Journal of Econometrics 200 (2017) 154–168

Contents lists available at ScienceDirect

Journal of Econometrics

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

# The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics[☆]

Yingyao Hu

*Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, United States*

ARTICLE INFO

*Article history:*
Available online 27 June 2017

*JEL classification:*
C01
C14
C22
C23
C26
C32
C33
C36
C57
C70
C78
D20
D31
D44
D83
D90
E24
I20
J21
J24
J60
L10

*Keywords:*
Measurement error model
Errors-in-variables
Latent variable
Unobserved heterogeneity
Unobserved state variable
Mixture model
Hidden Markov model
Dynamic discrete choice
Nonparametric identification
Conditional independence
Endogeneity
Instrument
Type
Unemployment rates
IPV auction
Multiple equilibria

ABSTRACT

This paper reviews the recent developments in nonparametric identification of measurement error models and their applications in applied microeconomics, in particular, in empirical industrial organization and labor economics. Measurement error models describe mappings from a latent distribution to an observed distribution. The identification and estimation of measurement error models focus on how to obtain the latent distribution and the measurement error distribution from the observed distribution. Such a framework is suitable for many microeconomic models with latent variables, such as models with unobserved heterogeneity or unobserved state variables and panel data models with fixed effects. Recent developments in measurement error models allow very flexible specification of the latent distribution and the measurement error distribution. These developments greatly broaden economic applications of measurement error models. This paper provides an accessible introduction of these technical results to empirical researchers so as to expand applications of measurement error models.


[☆] This paper was previously circulated under the title "Microeconomic models with latent variables: Applications of measurement error models in empirical industrial organization and labor economics".

*E-mail address:* yhu@jhu.edu.

http://dx.doi.org/10.1016/j.jeconom.2017.06.002

Incomplete information game
Belief
Learning model
Fixed effects
Panel data model
Cognitive and non-cognitive skills
Matching
Income dynamics

## 1. Introduction

This paper provides a concise introduction of recent developments in nonparametric identification of measurement error models and intends to invite empirical researchers to use these new results for measurement error models in the identification and estimation of microeconomic models with latent variables.

Measurement error models describe the relationship between latent variables, which are not observed in the data, and their measurements. Researchers only observe the measurements instead of the latent variables in the data. The goal is to identify the distribution of the latent variables and also the distribution of the measurement errors, which are defined as the difference between the latent variables and their measurements. In general, the parameter of interest is the joint distribution of the latent variables and their measurements, which can be used to describe the relationship between observables and unobservables in economic models.

This paper 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. I 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. I call this a 2.1-measurement model, where I use 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.

This paper also provides a brief introduction of empirical applications using these measurement error models. These studies cover auction models with unobserved heterogeneity, multiple equilibria in games, dynamic learning models with latent beliefs, misreporting errors in estimation of unemployment rates, dynamic models with unobserved state variables, fixed effects in panel data models, cognitive and non-cognitive skill formation, two-sided matching models, and income dynamics. This paper intends to be concise, informative, and heuristic. I refer to Wansbeek and Meijer (2000), Bound et al. (2001), Chen et al. (2011), Carroll et al. (2012), and Schennach (2016) for more complete reviews.

This paper is organized as follows. Section 2 introduces the nonparametric identification results for measurement error models, together with a few semiparametric and nonparametric estimators. Section 3 describes a few applications of the nonparametric identification results. Section 4 summarizes the paper.

## 2. Nonparametric identification of measurement error models

We start our discussion with a general definition of measurement. 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

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

where $card\left(\mathcal{X}\right)$ 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.

### 2.1. 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 $x \in \mathcal{X}$

$$f_X(x) = \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*, \tag{1}$$