x^*,x)\omega(x^*)dx^* + \int_{x^*} K_{B;\theta}(y,x^*)K_{D_2}(x^*,x)\omega(x^*)dx^*. \tag{40}$$

Define

$$\begin{aligned}
&L_{K_{C;\theta}} : \mathcal{L}^2(\mathcal{X}^*,\omega) \to \mathcal{L}^2(\mathcal{X}) \text{ with}\\
&(L_{K_{C;\theta}}h)(x) = \int \left(\left[\left(L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}\right)h_1(y,x)\right]\right)h(x^*)\omega(x^*)dx^*,\\
&L_{dK_{C;\theta_{\text{true}}}} : \mathcal{L}^2(\mathcal{X}^*,\omega) \to \mathcal{L}^2(\mathcal{X}) \text{ with}\\
&(L_{dK_{C;\theta_{\text{true}}}}h)(x) = \int \left[\left(\lim_{\theta\to\theta_{\text{true}}} \frac{L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}}{\theta-\theta_{\text{true}}}\right)h_1(y,x)\right]h(x^*)\omega(x^*)dx^*,\\
&L_{K_{D_2}} : \mathcal{L}^2(\mathcal{X}^*,\omega) \to \mathcal{L}^2(\mathcal{X}) \text{ with}\\
&(L_{K_{D_2}}h)(x) = \int \left(\omega(x^*)\left[L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}h_1(y,x)\right]\right)h(x^*)\omega(x^*)dx^*.
\end{aligned}$$

With these notations, similar to the derivation in Eq. (28), given $h \in \mathcal{L}^2(\mathcal{Y})$,

$$
\begin{aligned}
&(L_{K_{C;\theta}} \widetilde{L}_{f_{Y|X^*;\theta}} h)(x) \\
&= \int_{x^*} K_{C;\theta}(x^*, x)(\widetilde{L}_{f_{Y|X^*;\theta}} h)(x^*)\omega(x^*)dx^* \\
&= \int_{x^*} K_{C;\theta}(x^*, x) \int_y f_{Y|X^*}(y|x^*;\theta)h(y)dy\omega(x^*)dx^* \\
&= \int_y \left(\int_{x^*} f_{Y|X^*}(y|x^*;\theta) K_{C;\theta}(x^*, x)\omega(x^*)dx^* \right) h(y)dy \\
&= -\int_y \left(\int_{x^*} K_{B;\theta}(y, x^*) K_{D_2}(x^*, x)\omega(x^*)dx^* \right) h(y)dy \\
&= -\int_{x^*} K_{D_2}(x^*, x) \left(\int_y K_{B;\theta}(y, x^*)h(y)dy \right) \omega(x^*)dx^* \\
&= -(L_{K_{D_2}} L_{K_{B;\theta}} h)(x),
\end{aligned}
$$

where we have used Eq. (40). Therefore, $L_{K_{C;\theta}} \widetilde{L}_{f_{Y|X^*;\theta}} + L_{K_{D_2}} L_{K_{B;\theta}} = 0$. Divide the operator relationship by $\theta - \theta_{\text{true}}$ and let $\theta \to \theta_{\text{true}}$,

$$
L_{dK_{C;\theta_{\text{true}}}} \widetilde{L}_{f_{Y|X^*;\theta_{\text{true}}}} + L_{K_{D_2}} L_{dK_{B;\theta_{\text{true}}}} = 0, \tag{41}
$$

which is analogous to Eq. (34). Applying the invertibility of the operators $L_{K_{D_2}}$ to Eq. (41) along with invertibility of the operators $\widetilde{L}_{f_{Y|X^*;\theta_{\text{true}}}}$ and $L_{dK_{B;\theta_{\text{true}}}}$ from Assumption 2.2(ii) and Assumption 2.3 respectively, the operator $L_{dK_{C;\theta_{\text{true}}}}$ is also invertible. This invertibility implies that the family $\{ \left(\lim_{\theta \to \theta_{\text{true}}} \frac{L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}}{\theta - \theta_{\text{true}}} \right) h_1(y, x) : x \in \mathcal{X} \}$ is complete over $\mathcal{L}^2(\mathcal{X}^*, \omega)$.

Suppose that $f(x;\theta) = f(x;\theta_{\text{true}})$ for all $\theta \in \theta_{\text{true}}$. Write this out $\int f(x^*, x;\theta)dx^* = f(x)\int_{\mathcal{X}} \bar{f}(x;\theta)dx$. This suggests that for all x

$$
\int \frac{f(x^*, x;\theta)}{f(x)} dx^* = \int_{\mathcal{X}} \bar{f}(x;\theta)dx. \tag{42}
$$

The right-hand-side of the above equation is independent of x. Thus, given x_1 and x_2,

$$
\int \frac{f(x^*, x_1;\theta)}{f(x_1)} dx^* = \int \frac{f(x^*, x_2;\theta)}{f(x_2)} dx^*. \tag{43}
$$

It follows that

$$
\begin{aligned}
0 &= \int \left(\frac{f(x^*, x_1; \theta)}{f(x_1)} - \frac{f(x^*, x_2; \theta)}{f(x_2)} \right) dx^*, \\
&= \int \omega(x^*) \left(L^{-1}_{f_{Y|X^*;\theta}} \left(\frac{f_{Y,X}(y, x_1)}{f(x_1)} \right) - L^{-1}_{f_{Y|X^*;\theta}} \left(\frac{f_{Y,X}(y, x_2)}{f(x_2)} \right) \right) dx^*, \\
&= \int \omega(x^*) L^{-1}_{f_{Y|X^*;\theta}} \left(\frac{f_{Y,X}(y, x_1)}{f(x_1)} - \frac{f_{Y,X}(y, x_2)}{f(x_2)} \right) dx^*, \\
&= \frac{1}{f(x_1)} \int \omega(x^*) L^{-1}_{f_{Y|X^*;\theta}} \left([f_{Y,X}(y, x_1) - f_{Y,X}(y, x_2)] - \frac{f(x_1) - f(x_2)}{f(x_2)} f_{Y,X}(y, x_2) \right) dx^*.
\end{aligned}
$$

Set $x_1 = x + h$ and $x_2 = x$ in the above equation and divide it by h.

$$
0 = \int \omega(x^*) L^{-1}_{f_{Y|X^*;\theta}} \left(\frac{f_{Y,X}(y, x+h) - f_{Y,X}(y, x)}{h} - \frac{f(x+h) - f(x)}{h} \frac{f_{Y,X}(y, x)}{f(x)} \right) dx^*.
$$

Let $h \to 0$,

$$
\begin{aligned}
0 &= \int \omega(x^*) L^{-1}_{f_{Y|X^*;\theta}} \left(\frac{\partial}{\partial x} f_{Y,X}(y, x) - f'(x) \frac{f_{Y,X}(y, x)}{f(x)} \right) dx^* \\
&= \int \omega(x^*) L^{-1}_{f_{Y|X^*;\theta}} \left(h_1(y, x) \right) dx^* \text{ for all } \theta.
\end{aligned}
$$

Subtracting the above equation with parameters $\theta \neq \theta_{\text{true}}$ from the one with θ_{true} and then dividing it by $\theta - \theta_{\text{true}}$, we have

$$
0 = \int \left[\left(\frac{L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}}{\theta - \theta_{\text{true}}} \right) h_1(y, x) \right] \omega(x^*) dx^*,
$$

which is contradict to the completeness of $\{ \left(\lim_{\theta \to \theta_{\text{true}}} \frac{L^{-1}_{f_{Y|X^*;\theta}} - L^{-1}_{f_{Y|X^*;\theta_{\text{true}}}}}{\theta - \theta_{\text{true}}} \right) h_1(y, x) : x \in \mathcal{X} \}$. Therefore, $f(x; \theta)$ has a nontrivial variation in θ. *Q.E.D.*

To distinguish the true value of the parameter θ_{true} from a generic element of Θ, the Kullback information is introduced:

$$
K(\theta) = E \left[\log \left(\frac{f(x; \theta)}{f(x)} \right) \right] \tag{44}
$$

$$
= \int_{-\infty}^{\infty} \log \left(\frac{f(x; \theta)}{f(x)} \right) f(x) dx. \tag{45}
$$

Note that since $f_{Y|X^*}(y|x^*; \theta)$ is correctly specified at θ_{true}, $f(x; \theta)$ is also correctly specified at θ_{true} and $f(x;\theta_{\text{true}}) = f(x)$ if $L_{f_{Y|X^*;\theta_{\text{true}}}}$ is invertible.

Theorem 2.1. *Under Assumptions 2.1-2.5, given the distribution of the observable variable (y, x), the equation*

$$K(\theta) = 0 \tag{46}$$

has a unique solution at $\theta = \theta_{true}$ in Θ and $f_{Y|X^*}(y|x^*;\theta)$ is identified at θ_{true}.

Proof Using the strict concavity of $\log(\cdot)$ and applying Jensen's inequality, for $f(x;\theta) \neq f(x)$

$$K(\theta) = \int_{-\infty}^{\infty} \log\left(\frac{f(x;\theta)}{f(x)}\right) f(x)dx \tag{47}$$

$$< \log \int_{-\infty}^{\infty} \left(\frac{f(x;\theta)}{f(x)}\right) f(x)dx \tag{48}$$

$$= \log 1 \tag{49}$$

$$= 0. \tag{50}$$

and $K(\theta_{\text{true}}) = 0$. Since $f(x;\theta) \neq f(x;\theta_{\text{true}})$ if $\theta \neq \theta_{\text{true}}$, $K(\theta) < 0$ for $\theta \neq \theta_{\text{true}}$. $K(\theta) = 0$ has a unique solution at $\theta = \theta_{\text{true}}$. This implies that $f(x;\theta)$ is identified at $\theta = \theta_{\text{true}}$. It shows that no other $\theta \neq \theta_{\text{true}}$ generate a distribution indistinguishable from $f(x)$ on the basis of sample observations. Therefore, $f_{Y|X^*}(y|x^*;\theta)$ is identified at θ_{true}. *Q.E.D.*

Notice that identifiability in Theorem 2.1 is closely connected with the maximum of $K(\theta)$. If this maximum is global and attained only at $\theta = \theta_{\text{true}}$, then θ_{true} is globally identified. Thus, a sufficient condition for θ_{true} to be globally identified would be that $K'(\theta_{\text{true}}) = 0$ and $K''(\theta_{\text{true}}) < 0$ if K is differentiable. First, we investigate the form of these derivatives for a scalar θ. Since $\int_{-\infty}^{\infty} f(x;\theta)dx = 1$ for any θ, it follows that $\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x;\theta)dx = 0$ and $\int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} f(x;\theta)dx = 0$. Differentiating (45), we obtain

$$K'(\theta) = \int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)} f(x)dx,$$

and

$$K''(\theta) = \int_{-\infty}^{\infty} \left[\frac{\frac{\partial^2}{\partial\theta^2} f(x;\theta)}{f(x;\theta)} - \left(\frac{\frac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)} \right)^2 \right] f(x)dx$$
$$= -\int_{-\infty}^{\infty} \left(\frac{\frac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)} \right)^2 f(x)dx < 0.$$

When θ is extended to a vector of parameters, $K''(\theta)$ becomes the classical information matrix. If $K''(\theta_{\text{true}})$ is negative definite, standard maximization theory implies that $K''(\theta)$ has a unique maximum at θ_{true}. That is, if the information matrix at θ_{true} has full rank, then θ_{true} is locally identified

2.2. Examples

The identification of the parametric family relies on completeness of the conditional distribution of y given x^*, $f_{Y|X^*}(y|x^*;\theta)$. Discussions of the well-known completeness property of exponential families is introduced in this section. In addition, assumptions are illustrated using examples. The demonstration will be divided into two cases, a continuous case and a discrete case. The following results for completeness are adopted from Newey and Powell (2003).[4]

Theorem 2.2. *Consider $f(x|z) = s(x)t(z)\exp(\mu(z)\tau(x))$, where $s(x) > 0$, $\tau(x)$ is one-to-one in x, and support of $\mu(z)$ is an open set, then $E(h(x)|z) = 0$ for any z implies $h = 0$; equivalently, the family of conditional density functions $\{f(x|z) = s(x)t(z)\exp(\mu(z)\tau(x)) : z \in \mathcal{Z}\}$ is complete over $\mathcal{L}^2(\mathcal{X})$.*

The next result involves completeness in the conditional normal case.

Theorem 2.3. *Suppose that the distribution of x conditional on z is $N(a+bz,\sigma^2)$ for $\sigma^2 > 0$ and the support of z contains an open set, then $E(h(x)|z) = 0$ for any z implies $h = 0$; equivalently, $\{f(x|z) : z \in \mathcal{Z}\}$ is complete over $\mathcal{L}^2(\mathcal{X})$.*

[4]See Theorem 2.2 and 2.3 in Newey and Powell (2003) for details. There are more discussion of sufficient conditions for the completeness in Hu and Shiu (2011).

2.2.1. Continuous Case

Consider parametric models

$$Y = bX^* + \eta \tag{51}$$

$$X^* \perp \eta$$

We then have the relationship between the Fourier Transform

$$\phi_{Y,X}(t,x) = \int e^{ity} f_{Y,X}(y,x)dy \tag{52}$$

$$\phi_{Y,X}(t,x) = \phi_\eta(t)\phi_{X^*,X}(bt,x) \tag{53}$$

We assume ϕ_η is known and the unknowns are b and $\phi_{X^*,X}$. We have

$$\phi_{X^*,X}(t,x;b) = \frac{\phi_{Y,X}(t/b,x)}{\phi_\eta(t/b)}$$

Then

$$\begin{aligned} f_{X^*,X}(x^*,x;b) &= \frac{1}{2\pi}\int e^{-itx^*}\phi_{X^*,X}(t,x)dt \\ &= \frac{1}{2\pi}\int e^{-itx^*}\frac{\phi_{Y,X}(t/b,x)}{\phi_\eta(t/b)}dt \end{aligned}$$

Assumption 2.2 is needed to rule out independent case. Suppose that x and x^* are independent random variables, i.e., $f_{X^*,X}(x^*,x) = f_X(x)f_{X^*}(x^*)$. Eq. (53) becomes

$$\phi_{Y,X}(t,x) = \phi_\eta(t)\phi_{X^*}(bt)f(x) \tag{54}$$

Then

$$f_{X^*,X}(x^*,x;b) = \frac{1}{2\pi}\int e^{-itx^*}\frac{\phi_{Y,X}(t/b,x)}{\phi_\eta(t/b)}dt \tag{55}$$

$$= \frac{1}{2\pi}\int e^{-itx^*}\phi_{X^*}(t)f(x)dt \tag{56}$$

$$= f_{X^*}(x^*)f(x). \tag{57}$$

We lose the parameter b. It shows that although under Assumptions 2.1, $f(x^*,x;b) \equiv L^{-1}_{f_{Y|X^*;b}}(f_{Y,X}(y,x))$ exists it is not longer a function of b.

Integrate out x^* gives

$$\begin{aligned} f_X(x;b) &= \int f_{X^*,X}(x^*,x)dx^* \\ &= \int \left[\frac{1}{2\pi}\int e^{-itx^*}\frac{\phi_{Y,X}(t/b,x)}{\phi_\eta(t/b)}dt\right]dx^* \end{aligned}$$

Set $t/b = s$. Then,

$$f_X(x;b) = \frac{b}{2\pi}\int_{x^*}\left[\int_s e^{-ibsx^*}\frac{\phi_{Y,X}(s,x)}{\phi_\eta(s)}ds\right]dx^*. \tag{58}$$

Suppose that the domain of x^* is $\mathcal{X}^* = (-\infty,\infty)$. Eq. (58) becomes

$$f_X(x;b) = \frac{b}{2\pi}\int_{-\infty}^{\infty}\left[\int_s e^{-ibsx^*}\frac{\phi_{Y,X}(s,x)}{\phi_\eta(s)}ds\right]dx^* \tag{59}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_s e^{-is\tilde{x}}\frac{\phi_{Y,X}(s,x)}{\phi_\eta(s)}ds\right]d\tilde{x}, \tag{60}$$

where $bx^* = \tilde{x}$. In this case, $f_X(x;b)$ is independent of the parameter b after an integration over x^*. On the other hand, without loss of generality, assume $\mathcal{X}^* = [0,1]$. In the first case, Eq. (58) becomes

$$f_X(x;b) = \frac{b}{2\pi}\int_0^1\left[\int_s e^{-ibsx^*}\frac{\phi_{Y,X}(s,x)}{\phi_\eta(s)}ds\right]dx^*. \tag{61}$$

Using $bx^* = \tilde{x}$ to transform the above equation leads to

$$f_X(x;b) = \frac{1}{2\pi}\int_0^b\left[\int_s e^{-is\tilde{x}}\frac{\phi_{Y,X}(s,x)}{\phi_\eta(s)}ds\right]d\tilde{x}. \tag{62}$$

Denote $h(s,x) \equiv \frac{\phi_{Y,X}(s,x)}{\phi_\eta(s)}$. It follows that

$$f_X(x;b) = \frac{1}{2\pi}\int_0^b \phi_h(\tilde{x},x)d\tilde{x}. \tag{63}$$

Suppose that there exists (b_0, x_0) such that (1) $\phi_h(b_0,x_0) \neq 0$, and (2) $\phi_h(\tilde{x},x)$ is continuous at (b_0,x_0). The continuity of $\phi_h(\tilde{x},x)$ at (b_0,x_0) implies that there is some $\delta > 0$ and $\phi_h(\tilde{x},x_0) \neq 0$ for $\tilde{x} \in (b_0-\delta, b_0+\delta)$. Assume that $\phi_h(\tilde{x},x_0) > 0$ for $\tilde{x} \in (b_0-\delta, b_0+\delta)$. It follows that

$$\begin{aligned} f_X(x_0;b_0+\frac{1}{2}\delta) &= \frac{1}{2\pi}\int_0^{b_0+\frac{1}{2}\delta} \phi_h(\tilde{x},x_0)d\tilde{x} && (64)\\ &= \frac{1}{2\pi}\left(\int_0^{b_0} \phi_h(\tilde{x},x_0)d\tilde{x} + \int_{b_0}^{b_0+\frac{1}{2}\delta} \phi_h(\tilde{x},x_0)d\tilde{x}\right) && (65)\\ &> f_X(x_0;b_0). && (66) \end{aligned}$$

Thus, $f_X(x;b)$ depends on the parameter b. This explains the weighted space $\mathcal{L}^2(\mathcal{X}^*,\omega)$ is essential to the identification.

Assume that the true parameter $b_0 > 1$ and the η has truncated standard normal on $[-l_\eta, l_\eta]$ with $l_\eta > 0$ in the model (51). In addition, the domain of the true regressor $\mathcal{X}^* \equiv [l_{x^*}, u_{x^*}]$ and $l_{x^*} > l_\eta$. Hence, the true density is $f_{Y|X^*}(y|x^*) = c_\eta\phi_\eta(y - b_0x^*) = \frac{c_\eta}{\sqrt{2\pi}}e^{\frac{-(y-b_0x^*)^2}{2}}$ with $\mathcal{Y} = (b_0x^* - l_\eta, b_0x^* + l_\eta)$ where c_η is used to normalize the density. The model can be parameterized as $f_{Y|X^*}(y|x^*;b) = c_\eta\phi_\eta(y - bx^*) = \frac{c_\eta}{\sqrt{2\pi}}e^{\frac{-(y-bx^*)^2}{2}}$ with $\mathcal{Y}_b = (bx^* - l_\eta, bx^* + l_\eta)$. Appealing to Theorem 2.3 leads to that $\{f_{Y|X^*}(y|x^*;b) : x^* \in \mathcal{X}^*\}$ is complete in $\mathcal{L}^2(\mathcal{Y})$ and $\{f_{Y|X^*}(y|x^*;b) : y \in \mathcal{Y}\}$ is also complete in $\mathcal{L}^2(\mathcal{X}^*)$ if $b \neq 0$ and the support of x^* is an open set.[5] Thus, it's necessary to assume that $\mathcal{X}^*$ contains an open set in this parametric family.

The variation in this parametric family is included in $\{\frac{\partial}{\partial b}f_{Y|X^*}(y|x^*;b_0) : x^* \in \mathcal{X}^*\}$. Suppose that $\int \frac{\partial}{\partial b}f_{Y|X^*}(y|x^*;b_0)h(y)dy = 0$ for all $x^* \in \mathcal{X}^*$. Rewrite it as $\int \frac{\partial}{\partial y}f_{Y|X^*}(y|x^*;b_0)h(y)dy = 0\ \forall x^* \in \mathcal{X}^*$. Applying integration by part, $0 = \int \frac{\partial}{\partial y}f_{Y|X^*}(y|x^*;b_0)h(y)dy = (f_{Y|X^*}(y|x^*;b_0) - \frac{c_\eta}{\sqrt{2\pi}}e^{\frac{-l_\eta^2}{2}})h(y)\big|_{bx^*-l_\eta}^{bx^*+l_\eta} - \int(f_{Y|X^*}(y|x^*;b_0) - \frac{c_\eta}{\sqrt{2\pi}}e^{\frac{-l_\eta^2}{2}})\frac{\partial}{\partial y}h(y)dy \equiv \int(f_{Y|X^*}(y|x^*;b_0) - f(l_\eta))\frac{\partial}{\partial y}h(y)dy$

[5] When $\mathcal{X}^*$ is a finite domain, a choice of the weighted function ω can be a constant ratio, $\omega = \frac{1}{\int_{\mathcal{X}^*} 1dx^*}$. It is easy to extend a function in $\mathcal{L}^2(\mathcal{X}^*)$ to a function in $\mathcal{L}^2(\mathcal{R})$. Theorem 2.3 proves that $\{f_{Y|X^*}(y|x^*;\theta) : y \in \mathcal{Y}\}$ is complete in $\mathcal{L}^2(\mathcal{R})$ and it also implies that the family is complete in $\mathcal{L}^2(\mathcal{X}^*)$. Similarly, the argument works for the family $\{f_{Y|X^*}(y|x^*;\theta) : x^* \in \mathcal{X}^*\}$.

$\forall x^* \in \mathcal{X}^*$, where $f(l_\eta) = \frac{c_\eta}{\sqrt{2\pi}} e^{\frac{-l_\eta^2}{2}}$. It follows that

$$\int f_{Y|X^*}(y|x^*; b_0) \frac{\partial}{\partial y} h(y) dy = \int f(l_\eta) \frac{\partial}{\partial y} h(y) dy \equiv c_{f_\eta} = \int c_{f_\eta} f_{Y|X^*}(y|x^*; b_0) dy. \tag{67}$$

This implies that $\int f_{Y|X^*}(y|x^*; b_0)[\frac{\partial}{\partial y} h(y) - c_{f_\eta}]dy = 0$ for all $x^* \in \mathcal{X}^*$. By the completeness of $\{f_{Y|X^*}(y|x^*; b) : x^* \in \mathcal{X}^*\}$, $\frac{\partial}{\partial y} h(y) = c_{f_\eta}$ and then $h(y) = c_{f_\eta} y + c$. Plugging this form into the definition of c_{f_η}, $c_{f_\eta} = \int f(l_\eta) \frac{\partial}{\partial y} h(y) dy = 2c_{f_\eta} f(l_\eta) l_\eta$. If $c_{f_\eta} \neq 0$ then $2f(l_\eta) l_\eta = 1$ which contravenes $2f(l_\eta) l_\eta < 1$. Hence, $c_{f_\eta} = 0$ and $h(y) = c$. Since $\int \frac{\partial}{\partial y} f_{Y|X^*}(y|x^*; b_0) h(y) dy = 0$, $h(y) = 0$ and this shows the family $\{\frac{\partial}{\partial b} f_{Y|X^*}(y|x^*; b_0) : x^* \in \mathcal{X}^*\}$ is complete and Assumption 2.1 is fulfilled in this continuous case.

2.2.2. Discrete Case

The discrete case refers to that the variables y and x^* is discrete:

$$y \in \mathcal{Y} \equiv \{1, 2, ..., J_1\} \text{ and } x^* \in \mathcal{X}^* \equiv \{1, 2, ..., J_2\}.$$

The main difference between this discrete case and the previous continuous case is that the linear integral operators are replaced by matrices, which may be more straightforward. There are two missions for this subsection, the identification technique in Section 1 to a discrete case and more discussions of completeness in a discrete case.

First, we show how the identification works. For simplicity, assume $J_1 = J_2 = J$. The matrix expression of $f_{Y,X}(y, x) = \int f_{Y|X^*}(y|x^*; \theta) f(x^*, x; \theta) dx^*$ is

$$L_{f_{Y,X}(y,x)} = [f_{Y,X}(y, x)]_{l \in \mathcal{Y}, m \in \mathcal{X}} \tag{68}$$

$$= \begin{bmatrix} f_{Y|X^*}(1|1; \theta) & \dots & f_{Y|X^*}(1|J; \theta) \\ \vdots & \dots & \vdots \\ f_{Y|X^*}(J|1; \theta) & \dots & f_{Y|X^*}(J|J; \theta) \end{bmatrix} \begin{bmatrix} f(1, 1; \theta) & \dots & f(1, J; \theta) \\ \vdots & \dots & \vdots \\ f(J, 1; \theta) & \dots & f(J, J; \theta) \end{bmatrix} \tag{69}$$

$$\equiv L_{f_{Y|X^*}(y|x^*; \theta)} L_{f(x^*, x; \theta)}. \tag{70}$$

Assumption 2.1 ensures the invertibility of the square matrix $L_{f_{Y|X^*}(y|x^*;\theta)}$. Hence

$$L^{-1}_{f_{Y|X^*}(y|x^*;\theta)} L_{f_{Y,X}(y,x)} = L_{f(x^*,x;\theta)}, \tag{71}$$

and since LHS is observable, $L_{f(x^*,x;\theta)}$ is well defined and also observable.[6] Consider

$$\begin{aligned} L_{f_{Y,X}(y,x)} &= L_{f_{Y|X^*}(y|x^*;\theta_1)} L_{f(x^*,x;\theta_1)}, \\ L_{f_{Y,X}(y,x)} &= L_{f_{Y|X^*}(y|x^*;\theta_{\text{true}})} L_{f(x^*,x;\theta_{\text{true}})}. \end{aligned}$$

It follows that

$$\begin{aligned} 0 = {} & L_{f_{Y|X^*}(y|x^*;\theta_1)} \left(L_{f(x^*,x;\theta_1)} - L_{f(x^*,x;\theta_{\text{true}})}\right) \\ & + \left(L_{f_{Y|X^*}(y|x^*;\theta_1)} - L_{f_{Y|X^*}(y|x^*;\theta_{\text{true}})}\right) L_{f(x^*,x;\theta_{\text{true}})}. \end{aligned}$$

If $f(x^*,x;\theta)$ lost θ, then $0 = \left(L_{f_{Y|X^*}(y|x^*;\theta_1)} - L_{f_{Y|X^*}(y|x^*;\theta_{\text{true}})}\right) L_{f(x^*,x;\theta_{\text{true}})}$. Assumption 2.2 implies that $L_{f_{Y|X^*}(y|x^*;\theta_1)} = L_{f_{Y|X^*}(y|x^*;\theta_{\text{true}})}$, a contradiction. Thus, Assumption 2.2 makes sure that $L_{f(x^*,x;\theta)}$ still depends on the parameter θ. Let $I = (1, ..., 1)'$ be $J \times 1$ vector. Integrating out x^* on $f(x^*,x;\theta)$ is equal to summing over x^* on RHS. Then,

$$I' L^{-1}_{f_{Y|X^*}(y|x^*;\theta)} L_{f_{Y,X}(y,x)} = I' L_{f(x^*,x;\theta)} = (\bar{f}(1;\theta), ..., \bar{f}(J;\theta)). \tag{72}$$

If $\bar{f}(x;\theta)$ does not depend on θ, then $I'\left(L_{f(x^*,x;\theta)} - L_{f(x^*,x;\theta_{\text{true}})}\right) = 0$. This amounts to for $\theta \neq \theta_{\text{true}}$, $I'\left(\frac{L_{f(x^*,x;\theta)} - L_{f(x^*,x;\theta_{\text{true}})}}{\theta - \theta_{\text{true}}}\right) = 0$. Using the relationship

$$0 = \underbrace{L_{f_{Y|X^*}(y|x^*;\theta_1)}}_{\text{Assumption 2.1}} \frac{L_{f(x^*,x;\theta_1)} - L_{f(x^*,x;\theta_{\text{true}})}}{\theta - \theta_{\text{true}}} + \frac{L_{f_{Y|X^*}(y|x^*;\theta_1)} - L_{f_{Y|X^*}(y|x^*;\theta_{\text{true}})}}{\theta - \theta_{\text{true}}} \underbrace{L_{f(x^*,x;\theta_{\text{true}})}}_{\text{Assumption 2.2}},$$

the invertibility of the matrix $\frac{L_{f(x^*,x;\theta_1)} - L_{f(x^*,x;\theta_{\text{true}})}}{\theta - \theta_{\text{true}}}$ is equivalent to the invertibility of the matrix $\frac{L_{f_{Y|X^*}(y|x^*;\theta_1)} - L_{f_{Y|X^*}(y|x^*;\theta_{\text{true}})}}{\theta - \theta_{\text{true}}}$ under Assumption 2.1 & 2.2. Assumption 2.3 guarantees the invertibility of $\frac{L_{f_{Y|X^*}(y|x^*;\theta_1)} - L_{f_{Y|X^*}(y|x^*;\theta_{\text{true}})}}{\theta - \theta_{\text{true}}}$ and rules out $I'\left(\frac{L_{f(x^*,x;\theta)} - L_{f(x^*,x;\theta_{\text{true}})}}{\theta - \theta_{\text{true}}}\right) = 0$. The function $\{\bar{f}(l;\theta) : l \in \mathcal{X}\}$ still has a variation in θ. Finally, we normalize $\bar{f}(l;\theta)$ by

[6] As for the case, $J_1 \neq J_2$, the completeness of the family $\{f_{Y|X^*}(y|x^*;\theta) : y \in \mathcal{Y}_\theta\}$ is full rank condition of the matrix $L_{f_{Y|X^*}(y|x^*;\theta)}$ and the generalized inverse $(L^T_{f_{Y|X^*}(y|x^*;\theta)} L_{f_{Y|X^*}(y|x^*;\theta)})^{-1} L^T_{f_{Y|X^*}(y|x^*;\theta)}$ is used to recovery $L_{f(x^*,x;\theta)}$.

$f(l;\theta) \equiv \frac{\bar{f}(l;\theta)}{\sum_{l=1}^{J} \bar{f}(l;\theta)}$ and Assumption 2.4-2.5 guarantees that the family $\{f(l;\theta) : l \in \mathcal{X}\}$ is a legit parametric family of p.d.f's for $f(x)$. Therefore, we can apply standard MLE method to the parametric family $\{f(l;\theta) : l \in \mathcal{X}\}$ to search for the true value θ_{true}.

As for completeness, we conduct the discussion by two different categories. One is that the parametric family and the original family share the same domain, i.e., $\mathcal{Y}_\theta = \mathcal{Y}$ and the other one is $\mathcal{Y}_\theta \neq \mathcal{Y}$.

Category 1: $\mathcal{Y}_\theta = \mathcal{Y}$

Given y, and x^*, define J_1-by-J_2 matrices

$$L_{f_{Y|X^*}(y|x^*;\theta)} \tag{73}$$

$$= \left[f_{Y|X^*}(l|m;\theta)\right]_{1\le l\le J_1, 1\le m\le J_2} \tag{74}$$

$$= \begin{bmatrix} f_{Y|X^*}(1|1;\theta) & \ldots & f_{Y|X^*}(1|J_2;\theta) \\ f_{Y|X^*}(2|1;\theta) & \ldots & f_{Y|X^*}(2|J_2;\theta) \\ \vdots & \ldots & \vdots \\ 1 - \sum_{1=1}^{J_1-1} f_{Y|X^*}(l|1;\theta) & \ldots & 1 - \sum_{1=1}^{J_1-1} f_{Y|X^*}(l|J_2;\theta) \end{bmatrix}_{J_1\times J_2}. \tag{75}$$

In the case, $J_1 = J_2$, the completeness of the family $\{f_{Y|X^*}(y|x^*;\theta) : x^* \in \mathcal{X}^*\}$ is the invertibility of the square matrix $L_{f_{Y|X^*}(y|x^*;\theta)}$. The derivative of the above matrix with respect to θ is

$$L_{\frac{\partial}{\partial\theta} f_{Y|X^*}(y|x^*;\theta)} \tag{76}$$

$$= \left[\frac{\partial}{\partial\theta} f_{Y|X^*}(l|m;\theta)\right]_{1\le l\le J_1, 1\le m\le J_2} \tag{77}$$

$$= \begin{bmatrix} \frac{\partial}{\partial\theta} f_{Y|X^*}(1|1;\theta) & \ldots & \frac{\partial}{\partial\theta} f_{Y|X^*}(1|J_2;\theta) \\ \frac{\partial}{\partial\theta} f_{Y|X^*}(2|1;\theta) & \ldots & \frac{\partial}{\partial\theta} f_{Y|X^*}(2|J_2;\theta) \\ \vdots & \ldots & \vdots \\ -\sum_{1=1}^{J_1-1} \frac{\partial}{\partial\theta} f_{Y|X^*}(l|1;\theta) & \ldots & -\sum_{1=1}^{J_1-1} \frac{\partial}{\partial\theta} f_{Y|X^*}(l|J_2;\theta) \end{bmatrix}_{J_1\times J_2}, \tag{78}$$

which is not invertible and fails Assumption 2.3.

Category 2: $\mathcal{Y}_\theta \neq \mathcal{Y}$

Consider $\mathcal{Y} \equiv \{1,2\}$, $\mathcal{X}^* \equiv \{1,2\}$ and $\mathcal{Y}_\theta \equiv \{1,2,3\}$ with

$$\left[f_{Y|X^*}(l|m)\right]_{l\in\mathcal{Y},m\in\mathcal{X}^*} = \begin{bmatrix} f_{Y|X^*}(1|1) & f_{Y|X^*}(1|2) \\ f_{Y|X^*}(2|1) & f_{Y|X^*}(2|2) \end{bmatrix}_{2\times 2}, \tag{79}$$

and

$$\left[f_{Y|X^*}(l|m;\theta)\right]_{l\in\mathcal{Y}_\theta,m\in\mathcal{X}^*} = \begin{bmatrix} f_{Y|X^*}(1|1)+\theta & f_{Y|X^*}(1|2) \\ f_{Y|X^*}(2|1)-\theta & f_{Y|X^*}(2|2)-\theta \\ 0 & \theta \end{bmatrix}_{3\times 2}. \tag{80}$$

Suppose that $\int f_{Y|X^*}(y|x^*;\theta)h(y)dy = 0$ for all $x^* \in \mathcal{X}^*$. We have

$$\begin{bmatrix} f_{Y|X^*}(1|1)+\theta & f_{Y|X^*}(1|2) \\ f_{Y|X^*}(2|1)-\theta & f_{Y|X^*}(2|2)-\theta \end{bmatrix}' \begin{bmatrix} h(1) \\ h(2) \end{bmatrix} = 0. \tag{81}$$

A choice of $\left[f_{Y|X^*}(l|m)\right] = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$ and $\Theta = (-\delta,\delta)$ with $\delta < 0.2$ leads to $h = 0$. This shows that $\{f_{Y|X^*}(y|x^*;\theta) : x^* \in \mathcal{X}^*\}$ is complete over $\mathcal{L}^2(\mathcal{Y})$. A similar argument applies to the completeness of $\{f_{Y|X^*}(y|x^*;\theta) : y \in \mathcal{Y}\}$. On the other hand, the derivative

$$L_{\frac{\partial}{\partial\theta}f_{Y|X^*}(y|x^*;\theta)} = \begin{bmatrix} 1 & 0 \\ -1 & -1 \end{bmatrix} \tag{82}$$

satisfies Assumption 2.3.

3. Estimation

This section focuses on the estimation of parametric conditional density $f_{Y|X^*}(y|x^*;\theta)$. We start with a discrete case. Given an observed discrete data $\{(y_i,x_i) : i = 1,...,n\}$, we estimate $f_{Y,X}(y,x)$ by calculating the sample size of $\{(y_i,x_i) = (y,x)\}$ relative to the whole sample size, i.e., $\hat{f}_{Y,X}(y,x) \equiv \frac{\sum_{l=1}^{n} 1\{(y_i,x_i)=(y,x)\}}{n}$ where $1(\cdot)$ is an indicator function. Applying this estimated density function to the procedure discussed in subsection 2.2.2, a parametric family of p.d.f's $\hat{f}(x;\theta)$ for $f(x)$ in the discrete case can be easily constructed. A standard MLE method to the parametric family is

$$\hat{\theta}_n = \arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} \ln \hat{f}(x_i; \theta). \tag{83}$$

On the other hand, we propose semiparametric sieve maximum likelihood estimator (sieve MLE) for a continuous case. This method does not require finding a parametric family of p.d.f's for $f(x)$ like the discrete case. In Section 2, we have shown the following equation

$$f_{Y,X}(y,x) = \int f_{Y|X^*}(y|x^*;\theta) f(x^*,x;\theta) dx^*, \qquad \forall y \in \mathcal{Y}_\theta \cap \mathcal{Y}$$

uniquely determines $(\theta_{\text{true}}, f(x^*,x))$. Treating $f(x^*|x;\theta) \equiv \frac{f(x^*,x;\theta)}{f(x)}$ as a nonparametric nuisance functions suggests that

$$(\theta, f(x^*|x))^T = \arg \max_{(\theta,f_1)^T \in \mathcal{A}} E \ln \int f_{Y|X^*}(y|x^*;\theta) f_1(x^*|x;\theta) dx^*,$$

where $\mathcal{A}$ is a collection of functions containing the corresponding true densities. A corresponding semiparametric sieve MLE using an i.i.d. sample $\{(y_i, x_i) : i = 1, ..., n\}$ has the following form,

$$(\hat{\theta}_n, \hat{f}_1(x^*|x;\hat{\theta}_n))^T = \arg \max_{(\theta,f_1)^T \in \mathcal{A}^n} \frac{1}{n} \sum_{i=1}^{n} \ln \int f_{Y|X^*}(y_i|x^*;\theta) f_1(x^*|x_i;\theta) dx^*, \tag{84}$$

where $\mathcal{A}^n$ is a sequence of sieve spaces approximating $\mathcal{A}$.

The asymptotic theory of the proposed sieve MLE and the detailed development of sieve approximations of the nonparametric components can be found in Shen (1997), Chen and Shen (1998), and Ai and Chen (2003). The sieve specification of the nuisance parameter $f_1(x^*, x_i; \theta)$ is provided in Appendix A.

4. Monte Carlo Simulation

We now investigate the finite-sample properties of the proposed estimator via Monte Carlo simulations. First, consider the estimator developed in the discrete case to a probit model with a mismeasured 0-1 dichotomous explanatory variable. The data generating process (DGP) for this probit case:

$$Y_i = 1\left(\beta_0 + \beta_1 X_i^* + \beta_2 W_i + \varepsilon_i \geq 0\right) \quad \forall i = 1, ..., n,$$

with $p(X^* = 1) = p_1, W \sim U(0,1)$ and a measurement error probability

$$\left[f_{X|X^*}(l|m)\right]_{l\in\mathcal{X},m\in\mathcal{X}^*} = \begin{bmatrix} 1-p_2 & 1-p_3 \\ p_2 & p_3 \end{bmatrix}_{2\times 2}.$$

We consider two different values of $(\beta_0, \beta_1, \beta_2, p_1, p_2, p_3)$ with the same parameterizations,

$$\begin{aligned} \text{DGP I:} \quad (\beta_0, \beta_1, \beta_2, p_1, p_2, p_3) &= \quad (0.5, -0.7, 0, 0.5, 0.1, 0.9), \\ \text{DGP II:} \quad (\beta_0, \beta_1, \beta_2, p_1, p_2, p_3) &= \quad (-0.4, 0.6, -0.2, 0.5, 0.1, 0.9), \end{aligned}$$

with for each w,

$$\begin{aligned} &\left[f_{Y|X^*}(l|m, w; \theta)\right]_{l\in\mathcal{Y}_\theta, m\in\mathcal{X}^*} \\ &= \begin{bmatrix} 1-\Phi\left(\beta_0+\beta_2 w\right) & 1-\Phi\left(\beta_0+\beta_1+\beta_2 w\right) - 10(\beta_0+\beta_1)^2 + 0.4 \\ \Phi\left(\beta_0+\beta_2 w\right) - 0.5(\beta_0+\beta_1) + 0.1 & \Phi\left(\beta_0+\beta_1+\beta_2 w\right) \\ 0.5(\beta_0+\beta_1) - 0.1 & 10(\beta_0+\beta_1)^2 - 0.4 \end{bmatrix}_{3\times 2}. \end{aligned}$$

In DGP I & II, there are 10% of misreport rates conditional on $X^* = 0, 1$. In both cases, the measurement error probabilities are all invertible which satisfies Assumption 2.2 and the parametric conditional density also satisfies the requirement in Section 2.1.

The data generating process (DGP) for our continuous case is

$$Y_i = \beta_0 + \beta_1 X_i^* + \beta_2 W_i + \varepsilon_i \quad \forall i = 1, ..., n,$$

with $X^*, W \sim U(0,1)$ and the distribution ε is $N(0, \sigma^2)$ truncated on $(-1, 1)$.[7] In addition, assume $X = X^* + h(X^*)e$, where $h(x^*) = 0.2\exp(-x^*)$ and $e \sim N(0, \sigma^2)$ truncated on $(-1, 1)$.

[7]The truncation is necessary since it makes the domain of the dependent variable, $\mathcal{Y}_\theta$, varies with θ.

Two different values of $(\beta_0, \beta_1, \beta_2, \sigma^2)$ are considered,

$$\begin{aligned} \text{DGP III:} \quad (\beta_0, \beta_1, \beta_2, \sigma^2) &= \quad (0, 1, -1, 0.5), \\ \text{DGP IV:} \quad (\beta_0, \beta_1, \beta_2, \sigma^2) &= \quad (0, -1, 1, 0.5). \end{aligned}$$

The parametric family for this DGP is $f_{Y|X^*,W}(y|x^*, w; \theta) = \frac{c_\varepsilon}{\sqrt{2\pi}} e^{\frac{-(y-\beta_0-\beta_1 x^*-\beta_2 w)^2}{2}}$. Since the distributions of ε and e are chosen to be a normal, the density functions $f_{Y|X^*,W}(y|x^*, w; \theta)$ and $f_{X^*,X,W}$ in this Monte Carlo experiment fulfills the completeness assumptions of Theorem 2.1.

A sample size $N = 500$ are considered and 200 simulation replications are conducted at each estimation. Table 1 presents simulation results under the probit model. The simulation results of DGP I&II show different directions of bias in the model coefficients $(\beta_0, \beta_1, \beta_2)$. The coefficients exhibit downward bias in DGP I but there does not exists clear trend about bias in DGP II. In this sample size, the means and medians of the coefficients are not close in some estimated coefficients, reflecting some skewness in their respective distributions. In addition, the medians of the coefficients is more precisely estimated than the means in DGP II. The finite-sample properties of the proposed sieve MLE in the continuous case are reported in Table 2. The estimation results show the sieve MLE performs well with $N = 1000$ since means and medians of estimation values are close to the true values with appropriate standard errors. The approximation of the nonparametric nuisance function $f(x^*|x, w; \theta)$ is constructed by Fourier series. The numbers of term, $i_n = 4$, $j_n = 2$, and $k_n = 2$ are used as the length of three univariate Fourier series. See Appendix A for details.

5. Empirical Application

In this section, we illustrate the use of the sieve MLE method by estimating an empirical model of the demand for food. This empirical illustration is adopted from the specifications estimated in Blundell, Pistaferri, and Preston (2008). The paper examines the link between income and consumption inequality and find some partial insurance of permanent shocks, especially for the college educated and those near retirement. In this empirical application, we only focus on applying the proposed sieve MLE to an imputation procedure based on food demand estimates from the Consumer Expenditure Survey (CEX) in Blundell, Pistaferri, and

Preston (2008).[8]

To relate the level of food consumption to the level of nondurable consumption in the CEX data from 1980 to 1992, we also use the following demand equation for food:

$$o_{i,t} = \mathbf{W}'_{i,t}\mu + \mathbf{p}'_t\alpha + \beta(D_{i,t})c^*_{i,t} + \varepsilon_{i,t}, \tag{85}$$

where $o_{i,t}$ is the log of real food expenditure, $\mathbf{W}_{i,t}$ and $\mathbf{p}_t$ contain a set of demographic variables and relative prices respectively, $c^*_{i,t}$ is the log of the true unobserved nondurable expenditure, and $\varepsilon_{i,t}$ represents unobserved disturbance in food expenditure. Equation (85) is parameterized by assuming $\varepsilon_{i,t}|_{\mathbf{W}_{i,t},\mathbf{p}_t,c^*_{i,t}} \sim N(0,\sigma^2)$ truncated on (l,u) and the conditional density becomes $f(o_{i,t}|\mathbf{W}_{i,t},\mathbf{p}_t,c^*_{i,t};\theta) = \frac{c_\varepsilon}{\sqrt{2\pi}}e^{\frac{(o_{i,t}-\mathbf{W}'_{i,t}\mu-\mathbf{p}'_t\alpha-\beta(D_{i,t})c^*_{i,t})^2}{2}}$. The normality assumption together with the discussion in the end of subsection 2.2.1 ensure that this family of the conditional density functions $\{f(o_{i,t}|\mathbf{W}_{i,t},\mathbf{p}_t,c^*_{i,t};\theta) : c^*_{i,t} \in \mathcal{C}^*\}$ satisfies the identification assumptions in Section 2. Assume that the nondurable expenditure $c^*_{i,t}$ is measured with additive error such that $c_{i,t} = c^*_{i,t} + \nu_{i,t}$ and $c_{i,t}$ is the log of nondurable expenditure available in the CEX. Plugging the measurement error equation into Eq. (85) yields

$$\begin{aligned} o_{i,t} &= \mathbf{W}'_{i,t}\mu + \mathbf{p}'_t\alpha + \beta(D_{i,t})c_{i,t} + (\varepsilon_{i,t} - \beta(D_{i,t})\nu_{i,t}), \\ &\equiv \mathbf{W}'_{i,t}\mu + \mathbf{p}'_t\alpha + \beta(D_{i,t})c_{i,t} + e_{i,t}, \end{aligned}$$

which is the demand equation for food used in Blundell, Pistaferri, and Preston (2008). Note that the model reflects that the elasticity $\beta(D_{i,t})$ varies with time and with observable household characteristics (D).

The proposed sieve MLE method in Section 3 allows us to handle the measurement error in nondurable expenditure without using any instrument variable.[9] Table 3 presents the estimation results for our specification of Eq. (85). The point estimates of coefficients of

[8] Although the Panel Study of Income Dynamics (PSID) contains longitudinal income data, it has limited consumption data (limited to food expenditure and a few more items related to household consumption). Blundell, Pistaferri, and Preston (2008) use food expenditures to explore consumer behavior by combining existing PSID data with data from the repeated cross sections of the CEX. A demand function for food is estimated using CEX data and then the estimated coefficients are applied to the relevant PSID variables to create a measure of nondurable consumption in the PSID.

[9] Blundell, Pistaferri, and Preston (2008) tackle the endogeneity with two types of instruments including the average of the hourly wage of the husband and the average of the hourly wage of the wife both based on cohort, year, and education.

variables, $\ln c$ and $\ln p_{food}$, represent the budget elasticity and the price elasticity respectively. The estimated values of them are 0.868 and -0.966, with standard errors of 0.254 and 0.428 respectively. Thus, there is evidence of the positive relationship between food consumption and consumption of nondurables. The coefficient on family size is positive, as expected, but insignificant. The positive coefficient on age and the negative coefficient on age squared are consistent with age-consumption profile albeit insignificant. The estimate coefficient on race suggests that white households have significantly higher rates of food consumption. Compared with the estimates of Blundell, Pistaferri, and Preston (2008), the parameters have changed little and the standard errors have fallen for most coefficients, reflecting efficiency gain due to the normality assumption. The estimated coefficients can be used to invert the demand function and derive a new panel nondurable consumption series in the PSID. We refer to Blundell, Pistaferri, and Preston (2008) for more complete discussion.

6. Conclusion

The presence of measurement error can bias estimates of parameters of interest and it is sometimes questionable to assume the classical errors in variables model. We consider the identification problem of a parametric family in models where the data are measured with error and there exists arbitrary correlation between the true variable and the measurement error. The construction of consistent estimators of the parameters is particularly challenging for nonlinear model under these situations. Under the completeness assumptions of families of observable density functions, the parameter of the original parametric family can survive in several operations, applying an inverse operator, integrating out the unobserved true variable, and normalization. When the mismeasured variable and the parameter both continue to exist after these operations, the standard MLE argument can apply to the observed parametric family to reach the identification.

As shown in the literature of measurement error models, without additional information or functional form restrictions, a general nonlinear errors-in-variables model cannot be identified. Viewed from this perspective, there may exists some trade-off between functional form assumptions and additional data information to reach identification. This study assumes no any additional information of measurement error but requires the parametric conditional density is correctly specified. Therefore, one of the main advantages of this approach over other methods is that it is not restricted to extra data requirement if the parametric assumption of the distribution is not a major concern. For example, good valid instruments are uncorrelated with the error and strongly correlated with endogenous regressors. Thus, obtaining instruments can require considerable ingenuity or access to unusually rich data. This study provide a solution for this case, using a complete family of parametric density functions and applying our proposed econometric method can produce consistent estimates of the parameter of interest.

Appendix

A. The Sieve Specification

This appendix describes the sieve MLE method based on the likelihood function in Eq. (84). There are two parts $f_{Y|X^*,W}(y|x^*,w;\theta)$ and $f_1(x^*|x,w;\theta)$. Since $f_{Y|X^*,W}(y|x^*,w;\theta)$ is already parameterized, we only need to show show sieve approximation and the constraint of nonparametric component $f_1(x^*|x,w;\theta)$. Assume $f_{Y|X^*,W}(y|x^*,w;\theta) = \frac{c_\varepsilon}{\sqrt{2\pi}} e^{\frac{-(y-\beta_0-\beta_1 x^*-\beta_2 w)^2}{2}}$, where $\theta = (\beta_0, \beta_1, \beta_2)^T$.

A sieve specification of $f_1(x^*|x,w;\theta)$ is given by the following:

$$f_1(x^*|x,w;\theta) = \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\sum_{k=0}^{k_n} \hat{a}_{ijk} q_i(\beta_1 x^* - \beta_1 x) q_j(\beta_1 x - \beta_0 - \beta_2 w) q_k(\beta_1 x) \tag{86}$$

Our choice of $q_i's$ and $q_j's$ are the Fourier series:

$$q_{j_0}(\beta_1 x - \beta_0 - \beta_2 w) = 1, q_j(\beta_1 x - \beta_0 - \beta_2 w) = \cos(\frac{j\pi}{l_2}(\beta_1 x - \beta_0 - \beta_2 w)),$$
$$q_{k_0}(\beta_1 x) = 1 \text{ and } q_k(\beta_1 x) = \cos(\frac{k\pi}{l_3}\beta_1 x)$$
$$q_{i_0}(\beta_1 x^* - \beta_1 x) = 1 \text{ and } q_i(\beta_1 x^* - \beta_1 x) = \sin(\frac{i\pi}{l_1}(\beta_1 x^* - \beta_1 x)) \text{ or}$$
$$q_i(\beta_1 x^* - \beta_1 x) = \cos(\frac{i\pi}{l_1}(\beta_1 x^* - \beta_1 x)).$$

Set $\tilde{x} = \beta_1 x - \beta_0 - \beta_2 w$. We consider the case where $i_n = 4$, $j_n = 2$, and $k_n = 2$:

$$
\begin{aligned}
& f_1(x^*|x, w; \theta) \\
&= \left(\hat{c}_{00} + \hat{c}_{01} \cos \frac{\pi}{l_1}\tilde{x} + \hat{c}_{02} \cos \frac{2\pi}{l_1}\tilde{x}\right)\left(\hat{a}_{00} + \hat{a}_{01} \cos \frac{\pi}{l_1}\beta_1 x + \hat{a}_{02} \cos \frac{2\pi}{l_1}\beta_1 x\right) \\
&\quad + \sum_{i=1}^{4}\left(\hat{c}_{i0} + \hat{c}_{i1} \cos \frac{\pi}{l_1}\tilde{x} + \hat{c}_{i2} \cos \frac{2\pi}{l_1}\tilde{x}\right)\left(\hat{a}_{10} + \hat{a}_{11} \cos \frac{\pi}{l_1}\beta_1 x + \hat{a}_{12} \cos \frac{2\pi}{l_1}\beta_1 x\right) \\
&\qquad\qquad \times \cos \frac{i\pi}{l_2}(\beta_1 x^* - \beta_1 x) \\
&\quad + \sum_{i=1}^{4}\left(\hat{c}_{i0} + \hat{c}_{i1} \cos \frac{\pi}{l_1}\tilde{x} + \hat{c}_{i2} \cos \frac{2\pi}{l_1}\tilde{x}\right)\left(\hat{b}_{10} + \hat{b}_{11} \cos \frac{\pi}{l_1}\beta_1 x + \hat{b}_{12} \cos \frac{2\pi}{l_1}\beta_1 x\right) \\
&\qquad\qquad \times \sin \frac{i\pi}{l_2}(\beta_1 x^* - \beta_1 x) \\
&\quad + \left(\hat{c}_{i0} + \hat{c}_{i1} \cos \frac{\pi}{l_1}\tilde{x} + \hat{c}_{i2} \cos \frac{2\pi}{l_1}\tilde{x}\right)\left(\hat{d}_{00} + \hat{d}_{01} \cos \frac{\pi}{l_1}\beta_1 x + \hat{d}_{02} \cos \frac{2\pi}{l_1}\beta_1 x\right) \\
&\quad + \sum_{i=1}^{4}\left(\hat{c}_{i0} + \hat{c}_{i1} \cos \frac{\pi}{l_1}\tilde{x} + \hat{c}_{i2} \cos \frac{2\pi}{l_1}\tilde{x}\right)\left(\hat{d}_{20} + \hat{d}_{21} \cos \frac{\pi}{l_1}\beta_1 x + \hat{d}_{22} \cos \frac{2\pi}{l_1}\beta_1 x\right) \\
&\qquad\qquad \times \cos \frac{i\pi}{l_2}(\beta_1 x^* - \beta_1 x) \\
&\quad + \sum_{i=1}^{4}\left(\hat{c}_{00} + \hat{c}_{01} \cos \frac{\pi}{l_1}\tilde{x} + \hat{c}_{02} \cos \frac{2\pi}{l_1}\tilde{x}\right)\left(\hat{e}_{20} + \hat{e}_{21} \cos \frac{\pi}{l_1}\beta_1 x + \hat{e}_{22} \cos \frac{2\pi}{l_1}\beta_1 x\right) \\
&\qquad\qquad \times \sin \frac{i\pi}{l_2}(\beta_1 x^* - \beta_1 x).
\end{aligned}
$$

The density restriction $\int f_1(x^*|x, w; \theta)dx^*dx = 1$ for all θ and x, w amounts to

$$\hat{c}_{00}\hat{a}_{00} + \hat{c}_{00}\hat{d}_{00} = 1 \text{ and } \hat{c}_{01} = 0, \hat{c}_{02} = 0, \hat{a}_{0q} = 0, \hat{a}_{02} = 0, \hat{d}_{01} = 0, \hat{d}_{02} = 0$$

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

Andrews, D. (2011): "Examples of L^2-Complete and Boundedly-Complete Distributions," *Cowles Foundation for Research in Economics.*

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75(6), 1613.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5), 1887–1921.

Bollinger, C. (1998): "Measurement Error in the Current Population Survey: A Nonparametric Look," *Journal of Labor Economics*, 16(3), 576–594.

Bound, J., C. Brown, G. Duncan, and W. Rodgers (1994): "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12(3), 345–368.

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. Amsterdam: North-Holland.*, pp. 3705–3843.

Carrasco, M., J. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," *in Handbook of Econometrics, Vol. 6B, ed. by J.J. Heckman and E. Leamer. Amsterdam: North-Holland.*, pp. 5633–5751.

Chen, X., H. Hong, and E. Tamer (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72(2), 343–366.

Chen, X., H. Hong, and A. Tarozzi (2008): "Semiparametric Efficiency in GMM Models with Auxiliary Data," *The Annals of Statistics*, 36(2), 808–843.

Chen, X., Y. Hu, and A. Lewbel (2009): "Nonparametric Identification and Estimation of Nonclassical Errors-In-Variables Models without Additional Information," *Statistica Sinica*, 19, 949–968.

Chen, X., and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66(2), 289–314.

D'Haultfoeuille, X. (2011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, 1, 1–12.

Hu, Y. (2008): "Identification and Estimation of Nonlinear Models with Misclassification Error using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144(1), 27–61.

Hu, Y., and S. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

Hu, Y., and J. Shiu (2011): "Nonparametric Identification Using Instrumental Variables: Sufficient Conditions for Completeness," *Jonhs Hopkins University, Department of Economics Working Paper*, 581.

Hu, Y., and M. Shum (2008): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," *Jonhs Hopkins University, Dept. of Economics Working Paper*, 543.

Lewbel, A. (2007): "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75(2), 537–551.

Mahajan, A. (2006): "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74(3), 631–665.

Newey, W., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

Shen, X. (1997): "On Methods of Sieves and Penalization," *The Annals of Statistics*, 25, 2555–2591.

Shiu, J., and Y. Hu (2010): "Identification and Estimation of Nonlinear Dynamic Panel Data Models with Unobserved Covariates," *Economics Working Paper Archive.*

Table 1: Monte Carlo Simulation of Probit model (N=500)

DGP		Parameters β_0	β_1	β_2
DGP I:	true value	0.5	-0.7	0
	mean estimate	0.379	-0.777	-0.042
	median estimate	0.378	-0.815	0.015
	standard error	0.421	0.463	0.515
DGP II:	true value	-0.4	0.6	-0.2
	mean estimate	-0.477	0.686	-0.126
	median estimate	-0.385	0.595	-0.166
	standard error	0.684	0.669	0.370

Note: Standard errors of the parameters are computed by the standard deviation of the estimates across 200 simulations.

Table 2: Monte Carlo Simulation of Linear model (N=500)

DGP		Parameters β_0	β_1	β_1	σ^2
DGP III:	true value	0	1	-1	0.5
	mean estimate	0.004	0.809	-0.956	0.261
	median estimate	-0.076	0.849	-1.006	0.151
	standard error	0.287	0.237	0.202	0.362
DGP IV:	true value	0	-1	1	0.5
	mean estimate	-0.041	-1.189	0.985	0.470
	median estimate	-0.043	-1.225	1.004	0.509
	standard error	0.076	0.287	0.097	0.156

Note: Standard errors of the parameters are computed by the standard deviation of the estimates across 100 simulations.

Table 3: The Demand for Food

Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
$\ln c$	0.869 (0.254)	$\ln c \times 1992$	0.004 (0.011)	Family size	0.027 (0.034)
$\ln c \times$H.s. dropout	0.074 (0.101)	$\ln c \times$one child	0.021 (0.023)	$\ln p_{food}$	-0.966 (0.428)
$\ln c \times$H.s. graduate	0.081 (0.089)	$\ln c \times$two children	-0.025 (0.011)	$\ln p_{transports}$	5.900 (7.031)
$\ln c \times 1981$	0.116 (0.019)	$\ln c \times$three children	0.009 (0.033)	$\ln p_{fuel+utils}$	-0.663 (0.156)
$\ln c \times 1982$	0.063 (0.027)	One child	-0.156 (0.028)	$\ln p_{alcohol+tobacco}$	-1.834 (0.582)
$\ln c \times 1983$	0.051 (0.007)	Two children	0.325 (0.120)	Born 1955–59	-0.038 (0.010)
$\ln c \times 1984$	0.048 (0.016)	Three children+	0.013 (0.058)	Born 1950–54	-0.009 (0.011)
$\ln c \times 1985$	0.031 (0.003)	H.s. dropout	-0.699 (0.163)	Born 1945–49	-0.006 (0.004)
$\ln c \times 1986$	0.022 (0.021)	H.s. graduate	-0.819 (0.683)	Born 1940–44	-0.005 (0.016)
$\ln c \times 1987$	0.053 (0.033)	Age	0.013 (0.021)	Born 1935–39	-0.004 (0.003)
$\ln c \times 1988$	0.042 (0.015)	Age^2	-0.001 (0.007)	Born 1930–34	0.003 (0.009)
$\ln c \times 1989$	0.037 (0.018)	Northeast	0.009 (0.014)	Born 1925–29	-0.005 (0.018)
$\ln c \times 1990$	0.019 (0.008)	Midwest	-0.021 (0.010)	White	0.077 (0.023)
$\ln c \times 1991$	-0.001 (0.005)	South	-0.027 (0.006)	Constant	-0.619 (0.789)
σ^2	0.258 (0.189)				

Note: Standard errors of the parameters are computed by the standard deviation of the estimates across 100 simulations.