The Econometrics of Unobservables

– Latent Variable and Measurement Error Models and Their Applications in Empirical Industrial Organization and Labor Economics

by

YINGYAO HU

The Johns Hopkins University
Department of Economics
yhu@jhu.edu

Incomplete. Comments welcome. [1]
The latest version is available at
`http://www.econ.jhu.edu/people/hu/`

May 10, 2018

Contents

Preface

This manuscript is designed for an advanced micro-econometrics course for graduate students. For empirical researchers, it provides a tool kit to tackle latent variables, such as unobserved heterogeneity, belief, effort, ability, and misreporting errors, in applied microeconomics, especially empirical industrial organization and labor economics. It focuses on nonparametric identification and parametric or semiparametric estimation, and presents specific empirical applications.

I plan to keep updating this manuscript, not necessarily for publication, but for the enjoyment of research. Any comments are highly appreciated. Especially, one should feel free to contact me if she or he wants me to cite or discuss her or his work in this manuscript.

This manuscript is also written for my three kids. From here, they will find out what daddy was doing while they were skating, playing soccer, taking piano, violin, and karate lessons, studying at AOPS, Kumon, Spidersmart, ...

Structure of book

This first half of the manuscript presents flexible nonparametric identification and parametric or semiparametric estimation methods for nonlinear models with latent variables. The key methods are extended from the nonclassical measurement error literature.

The second half provides applications of these methods in structural and reduced-form econometrics and in empirical industrial organization and labor economics. These applications involve errors-in-variables, latent variable, unobserved heterogeneity, unobserved state variable, mixture model, hidden Markov model, dynamic discrete choice, unemployment rates, IPV auction, multiple equilibria in incomplete information games, belief, learning model, fixed effects, panel data model, cognitive and non-cognitive skills, matching, income dynamics.

About the companion website

The website[1] for this file contains:

- A link to (freely downlodable) latest version of this document.
- Miscellaneous materials.

[1] `https://http://www.econ.jhu.edu/people/hu/`

Acknowledgements

I am grateful to Yonghong An, Yajing Jiang, Zhongjian Lin, Jian Ni, Katheryn Russ, Yuya Sasaki, Tom Wansbeek, Ruli Xiao, Yi Xin for suggestions and comments. All errors are mine.

Yingyao Hu

1

Introduction

1.1 Observables and unobservables in applied microeconomics

Researchers in applies microeconomics study behavior of economic agents, such as consumers and firms, from observed information in the data. Researchers in this area usually start with an existing microeconomic theoretical models or an intuitive microeconomic relationship, which contains a set of variables to describe an economic agent' behavior. This set of variables is composed of two subsets: what the agent knows and what is unknown to the agent. In the meanwhile, an empirical researcher also observes part of variables in this set. In most applications, researchers estimate a model based on what they observe in the data, which may contain dependent or endogenous variables and exogenous variables, and treat those unobserved in the data as shocks or error terms.

In this manuscript, we are interested in those variables which are in the agents' information set but unobserved to researchers, in particular, those that can't be left in the error terms. If a complete model can fully explain an agent's behavior, the main reason for endogeneity in empirical research is due to these unobservables which we focus on. Therefore, the methods we provide here for unobservables also provide a solution to the endogeneity problem, which is arguably the most important problem in econometrics.

variables in a complete model		to economic agent:	
		known	unknown
researcher:	observes	endog. & exog. variables	shocks
	unobserves	*unobservables of interest*	error terms

Our methods rely on additional data information, i.e., those variables which are not in the complete model of interest, but are associated with those variables in the model. These variables may be a measurement of a latent true variable in the model or an instrumental variable correlated with an endogenous variable. Obviously, how these observables are associated with those unobservables is the key issue in this framework.

1.2 Why identification is important and challenging

Under the ideal condition, what researchers observe in the data coincides with the information set of economic agents. One may directly estimate the model, structural or reduced-form, using a random sample of the variables in the complete model. In many empirical applications, however, there are important variables in the agents' information set but unobserved by the researchers, such as belief, ability, mood, and effort.

A simple approach to deal with such lack of data information is to assume that these unobservables are independent of the observables and have a known or partially known distribution. The model of interest may then be specified in a likelihood or moment function, in which the unobservables are integrated out.

A reasonable approach is to use additional data information or additional assumptions to identify and estimate the complete model using observed variables by researchers, which may be a subset of the information set of economic agents. Ideally, these additional assumptions, e.g., conditional independence, can be motivated by the economic model. This task is quite challenging due to the existence of unobservables. The identification of complete models with incomplete data information is interesting and important because it lies at the intersection of economic theory and econometric methodology.

Furthermore, we prefer to establish identification before the model of interest is parameterized, which usually leads to local identification and is inherently subject to misspecification. Nonparametric identification allows researchers to answer the following question: Can the economic relationship be revealed by incomplete data information? In the meanwhile, identification of a nonparametric model becomes much more challenging than parametric identification.

1.3 Latent variable and measurement error models

Latent variable and measurement error models describe the relationship between unobservables and observables. The goal is to identify the distribution of unobservables and also the distribution of observables conditional on unobservables, which corresponds to the distribution of measurement errors. In general, the parameter of interest is the joint distribution, which can be used to describe the relationship between observables and unobservables in economic models.

Early studies on measurement errors in the econometric literature started with the so-called classical measurement error, where the errors are usually assumed to be independent of the true values, arguably because the measurement error models were borrowed from the relevant statistical literature, where the independence assumption is quite reasonable when the measurement error is caused by using an instrument to measure a certain property of an object. The additivity and independence in the classical measurement error models lead to important and fruitful results. In the econometric literature, the classical measurement error framework is adopted mainly for the parsimony of the measurement error part of the model and for the convenience of using existing results. In empirical macroeconomics and some applied microeconomic research, the classical measurement error framework is usually embedded into linear models, such as factor models, linear dynamic models, and linear panel data models. In microeconometrics, identification and estimation of nonlinear models, such as nonlinear regressions and limited dependent models, with classical measurement errors, had been a difficult problem for many years.

In recent years, econometricians have been leading the studies on the nonclassical measurement error model because of the need of handling measurement errors in economic survey data, where the measurement errors are usually caused by self-reporting behaviors. Such a need exists in most disciplines in social sciences. Instead of measuring certain properties of an object, many economic data are from surveys, where interviewees self-report their information. The classical measurement error assumption is unlikely to hold in these scenarios. Econometricians are, therefore, on the frontier of identification and estimation of the so-called nonclassical measurement errors models, where the errors may be correlated with the latent true values. In particular, the presence of nonclassical measurement errors makes the identification of nonlinear models containing the latent true values extremely difficult, that is, whether the models can be uniquely determined from the joint distribution of observed variables.

Based on conditional independence assumptions, which widely exist in economic theories, a breakthrough in the measurement error models literature has been the realization that the joint distribution of three observables may uniquely determine the joint distribution of four variables including the three observables and the latent variable. Hu (2008) uses a matrix eigenvalue-eigenvector decomposition to show this pathbreaking result for the case where the latent variable is a general discrete variable. The Hu-Schennach Theorem in Hu and Schennach (2008) nontrivially extends this result to the general continuous case using a unique representation of bounded linear operators. In addition, one of the three observables may contain as few information as a binary indicator. Such an identification result is nonparametric and global and leads to a closed-form estimation procedure in the discrete case. The flexibility of these results greatly extend applications of measurement error models to various areas in empirical economic research. This manuscript follows Hu (2017), organizes the existing technical results in terms of the number of measurements, and shows that these technical results may not only apply to measurement error models, but also many economic models with latent variables. For more reviews of this extensive literature, we refer to Wansbeek and Meijer (2000), Bound et al. (2001), Fuller (2009), Chen et al. (2011), Carroll et al. (2012), and Schennach (2016).

2

Nonparametric identification with unobservables

This chapter starts with a general framework, where “a measurement" can be simply an observed variable with an informative support. The measurement error distribution contains the information about a mapping from the distribution of the latent variables to the observed measurements. We organize the technical results by the number of measurements needed for identification. In the first example, there are two measurements, which are mutually independent conditioning on the latent variable. With such limited information, strong restrictions on measurement errors are needed to achieve identification in this 2-measurement model. Nevertheless, there are still well known useful results in this framework, such as Kotlarski's identity.

However, when a 0-1 dichotomous indicator of the latent variable is available together with two measurements, nonparametric identification is feasible under a very flexible specification of the model. Hu (2017) names this a 2.1-measurement model, where he uses 0.1 measurement to refer to a 0-1 binary variable. A major breakthrough in the measurement error literature is that the 2.1-measurement model can be non-parametrically identified under mild restrictions (see Hu (2008) and Hu and Schennach (2008)). Since it allows very flexible specifications, the 2.1-measurement model is widely applicable to microeconomic models with latent variables even beyond many existing applications.

Given that any observed random variable can be manually transformed to a 0-1 binary variable, the results for a 2.1-measurement model can be easily extended to a 3-measurement model. A 3-measurement model is useful because many dynamic models involve multiple measurements of a latent variable. A typical example is the hidden Markov model. Results for the 3-measurement model show the exchangeable roles which each measurement may play. In particular, in many cases, it does not matter which one of the three measurements is called a dependent variable, a proxy, or an instrument.

One may also interpret the identification strategy of the 2.1-measurement model as a nonparametric instrumental approach. In that sense, a nonparametric difference-in-differences version of this strategy may help identify more general dynamic processes with more measurements. As shown in Hu and Shum (2012) , four measurements or four periods of data are enough to identify a rather general partially observed first-order Markov process. Such an identification result is directly applicable to the nonparametric identification of dynamic models with unobserved state variables.

2.1 Definition of a measurement

In the measurement error literature, researchers usually use the term "measurement" without a formal definition. Here, we adopt the general definition of measurement in Hu (2017). Such a definition is a helpful concept to organize the literature.

Let X denote an observed random variable and X^* be a latent random variable of interest. We define a measurement of X^* as follows:

Definition 1 *A random variable X with support $\mathcal{X}$ is called* ***a measurement*** *of a latent random variable X^* with support $\mathcal{X}^*$ if*

$$card\,(\mathcal{X}) \geq card\,(\mathcal{X}^*)\,,$$

where $card\,(\mathcal{X})$ stands for the cardinality of set $\mathcal{X}$.

The support condition in Definition 1 implies that there exists an injective function from $\mathcal{X}^*$ into $\mathcal{X}$. When X is continuous, the support condition is not restrictive whether X^* is discrete or continuous. When X is discrete, the support condition implies that the number of possible values of one measurement is larger than or equal to that of the latent variable. In addition, the possible values in $\mathcal{X}^*$ are unknown and usually normalized to be the same as those of one measurement with an equal cardinality of the support.

2.2 A general framework

In a random sample, we observe measurement X, while the variable of interest X^* is unobserved. The measurement error is defined as the difference $X - X^*$. We can identify the distribution function f_X of measurement X directly from the sample, but our main interest is to identify the distribution of the latent variable f_{X^*}, together with the measurement error distribution described by $f_{X|X^*}$. The observed measurement and the latent variable are associated as follows: for all $x \in \mathcal{X}$

$$f_X(x) = \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*, \tag{2.1}$$

when X^* is continuous and f_{X^*} is the probability density function of X^*, and for all $x \in \mathcal{X} = \{x_1, x_2, \ldots, x_L\}$

$$f_X(x) = \sum_{x^* \in \mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*), \tag{2.2}$$

when X^* is discrete with support $\mathcal{X}^* = \{x_1^*, x_2^*, \ldots, x_K^*\}$ and $f_{X^*}(x^*) = Pr(X^* = x^*)$ is the probability mass function of X^* and $f_{X|X^*}(x|x^*) = Pr(X = x|X^* = x^*)$. Definition 1 of measurement requires $L \geq K$. We omit arguments of the functions when it does not cause any confusion. This general framework can be used to describe a wide range of economic relationships between observables and unobservables in the sense that the latent variable X^* can be interpreted as unobserved heterogeneity, fixed effects, random coefficients, or latent types in mixture models, etc.

For simplicity, we start with the discrete case and define

$$\begin{aligned} \overrightarrow{p}_X &= [f_X(x_1), f_X(x_2), \ldots, f_X(x_L)]^T \\ \overrightarrow{p}_{X^*} &= [f_{X^*}(x_1^*), f_{X^*}(x_2^*), \ldots, f_{X^*}(x_K^*)]^T \\ M_{X|X^*} &= \left[f_{X|X^*}(x_l|x_k^*) \right]_{l=1,2,\ldots,L;k=1,2,\ldots,K}. \end{aligned} \tag{2.3}$$

The notation M^T stands for the transpose of M. Note that $\overrightarrow{p}_X$, $\overrightarrow{p}_{X^*}$, and $M_{X|X^*}$ contain the same information as distributions f_X, f_{X^*}, and $f_{X|X^*}$, respectively. Equation (2.2) is then equivalent to

$$\overrightarrow{p}_X = M_{X|X^*}\overrightarrow{p}_{X^*}. \tag{2.4}$$

The matrix $M_{X|X^*}$ describes the linear transformation from $\mathbb{R}^K$, a vector space containing $\overrightarrow{p}_{X^*}$, to $\mathbb{R}^L$, a vector space containing $\overrightarrow{p}_X$. Suppose that the measurement error distribution, i.e., $M_{X|X^*}$, is known. The identification of the latent distribution f_{X^*} means that if two possible marginal distributions $\overrightarrow{p}^a_{X^*}$ and $\overrightarrow{p}^b_{X^*}$ are observationally equivalent, i.e.,

$$\overrightarrow{p}_X = M_{X|X^*}\overrightarrow{p}^a_{X^*} = M_{X|X^*}\overrightarrow{p}^b_{X^*}, \tag{2.5}$$

then the two distributions are the same, i.e., $\overrightarrow{p}^a_{X^*} = \overrightarrow{p}^b_{X^*}$. Let $h = \overrightarrow{p}^a_{X^*} - \overrightarrow{p}^b_{X^*}$. Equation (2.5) implies that $M_{X|X^*}h = 0$. The identification of f_{X^*} then requires that $M_{X|X^*}h = 0$ implies $h = 0$ for any $h \in \mathbb{R}^K$, or that matrix $M_{X|X^*}$ has rank K, i.e., $Rank\left(M_{X|X^*}\right) = K$. This is a necessary rank condition for the nonparametric identification of the latent distribution f_{X^*}.

In the continuous case, we need to define the linear operator corresponding to $f_{X|X^*}$, which maps f_{X^*} to f_X. Suppose that we know both f_{X^*} and f_X are bounded and integrable. We define $\mathcal{L}^1_{bnd}(\mathcal{X}^*)$ as the set of bounded and integrable functions defined on $\mathcal{X}^*$, i.e.,[1]

$$\mathcal{L}^1_{bnd}(\mathcal{X}^*) = \left\{h : \int_{\mathcal{X}^*} |h(x^*)|\, dx^* < \infty \text{ and } \sup_{x^*\in\mathcal{X}^*} |h(x^*)| < \infty\right\}. \tag{2.6}$$

The linear operator can be defined as

$$\begin{aligned} L_{X|X^*} &: \mathcal{L}^1_{bnd}(\mathcal{X}^*) \to \mathcal{L}^1_{bnd}(\mathcal{X}) \\ \left(L_{X|X^*}h\right)(x) &= \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*)h(x^*)dx^*. \end{aligned} \tag{2.7}$$

Equation (2.1) is then equivalent to

$$f_X = L_{X|X^*}f_{X^*}. \tag{2.8}$$

Following a similar argument, we can show that a necessary condition for the identification of f_{X^*} in the functional space $\mathcal{L}^1_{bnd}(\mathcal{X}^*)$ is that the linear operator $L_{X|X^*}$ is injective, i.e., $L_{X|X^*}h = 0$ implies $h = 0$ for any $h \in \mathcal{L}^1_{bnd}(\mathcal{X}^*)$. This condition can also be interpreted as completeness of conditional density $f_{X|X^*}$ in $\mathcal{L}^1_{bnd}(\mathcal{X}^*)$. We refer to Hu and Schennach (2008) for detailed discussion on this injectivity condition.

Since both the measurement error distribution $f_{X|X^*}$ and the marginal distribution f_{X^*} are unknown, we have to rely on additional restrictions or additional data information to achieve identification. On the one hand, parametric identification may be feasible if $f_{X|X^*}$ and f_{X^*} belong to parametric families (see Fuller (2009)). On the other hand, we can use additional data information to achieve nonparametric identification. For example, if we observe the joint distribution of X and X^* in a validation sample, we can identify $f_{X|X^*}$ from the validation sample and then identify f_{X^*} in the primary sample (see Chen et al. (2005)). In this paper, we focus on methodologies using additional measurements in a single sample.

[1]We may also define the operator on other functional spaces containing f_{X^*}.

2.3 A 2-measurement model

Given very limited identification results which one may obtain from equations (2.1)-(2.2), a direct extension is to use more data information, i.e., an additional measurement. Define a 2-measurement model as follows:

Definition 2 *A **2-measurement model** contains two measurements, as in Definition 1, $X \in \mathcal{X}$ and $Z \in \mathcal{Z}$ of the latent variable $X^* \in \mathcal{X}^*$ satisfying*

$$X \perp Z \mid X^*, \tag{2.9}$$

i.e., X and Z are independent conditional on X^*.

The 2-measurement model implies that two measurements X and Z not only have distinctive information on the latent variable X^*, but also are mutually independent conditional on the latent variable.

In the case where all the variables X, Z, and X^* are discrete with $\mathcal{Z} = \{z_1, z_2, \ldots, z_J\}$, we define

$$\begin{aligned} M_{X,Z} &= [f_{X,Z}(x_l, z_j)]_{l=1,2,...,L;j=1,2,...,J} \\ M_{Z|X^*} &= \left[f_{Z|X^*}(z_j|x_k^*)\right]_{j=1,2,...,J;k=1,2,...,K} \end{aligned} \tag{2.10}$$

and a diagonal matrix

$$D_{X^*} = diag\left\{f_{X^*}(x_1^*), f_{X^*}(x_2^*), \ldots, f_{X^*}(x_K^*)\right\}, \tag{2.11}$$

where $f_{X^*}(x_i^*) > 0$ for $i = 1, 2, ..., K$ by the definition of the discrete support $\mathcal{X}^*$. Definition 1 implies that $K \leq L$ and $K \leq J$. Equation (2.9) means

$$f_{X,Z}(x, z) = \sum_{x^* \in \mathcal{X}^*} f_{X|X^*}(x|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*), \tag{2.12}$$

which is equivalent to

$$M_{X,Z} = M_{X|X^*} D_{X^*} M_{Z|X^*}^T . \tag{2.13}$$

Without further restrictions to reduce the number of unknowns on the right hand side, point identification of $f_{X|X^*}$, $f_{Z|X^*}$, and f_{X^*} may not be feasible. But one element that can be identified from observed $M_{X,Z}$ is the dimension K of the latent variable X^*, as elucidated in the following Lemma:

Lemma 1 *In the 2-measurement model in Definition 2 with support $\mathcal{X}^* = \{x_1^*, x_2^*, \ldots, x_K^*\}$, suppose that matrices $M_{X|X^*}$ and $M_{Z|X^*}$ both have rank K. Then $K = \text{rank}(M_{X,Z})$.*

Proof. In the 2-measurement model, Definition 1 requires that $K \leq L$ and $K \leq J$. The definition of the discrete support $\mathcal{X}^*$ implies that $f_{X^*}(x_i^*) > 0$ for $i = 1, 2, ..., K$ and D_{X^*} has rank K. Using the rank inequality: for any p-by-m matrix A and m-by-q matrix B, $\text{rank}(A) + \text{rank}(B) - m \leq \text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$, we may first show $M_{X|X^*} D_{X^*}$ has rank K, then use the inequality again to show the right hand side of Equation (2.13) has rank K. Thus, we have $\text{rank}(M_{X,Z}) = K$. ■

Point identification of this model requires further restrictions. For example, if $M_{X|X^*}$ and $M_{Z|X^*}^T$ are lower and upper triangular matrices, respectively, point identification is feasible through the so-called LU decomposition (See Hu and Sasaki (forthcoming, 2017b) for a generalization of such a result). In general, this is also related to the literature on non-negative matrix factorization, which focuses more on existence and approximation, instead of uniqueness.

Although point identification may not be feasible without further assumptions, we can still have some partial identification results. Consider a linear regression model with a discrete regressor X^* as follows:

$$\begin{aligned} Y &= X^*\beta + \eta \\ Y &\perp X \mid X^* \end{aligned} \tag{2.14}$$

where $X^* \in \{0,1\}$ and $E[\eta|X^*] = 0$. Here the dependent variable Y takes the place of Z as a measurement of X^*.[2] We observe (Y,X) with $X \in \{0,1\}$ in the data as two measurements of the latent X^*. Since Y and X are independent conditional on X^*, the two observed distributions with $x = 0,1$ are different weighted averages of the same two latent distributions, i.e.,

$$f_{Y|X}(y|x) = f_{Y|X^*}(y|0)f_{X^*|X}(0|x) + f_{Y|X^*}(y|1)f_{X^*|X}(1|x). \tag{2.15}$$

Taking the difference with respect to $x = 0,1$ leads to

$$\begin{aligned} & |E[Y|X^*=1] - E[Y|X^*=0]| \\ \geq\ & |E[Y|X=1] - E[Y|X=0]|\,. \end{aligned} \tag{2.16}$$

That means the observed difference provides a lower bound on the parameter of interest $|\beta|$. This is the so-called attenuation phenomenon. Such a lower bound is useful for testing the hypothesis $\beta = 0$ without further restrictions on the misclassification probability. More partial identification results can be found in Bollinger (1996) and Molinari (2008).

Furthermore, the model can be point identified under the assumption that the regression error η is independent of the regressor X^*. Chen et al. (2009) consider a nonlinear regression model with a general discrete X^* as follows:

$$Y = m(X^*) + \eta \tag{2.17}$$

where the regression function m is the unknown of interest. They provide sufficient conditions for identification of m from the joint distribution $f_{Y,X}$ when X^* is independent of η, i.e., $X^* \perp \eta$. In particular, when X^* is 0-1 dichotomous, we have

$$Y = m(0) + [m(1) - m(0)]X^* + \eta. \tag{2.18}$$

Chen et al. (2008) show that the model can be identified with closed-form expressions. Define

$$\begin{aligned} \mu_j &= E[Y|X=j] \\ v_j &= E[(Y-\mu_j)^2|X=j] \\ s_j &= E[(Y-\mu_j)^3|X=j] \\ C_1 &= \frac{(v_1+\mu_1^2)-(v_0+\mu_0^2)}{\mu_1-\mu_0} \end{aligned}$$

[2]We follow the routine to use Y to denote a dependent variable instead of Z.

$$C_2 = \frac{1}{2}(\mu_1 - \mu_0)^2 + \frac{3}{2}\left(\frac{v_1 - v_0}{\mu_1 - \mu_0}\right)^2 - \frac{s_1 - s_0}{\mu_1 - \mu_0}.$$

Under assumptions that $\mu_1 > \mu_0$ and $f_{X^*|X}(1|0) + f_{X^*|X}(0|1) < 1$, they show the unknown elements of the model can be expressed as closed-form functions of observables as follows:

$$m(0) = \frac{1}{2}C_1 - \sqrt{\frac{1}{2}C_2}$$

$$m(1) = \frac{1}{2}C_1 + \sqrt{\frac{1}{2}C_2}$$

$$f_{X^*|X}(1|0) = \frac{\mu_0 - \frac{1}{2}C_1}{\sqrt{2C_2}} - \frac{1}{2}$$

$$f_{X^*|X}(0|1) = \frac{\frac{1}{2}C_1 - \mu_1}{\sqrt{2C_2}} - \frac{1}{2}$$

$$f_{Y|X^*}(y|j) = \frac{\mu_1 - m(j)}{\mu_1 - \mu_0} f_{Y|X}(y|0) + \frac{m(j) - \mu_0}{\mu_1 - \mu_0} f_{Y|X}(y|1).$$

Such closed-form identification results may be very convenient for empirical researchers.

In the case where all the variables X, Z, and X^* are continuous, a widely-used setup is

$$\begin{aligned} X &= X^* + \epsilon \\ Z &= X^* + \epsilon' \end{aligned} \tag{2.19}$$

where X^*, ϵ, and ϵ' are mutually independent with $E[\epsilon] = 0$. When the error $\epsilon := X - X^*$ is independent of the latent variable X^*, it is called a classical measurement error. This setup is well known because the density of the latent variable X^* can be written as a closed-form function of the observed distribution $f_{X,Z}$. Define $\phi_{X^*}(t) = E\left[e^{itX^*}\right]$ with $i = \sqrt{-1}$ as the characteristic function of X^*. Under the assumption that $\phi_Z(t)$ is absolutely integrable and does not vanish on the real line, we have

$$\begin{aligned} f_{X^*}(x^*) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ix^*t}\phi_{X^*}(t)\,dt \\ \phi_{X^*}(t) &= \exp\left[\int_0^t \frac{iE\left[Xe^{isZ}\right]}{E\left[e^{isZ}\right]}ds\right]. \end{aligned} \tag{2.20}$$

This is the so-called Kotlarski's identity (See Kotlarski (1965) and Rao (1992)). Note that the independence between ϵ and (X^*, ϵ') can be relaxed to a mean independence condition $E[\epsilon|X^*, \epsilon'] = E[\epsilon]$. This identity was first introduced to econometric research by Li and Vuong (1998). Li (2002) first used this result to consistently estimate nonlinear regression models with classical measurement errors. The Kotlarski's identity has been used in many empirical and theoretical studies, including Li et al. (2000) , Krasnokutskaya (2011) , Schennach (2004a), and Evdokimov (2010) .

The intuition of Kotlarski's identity is that the variance of X^* is revealed by the covariance of X and Z, i.e., $var(X^*) = cov(X, Z)$. Therefore, the higher order moments between X and Z can reveal more moments of X^*. If one can pin down all the moments of X^*

from the observed moments, the distribution of X^* is then identified under some regularity assumptions. A similar argument also applies to an extended model as follows:

$$\begin{aligned} X &= X^*\beta + \epsilon \\ Z &= X^* + \epsilon'. \end{aligned} \tag{2.21}$$

Suppose $\beta > 0$. A naive OLS estimator obtained by regressing X on Z converges in probability to $\frac{cov(X,Z)}{var(Z)}$, which provides a lower bound on the regression coefficient β. In fact, we have explicit bounds as follows:

$$\frac{cov(X,Z)}{var(Z)} \leq \beta \leq \frac{var(X)}{cov(X,Z)}. \tag{2.22}$$

Furthermore, additional assumptions, such as the joint independence of X^*, ϵ, and ϵ', can lead to point identification of β. Reiersøl (1950) shows that such point identification is feasible when X^* is not normally distributed. A more general extension is to consider

$$\begin{aligned} X &= g(X^*) + \epsilon \\ Z &= X^* + \epsilon', \end{aligned} \tag{2.23}$$

where function g is nonparametric and unknown. Schennach and Hu (2013) generalize Reiersøl's result and show that function g and distribution of X^* are nonparametrically identified except for a particular functional form of g or f_{X^*}. The only difference between the model in equation (2.23) and a nonparametric regression model with a classical measurement error is that the regression error ϵ needs to be independent of the regressor X^*.

Schennach and Hu (2013) assume that

Assumption 1 *The variables X^*, ϵ', ϵ, are mutually independent, $E[\epsilon'] = 0$ and $E[\epsilon] = 0$.*

Assumption 2 $E\left[e^{i\xi\epsilon'}\right]$ *and* $E\left[e^{i\gamma\epsilon}\right]$ *do not vanish for any $\xi, \gamma \in \mathbb{R}$, where $i = \sqrt{-1}$.*

Assumption 3 $E\left[e^{i\xi X^*}\right]$ *and* $E\left[e^{i\xi g(X^*)}\right]$ *do not vanish for all ξ in a dense subset of $\mathbb{R}$.*

Assumption 4 *The distribution of X^* admits a uniformly bounded density $f_{X^*}(x^*)$ with respect to the Lebesgue measure.*

Assumption 5 *The regression function $g(x^*)$ is continuously differentiable over the support of X^*.*

Assumption 6 *The set $\mathcal{Z} = \{x^* : g'(x^*) = 0\}$ has at most a finite number of elements $x_1^*, \ldots, x_m^*$. If $\mathcal{Z}$ is nonempty, $f_{X^*}(x^*)$ is continuous and nonvanishing in a neighborhood of each x_k^*, $k = 1, \ldots, m$.*

Their main result can then be stated as follows:

Theorem 1 *(Schennach and Hu (2013)) Let Assumptions 1-6 hold.*

1. *If $g(x^*)$ is not of the form*

$$g(x^*) = a + b\ln\left(e^{cx^*} + d\right) \tag{2.24}$$

for some constants $a, b, c, d \in \mathbb{R}$ then $f_{X^*}(x^*)$ and $g(x^*)$ (over the support of $f_{X^*}(x^*)$) in equation 2.23 are identified.

2. *If $g(x^*)$ is of the form (2.24) with*[3] $d > 0$, *then neither $f_{X^*}(x^*)$ nor $g(x^*)$ in equation 2.23 are identified iff X^* has a density of the form*

$$f_{X^*}(x^*) = A \exp\left(-Be^{Cx^*} + CDx^*\right)\left(e^{Cx^*} + E\right)^{-F} \tag{2.25}$$

with[4] $C \in \mathbb{R}$, $A, B, D, E, F \in [0, \infty[$ *and ϵ is decomposable with a type I extreme value factor.*[5]

3. *If $g(x^*)$ is linear (i.e. of the form (2.24) with $d = 0$), then neither $f_{X^*}(x^*)$ nor $g(x^*)$ in equation 2.23 are identified iff X^* is normally distributed and either ϵ' or ϵ is decomposable with a normal factor.* [6]

This is the most general identification result for a 2-measurement model in the continuous case, which has been published so far.

2.4 A 2.1-measurement model

An arguably surprising result is that we can achieve quite general nonparametric identification of a measurement error model if we observe a little more data information, i.e., an extra binary indicator, than in the 2-measurement model. Define a 2.1-measurement model as follows:[7]

Definition 3 *A **2.1-measurement model** contains two measurements, as in Definition 1, $X \in \mathcal{X}$ and $Z \in \mathcal{Z}$ and a 0-1 dichotomous indicator $Y \in \mathcal{Y} = \{0, 1\}$ of the latent variable $X^* \in \mathcal{X}^*$ satisfying*

$$X \perp Y \perp Z \mid X^*, \tag{2.26}$$

i.e., (X, Y, Z) are jointly independent conditional on X^*.

2.4.1 The discrete case

In the case where X, Z, and X^* are discrete, Definition 1 implies that the supports of observed X and Z are larger than or equal to that of the latent X^*. We start our discussion with the case where the three variables share the same support. We assume

Assumption 7 *The two measurements X and Z and the latent variable X^* share the same support $\mathcal{X}^* = \{x_1^*, x_2^*, \ldots, x_K^*\}$.*

[3] A case where $d < 0$ can be converted into a case with $d > 0$ by permuting the roles of Z and X.

[4] The constants A, B, C, D, E, F depend on a, b, c, d, although this dependence is omitted here for simplicity. Constants yielding a valid density can be found for any a, b, c, d (with $d > 0$).

[5] A type I extreme value distribution has a density of the general form $f(u) = K_1 \exp(K_2 \exp(K_3 u) + K_4 u)$. Here, the constant K_1, K_2, K_3, K_4 are such that $f(u)$ integrates to 1 and has zero mean and may depend on a, b, c, d, although this dependence is omitted here for simplicity.

[6] We say that a random variable r is *decomposable with F factor* if r can be written as the sum of two independent random variables (which may be degenerate), one of which has the distribution F.

[7] I use “0.1 measurement" to refer to a 0-1 dichotomous indicator of the latent variable. I name it the 2.1-measurement model instead of 3-measurement one in order to emphasize the fact that we only need slightly more data information than the 2-measurement model, given that a binary variable is arguably the least informative measurement, except a constant measurement, of a latent random variable.

This condition is not restrictive because the number of possible values in $\mathcal{X}^*$ can be identified, as shown in Lemma 1, and one can always transform a discrete variable into one with less possible values. We will later discuss that case where supports of measurements X and Z are larger than that of X^*.

The conditional independence in equation (2.26) implies[8]

$$f_{X,Y,Z}(x,y,z) = \sum_{x^* \in \mathcal{X}^*} f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*). \tag{2.27}$$

For each value of $Y = y$, we define

$$\begin{aligned} M_{X,y,Z} &= [f_{X,Y,Z}(x_i, y, z_j)]_{i=1,2,\ldots,K;j=1,2,\ldots,K} \\ D_{y|X^*} &= diag\left\{ f_{Y|X^*}(y|x_1^*), f_{Y|X^*}(y|x_2^*), \ldots, f_{Y|X^*}(y|x_K^*) \right\}. \end{aligned} \tag{2.28}$$

Equation (2.27) is then equivalent to

$$M_{X,y,Z} = M_{X|X^*} D_{y|X^*} D_{X^*} M_{Z|X^*}^T . \tag{2.29}$$

Next, we assume

Assumption 8 *Matrix $M_{X,Z}$ has rank K.*

This assumption is imposed on observed probabilities, and therefore, is directly testable. Equation (2.13) then implies $M_{X|X^*}$ and $M_{Z|X^*}$ both have rank K. We then eliminate $D_{X^*} M_{Z|X^*}^T$ to obtain

$$M_{X,y,Z} M_{X,Z}^{-1} = M_{X|X^*} D_{y|X^*} M_{X|X^*}^{-1}. \tag{2.30}$$

This equation implies that the observed matrix on the left hand side has an inherent eigenvalue-eigenvector decomposition, where each column in $M_{X|X^*}$ corresponding to $f_{X|X^*}(\cdot|x_k^*)$ is an eigenvector and the corresponding eigenvalue is $f_{Y|X^*}(y|x_k^*)$. In order to achieve a unique decomposition, we require that the eigenvalues are distinctive, and that certain location of distribution $f_{X|X^*}(\cdot|x_k^*)$ reveals the value of x_k^*. We assume

Assumption 9 *There exists a function $\omega(\cdot)$ such that $E[\omega(Y)|X^* = \overline{x}^*] \neq E[\omega(Y)|X^* = \widetilde{x}^*]$ for any $\overline{x}^* \neq \widetilde{x}^*$ in $\mathcal{X}^*$.*

Assumption 10 *One of the following conditions holds:*
1) $f_{X|X^*}\left(x_1|x_j^*\right) > f_{X|X^*}\left(x_1|x_{j+1}^*\right)$ for $j = 1, 2, \ldots, K-1$;
2) $f_{X|X^*}(x^*|x^*) > f_{X|X^*}(\widetilde{x}^*|x^*)$ for any $\widetilde{x}^* \neq x^* \in \mathcal{X}^*$;
3) There exists a function $\omega(\cdot)$ such that $E\left[\omega(Y)|X^* = x_j^*\right] > E\left[\omega(Y)|X^* = x_{j+1}^*\right]$.

The function $\omega(\cdot)$ may be user-specified, such as $\omega(y) = y$, $\omega(y) = 1(y > y_0)$, or $\omega(y) = \delta(y - y_0)$ for some given y_0.[9] When estimating the model using the eigenvalue-eigenvector decomposition, especially with a continuous Y as later in the paper, it is more convenient

[8] Hui and Walter (1980) first consider the case where the latent variable X^* is binary and show that this identification problem can be reduced to solving a quadratic equation. Mahajan (2006) and Lewbel (2007) also consider this binary case in regression models and treatment effect models.

[9] When Y is binary, the choice of function $\omega(\cdot)$ does not matter. I state the assumptions in this way so that there is no need to rephrase them later for a general Y.

to average over Y and use the equation below than directly using Equation (2.27) with a fixed y

$$E\left[\omega\left(Y\right)|X=x,Z=z\right]f_{X,Z}\left(x,z\right)=\sum_{x^{*}\in\mathcal{X}^{*}}f_{X|X^{*}}(x|x^{*})E\left[\omega\left(Y\right)|x^{*}\right]f_{Z|X^{*}}(z|x^{*})f_{X^{*}}(x^{*}). \tag{2.31}$$

If the conditional mean $E\left[Y|X^{*}\right]$ is an object of interest instead of $f_{Y|X^{*}}$ as in a regression model, we can consider the equation above with $\omega\left(y\right)=y$ and relax the conditional independence assumption $f_{Y|X^{*},X,Z}=f_{Y|X^{*}}$ implied in the 2.1-measurement model to a conditional mean independence assumption $E\left[Y|X^{*},X,Z\right]=E\left[Y|X^{*}\right]$.

We summarize the identification result as follows:

Theorem 2 *(Hu (2008)) Under assumptions 7, 8, 9, and 10, the 2.1-measurement model in Definition 3 is non-parametrically identified in the sense that the joint distribution of the three variables (X,Y,Z), i.e., $f_{X,Y,Z}$, uniquely determines the joint distribution of the four variables (X,Y,Z,X^{*}), i.e., $f_{X,Y,Z,X^{*}}$, which satisfies*

$$f_{X,Y,Z,X^{*}}=f_{X|X^{*}}f_{Y|X^{*}}f_{Z|X^{*}}f_{X^{*}}. \tag{2.32}$$

A brief proof: The conditional independence in Definition 3 of the 2.1-measurement model implies that Equation (2.29) holds. Assumption 8 leads to an inherent eigenvalue-eigenvector decomposition in Equation (2.30). Assumption 9 guarantees that there are K linearly independent eigenvectors. These eigenvectors are conditional distributions, and therefore, are normalized automatically because the column sum of each eigenvector is equal to one. Assumption 10 pins down the ordering of the eigenvectors or the eigenvalues, i.e., the value of the latent variable corresponding to each eigenvector. Assumption 10(i) implies that the first row of matrix $M_{X|X^{*}}$ is decreasing in x_{j}^{*} and Assumption 10(ii) implies that x^{*} is the mode of distribution $f_{X|X^{*}}\left(\cdot|x^{*}\right)$. Assumption 10(i) directly implies an ordering of the eigenvalues. Therefore, each element on the right hand side of Equation (2.30) is uniquely determined by the observed matrix on the left hand side. The eigenvectors reveal the conditional distribution $f_{X|X^{*}}$ and the identification of other distributions then follows. ■

Theorem 2, particularly under Assumption 7, provides an exact identification result in the sense that the number of unknown probabilities is equal to the number of observed probabilities in equation (2.27). Assumption 7 implies that there are $2K^{2}-1$ observed probabilities in $f_{X,Y,Z}\left(x,y,z\right)$ on the left hand side of equation (2.27). On the right hand side, there are $K^{2}-K$ unknown probabilities in each of $f_{X|X^{*}}(x|x^{*})$ and $f_{Z|X^{*}}(z|x^{*})$, $K-1$ in $f_{X^{*}}(x^{*})$, and K in $f_{Y|X^{*}}(y|x^{*})$ when Y is binary, which sum up to $2K^{2}-1$. More importantly, this point identification result is nonparametric, global, and constructive. It is constructive in the sense that an estimator can directly mimic the identification procedure.

When supports of measurements X and Z are larger than that of X^{*}, we can still achieve the identification with minor modification of the conditions. Suppose supports $\mathcal{X}$ and $\mathcal{Z}$ are larger than $\mathcal{X}^{*}$, i.e., $\mathcal{X}=\{x_{1},x_{2},\ldots,x_{L}\}$, $\mathcal{Z}=\{z_{1},z_{2},\ldots,z_{J}\}$, and $\mathcal{X}^{*}=\{x_{1}^{*},x_{2}^{*},\ldots,x_{K}^{*}\}$ with $L>K$ and $J>K$. By combining some values in the supports of X and Z, we first transform X and Z to $\widetilde{X}$ and $\widetilde{Z}$ so that they share the same support $\mathcal{X}^{*}$ as X^{*}. We then identify $f_{\widetilde{X}|X^{*}}$ and $f_{\widetilde{Z}|X^{*}}$ by Theorem 2 with those assumptions imposed on $\left(\widetilde{X},Y,\widetilde{Z},X^{*}\right)$. However, the joint distribution $f_{X,Y,Z,X^{*}}$ may still be of interest. In order to identify $f_{Z|X^{*}}$ or $M_{Z|X^{*}}$, we consider the joint distribution

$$f_{\widetilde{X},Z}=\sum_{x^{*}\in\mathcal{X}^{*}}f_{\widetilde{X}|X^{*}}f_{Z|X^{*}}f_{X^{*}}, \tag{2.33}$$

which is equivalent to

$$M_{\widetilde{X},Z} = M_{\widetilde{X}|X^*} D_{X^*} M^T_{Z|X^*} \,. \tag{2.34}$$

Since we have identified $M_{\widetilde{X}|X^*}$ and D_{X^*}, we can identify $M_{Z|X^*}$, i.e., $f_{Z|X^*}$, by inverting $M_{\widetilde{X}|X^*}$. Similar argument holds for identification of $f_{X|X^*}$. This discussion implies that Assumptions 7 is not necessary. We keep it in Theorem 2 in order to show minimum data information needed for nonparametric identification of the 2.1-measurement model.

2.4.2 Misclassification versus finite mixture

Misclassification models and finite mixture models are closely related, if not equivalent. The literature on finite mixture models started with a setup as follows:

$$f_D = \sum_{\tau \in \{1,2,...,T\}} f_{D|\tau} f_\tau. \tag{2.35}$$

where $\tau \in \{1, 2, ..., T\}$ for some finite T and D stands for observed variables in the data. Researchers are interested in the distribution $f_{D|\tau}$ while only f_D is observed.

An approach of finite mixture models considers what restrictions can be imposed on the distributions $f_{D|\tau}$ for a small T, e.g., $T = 2$, so that $f_{D|\tau}$ can be uniquely determined by f_D. For example, one of such restrictions may be that $f_{D|\tau}$ is symmetric.

Since such restrictions may be too restrictive for empirical research, another approach of finite mixture models impose conditional independence restrictions such as $D = (X, Y, Z)$ and

$$f_D = \sum_\tau f_{X|\tau} f_{Y|\tau} f_{Z|\tau} f_\tau. \tag{2.36}$$

Such a setup is literally equivalent to a misclassification model, where many existing identification results apply.

A general local identification result, without the ordering conditions in Assumption 10 and the support condition in Definition 1, may be found in Allman et al. (2009). In our 2.1-measurement model, the equality in the rank condition in their Theorem 1 holds. To be specific, Assumption 9, which guarantees distinctive eigenvalues, holds if and only if the so-called Kruskal rank of their matrix corresponding to the binary Y is equal to 2. The Kruskal ranks of their other two matrices are equal to the regular matrix rank K, and therefore, the total Kruskal rank equals $2K + 2$. In addition, for a general discrete Y, Assumption 9 implies that the Kruskal rank of their matrix corresponding to Y is at least 2.

We prove the claims above as follows. We may define the matrix corresponding to the variable Y in the same way as in Allman et al. (2009) as follows:

$$M^T_{Y|X^*} \quad = \quad \begin{pmatrix} f_{Y|X^*}(0|x_1^*) & f_{Y|X^*}(1|x_1^*) \\ f_{Y|X^*}(0|x_2^*) & f_{Y|X^*}(1|x_2^*) \\ ... & ... \\ f_{Y|X^*}(0|x_K^*) & f_{Y|X^*}(1|x_K^*) \end{pmatrix}$$

For a matrix M, the Kruskal rank of M will mean the largest number I such that every set of I rows of M are independent. The Kruskal rank is smaller than or equal to the regular rank of the same matrix. In the case where a matrix M of size K-by-L has rank K, it also has Kruskal rank K. That means the Kruskal rank of $M^T_{X|X^*}$ and $M^T_{Z|X^*}$ is K.

We may then show that Assumption 9 holds if and only if the Kruskal rank of $M^T_{Y|X^*}$ is equal to 2. Let $\omega(x) = x$. Assumption 9 becomes $f_{Y|X^*}(1|\widetilde{x}^*) - f_{Y|X^*}(1|\overline{x}^*) \neq 0$ for any $\overline{x}^* \neq \widetilde{x}^*$ in $\mathcal{X}^*$. For any 2 rows of $M^T_{y|x^*}$ corresponding to $\overline{x}^* \neq \widetilde{x}^*$, we consider the following matrix

$$\begin{pmatrix} f_{Y|X^*}(0|\overline{x}^*) & f_{Y|X^*}(1|\overline{x}^*) \\ f_{Y|X^*}(0|\widetilde{x}^*) & f_{Y|X^*}(1|\widetilde{x}^*) \end{pmatrix}$$

The determinant of this matrix is

$$\begin{aligned} & \left(1 - f_{Y|X^*}(1|\overline{x}^*)\right) f_{Y|X^*}(1|\widetilde{x}^*) - f_{Y|X^*}(1|\overline{x}^*) \left(1 - f_{Y|X^*}(1|\widetilde{x}^*)\right) \\ = \; & f_{Y|X^*}(1|\widetilde{x}^*) - f_{Y|X^*}(1|\overline{x}^*) \end{aligned}$$

Therefore, Assumption 9 implies that any 2 rows of $M^T_{Y|X^*}$ are independent. Since Y is binary, the largest number of independent rows is 2. Therefore, the Kruskal rank of $M^T_{Y|X^*}$ is 2. The reverse argument also holds. If the Kruskal rank of $M^T_{Y|X^*}$ is 2, any two rows of that matrix are independent. Therefore, the determinant of the 2-by-2 matrix formed by these two rows is not equal to zero, which implies Assumption 9 with $\omega(x) = x$.

For a general discrete Y with support $\{y_1, y_2, \ldots, y_m\}$. We may consider

$$M^T_{Y|X^*} = \begin{pmatrix} f_{Y|X^*}(y_1|x_1^*) & f_{Y|X^*}(y_2|x_1^*) & \ldots & f_{Y|X^*}(y_m|x_1^*) \\ f_{Y|X^*}(y_1|x_2^*) & f_{Y|X^*}(y_2|x_2^*) & \ldots & f_{Y|X^*}(y_m|x_2^*) \\ \ldots & \ldots & \ldots & \ldots \\ f_{Y|X^*}(y_1|x_K^*) & f_{Y|X^*}(y_2|x_K^*) & \ldots & f_{Y|X^*}(y_m|x_K^*) \end{pmatrix}$$

We can show that the Kruskal rank of $M^T_{Y|X^*}$ is at least 2 if and only if for any $\overline{x}^* \neq \widetilde{x}^*$ there exists a y_j such that $f_{Y|X^*}(y_j|\widetilde{x}^*) - f_{Y|X^*}(y_j|\overline{x}^*) \neq 0$. For any two rows with $\overline{x}^* \neq \widetilde{x}^*$, we consider the following matrix

$$M_2 = \begin{pmatrix} f_{Y|X^*}(y_1|\overline{x}^*) & f_{Y|X^*}(y_2|\overline{x}^*) & \ldots & f_{Y|X^*}(y_m|\overline{x}^*) \\ f_{Y|X^*}(y_1|\widetilde{x}^*) & f_{Y|X^*}(y_2|\widetilde{x}^*) & \ldots & f_{Y|X^*}(y_m|\widetilde{x}^*) \end{pmatrix}$$

Let $\mathbf{1} = (1, 1, ..., 1)^T$ and $e_j = (0, ..., 0, 1, 0, ..., 0)^T$, where 1 is at the j-th coordinate. We consider

$$M_2 \times (e_j \quad \mathbf{1}) = \begin{pmatrix} f_{Y|X^*}(y_j|\overline{x}^*) & 1 \\ f_{Y|X^*}(y_j|\widetilde{x}^*) & 1 \end{pmatrix}$$

Therefore, the rank of M_2 equals 2 if $f_{Y|X^*}(y_j|\widetilde{x}^*) - f_{Y|X^*}(y_j|\overline{x}^*) \neq 0$. That means the Kruskal rank of $M^T_{Y|X^*}$ is at least 2.

There reverse argument can also be shown similarly. If the Kruskal rank of $M^T_{Y|X^*}$ is at least 2, the rank of matrix M_2 equals 2. That means there must exist a column, say j, in M_2 such that $f_{Y|X^*}(y_j|\widetilde{x}^*) - f_{Y|X^*}(y_j|\overline{x}^*) \neq 0$.

2.4.3 A geometric illustration

Given that a matrix is a linear transformation from one vector space to another, we provide a geometric interpretation of the identification strategy. Consider $K = 3$ and define

$$\begin{aligned} \overrightarrow{p}_{X|x_i^*} &= \left[f_{X|X^*}(x_1|x_i^*), f_{X|X^*}(x_2|x_i^*), f_{X|X^*}(x_3|x_i^*)\right]^T \qquad (2.37) \\ \overrightarrow{p}_{X|z} &= \left[f_{X|Z}(x_1|z), f_{X|Z}(x_2|z), f_{X|Z}(x_3|z)\right]^T. \end{aligned}$$

We have for each z

$$\overrightarrow{p}_{X|z} = \sum_{i=1}^{3} w_i^z \left(\overrightarrow{p}_{X|x_i^*}\right) \tag{2.38}$$

with $w_i^z = f_{X^*|Z}(x_i^*|z)$ and $w_1^z + w_2^z + w_3^z = 1$. That means each observed distribution of X conditional on $Z = z$ is a weighted average of $\overrightarrow{p}_{X|x_1^*}$, $\overrightarrow{p}_{X|x_2^*}$, and $\overrightarrow{p}_{X|x_3^*}$. Similarly, if we consider the subsample with $Y = 1$, we have

$$\overrightarrow{p}_{y_1,X|z} = \sum_{i=1}^{3} w_i^z \left(\lambda_i \overrightarrow{p}_{X|x_i^*}\right) \tag{2.39}$$

where $\lambda_i = f_{Y|X^*}(1|x_i^*)$ and

$$\overrightarrow{p}_{y_1,X|z} = \left[f_{Y,X|Z}(1, x_1|z), f_{Y,X|Z}(1, x_2|z), f_{Y,X|Z}(1, x_3|z)\right]^T. \tag{2.40}$$

That means vector $\overrightarrow{p}_{y_1,X|z}$ is a weighted average of $\left(\lambda_i \overrightarrow{p}_{X|x_i^*}\right)$ for $i = 1, 2, 3$, where weights w_i^z are the same as in equation (2.38) from the whole sample. Notice that the direction of basis vectors $\left(\lambda_i \overrightarrow{p}_{X|x_i^*}\right)$ corresponding to the subsample with $Y = 1$ is the same as the direction of basis vectors $\overrightarrow{p}_{X|x_i^*}$ corresponding to the whole sample. The only difference is the length of the basis vectors. Therefore, if we consider a mapping from the vector space spanned by $\overrightarrow{p}_{X|z}$ to one spanned by $\overrightarrow{p}_{y_1,X|z}$, the basis vectors do not vary in direction so that they are called eigenvectors, and the variation in the length of these basis vectors is given by the corresponding eigenvalues, i.e., λ_i. This mapping is in fact $M_{X,y,Z}M_{X,Z}^{-1}$ on the left hand side of equation (2.30). The variation in variable Z guarantees that such a mapping exists. Figure 1 illustrates this framework.

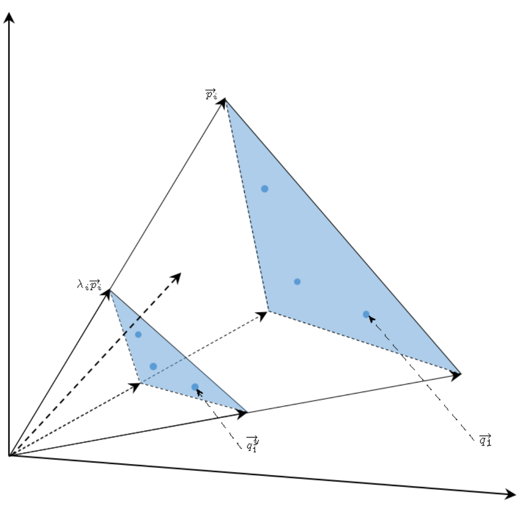

Figure 1: Eigenvalue-eigenvector decomposition in the 2.1-measurement model.
Eigenvalue: $\lambda_i = f_{Y|X^*}(1|x_i^*)$.
Eigenvector: $\overrightarrow{p_i} = \overrightarrow{p}_{X|x_i^*} = \left[f_{X|X^*}(x_1|x_i^*), f_{X|X^*}(x_2|x_i^*), f_{X|X^*}(x_3|x_i^*)\right]^T$.
Observed distribution in the whole sample:

$\overrightarrow{q}_1 = \overrightarrow{p}_{X|z_1} = \left[f_{X|Z}(x_1|z_1), f_{X|Z}(x_2|z_1), f_{X|Z}(x_3|z_1)\right]^T$.

Observed distribution in the subsample with $Y = 1$:

$\overrightarrow{q}_1^y = \overrightarrow{p}_{y_1,X|z_1} = \left[f_{Y,X|Z}(1, x_1|z_1), f_{Y,X|Z}(1, x_2|z_1), f_{Y,X|Z}(1, x_3|z_1)\right]^T$.

2.4.4 The continuous case

In the case where X, Z, and X^* are continuous, the identification strategy still work by replacing matrices with integral operators. We state assumptions as follows:

Assumption 11 *The joint distribution of (X, Y, Z, X^*) admits a bounded density with respect to the product measure of some dominating measure defined on $\mathcal{Y}$ and the Lebesgue measure on $\mathcal{X} \times \mathcal{X}^* \times \mathcal{Z}$. All marginal and conditional densities are also bounded.*

Assumption 12 *The operators $L_{X|X^*}$ and $L_{Z|X}$ are injective.*[10]

Assumption 13 *For all $\overline{x}^* \neq \widetilde{x}^*$ in $\mathcal{X}^*$, the set $\left\{y : f_{Y|X^*}(y|\overline{x}^*) \neq f_{Y|X^*}(y|\widetilde{x}^*)\right\}$ has positive probability.*

Assumption 14 *There exists a known functional M such that $M\left[f_{X|X^*}(\cdot|x^*)\right] = x^*$ for all $x^* \in \mathcal{X}^*$.*

Assumption 12 is a high-level technical condition. A sufficient condition for the injectivity of $L_{Z|X}$ is that the only function $h(\cdot)$ satisfying $E[h(X)|Z = z] = 0$ for any $z \in \mathcal{Z}$ is $h(\cdot) = 0$ over $\mathcal{X}$. This condition is also equivalent to the completeness of the density $f_{X|Z}$ over certain functional space. Assumption 13 requires that each possible value of the latent variable X^* affects the distribution of Y. The functional $M[\cdot]$ in Assumption 14 may be mean, mode, median, or another quantile, which maps a probability distribution to a point on the real line. We summarize the results as follows:

Theorem 3 *(Hu and Schennach (2008)) Under assumptions 11, 12, 13, and 14, the 2.1-measurement model in Definition 3 with a continuous X^* is non-parametrically identified in the sense that the joint distribution of the three variables (X, Y, Z), $f_{X,Y,Z}$, uniquely determines the joint distribution of the four variables (X, Y, Z, X^*), f_{X,Y,Z,X^*}, which satisfies equation (2.32).*

This result implies that if we observe an additional binary indicator of the latent variable together with two measurements, we can relax the additivity and the independence assumptions in equation (2.19) and achieve nonparametric identification of very general models. Comparing the model in equation (2.19) and the 2.1-measurement model, which are both point identified, the latter is much more flexible to accommodate various economic models with latent variables. For example, Theorem 3 identifies the joint distribution of X^* and Z, and therefore, applies to both the case where $Z = X^* + \epsilon'$ and the case where the relationship between Z and X^* is specified as $X^* = Z + \epsilon'$. The latter case is related to the so-called Berkson-type measurement error models (Schennach (2013)).

[10] $L_{Z|X}$ is defined in the same way as $L_{X|X^*}$ in equation (2.7).

2.4.5 An illustrative example

Here we use a simple example to illustrate the intuition of the identification results. Consider a labor supply model for college graduates, where Y is the 0-1 dichotomous employment status, X is the college GPA, Z is the SAT scores, and X^* is the latent ability type. We are interested in the probability of being employed given different ability, i.e., $\Pr(Y = 1|X^*)$, and the marginal probability of the latent ability f_{X^*}.

We consider a simplified version of the 2.1-measurement model with

$$\begin{aligned} \Pr(Y = 1|X^*) &\neq \Pr(Y = 1) \\ X &= X^*\gamma + \epsilon \\ Z &= X^*\gamma' + \epsilon' \end{aligned} \tag{2.41}$$

where $(X^*, \epsilon, \epsilon')$ are mutually independent. We may interpret the error term ϵ' as a performance shock in the SAT test. If coefficients γ and γ' are known, we can use X/γ and Z/γ' as the two measurements in equation (2.19) to identify the marginal distribution of ability without using the binary measurement Y. As shown in Hu and Sasaki (2015) , we can identify all the elements of interest in this model. Here we focus on the identification of the coefficients γ and γ' to illustrate the intuition of the identification results.

Since X^* is unobserved, we normalize $\gamma' = 1$ without loss of generality. A naive estimator for γ may be from the following regression equation

$$X = Z\gamma + (\epsilon - \epsilon'\gamma). \tag{2.42}$$

The OLS estimator corresponds to $\frac{cov(X,Z)}{var(Z)} = \gamma\frac{var(X^*)}{var(X^*)+var(\epsilon')}$, which is the well-known attenuation result with $|\frac{cov(X,Z)}{var(Z)}| < |\gamma|$. This regression equation suffers an endogeneity problem because the regressor, the SAT scores Z, does not perfectly reflect the ability X^* and is negatively correlated with the performance shock ϵ' in the regression error $(\epsilon - \epsilon'\gamma)$. When an additional variable Y is available, even if it is binary, we can use Y as an instrument to solve the endogeneity problem and identify γ as

$$\gamma = \frac{E[X|Y = 1] - E[X|Y = 0]}{E[Z|Y = 1] - E[Z|Y = 0]}. \tag{2.43}$$

This is literally the two-stage least square estimator. The regressor, SAT scores Z, is endogenous in both the employed subsample and the unemployed subsample. But the difference between the two subsamples may reveal how the observed GPA X is associated with ability X^* through γ.

The intuition of this identification strategy is that when we compare the employed $(Y = 1)$ subsample with the unemployed $(Y = 0)$ subsample, the only different element on the right hand side of the equation below is the marginal distribution of ability, i.e., $f_{X^*|Y=1}$ and $f_{X^*|Y=0}$ in

$$f_{X,Z|Y=y} = \int_{\mathcal{X}^*} f_{X|X^*} f_{Z|X^*} f_{X^*|Y=y} dx^*. \tag{2.44}$$

If we naively treat SAT scores Z as latent ability X^* to study the relationship between college GPA X and latent ability X^*, we may end up with a model with an endogeneity problem as in equation (2.42). However, the conditional independence assumption guarantees that the change in the employment status Y "exogenously" varies with latent ability X^*, and therefore, with the observed SAT scores Z, but does not vary with the performance shock ϵ', which is the cause of the endogeneity problem. Therefore, the employment status Y may serve as an instrument to achieve identification. Notice that this argument still holds if we compare the employed subsample with the whole sample, which is what we use in equations (2.38) and (2.39) in Section 2.4.3.[11]

Furthermore, an arguably surprising result is that such identification of the 2.1 measurement model is still nonparametric and global even if the instrument Y is binary. This is because the conditional independence assumption reduces the joint distribution f_{X,Y,Z,X^*} to distributions of each measurement conditional the latent variable $\left(f_{X|X^*}, f_{Y|X^*}, f_{Z|X^*}\right)$, and the marginal distribution f_{X^*} as in equation (2.32). The joint distribution f_{X,Y,Z,X^*} is a four-dimensional function, while $\left(f_{X|X^*}, f_{Y|X^*}, f_{Z|X^*}\right)$ are three two-dimensional functions. Therefore, the number of unknowns are greatly reduced under the conditional independence assumption.

2.5 A 3-measurement model

We introduce the 2.1-measurement model to show the least data information needed for nonparametric identification of a measurement error model. Given that a random variable can always be transformed to a 0-1 dichotomous variable, the identification result can still hold when there are three measurements of the latent variable. In this section, we introduce the 3-measurement model to emphasize that three observables may play exchangeable roles so that it does not matter which measurement is called a dependent variable, a measurement, or an instrument variable. We define this case as follows:

Definition 4 *A* ***3-measurement model*** *contains three measurements, as in Definition 1, $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, and $Z \in \mathcal{Z}$ of the latent variable $X^* \in \mathcal{X}^*$ satisfying*

$$X \perp Y \perp Z \mid X^*, \tag{2.45}$$

i.e., (X, Y, Z) are jointly independent conditional on X^*.

Based on the results for the 2.1-measurement model, nonparametric identification of the joint distribution f_{X,Y,Z,X^*} in the 3-measurement model is feasible because one can always replace Y with a 0-1 binary indicator, e.g., $I(Y > E[Y])$. In fact, we intentionally write the results in section 2.4 in such a way that the assumptions and the theorems remain the same after replacing the binary support $\{0, 1\}$ with a general support $\mathcal{Y}$ for variable Y. An important observation here is that the three measurements (X, Y, Z) play exchangeable roles in the 3-measurement model. We can impose different restrictions on different measurements, which makes one look like a dependent variable, one like a measurement, and another like an instrument. But these “assignments" are arbitrary. On the one hand, the researcher may decide which “assignments" are reasonable based on the economic model. On the other hand, it does not matter which variable is called a dependent variable, a measurement, or an instrument variable in terms of identification. We summarize the results as follows:

[11]Another way to look at this is that γ can also be expressed as

$$\gamma = \frac{E[X|Y=1] - E[X]}{E[Z|Y=1] - E[Z]}.$$

Corollary 1 *Theorems 2 and 3 both hold for the 3-measurement model in Definition 4.*

For example, we consider a hidden Markov model containing $\{X_t, X_t^*\}$, where $\{X_t^*\}$ is a latent first-order Markov process, i.e.,

$$X_{t+1}^* \perp \{X_s^*\}_{s\leq t-1} \mid X_t^*. \tag{2.46}$$

In each period, we observe a measurement X_t of the latent X_t^* satisfying

$$X_t \perp \{X_s, X_s^*\}_{s\neq t} \mid X_t^*. \tag{2.47}$$

This is the so-called local independence assumption, where a measurement X_t is independent of everything else conditional the latent variable X_t^* in the sample period. The relationship among the variables can be shown in the flow chart as follows.

$$\begin{array}{ccccccc} & X_{t-1} & & X_t & & X_{t+1} & \\ & \uparrow & & \uparrow & & \uparrow & \\ \longrightarrow & X_{t-1}^* & \longrightarrow & X_t^* & \longrightarrow & X_{t+1}^* & \longrightarrow \end{array}$$

Consider a panel data set, where we observed three periods of data $\{X_{t-1}, X_t, X_{t+1}\}$. The conditions in equations (2.46) and (2.47) imply

$$X_{t-1} \perp X_t \perp X_{t+1} \mid X_t^*, \tag{2.48}$$

i.e., (X_{t-1}, X_t, X_{t+1}) are jointly independent conditional on X_t^*. Although the original model is dynamic, it can be reduced to a 3-measurement model as in equation (2.48). Corollary 1 then non-parametrically identifies $f_{X_{t+1}|X_t^*}$, $f_{X_t|X_t^*}$, $f_{X_{t-1}|X_t^*}$, and $f_{X_t^*}$. Under a stationarity assumption that $f_{X_{t+1}|X_{t+1}^*} = f_{X_t|X_t^*}$, we can then identify the Markov kernel $f_{X_{t+1}^*|X_t^*}$ from

$$f_{X_{t+1}|X_t^*} = \int_{\mathcal{X}^*} f_{X_{t+1}|X_{t+1}^*} f_{X_{t+1}^*|X_t^*} dx_{t+1}^*, \tag{2.49}$$

by inverting the integral operator corresponding to $f_{X_{t+1}|X_{t+1}^*}$.[12] Therefore, it does not really matter which one of $\{X_{t-1}, X_t, X_{t+1}\}$ is treated as measurement or instrument for X_t^*. Applications of nonparametric identification of such a hidden Markov model or, in general, the 3-measurement model can be found in Hu et al. (2013b) , Feng and Hu (2013) , Wilhelm (2013) , and Hu and Sasaki (forthcoming, 2017a) , etc.

2.6 A dynamic measurement model

A natural extension to the hidden Markov model in equations (2.46)-(2.47) is to relax the local independence assumption in equation (2.47) when more periods of data are available. For example, we may allow direct serial correlation among observed measurements $\{X_t\}$ of latent variables $\{X_t^*\}$. To this end, we assume the following:

Assumption 15 *The joint process $\{X_t, X_t^*\}$ is a first-order Markov process. Furthermore, the Markov kernel satisfies*

$$f_{X_t,X_t^*|X_{t-1},X_{t-1}^*} = f_{X_t|X_t^*,X_{t-1}} f_{X_t^*|X_{t-1},X_{t-1}^*}. \tag{2.50}$$

[12] Without stationarity, one can use one more period of data, i.e., X_{t+2}, to identify $f_{X_{t+1}|X_{t+1}^*}$ from the joint distribution of (X_t, X_{t+1}, X_{t+2}) .

Equation (2.50) is the so-called limited feedback assumption in Hu and Shum (2012) . It implies that the latent variable in current period has summarized all the information on the latent part of the process. The relationship among the variables may be described as follows:

$$\begin{array}{ccccccccc} \longrightarrow & X_{t-2} & \longrightarrow & X_{t-1} & \longrightarrow & X_t & \longrightarrow & X_{t+1} & \longrightarrow \\ \searrow & \updownarrow & \searrow & \updownarrow & \searrow & \updownarrow & \searrow & \updownarrow & \searrow \\ \longrightarrow & X_{t-2}^* & \longrightarrow & X_{t-1}^* & \longrightarrow & X_t^* & \longrightarrow & X_{t+1}^* & \longrightarrow \end{array}$$

For simplicity, we focus on the discrete case and assume

Assumption 16 *X_t and X_t^* share the same support $\mathcal{X}^* = \{x_1^*, x_2^*, \ldots, x_K^*\}$.*

The observed distribution is associated with unobserved ones as follows:

$$f_{X_{t+1},X_t,X_{t-1},X_{t-2}} = \sum_{x^*} f_{X_{t+1}|X_t,X_t^*} f_{X_t|X_t^*,X_{t-1}} f_{X_t^*,X_{t-1},X_{t-2}}. \tag{2.51}$$

We define for any fixed (x_t, x_{t-1})

$$\begin{aligned} M_{X_{t+1},x_t|x_{t-1},X_{t-2}} &= \left[f_{X_{t+1},X_t|X_{t-1},X_{t-2}}(x_i, x_t|x_{t-1}, x_j)\right]_{i=1,2,\ldots,K;j=1,2,\ldots,K} \\ M_{X_t|x_{t-1},X_{t-2}} &= \left[f_{X_t|X_{t-1},X_{t-2}}(x_i|x_{t-1}, x_j)\right]_{i=1,2,\ldots,K;j=1,2,\ldots,K}. \end{aligned} \tag{2.52}$$

Assumption 17 *(i) for any $x_{t-1} \in \mathcal{X}$, $M_{X_t|x_{t-1},X_{t-2}}$ is invertible.*

(ii) for any $x_t \in \mathcal{X}$, there exists a $(x_{t-1}, \overline{x}_{t-1}, \overline{x}_t)$ such that $M_{X_{t+1},x_t|x_{t-1},X_{t-2}}$, $M_{X_{t+1},x_t|\overline{x}_{t-1},X_{t-2}}$, $M_{X_{t+1},\overline{x}_t|x_{t-1},X_{t-2}}$, and $M_{X_{t+1},\overline{x}_t|\overline{x}_{t-1},X_{t-2}}$ are invertible and that for all $x_t^* \neq \widetilde{x}_t^*$ in $\mathcal{X}^*$

$$\Delta_{x_t}\Delta_{x_{t-1}} \ln f_{X_t|X_t^*,X_{t-1}}(x_t^*) \neq \Delta_{x_t}\Delta_{x_{t-1}} \ln f_{X_t|X_t^*,X_{t-1}}(\widetilde{x}_t^*)$$

where $\Delta_{x_t}\Delta_{x_{t-1}} \ln f_{X_t|X_t^*,X_{t-1}}(x_t^*)$ is defined as

$$\begin{aligned} \Delta_{x_t}\Delta_{x_{t-1}} \ln f_{X_t|X_t^*,X_{t-1}}(x_t^*) \quad : \quad &= \left[\ln f_{X_t|X_t^*,X_{t-1}}(x_t|x_t^*, x_{t-1}) - \ln f_{X_t|X_t^*,X_{t-1}}(x_t|x_t^*, \overline{x}_{t-1})\right] \\ &- \left[\ln f_{X_t|X_t^*,X_{t-1}}(\overline{x}_t|x_t^*, x_{t-1}) - \ln f_{X_t|X_t^*,X_{t-1}}(\overline{x}_t|x_t^*, \overline{x}_{t-1})\right]. \end{aligned}$$

Assumption 18 *For any $x_t \in \mathcal{X}$, there exists a known functional M such that $M\left[f_{X_{t+1}|X_t,X_t^*}(\cdot|x_t, x_t^*)\right]$ is strictly increasing in x_t^*.*

Assumption 19 *The Markov kernel is stationary, i.e.,*

$$f_{X_t,X_t^*|X_{t-1},X_{t-1}^*} = f_{X_2,X_2^*|X_1,X_1^*}. \tag{2.53}$$

The invertibility in Assumption 17 is testable because it imposes a rank condition on observed matrices. The invertibility guarantees that a directly estimable matrix has an eigenvalue-eigenvector decomposition, where the eigenvalues are associated with $\Delta_{x_t}\Delta_{x_{t-1}} \ln f_{X_t|X_t^*,X_{t-1}}$ and the eigenvectors are related to $f_{X_{t+1}|X_t,X_t^*}(\cdot|x_t, x_t^*)$ for a fixed x_t. Assumption 17(ii) is needed for the distinctiveness of the eigenvalues. And Assumption 18 reveals the ordering of the eigenvectors as Assumption 14. Assumption 19 is a stationarity assumption, which is not needed with one more periods of data. We summarize the results as follows:

Theorem 4 *(Hu and Shum (2012)) Under assumptions 15, 16, 17, 18, and 19, the joint distribution of four periods of data $f_{X_{t+1},X_t,X_{t-1},X_{t-2}}$ uniquely determines the Markov transition kernel $f_{X_t,X_t^*|X_{t-1},X_{t-1}^*}$ and the initial condition f_{X_{t-2},X_{t-2}^*}.*

For the continuous case and other variations of the assumptions, such as non-stationarity, I refer to Hu and Shum (2012) for details. A simple extension of this result is the case where X_t^* is discrete and X_t is continuous. As in the discussion following Theorem 2, the identification results still apply with minor modification of the assumptions.

In the case where $X_t^* = X^*$ is time-invariant, the condition in equation (2.50) is not restrictive and the Markov kernel becomes $f_{X_t|X_{t-1},X^*}$. For such a first-order Markov model, Kasahara and Shimotsu (2009) suggest using two periods of data to break the interdependence and use six periods of data to identify the transition kernel. For fixed $X_t = x_t$, $X_{t+2} = x_{t+2}$, $X_{t+4} = x_{t+4}$, it can be shown that $X_{t+1}, X_{t+3}, X_{t+5}$ are independent conditional on X^* as follows:

$$f_{X_{t+5},x_{t+4},X_{t+3},x_{t+2},X_{t+1},x_t} = \sum_{x^*\in\mathcal{X}^*} f_{X_{t+5}|x_{t+4},X^*} f_{x_{t+4},X_{t+3}|x_{t+2},X^*} f_{x_{t+2}.X_{t+1},x_t,X^*}.$$

The model then falls into the framework of the 3-measurement model, where $(X_{t+1}, X_{t+3}, X_{t+5})$ may serve as three measurements for each fixed (x_t, x_{t+2}, x_{t+4}) to achieve identification. This similarity to the 3-measurement model can also be seen in Bonhomme et al. (2015) and Bonhomme et al. (2016) . However, the 2.1-measurement model implies that minimum data information for nonparametric identification can be “2.1 measurements" instead of “3 measurements". Hu and Shum (2012) shows that the interaction between observables in the middle two periods may play the role of the binary measurement in the 2.1-measurement model so that such a model, even with a time-varying unobserved state variable, can be identified using only four periods of data.

2.6.1 Illustrative Examples

In this section, we use a simple example to illustrate the identification strategy in Theorem 4, which is based on Carroll et al. (2010) . Consider estimation of a consumption equation using two samples. Let Y be the consumption, X^* be the latent true income, Z be the family size, and $S \in \{s_1, s_2\}$ be a sample indicator. The data structure can be described as follows:

$$f_{Y,X|Z,S} = \int f_{Y|X^*,Z} f_{X|X^*,S} f_{X^*|Z,S} dx^*. \tag{2.54}$$

The consumption model is described by $f_{Y|X^*,Z}$, where consumption depends on income and family size. The self-reported income X may have different distributions in the two samples. The income X^* may be correlated with the family size Z and the income distribution may also be different in the two samples. Carroll et al. (2010) provide sufficient conditions for nonparametric identification of all the densities on the right hand side of equation (2.54). To illustrate the identification strategy, we consider the following parametric specification

$$\begin{aligned} Y &= \beta X^* + \gamma Z + \eta \qquad (2.55) \\ X &= X^* + \gamma' S + \epsilon \\ X^* &= \delta_1 S + \delta_2 Z + \delta_3 (S \times Z) + u, \end{aligned}$$

where $(\beta, \gamma, \gamma', \delta_1, \delta_2, \delta_3)$ are unknown constants with $\delta_3 \neq 0$.

We focus on the identification of β. If we naively treat X as the latent true income X^*, we have a model with endogeneity as follows:

$$\begin{aligned} Y &= \beta (X - \gamma' S - \epsilon) + \gamma Z + \eta \qquad (2.56) \\ &= \beta X + \gamma Z - \beta\gamma' S + (\eta - \beta\epsilon). \end{aligned}$$

The regressor X is endogenous because it is correlated with the measurement error ϵ. Note that the income X^* may vary with the family size Z and the sample indicator S, which are independent of ϵ, the source of the endogeneity. The fact that there is no interaction term of Z and S on the right hand side of equation (2.56) is consistent with the conditional independence implied in equation (2.54). Let (z_0, z_1) and (s_0, s_1) be possible values of Z and S, respectively. Assuming $E[\eta|Z, S, X^*] = E[\epsilon|Z, S] = E[u|Z, S] = 0$, we estimate β as follows

$$\beta = \frac{[E(Y|z_1, s_1) - E(Y|z_0, s_1)] - [E(Y|z_1, s_0) - E(Y|z_0, s_0)]}{[E(X|z_1, s_1) - E(X|z_0, s_1)] - [E(X|z_1, s_0) - E(X|z_0, s_0)]}. \tag{2.57}$$

This is a 2SLS estimator using $(S \times Z)$ as an IV in the first stage, in which the numerator is a difference-in-differences estimator for $\beta\delta_3 (z_1 - z_0)(s_1 - s_0)$ and the denominator is a difference-in-differences estimator for $\delta_3 (z_1 - z_0)(s_1 - s_0)$.

In the dynamic model in Theorem 4, we can re-write equation (2.51) as

$$f_{X_{t+1}, X_{t-2}|X_t, X_{t-1}} = \sum_{x^*} f_{X_{t+1}|X_t^*, X_t} f_{X_{t-2}|X_t^*, X_{t-1}} f_{X_t^*|X_t, X_{t-1}}, \tag{2.58}$$

which is analogical to equation (2.54). Similar to the previous example on consumption, suppose we naively treat X_{t-2} as X_t^* to study the relationship between X_{t+1} and (X_t, X_t^*), say $X_{t+1} = H(X_t^*, X_t, \eta)$, where η is an independent error term. And suppose the conditional density $f_{X_{t-2}|X_t^*, X_{t-1}}$ implies $X_{t-2} = G(X_t^*, X_{t-1}, \epsilon)$, where ϵ represents an independent error term. Suppose we can replace X_t^* by $G^{-1}(X_{t-2}, X_{t-1}, \epsilon)$ to obtain

$$X_{t+1} = H\left(G^{-1}(X_{t-2}, X_{t-1}, \epsilon), X_t, \eta\right), \tag{2.59}$$

where X_{t-2} is endogenous and correlated with ϵ. The conditional independence in equation (2.58) implies that the variation in X_t and X_{t-1} may vary with X_t^*, but not with the error ϵ. However, the variation in X_t may change the relationship between the future X_{t+1} and the latent variable X_t^*, while the variation in X_{t-1} may change the relationship between the early X_{t-2} and the latent X_t^*. Therefore, a "joint" second-order variation in (X_t, X_{t-1}) may lead to an "exogenous" variation in X^*, which may solve the endogeneity problem. Thus, our identification strategy may be considered as a nonparametric version of a difference-in-differences argument.

For example, let X_t stand for the choice of health insurance between a high coverage plan and a low coverage plan. And X_t^* stands for the good or bad health status. The Markov process $\{X_t, X_t^*\}$ describes the interaction between insurance choices and health status. We consider the joint distribution of four periods of insurance choices $f_{X_{t+1}, X_t, X_{t-1}, X_{t-2}}$. If we compare a subsample with (X_t, X_{t-1}) = (high, high) and a subsample with and (X_t, X_{t-1}) = (high, low), we should be able to "difference out" the direct impact of health insurance choice X_t on the choice X_{t+1} in next period in $f_{X_{t+1}|X_t^*, X_t}$. Then, we may repeat such a comparison again with (X_t, X_{t-1}) = (low, high) and (X_t, X_{t-1}) = (low, low). In both comparisons, the impact of changes in insurance choice X_{t-1} described in $f_{X_{t-2}|X_t^*, X_{t-1}}$ is independent of the choice X_t. Therefore, the difference in the differences from those two comparisons above may lead to exogenous variation in X_t^* as described in $f_{X_t^*|X_t, X_{t-1}}$, which is independent of the endogenous error due to naively using X_{t-2} as X_t^*. Therefore, the second-order joint variation in observed insurance choices (X_t, X_{t-1}) may serve as an instrument to solve the endogeneity problem caused by using the observed insurance choice X_{t-2} as a proxy for the unobserved health condition X_t^*.

3

Estimation

This paper focuses on nonparametric identification of models with latent variables and its applications in applied microeconomic models. Given the length limit of the paper, I only provide a brief description of estimators proposed for the models above. All the identification results above are at the distribution level in the sense that probability distribution functions involving latent variables are uniquely determined by probability distribution functions of observables, which are directly estimable from a random sample of observables. Therefore, a maximum likelihood estimator is a straightforward choice for these models.

3.1 Sieve maximum likelihood estimators

Consider the 2.1-measurement model in Theorem 3, where the observed density is associated with the unobserved ones as follows:

$$f_{X,Y,Z}(x,y,z) = \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*) dx^*. \tag{3.1}$$

Our identification results provide conditions under which this equation has a unique solution $\left(f_{X|X^*}, f_{Y|X^*}, f_{Z|X^*}, f_{X^*}\right)$. Suppose that Y is the dependent variable and the model of interest is described by a parametric conditional density function as

$$f_{Y|X^*}(y|x^*) = f_{Y|X^*}(y|x^*;\theta). \tag{3.2}$$

With an i.i.d. sample $\{X_i, Y_i, Z_i\}_{i=1,2,...,N}$, we can use a sieve maximum likelihood estimator (Shen (1997) and Chen and Shen (1998)) based on

$$\left(\widehat{\theta}, \widehat{f}_{X|X^*}, \widehat{f}_{Z|X^*}, \widehat{f}_{X^*}\right) = \underset{(\theta,f_1,f_2,f_3)\in\mathcal{A}_N}{\arg\max} \frac{1}{N}\sum_{i=1}^{N} \ln \int_{\mathcal{X}^*} f_1(X_i|x^*) f_{Y|X^*}(Y_i|x^*;\theta) f_2(Z_i|x^*) f_3(x^*) dx^*, \tag{3.3}$$

where $\mathcal{A}_N$ is approximating sieve spaces which contain truncated series as parametric approximations to densities $\left(f_{X|X^*}, f_{Z|X^*}, f_{X^*}\right)$. For example, function $f_1(x|x^*)$ in the sieve space $\mathcal{A}_N$ can be as follows:

$$f_1(x|x^*) = \sum_{j=1}^{J_N}\sum_{k=1}^{K_N} \beta_{jk} p_j(x - x^*) p_k(x^*),$$

where $p_j(\cdot)$ is a known basis function, such as power series, splines, Fourier series, etc. and J_N and K_N are smoothing parameters. The choice of a sieve space depends on how well it can approximate the original functional space and how much computation burden it may lead to (See section 2.3.6 of Chen (2007) for details). One advantage of a sieve estimator is that it is relatively convenient to impose restrictions on the sieve space $\mathcal{A}_N$. To be specific, Assumption 14 can be imposed on the sieve coefficients β_{jk} (See section S4 of supplementary materials of Hu and Schennach (2008) for details). Since the coefficients are treated as unknown parameters in the likelihood function, the parameters of interest in Equation (3.3) can be estimated just as a parametric MLE. The number of coefficients $J_N \times K_N$ diverges at a given speed with the sample size N, which makes the approximation more flexible with a larger sample size. A useful result worth mentioning is that the parametric part of the model can converge at a fast rate, i.e., $\widehat{\theta}$ can be $\sqrt{n}$ consistent and asymptotically normally distributed under suitable assumptions (Shen (1997)). We refer to Hu and Schennach (2008) , Carroll et al. (2010) and supplementary materials for more discussion on this semi-nonparametric extremum estimator.

3.2 Closed-form estimators

Although the sieve MLE in (3.3) is quite general and flexible, a few identification results in this section provide closed-form expressions for the unobserved components as functions of observed distribution functions, which can lead to straightforward closed-form estimators. In the case where X^* is continuous, for example, Li and Vuong (1998) suggest that the distribution of the latent variable f_{X^*} in equation (2.20) can be estimated using Kotlarski's identity with characteristic functions being replaced by corresponding empirical characteristic functions. In general, one can consider a nonlinear regression model in the framework of the 3-measurement model as

$$\begin{aligned} Y &= g_1(X^*) + \eta \\ X &= g_2(X^*) + \epsilon \\ Z &= g_3(X^*) + \epsilon' \end{aligned} \tag{3.4}$$

where ϵ and ϵ' are independent of X^* and η with $E[\eta|X^*] = 0$. Since X^* is unobserved, we may normalize $g_3(X^*) = X^*$. Schennach (2004b) provides a closed-form estimator of $g_1(\cdot)$ in the case where $g_2(X^*) = X^*$ using Kotlarski's identity as follows:[1]

$$g_1(x^*) = \frac{\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-it_1 x^*}\left(\frac{\left[\frac{\partial}{\partial s}\phi_{X,Y}(t_1,s)\right]_{s=0}}{i\phi_X(t_1)}\phi_{X^*}(t_1)\right)dt_1}{\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx^*}\phi_{X^*}(t)\,dt}$$

where f_{X^*} is identified from the characteristic function

$$\phi_{X^*}(t) = \exp\left(\int_0^t \frac{\left[\frac{\partial}{\partial t_2}\phi_{X,Z}(s,t_2)\right]_{t_2=0}}{i\phi_X(s)}ds\right).$$

Hu and Sasaki (2015) generalize that estimator to the case where $g_2(\cdot)$ is a polynomial. Whether a closed-form estimator of $g_1(\cdot)$ exists or not with a general $g_2(\cdot)$ is a challenging and open question for future research.

[1] Schennach (2007) also provides a closed-form estimator for a similar nonparametric regression model using a generalized function approach.

In the case where X^* is discrete as in Theorem 2 and Corollary 1, the sieve MLE is still applicable. Nevertheless, the identification strategy in the discrete case also leads to a closed-form estimator for the unknown probabilities in the sense that one can mimic the identification procedure to solve for the unknowns. In estimation, it is more convenient to use the equation below than directly using Equation (2.27)

$$E\left[\omega\left(Y\right)|X=x,Z=z\right]f_{X,Z}\left(x,z\right)=\sum_{x^*\in\mathcal{X}^*}f_{X|X^*}(x|x^*)E\left[\omega\left(Y\right)|x^*\right]f_{Z|X^*}(z|x^*)f_{X^*}(x^*), \tag{3.5}$$

which leads to an eigenvalue-eigenvector decomposition

$$M_{X,\omega,Z}M_{X,Z}^{-1}=M_{X|X^*}D_{\omega|X^*}M_{X|X^*}^{-1} \tag{3.6}$$

with

$$\begin{aligned} M_{X,\omega,Z} &= \left[E\left[\omega\left(Y\right)|X=x_k,Z=z_l\right]f_{X,Z}\left(x_k,z_l\right)\right]_{k=1,2,...,K;l=1,2,...,K} \\ D_{\omega|X^*} &= diag\left\{E\left[\omega\left(Y\right)|x_1^*\right],E\left[\omega\left(Y\right)|x_2^*\right],\ldots,E\left[\omega\left(Y\right)|x_K^*\right]\right\}. \end{aligned} \tag{3.7}$$

The matrix $M_{X,\omega,Z}$ can be directly estimated as

$$\widehat{M_{X,\omega,Z}}=\left[\frac{1}{N}\sum_{i=1}^{N}\omega\left(Y_i\right)\mathbf{1}\left(X_i=x_k,Z_i=z_l\right)\right]_{k=1,2,...,K;l=1,2,...,K}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Similarly, matrix $M_{X,Z}$ can be estimated as $\widehat{M_{X,Z}}=\left.\widehat{M_{X,\omega,Z}}\right|_{\omega(\cdot)=1}$. Solving for eigenvectors and eigenvalues in Equation (3.6) can be considered as a procedure to minimize the Euclidean distance $\|\cdot\|$ between the left hand side and the right hand side of that equation, in fact, to zero. Moreover, Assumption 10 can be directly used to order the eigenvectors or the eigenvalues. With a finite sample, estimated probabilities might be outside $[0,1]$ or even a complex number. One remedy is to use Equation (3.6) as a moment condition to estimate the unknown probabilities under suitable restrictions. To be specific, matrices $M_{X|X^*}$ and $D_{\omega|X^*}$ can be estimated as follows:

$$\begin{aligned} \left(\widehat{M_{X|X^*}},\widehat{D_{\omega|X^*}}\right) &= \arg\min_{M,D}\left\|\widehat{M_{X,\omega,Z}}\left(\widehat{M_{X,Z}}\right)^{-1}M-M\times D\right\| \\ & \quad \text{such that} \\ & \quad \text{1) each entry in } M \text{ is in } [0,1]\ ; \\ & \quad \text{2) each column sum of } M \text{ equals 1 and } D \text{ is diagonal;} \\ & \quad \text{3) entries in } M \text{ and } D \text{ satisfy Assumptions 9 and 10.} \end{aligned}$$

This closed-form estimator performs well in empirical studies, such as An et al. (2015) , An et al. (2010) , Feng and Hu (2013) , and Hu et al. (2013b) .

Such closed-form estimators may not be as efficient as the sieve MLE, but they have their advantages that there are much fewer nuisance parameters involved than indirect estimators and that the computation of closed-form estimators may not rely on optimization algorithms, which usually involve many iterations and are time-consuming. An optimization algorithm can only guarantee a local maximum or minimum, while a closed-form estimator is a global one by construction. Although a closed-form estimator may not always exist, it is much more straightforward and transparent, if available, than an indirect estimator. Such closed-form estimation may be a challenging but useful approach for future research.

4

Unobservables in empirical industrial organization

A major breakthrough in the measurement error literature is the nonparametric identification of the 2.1-measurement model in section 2.4, which allows a very flexible relationship between observables and unobservables. The generality of these results enables researchers to tackle many important problems involving latent variables, such as belief, productivity, unobserved heterogeneity, and fixed effects, in the field of empirical industrial organization and labor economics.

4.1 Unobserved heterogeneity in Auctions

Unobserved heterogeneity has been a concern in the estimation of auction models for a long time. Li et al. (2000) and Krasnokutskaya (2011) use the identification result of 2-measurement model in equation (2.19) to estimate auction models with separable unobserved heterogeneity. In a first-price auction indexed by t for $t = 1, 2, \ldots, T$ without a reserve price, there are N symmetric risk-neutral bidders. For $i = 1, 2, \ldots, N$, each bidder i's cost is assumed to be decomposed into two independent factors as $s_t^* \times x_i$, where x_i is her private value and s_t^* is an auction-specific state or unobserved heterogeneity. With this decomposition of the cost, it can be shown that equilibrium bidding strategies b_{it} can also be decomposed as follows

$$b_{it} = s_t^* a_i, \tag{4.1}$$

where $a_i = a_i(x_i)$ represents equilibrium bidding strategies in the auction with $s_t^* = 1$. This falls into the 2-measurement model given that

$$b_{1t} \perp b_{2t} \mid s_t^*. \tag{4.2}$$

With such separable unobserved heterogeneity, one can consider the joint distribution of two bids as follows:

$$\begin{aligned} \ln b_{1t} &= \ln s_t^* + \ln a_1 \\ \ln b_{2t} &= \ln s_t^* + \ln a_2, \end{aligned} \tag{4.3}$$

where Kotlarski's identity is applicable for nonparametric identification of the distributions of $\ln s_t^*$ and $\ln a_i$. Further estimation of the value distribution from the distribution of $a_i(x_i)$ can be found in Guerre et al. (2000) .

Hu et al. (2013a) consider auction models with non-separable unobserved heterogeneity. They assume that the private values x_i are independent conditional on an auction-specific state or unobserved heterogeneity s_t^*. Based on the conditional independence of the values, the conditional independence of the bids holds, i.e.,

$$b_{1t} \perp b_{2t} \perp b_{3t} \mid s_t^*. \tag{4.4}$$

This falls into a 3-measurement model, where the three measurements, i.e., bids, are independent conditional on the unobserved heterogeneity. Nonparametric identification of the model then follows.

4.2 Auctions with unknown number of bidders

Since the earliest papers in the structural empirical auction literature, researchers have had to grapple with a lack of information on N^*, the number of potential bidders in the auction, which is an indicator of market competitiveness. The number of potential bidders may be different from the observed number of bidders A due to binding reserve prices, participation costs, or misreporting errors. For example, when reserve prices are binding, the number of potential bidders N^* would be observed by bidders and affect their bidding behavior. However, the observed number of bidders A, which is the number of participants whose bids exceed the reserve price, would be less than or equal to N^*.

In first-price sealed-bid auctions under the symmetric independent private values (IPV) paradigm, each of N^* potential bidders draws a private valuation from the distribution $F_{N^*}(x)$ with support $[\underline{x}, \overline{x}]$. The bidders observe N^*, which is latent to researchers. The reserve price r is assumed to be known and fixed across all auctions with $r > \underline{x}$. For each bidder i with valuation x_i, the equilibrium bidding function $b(x_i, N^*)$ can be shown as follows:

$$b(x_i; N^*) = \begin{cases} x_i - \frac{\int_r^{x_i} F_{N^*}(s)^{N^*-1} ds}{F_{N^*}(x_i)^{N^*-1}} & \text{for} \quad x_i \geq r \\ 0 & \text{for} \quad x_i < r. \end{cases} \tag{4.5}$$

The observed number of bidders is $A = \sum_{i=1}^{N^*} \mathbf{1}(x_i > r)$. In a random sample, we observe $\{A_t, b_{1t}, b_{2t}, \ldots, b_{A_t t}\}$ for each auction $t = 1, 2, \ldots, T$. One can show that

$$\begin{aligned} & f(A_t, b_{1t}, b_{2t} | b_{1t} > r, b_{2t} > r) \qquad (4.6) \\ = & \sum_{N^*} f(A_t | A_t \geq 2, N^*) f(b_{1t} | b_{1t} > r, N^*) f(b_{2t} | b_{2t} > r, N^*) f(N^* | b_{1t} > r, b_{2t} > r). \end{aligned}$$

That means that the two bids and the observed number of bidders are independent conditional on the number of potential bidders, which forms a 3-measurement model. In addition, the fact that $A_t \leq N^*$ provides an ordering of the eigenvectors corresponding to $f_{A_t|N_t^*}$. As shown in An et al. (2010) , the bid distribution, and therefore, the value distribution, can be non-parametrically identified. Furthermore, such identification is constructive and directly leads to an estimator.

4.3 Beliefs in learning models

How economic agents learn from past experience has been an important issue in both empirical industrial organization and labor economics. The key difficulty in the estimation of learning models is that beliefs are time-varying and unobserved in the data. Hu et al. (2013b) use bandit experiments to non-parametrically estimate the learning rule using auxiliary measurements of beliefs. In each period, an economic agent is asked to choose between two slot machines, which have different winning probabilities. Based on her own belief on which slot machine has a higher winning probability, the agent makes her choice of slot machine and receives rewards according to its winning probability. Although she does not know which slot machine has a higher winning probability, the agent is informed that the winning probabilities may switch between the two slot machines.

In addition to choices Y_t and rewards R_t, researchers also observe a proxy Z_t for the agent's belief X_t^*. Recorded by an eye-tracker machine, the proxy describes how much more time the agent looks at one slot machine than at the other. Under a first-order Markovian assumption, the learning rule is described by the distribution of the next period's belief conditional on previous belief, choice, and reward, i.e., $\Pr\left(X_{t+1}^*|X_t^*, Y_t, R_t\right)$. They assume that the choice only depends the belief and that the proxy Z_t is also independent of other variables conditional on the current belief X_t^*. The former assumption is motivated by a fully-rational Bayesian belief-updating rule, while the latter is a local independence assumption widely-used in the measurement error literature. These assumptions imply a 2.1-measurement model with

$$Z_t \perp Y_t \perp Z_{t-1}|X_t^*. \tag{4.7}$$

Therefore, the proxy rule $\Pr\left(Z_t|X_t^*\right)$ is non-parametrically identified. Under the local independence assumption, one can identify distribution functions containing the latent belief X_t^* from the corresponding distribution functions containing the observed proxy Z_t. That means the learning rule $\Pr\left(X_{t+1}^*|X_t^*, Y_t, R_t\right)$ can be identified from the observed distribution $\Pr\left(Z_{t+1}, Y_t, R_t, Z_t\right)$ through

$$\begin{aligned} & \Pr\left(Z_{t+1}, Y_t, R_t, Z_t\right) \\ = & \sum_{X_{t+1}^*}\sum_{X_t^*} \Pr\left(Z_{t+1}|X_{t+1}^*\right)\Pr\left(Z_t|X_t^*\right)\Pr\left(X_{t+1}^*, X_t^*, Y_t, R_t\right). \end{aligned} \tag{4.8}$$

The nonparametric learning rule they found implies that agents are more reluctant to "update down" following unsuccessful choices, than "update up" following successful choices. That leads to the sub-optimality of this learning rule in terms of profits.

4.4 Effort and types in online credit market

Xin (2018) studies the impact of reputation/feedback systems on the operation of online credit markets using data from Prosper.com. A major concern in markets of unsecured loans is the ability of lenders to recover their loan due to the problems of asymmetric information. On the one hand, borrowers differ in their inherent default costs c, which is hidden information; on the other hand, borrowers' efforts e_t to repay debts are hidden as well, so additional incentives are necessary to motivate them.

Xin (2018) is the first to quantify the extent to which reputation/feedback systems improve the total welfare of market participants when both hidden information (adverse selection) and hidden actions (moral hazard) are present. She identifies and estimates a finite-horizon dynamic model of a credit market in which borrowers and lenders interact repeatedly over time. The observables include the outcome variables O_t, including default and late payment performances, and individual characteristics X_t, such as debt-to-income ratio and credit grade.

The dynamic structure implies that

$$f(O_t, X_t, O_{t-1}, X_{t-1}) = \sum_c f(c, X_{t-1}, O_{t-1}) f(X_t|X_{t-1}, O_{t-1}, c) f(O_t|c, X_t)$$

The type distribution $f(c|X_{t-1}, O_{t-1})$ is identified for borrowers with multiple loans using the identification results in Hu and Shum (2012).

Furthermore, loan outcomes include borrowers' default and late payment performances, $O_t = \{D_t, L_t\}$. The model implies that default and late payment are independent conditional on effort, i.e.,

$$f(O_t|c, X_t) = \sum_{e_t} f(D_t|e_t) f(L_t|e_t) f(e_t|c, X_t)$$

Following the results in Hu (2008), effort choice probabilities and outcome realization process are identified.

These results lead to identification of utility parameters in borrowers' payoff functions and the outside option distributions for borrowers and lenders using variations in interest rates. In the last step, given other primitives that have been recovered, she identifies the original type distribution for all borrowers before any selection occurs. Using these structural estimates, the paper also conducts counterfactual experiments.

5

Unobservables in labor economics

5.1 Unemployment and labor force participation

Unemployment rates may be one of the most important economic indicators. The official US unemployment rates are estimated using self-reported labor force statuses in the Current Population Survey (CPS). It is known that ignoring misreporting errors in the CPS may lead to biased estimates. Feng and Hu (2013) use a hidden Markov approach to identify and estimate the distribution of the true labor force status. Let X_t^* and X_t denote the true and self-reported labor force status in period t. They merge monthly CPS surveys and are able to obtain a random sample $\{X_{t+1}, X_t, X_{t-9}\}_i$ for $i = 1, 2, \ldots, N$. Using X_{t-9} instead of X_{t-1} may provide more variation in the observed labor force status. They assume that the misreporting error only depends on the true labor force status in the current period, and therefore,

$$\begin{aligned} & \Pr\left(X_{t+1}, X_t, X_{t-9}\right) \\ = & \sum_{X_{t+1}^*}\sum_{X_t^*}\sum_{X_{t-9}^*} \Pr\left(X_{t+1}|X_{t+1}^*\right)\Pr\left(X_t|X_t^*\right)\Pr\left(X_{t-9}|X_{t-9}^*\right)\Pr\left(X_{t+1}^*, X_t^*, X_{t-9}^*\right). \end{aligned} \tag{5.1}$$

With three unobservables and three observables, nonparametric identification is not feasible without further restrictions. They then assume that $\Pr\left(X_{t+1}^*|X_t^*, X_{t-9}^*\right) = \Pr\left(X_{t+1}^*|X_t^*\right)$, which is similar to a first-order Markov condition. Under these assumptions, they obtain

$$\begin{aligned} & \Pr\left(X_{t+1}, X_t, X_{t-9}\right) \\ = & \sum_{X_t^*} \Pr\left(X_{t+1}|X_t^*\right)\Pr\left(X_t|X_t^*\right)\Pr\left(X_t^*, X_{t-9}\right), \end{aligned} \tag{5.2}$$

which implies a 3-measurement model. This model can be considered as an application of Theorem 2 to a hidden Markov model.

Feng and Hu (2013) found that the official U.S. unemployment rates substantially underestimate the true level of unemployment, due to misreporting errors in the labor force status in the Current Population Survey. From January 1996 to August 2011, the corrected monthly unemployment rates are 2.1 percentage points higher than the official rates on average, and are more sensitive to changes in business cycles. The labor force participation rates, however, are not affected by this correction.

5.2 Cognitive and noncognitive skill formation

Cunha et al. (2010) consider a model of cognitive and non-cognitive skill formation, where for multiple periods of childhood $t \in \{1, 2, \ldots, T\}$, $X_t^* = \left(X_{C,t}^*, X_{N,t}^*\right)$ stands for cognitive and non-cognitive skill stocks in period t, respectively. The T childhood periods are divided into $s \in \{1, 2, \ldots, S\}$ stages of childhood development with $S \leq T$. Let $I_t = (I_{C,t}, I_{N,t})$ be parental investments at age t in cognitive and non-cognitive skills, respectively. For $k \in \{C, N\}$, they assume that skills evolve as follows:

$$X_{k,t+1}^* = f_{k,s}\left(X_t^*, I_t, X_P^*, \eta_{k,t}\right), \tag{5.3}$$

where $X_P^* = \left(X_{C,P}^*, X_{N,P}^*\right)$ are parental cognitive and non-cognitive skills and $\eta_t = \left(\eta_{C,t}, \eta_{N,t}\right)$ is random shocks. If one observes the joint distribution of X^* defined as

$$X^* = \left(\left\{X_{C,t}^*\right\}_{t=1}^T, \left\{X_{N,t}^*\right\}_{t=1}^T, \{I_{C,t}\}_{t=1}^T, \{I_{N,t}\}_{t=1}^T, X_{C,P}^*, X_{N,P}^*\right), \tag{5.4}$$

one can estimate the skill production function $f_{k,s}$.

However, the vector of latent factors X^* is not directly observed in the sample. Instead, they use measurements of these factors satisfying

$$X_j = g_j\left(X^*, \varepsilon_j\right) \tag{5.5}$$

for $j = 1, 2, \ldots, M$ with $M \geq 3$. The variables X_j and ε_j are assumed to have the same dimension as X^*. Under the assumption that

$$X_1 \perp X_2 \perp X_3 \mid X^*, \tag{5.6}$$

this leads to a 3-measurement model and the distribution of X^* can then be identified from the joint distribution of the three observed measurements. The measurements X_j in their application include test scores, parental and teacher assessments of skills, and measurements on investment and parental endowments. While estimating the empirical model, they assume a linear function g_j and use Kotlarski's identity to directly estimate the latent distribution.

5.3 Income dynamics

The literature on income dynamics has been focusing mostly on linear models, where identification is usually not a major concern. When income dynamics have a nonlinear transmission of shocks, however, it is not clear how much of the model can be identified. Arellano et al. (2014) investigate the nonlinear aspect of income dynamics and also assess the impact of nonlinear income shocks on household consumption.

They assume that the pre-tax labor income Y_{it} of household i at age t satisfies

$$Y_{it} = \eta_{it} + \varepsilon_{it} \tag{5.7}$$

where η_{it} is the persistent component of income and ε_{it} is the transitory one. Furthermore, they assume that ε_{it} has a zero mean and is independent over time, and that the persistent component η_{it} follows a first-order Markov process satisfying

$$\eta_{it} = Q_t\left(\eta_{i,t-1}, u_{it}\right) \tag{5.8}$$

where Q_t is the conditional quantile function and u_{it} is uniformly distributed and independent of $\left(\eta_{i,t-1}, \eta_{i,t-2}, \ldots\right)$. Such a specification is without loss of generality under the assumption that the conditional CDF $F\left(\eta_{it}|\eta_{i,t-1}\right)$ is invertible with respect to η_{it}.

The dynamic process $\{Y_{it}, \eta_{it}\}$ can be considered as a hidden Markov process as $\{X_t, X_t^*\}$ in equations (2.46) and (2.47). As we discussed before, the nonparametric identification is feasible with three periods of observed income $(Y_{i,t-1}, Y_{it}, Y_{i,t+1})$ satisfying

$$Y_{i,t-1} \perp Y_{it} \perp Y_{i,t+1} \mid \eta_{it} \tag{5.9}$$

which forms a 3-measurement model. Under the assumptions in Theorem 3, the distribution of ε_{it} is identified from $f(Y_{it}|\eta_{it})$ for $t = 2, \ldots, T-1$. The joint distribution of η_{it} for all $t = 2, \ldots, T-1$ can then be identified from the joint distribution of Y_{it} for all $t = 2, \ldots, T-1$. This leads to the identification of the conditional quantile function Q_t.

For a non-Markovian process, Hu et al. (2018) consider the canonical model of earnings dynamics developed in the 1970s and 1980s, which includes a random walk permanent component and an ARMA transitory component, with the un- derlying permanent and transitory unobservable shocks assumed to be i.i.d. but otherwise unspecified. The observed earnings Y_t in year t is decomposed into two independent components:

$$Y_t = U_t + V_t. \tag{5.10}$$

The first one, U_t is the permanent component which follows the unit root process:

$$U_t = U_{t-1} + u_t, \tag{5.11}$$

where u_t is the permanent shock. The second one, V_t is the transitory component which follows the ARMA(p,q) process:

$$V_t = \rho_{t,1}V_{t-1} + \rho_{t,2}V_{t-2} \cdots + \rho_{t,p}V_{t-p} + G_t\left(\epsilon_t, \epsilon_{t-1}, \cdots, \epsilon_{t-q}\right), \tag{5.12}$$

Define $Y_{t+1} = Y_{t+1} - Y_t$. For an ARMA(1,1) process, they show that the AR coefficient can be directly estimated up to a normalization as follows:

$$\rho_{t+1}\frac{1-\rho_{t+2}}{1-\rho_{t+1}} = \frac{cov\left(\Delta Y_{t+2}, Y_{t-1}\right)}{cov\left(\Delta Y_{t+1}, Y_{t-1}\right)}$$

Furthermore, they show

$$\begin{aligned} Y_t &= V_t + U_t \\ \frac{\Delta Y_{t+2}}{\rho_{t+2}-1} - \Delta Y_{t+1} &= V_t + \frac{G_{t+2}(\epsilon_{t+2}, \epsilon_{t+1}) + u_{t+2}}{\rho_{t+2}-1} - u_{t+1} \end{aligned}$$

The Kotlarski's identity then implies that the distribution of V_t can be identified with a closed-form. In the end, they show that the joint distribution of $\{Y_t\}_{t=1,\ldots,T\geqslant 3}$ uniquely determines distributions of latent variables u_t, ϵ_t, U_t, and V_t. Although this model imposes parametric restrictions, such as random work and ARMA structures, the distributions of shocks are left nonparametrically. The identification of such semiparametric dynamic models with latent variables is complimentary to the existing results, which all heavily rely on a Markovian property of the dynamic structure. Hu et al. (2018) is the first to show the identification of a non-Markovian process with latent variables. Their identification results open the possibility of identification of more general non-Markovian processes with latent variables, which could have broad applications in empirical research.

6

Unobservables in reduced-form and structural econometrics

6.1 Fixed effects in panel data models

Evdokimov (2010) considers a panel data model as follows: for individual i in period t

$$Y_{it} = g\left(X_{it}, \alpha_i\right) + \xi_{it}, \tag{6.1}$$

where X_{it} is a explanatory variable, Y_{it} is the dependent variable, ξ_{it} is an independent error term, and α_i represents fixed effects. In order to use Kotlarski's identity, he considers the event where $\{X_{i1} = X_{i2} = x\}$ for two periods of data to obtain

$$\begin{aligned} Y_{i1} &= g\left(x, \alpha_i\right) + \xi_{i1}, \\ Y_{i2} &= g\left(x, \alpha_i\right) + \xi_{i2}. \end{aligned} \tag{6.2}$$

Under the assumption that ξ_{it} and α_i are independent conditional on X_{it}, the paper is able to identify the distributions of $g\left(x, \alpha_i\right)$, ξ_{i1} and ξ_{i2} conditional on $\{X_{i1} = X_{i2} = x\}$. That means this identification strategy relies on the static aspect of the panel data model. Assuming that ξ_{i1} is independent of X_{i2} conditional X_{i1}, he then identifies $f\left(\xi_{i1} | X_{i1} = x\right)$, and similarly $f\left(\xi_{i2} | X_{i2} = x\right)$, which leads to identification of the regression function $g\left(x, \alpha_i\right)$ under a normalization assumption.

Shiu and Hu (2013) consider a dynamic panel data model

$$Y_{it} = g\left(X_{it}, Y_{i,t-1}, U_{it}, \xi_{it}\right), \tag{6.3}$$

where U_{it} is a time-varying unobserved heterogeneity or an unobserved covariate, and ξ_{it} is a random shock independent of $(X_{it}, Y_{i,t-1}, U_{it})$. They impose the following Markov-type assumption

$$X_{i,t+1} \perp \left(Y_{it}, Y_{i,t-1}, X_{i,t-1}\right) | \left(X_{it}, U_{it}\right) \tag{6.4}$$

to obtain

$$f_{X_{i,t+1}, Y_{it}, X_{it}, Y_{i,t-1}, X_{i,t-1}} = \int f_{X_{i,t+1}|X_{it}, U_{it}} f_{Y_{it}|X_{it}, Y_{i,t-1}, U_{it}} f_{X_{it}, Y_{i,t-1}, X_{i,t-1}, U_{it}} dU_{it}. \tag{6.5}$$

Notice that the dependent variable Y_{it} may represent a discrete choice. With a binary Y_{it} and fixed $(X_{it}, Y_{i,t-1})$, equation (6.5) implies a 2.1-measurement model. Their identification results require users to carefully check conditional independence assumptions in their model because the conditional independence assumption in equation (6.4) is not directly motivated by economic structure.

Freyberger (2012) embeds a factor structure into a panel data model as follows:

$$Y_{it} = g\left(X_{it}, \alpha_i' F_t + \xi_{it}\right), \tag{6.6}$$

where $\alpha_i \in \mathbb{R}^m$ stands for a vector of unobserved individual effects and F_t is a vector of constants. Under the assumption that ξ_{it} for $t = 1, 2, \ldots, T$ are jointly independent conditional on α_i and $X_i = (X_{i1}, X_{i2}, \ldots, X_{iT})$, he obtains

$$Y_{i1} \perp Y_{i2} \perp \ldots \perp Y_{iT} | \left(\alpha_i, X_i\right), \tag{6.7}$$

which forms a 3-measurement model. A useful feature of this model is that the factor structure $\alpha_i' F_t$ provides a more specific identification of the model with a multi-dimensional individual effects α_i than a general argument as in Theorem 3.

Sasaki (2015) considers a dynamic panel with unobserved heterogeneity α_i and sample attrition as follows:

$$\begin{aligned} Y_{it} &= g\left(Y_{i,t-1}, \alpha_i, \xi_{it}\right) \\ D_{it} &= h\left(Y_{it}, \alpha_i, \eta_{it}\right) \\ Z_i &= \varsigma\left(\alpha_i, \epsilon_i\right) \end{aligned} \tag{6.8}$$

where Z_i is a noisy signal of α_i and $D_{it} \in \{0, 1\}$ is a binary indicator for attrition, i.e., Y_{it} is observed if $D_{it} = 1$. Under suitable restrictions on the error terms, the following conditional independence holds

$$Y_{i3} \perp Z_i \perp Y_{i1} \ \mid \ \left(\alpha_i, Y_2 = y_2, D_2 = D_1 = 1\right). \tag{6.9}$$

In the case where α_i is discrete, the model is identified using the results in Theorem 2. Sasaki (2015) also extends this identification result to more general settings.

6.2 Misclassification in treatment effect models

Lewbel (2007) considers identification and estimation of the effect of a mismeasured binary regressor in a nonparametric or semiparametric regression, or the conditional average effect of a binary treatment or policy on some outcome where treatment may be misclassified. Let's consider a simplified version of the model without covariates. Define Y as the outcome variable, $T*$ is the true binary treatment, and V is an exogenous variable. The research observes a mismeasured binary treatment T instead of the true treatment. The key assumption is that the variable V only affects the true treatment probability but not the treatment effect nor misclassification probability. Lewbel (2007) imposed restrictions on the support of V, such as at least three values in the support, and the relationship between V and an explicit function of mislassification probability to avoid directly imposing conditional independence. In fact, the intuition is better captured by a 3-measurement model satisfying

$$f(Y, T, |V) = \sum_{T^* \in \{0,1\}} f(Y|T^*) f(T|T^*) f(T^*|V)$$

This is similar to the mean regression case in Mahajan (2006), where the key relationship can be described as

$$E(Y|T,V) = \frac{1}{f(T|V)} \sum_{T^* \in \{0,1\}} E(Y|T^*)f(T|T^*)f(T^*|V).$$

In fact, Hui and Walter (1980) considers the same theoretical framework with different interpretation, where T^* is the true binary indictor of whether an individual has certain disease, Y and T are two separate diagnostic tests' binary outcome, and V stands for different subpopulations. In this case, the conditional independence assumptions seem very reasonable, i.e.,

$$Y \perp T \perp V \mid T^*.$$

Let $Y, T, V \in \{0,1\}$. They further specify the likelihood function as follows:

$$\begin{aligned} & f(Y,T,|V) \qquad (6.10) \\ = & \sum_{T^* \in \{0,1\}} f(Y|T^*)f(T|T^*)f(T^*|V) \\ \equiv & [f_{Y|T^*}(1|0)]^Y [1 - f_{Y|T^*}(1|0)]^{1-Y} [f_{T|T^*}(1|0)]^T [1 - f_{T|T^*}(1|0)]^{1-T} [1 - f_{T^*|V}(1|v)] \\ + & [1 - f_{Y|T^*}(0|1)]^Y [f_{Y|T^*}(0|1)]^{1-Y} [1 - f_{T|T^*}(0|1)]^T [f_{T|T^*}(0|1)]^{1-T} f_{T^*|V}(1|v) \end{aligned}$$

Notice that $f_{T^*|V}(1|v)$ stands for the probability of a diseased individual in subpopulation $V = v$, $f_{Y|T^*}(1|0)$ is the false positive rate of test Y, and $f_{Y|T^*}(0|1)$ is the false negative rate of test Y. Similarly, $f_{T|T^*}(1|0)$ and $f_{T|T^*}(0|1)$ are the false positive and false negative rates of test T.

Using their notation, we define

$$p_{gij} = f(Y = i, T = j|V = g)$$

with for $g, i, j \in \{0,1\}$ (their paper uses $g, i, j \in \{1,2\}$) the true probability of test outcomes i in test Y and j in test T. Let the notation "." in $p_{g.j}$ or $p_{gi.}$ denote summation over an index. Hui and Walter (1980) first show that this identification problem can be reduced to solving a quadratic equation and provide closed-form solutions as follows:
The false positive rates are

$$f_{Y|T^*}(1|0) = (p_{00.}p_{1.0} - p_{0.0}p_{10.} + p_{100} - p_{000} + D)/2E_0$$

$$f_{T|T^*}(1|0) = (p_{10.}p_{0.0} - p_{1.0}p_{00.} + p_{100} - p_{000} + D)/2E_1$$

The false negative rates are

$$f_{Y|T^*}(0|1) = (p_{0.1}p_{11.} - p_{01.}p_{1.1} + p_{011} - p_{111} + D)/2E_0$$

$$f_{T|T^*}(0|1) = (p_{1.1}p_{01.} - p_{11.}p_{0.1} + p_{011} - p_{111} + D)/2E_1$$

The probability of being diseased is in subpopulation $g \in \{0,1\}$ is

$$f_{T^*|V}(1|g) = \frac{1}{2} + \{p_{g0.}(p_{0.0} - p_{1.0}) + p_{g.0}(p_{00.} - p_{10.}) + p_{100} - p_{000}\}/2D$$

where

$$E_0 = p_{1.0} - p_{0.0}$$

$$E_1 = p_{10.} - p_{00.}$$

$$D = \pm\{(p_{00.}p_{1.0} - p_{10.}p_{0.0} + p_{000} - p_{100})^2 - 4(p_{00.} - p_{10.})(p_{000}p_{1.0} - p_{100}p_{0.0})\}^{1/2}.$$

The sign of D is not determined because they don't impose the ordering assumption summarized in Hu (2008).

6.3 Dynamic discrete choice with unobserved state variables

Hu and Shum (2012) show that the transition kernel of a Markov process $\{W_t, X_t^*\}$ can be uniquely determined by the joint distribution of four periods of data $\{W_{t+1}, W_t, W_{t-1}, W_{t-2}\}$. This result can be directly applied to identification of dynamic discrete choice model with unobserved state variables. Such a Markov process may characterize the optimal path of the decision and the state variables in Markov dynamic optimization problems. Let $W_t = (Y_t, M_t)$, where Y_t is the agent's choice in period t, and M_t denotes the period-t observed state variable, while X_t^* is the unobserved state variable. For Markovian dynamic optimization models, the transition kernel can be decomposed as follows:

$$f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} = f_{Y_t|M_t,X_t^*} f_{M_t,X_t^*|Y_{t-1},M_{t-1},X_{t-1}^*}. \tag{6.11}$$

The first term on the right hand side is the conditional choice probability for the agent's optimal choice in period t. The second term is the joint law of motion of the observed and unobserved state variables. As shown in Hotz and Miller (1993) , the identified Markov law of motion may be a crucial input in the estimation of Markovian dynamic models. One advantage of this conditional choice probability approach is that a parametric specification of the model leads to a parametric GMM estimator. That implies an estimator for a dynamic discrete choice model with unobserved state variables, where one can identify the Markov transition kernel containing unobserved state variables, and then apply the conditional choice probability estimator to estimate the model primitives. Hu and Shum (2013) extend this result to dynamic games with unobserved state variables.

Although the nonparametric identification is quite general, it is still useful for empirical research to provide a relatively simple estimator for a particular specification of the model as long as such a specification can capture the key economic causality in the model. Given the difficulty in the estimation of dynamic discrete choice models with unobserved state variables, Hu and Sasaki (forthcoming, 2017a) consider a popular parametric specification of the model and provide a closed-form estimator for the inputs of the conditional choice probability estimator. Let Y_t denote firms' exit decisions based on their productivity X_t^* and other covariates M_t. The law of motion of the productivity is

$$X_t^* = \alpha^d + \beta^d X_{t-1}^* + \eta_t^d \text{ if } Y_{t-1} = d \in \{0, 1\}. \tag{6.12}$$

In addition, they use residuals from the production function as a proxy X_t for latent X_t^* satisfying

$$X_t = X_t^* + \epsilon_t. \tag{6.13}$$

Therefore, they obtain

$$X_{t+1} = \alpha^d + \beta^d X_t^* + \eta_{t+1}^d + \epsilon_{t+1} \tag{6.14}$$

Under the assumption that the error terms η_t^d and ϵ_t are random shocks, they first estimate the coefficients $\left(\alpha^d, \beta^d\right)$ using other covariates M_t as instruments. The distribution of the error term ϵ_t can then be estimated using Kotlarski's identity. Furthermore, they are able to provide a closed-form expression for the conditional choice probability $\Pr\left(Y_t|X_t^*, M_t\right)$ as a function of observed distribution functions.

6.4 Multiple equilibria in incomplete information games

Xiao (forthcoming) considers a static simultaneous move game, in which player i for $i = 1, 2, \ldots, N$ chooses an action a_i from a choice set $\{0, 1, \ldots, K\}$. Let a_{-i} denote actions of the other players, i.e., $a_{-i} = \{a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N\}$. The player i's payoff is specified as

$$u_i\left(a_i, a_{-i}, \epsilon_i\right) = \pi_i\left(a_i, a_{-i}\right) + \epsilon_i\left(a_i\right), \tag{6.15}$$

where $\epsilon_i(k)$ for $k \in \{0, 1, \ldots, K\}$ is a choice-specific payoff shock for player i. The object of interest contains the payoff primitives and the equilibrium selection probability. Here we omit other observed state variables. These shocks $\epsilon_i(k)$ are assumed to be private information to player i, while the distribution of $\epsilon_i(k)$ is common knowledge to all the players. A widely used assumption is that the payoff shocks $\epsilon_i(k)$ are independent across all the actions k and all the players i. Let $\Pr(a_{-i})$ be player i's belief of other player's actions. The expected payoff of player i from choosing action a_i is then

$$\sum_{a_{-i}} \pi_i\left(a_i, a_{-i}\right) \Pr\left(a_{-i}\right) + \epsilon_i\left(a_i\right) \equiv \Pi_i\left(a_i\right) + \epsilon_i\left(a_i\right) \tag{6.16}$$

The Bayesian Nash Equilibrium is defined as a set of choice probabilities $\Pr(a_i)$ such that

$$\Pr\left(a_i = k\right) = \Pr\left(\left\{\Pi_i(k) + \epsilon_i(k) > \max_{j \neq k} \Pi_i(j) + \epsilon_i(j)\right\}\right). \tag{6.17}$$

The existence of such an equilibrium is guaranteed by the Brouwer's fixed point theorem. Given an equilibrium, the mapping between the choice probabilities and the expected payoff function has also be established by Hotz and Miller (1993) .

However, multiple equilibria may exist for this game, which means the observed choice probabilities are a mixture from different equilibria. Let e^* denote the index of equilibria. Under each equilibrium e^*, the players' actions a_i are independent because of the independence assumption of private information, i.e.,

$$a_1 \perp a_2 \perp \ldots \perp a_N | e^*. \tag{6.18}$$

Therefore, the observed correlation among the actions contains information on multiple equilibria. If the support of actions is larger than that of e^*, one can use three players' actions as three measurements for e^*. Otherwise, if there are enough players, one can partition the players into three groups and use the group actions as the three measurements. Comparing with many existing studies on multiple equilibria, using the results for measurement error models makes the nonparametric identification in Xiao (forthcoming) more transparent on why and where the assumptions are imposed and what can and can not be identified.

6.5 Two-sided matching models

Agarwal and Diamond (2013) consider an economy containing n workers with characteristics (X_i, ε_i) and n firms described by $\left(Z_j, \eta_j\right)$ for $i, j = 1, 2, \ldots, n$. For example, wages offered by a firm is public information in Z_j or η_j. They assume that the observed characteristics X_i and Z_i are independent of other characteristics ε_i and η_j unobserved to researchers. A firm ranks workers by a human capital index as

$$v\left(X_i, \varepsilon_i\right) = h\left(X_i\right) + \varepsilon_i. \tag{6.19}$$

The workers' preference for firm j is described by

$$u\left(Z_j, \eta_j\right) = g\left(Z_j\right) + \eta_j. \tag{6.20}$$

The preferences on both sides are public information in the market. Researchers are interested in the preferences, including functions h, g, and distributions of ε_i and η_j.

A match is a set of pairs that show which firm hires which worker. The observed matches are assumed as outcomes of a pairwise stable equilibrium, where no two agents on opposite sides of the market prefer each other over their matched partners. When the numbers of firms and workers are both large, it can be shown that in the unique pairwise stable equilibrium the firm with the q-th quantile position of preference value, i.e., $F_U\left(u\left(Z_j, \eta_j\right)\right) = q$ is matched with the worker with the q-th quantile position of the human capital index, i.e., $F_V\left(v\left(X_i, \varepsilon_i\right)\right) = q$, where F_U and F_V are cumulative distribution functions of u and v.

The joint distribution of (X, Z) from observed pairs then satisfies

$$f\left(X, Z\right) = \int_0^1 f\left(X|q\right) f\left(Z|q\right) dq, \tag{6.21}$$

This forms a 2-measurement model. Under the specification of the preferences above, i.e.,

$$\begin{aligned} f\left(X|q\right) &= f_\varepsilon\left(F_V^{-1}\left(q\right) - h(X)\right) \\ f\left(Z|q\right) &= f_\eta\left(F_U^{-1}\left(q\right) - g(Z)\right), \end{aligned} \tag{6.22}$$

the functions h and g can be identified up to a monotone transformation. The intuition is that under suitable conditions if two workers with different characteristics x_1 and x_2 are hired by firms with the same characteristics, i.e., $f_{Z|X}\left(z|x_1\right) = f_{Z|X}\left(z|x_2\right)$ for all z, then the two workers must have the same observed part of the human capital index, i.e., $h\left(x_1\right) = h\left(x_2\right)$. A similar argument also holds for function g. In order to further identify the model, Agarwal and Diamond (2013) consider many-to-one matching where one firm may have two or more identical slots for workers. In such a sample, they can observe the joint distribution of (X_1, X_2, Z), where (X_1, X_2) are observed characteristics of the two matched workers. Therefore, they obtain

$$f\left(X_1, X_2, Z\right) = \int_0^1 f\left(X_1|q\right) f\left(X_2|q\right) f\left(Z|q\right) dq. \tag{6.23}$$

This is a 3-measurement model, for which nonparametric identification is feasible under suitable conditions.

7

Retrospect and prospect

This manuscript reviews recent developments in nonparametric identification of measurement error models and their applications in microeconomic models with latent variables. The powerful identification results promote a close integration of microeconomic theory and econometric methodology, especially when latent variables are involved. With econometricians developing more application-oriented methodologies, we expect such an integration to deepen in the future research.

Besides the methodologies and the applications of measurement error models presented here, we expect this literature to advance further, with more important results. For example, the flexible nonclassical measurement error models may also provide new and convincing solutions to the endogeneity problem, a fundamental problem in econometrics. Presumably, a complete economic model should explain the causality among all the variables in the model. Endogeneity then occurs when some of the variables in the model are unobserved by the researcher. Nonclassical measurement error models may then be used to handle the unobservables, and therefore, solve the endogeneity problem under certain assumptions.

With more and more data available for researchers, we look forward to more extensive applications of the measurement error models. Given the nonparametric identification, nonparametric or semiparametric estimation of the models with latent variables may become easier than before. On the one hand, sample sizes will become much larger than before with the abundance of observations; on the other hand, researchers may observe more measurements of the latent variables. Therefore, we expect that the literature of measurement error models and their applications will keep thriving.

Bibliography

Agarwal, Nikhil and William Diamond, "Identification and Estimation in Two-Sided Matching Markets," Technical Report, Cowles foundation 2013.

Allman, Elizabeth S, Catherine Matias, and John A Rhodes, "Identifiability of Parameters in Latent Structure Models with Many Observed Variables," *The Annals of Statistics*, 2009, pp. 3099–3132.

An, Yonghong, Michael R Baye, Yingyao Hu, John Morgan, and Matthew Shum, "Identification and Estimation of Online Price Competition with an Unknown Number of Firms," 2015. forthcoming, *Journal of Applied Econometrics.*

_ , Yingyao Hu, and Matthew Shum, "Estimating First-Price Auctions with an Unknown Number of Bidders: A Misclassification Approach," *Journal of Econometrics*, 2010, *157*, 328–341.

Arellano, Manuel, Richard Blundell, and Stephane Bonhomme, "Household Earnings and Consumption: A Nonlinear Framework," 2014.

Bollinger, Christopher, "Bounding Mean Regressions When a Binary Regressor is Mismeasured," *Journal of Econometrics*, 1996, *73*, 387–399.

Bonhomme, Stéphane, Koen Jochmans, and Jean-Marc Robin, "Estimating Multivariate Latent-Structure Models," 2015. forthcoming, *Annals of Statistics.*

_ , _ , and _ , "Non-parametric Estimation of Finite Mixtures from Repeated Measurements," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016, *78* (1), 211–229.

Bound, John, Charles Brown, and Nancy Mathiowetz, "Measurement Error in Survey Data," *Handbook of econometrics*, 2001, *5*, 3705–3843.

Carroll, Raymond J, David Ruppert, Leonard A Stefanski, and Ciprian M Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, CRC press, 2012.

_ , Xiaohong Chen, and Yingyao Hu, "Identification and Estimation of Nonlinear Models Using Two Samples with Nonclassical Measurement Errors," *Journal of nonparametric statistics*, 2010, *22* (4), 379–399.

Chen, Xiaohong, "Large Sample Sieve Estimation of Semi-nonparametric Models. The Handbook of Econometrics, JJ Heckman and EE Leamer (eds.), 6B," 2007.

_ **and Xiaotong Shen**, "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 1998, pp. 289–314.

_ **, Han Hong, and Denis Nekipelov**, "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 2011, *49* (4), 901–937.

_ **,** _ **, and Elie Tamer**, "Measurement Error Models with Auxiliary Data," *The Review of Economic Studies*, 2005, *72* (2), 343–366.

_ **, Yingyao Hu, and Arthur Lewbel**, "A Note on the Closed-form Identification of Regression Models with a Mismeasured Binary Regressor," *Statistics and Probability Letters*, 2008, *78*, 1473–1479.

_ **,** _ **, and** _ , "Nonparametric Identification and Estimation of Nonclassical Errors-in-variables Models without Additional Information," *Statistica Sinica*, 2009, *19*, 949–968.

Cunha, Flavio, James J Heckman, and Susanne M Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

Evdokimov, Kirill, "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," 2010. Working paper, Princeton University.

Feng, Shuaizhang and Yingyao Hu, "Misclassification Errors and the Underestimation of the US Unemployment Rates," *The American Economic Review*, 2013, *103* (2), 1054–1070.

Freyberger, Joachim, "Nonparametric Panel Data Models with Interactive Fixed Effects," Technical Report, University of Wisconsin, Madison 2012.

Fuller, Wayne A, *Measurement Error Models*, Vol. 305, John Wiley & Sons, 2009.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong, "Optimal Nonparametric Estimation of First-price Auctions," *Econometrica*, 2000, *68* (3), 525–574.

Hotz, V. Joseph and Robert A. Miller, "Conditional Choice Probabilties and the Estimation of Dynamic Models," *Review of Economic Studies*, 1993, *60*, 497–529.

Hu, Yingyao, "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 2008, *144*, 27–61.

_ , "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics," *Journal of Econometrics*, 2017, *200*, 154–168.

_ **and Matthew Shum**, "Nonparametric Identification of Dynamic Models with Unobserved State Variables," *Journal of Econometrics*, 2012, *171* (1), 32–44.

_ **and** _ , "Identifying Dynamic Games with Serially-correlated Unobservables," *Advances in Econometrics*, 2013, *31*, 97–113.

_ **and Susanne Schennach**, "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 2008, *76*, 195–216.

_ **and Yuya Sasaki**, "Closed-Form Estimation of Nonparametric Models with Non-Classical Measurement Errors," *Journal of Econometrics*, 2015, *185* (2), 392–408.

_ **and** _ , "Closed-Form Identification of Dynamic Discrete Choice Models with Proxies for Unobserved State Variables," *Econometric Theory*, forthcoming, 2017.

_ **and** _ , "Identification of Paired Nonseparable Measurement Error Models," *Econometric Theory*, forthcoming, 2017.

_ **, David McAdams, and Matthew Shum**, "Nonparametric Identification of First-price Auction Models with Non-Separable Unobserved Heterogeneity," *Journal of Econometrics*, 2013, *174*, 186–193.

_ **, Robert Moffitt, and Yuya Sasaki**, "Semiparametric Estimation of the Canonical Permanent-Transitory Model of Earnings Dynamics," *working paper*, 2018.

_ **, Yutaka Kayaba, and Matthew Shum**, "Nonparametric Learning Rules from Bandit Experiments: The Eyes Have It!," *Games and Economic Behavior*, 2013, *81*, 215–231.

Hui, S.L. and S.D. Walter, "Estimating the Error Rates of Diagnostic Tests," *Biometrics*, 1980, *36*, 167–171.

Kasahara, Hiroyuki and Katsumi Shimotsu, "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 2009, *77* (1), 135–175.

Kotlarski, Ignacy, "On Pairs of Independent Random Variables Whose Product Follows the Gamma Distribution," *Biometrika*, 1965, pp. 289–294.

Krasnokutskaya, Elena, "Identification and Estimation in Procurement Auctions under Unobserved Auction Heterogeneity," *Review of Economic Studies*, 2011, *28*, 293–327.

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

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

Li, Tong and Quang Vuong, "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 1998, *65* (2), 139–165.

_ **, Isabelle Perrigne, and Quang Vuong**, "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics*, 2000, *98*, 129–161.

Mahajan, Aprajit, "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 2006, pp. 631–665.

Molinari, Francesca, "Partial Identification of Probability Distributions with Misclassified Data," *Journal of Econometrics*, 2008, *144*, 81–117.

Rao, B.L.S. Prakasa, *Identifiability in Stochastic Models: Characterization of Probability Distributions*, Academic Press, Inc., 1992.

Reiersøl, Olav, "Identifiability of a Linear Relation Between Variables Which are Subject to Error," *Econometrica*, 1950, pp. 375–389.

Sasaki, Yuya, "Heterogeneity and Selection in Dynamic Panel Data," *Journal of Econometrics*, 2015, *188*, 236–249.

Schennach, Susanne M, "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 2004, *72* (1), 33–75.

_ , "Nonparametric Regression in the Presence of Measurement Error," *Econometric Theory*, 2004, *20* (06), 1046–1093.

_ , "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 2007, *75* (1), 201–239.

_ , "Regressions with Berkson Errors in Covariates: A Nonparametric Approach," *The Annals of Statistics*, 2013, *41* (3), 1642–1668.

_ , "Recent Advances in the Measurement Error Literature," *Annual Review of Economics*, 2016, *8*, 341–377.

_ **and Yingyao Hu**, "Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models without Side Information," *Journal of the American Statistical Association*, 2013, *108* (501), 177–186.

Shen, Xiaotong, "On Methods of Sieves and Penalization," *The Annals of Statistics*, 1997, *25* (6), 2555–2591.

Shiu, Ji-Liang and Yingyao Hu, "Identification and Estimation of Nonlinear Dynamic Panel Data Models with Unobserved Covariates," *Journal of Econometrics*, 2013, *175* (2), 116–131.

Wansbeek, Thomas J. and Erik Meijer, *Measurement Error and Latent Variables in Econometrics* Advanced textbooks in economics, Elsevier, 2000.

Wilhelm, Daniel, "Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error," Technical Report, University College London 2013.

Xiao, Ruli, "Identification and Estimation of Incomplete Information Games with Multiple Equilibria," *Journal of Econometrics*, forthcoming.

Xin, Yi, "Asymmetric Information, Reputation, and Welfare in Online Credit Markets," *Working paper, Johns Hopkins University*, 2018.