

# Microeconomic models with latent variables: applications of measurement error models in empirical industrial organization and labor economics

---

Yingyao Hu

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP03/15



An ESRC Research Centre

# Microeconomic Models with Latent Variables: Applications of Measurement Error Models in Empirical Industrial Organization and Labor Economics

Yingyao Hu\*  
Johns Hopkins University

January 23, 2015

## Abstract

This paper reviews 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 may be 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.

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, incomplete information game, belief, learning model, fixed effect, panel data model, cognitive and noncognitive skills, matching, income dynamics.*

---

\*I am grateful to Tom Wansbeck for encouraging me to write this paper. I also thank Yonghong An, Yajing Jiang, Jian Ni, Katheryn Russ, Yuya Sasaki, Ruli Xiao, and Yi Xin for suggestions and comments. All errors are mine. Contact information: Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218. Tel: 410-516-7610. Email: yhu@jhu.edu.

# 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 "measurements" can be simply observed variables with an informative support. The measurement error distribution contains the information on 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 may be nonparametrically 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 symmetric roles which each measurement may play. In particular, in many cases, it doesn't 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 non-

parametric instrumental approach. In that sense, a nonparametric difference-in-difference 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 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, dynamic learning models with latent beliefs, fixed effects in panel data models, misreporting errors in estimation of unemployment rates, cognitive and noncognitive skill formation, and two-sided matching models. This paper intends to be concise, informative, and heuristic. I refer to Bound, Brown, and Mathiowetz (2001), Chen, Hong, and Nekipelov (2011), Schenach (2012), and Carroll, Ruppert, Stefanski, and Crainiceanu (2012) for more complete reviews.

This paper is organized as follows. Section 2 introduces the nonparametric identification results for measurement error models. 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 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 the number of possible values in  $\mathcal{X}$  is larger than or equal to that in  $\mathcal{X}^*$ .*

When  $X$  is continuous, the support condition in Definition 1 is not restrictive whether  $X^*$  is discrete or continuous. When  $X$  is discrete, the support condition implies that  $X$  can only be a measurement of a discrete random variable with a smaller or equal number of possible values. In particular, we don't consider a discrete variable as a measurement of a continuous variable.

### 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 may 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  $x \in \mathcal{X}$

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

when  $X^*$  is continuous, and for all  $x \in \mathcal{X} = \{x_1, x_2, \dots, x_L\}$

$$f_X(x) = \sum_{x^* \in \mathcal{X}^*} f_{X|X^*}(x|x^*)f_{X^*}(x^*), \quad (2)$$

when  $X^*$  is discrete with support  $\mathcal{X}^* = \{x_1^*, x_2^*, \dots, x_K^*\}$ . The definition of measurement requires  $L \geq K$ . We omit arguments of the functions when it doesn't cause any confusion. This general framework may be used to describe a wide range of economic relationships between observables and unobservables in the sense that the latent variable  $X^*$  may 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

$$\begin{aligned} \vec{p}_X &= [f_X(x_1), f_X(x_2), \dots, f_X(x_L)]^T \\ \vec{p}_{X^*} &= [f_{X^*}(x_1^*), f_{X^*}(x_2^*), \dots, f_{X^*}(x_K^*)]^T \\ M_{X|X^*} &= [f_{X|X^*}(x_l|x_k^*)]_{l=1,2,\dots,L;k=1,2,\dots,K} \end{aligned} \quad (3)$$

The notation  $M^T$  stands for the transpose of  $M$ . Note that  $\vec{p}_X$ ,  $\vec{p}_{X^*}$ , and  $M_{X|X^*}$  contain the same information as distributions  $f_X$ ,  $f_{X^*}$ , and  $f_{X|X^*}$ , respectively. Equation (2) is then equivalent to

$$\vec{p}_X = M_{X|X^*} \vec{p}_{X^*}. \quad (4)$$

The matrix  $M_{X|X^*}$  describes the linear transformation from  $\mathbb{R}^K$ , a vector space containing  $\vec{p}_{X^*}$ , to  $\mathbb{R}^L$ , a vector space containing  $\vec{p}_X$ . 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  $\vec{p}_{X^*}^a$  and  $\vec{p}_{X^*}^b$  are observationally equivalent, i.e.,

$$\vec{p}_X = M_{X|X^*} \vec{p}_{X^*}^a = M_{X|X^*} \vec{p}_{X^*}^b, \quad (5)$$

then the two distributions are the same, i.e.,  $\vec{p}_{X^*}^a = \vec{p}_{X^*}^b$ . Let  $h = \vec{p}_{X^*}^a - \vec{p}_{X^*}^b$ . Equation (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 the matrix  $M_{X|X^*}$  has a full rank, i.e.,  $\text{Rank}(M_{X|X^*}) = K$ . 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  $\mathcal{L}_{bnd}^1(\mathcal{X}^*)$  as the set of bounded and integrable functions defined on  $\mathcal{X}^*$ , i.e.,<sup>1</sup>

$$\mathcal{L}_{bnd}^1(\mathcal{X}^*) = \left\{ h : \int_{\mathcal{X}^*} |h(x^*)| dx^* < \infty \text{ and } \sup_{x^* \in \mathcal{X}^*} |h(x^*)| < \infty \right\}. \quad (6)$$

The linear operator may be defined as

$$\begin{aligned} L_{X|X^*} &: \mathcal{L}_{bnd}^1(\mathcal{X}^*) \rightarrow \mathcal{L}_{bnd}^1(\mathcal{X}) \\ (L_{X|X^*}h)(x) &= \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*)h(x^*)dx^*. \end{aligned} \quad (7)$$

Equation (1) is then equivalent to

$$f_X = L_{X|X^*}f_{X^*}. \quad (8)$$

Following a similar argument, we may show that a necessary condition for the identification of  $f_{X^*}$  in the functional space  $\mathcal{L}_{bnd}^1(\mathcal{X}^*)$  is that the linear operator  $L_{X|X^*}$  is injective, i.e.,  $L_{X|X^*}h = 0$  implies  $h = 0$  for any  $h \in \mathcal{L}_{bnd}^1(\mathcal{X}^*)$ . This condition can also be interpreted as completeness of conditional density  $f_{X|X^*}$  in  $\mathcal{L}_{bnd}^1(\mathcal{X}^*)$ . 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 may 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 may identify  $f_{X|X^*}$  from the validation sample and then identify  $f_{X^*}$  in the primary sample (see Chen, Hong, and Tamer (2005)). In this paper, we focus on methodologies using additional measurements in a single sample.

## 2.2 A 2-measurement model

Given very limited identification results which one may obtain from equations (1)-(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  $X \in \mathcal{X}$  and  $Z \in \mathcal{Z}$  of*

---

<sup>1</sup>We may also define the operator on other functional spaces containing  $f_{X^*}$ .

the latent variable  $X^* \in \mathcal{X}^*$  satisfying

$$X \perp Z \mid X^*, \quad (9)$$

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  $\mathcal{Z} = \{z_1, z_2, \dots, z_J\}$ , we may define

$$\begin{aligned} M_{X,Z} &= [f_{X,Z}(x_l, z_j)]_{l=1,2,\dots,L; j=1,2,\dots,J} \\ M_{Z|X^*} &= [f_{Z|X^*}(z_j|x_k^*)]_{j=1,2,\dots,J; k=1,2,\dots,K} \end{aligned} \quad (10)$$

and a diagonal matrix

$$D_{X^*} = \text{diag}\{f_{X^*}(x_1^*), f_{X^*}(x_2^*), \dots, f_{X^*}(x_K^*)\}. \quad (11)$$

Definition 1 implies that  $K \leq L$  and  $K \leq J$ . Equation (9) means

$$f_{X,Z}(x, z) = \sum_{x^* \in \mathcal{X}^*} f_{X|X^*}(x|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*), \quad (12)$$

which is equivalent to

$$M_{X,Z} = M_{X|X^*} D_{X^*} M_{Z|X^*}^T. \quad (13)$$

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.<sup>2</sup> But one element that may be identified from observed  $M_{X,Z}$  is the dimension  $K$  of the latent variable  $X^*$ , as elucidated in the following Lemma:

**Lemma 1** *Suppose that matrices  $M_{X|X^*}$  and  $M_{Z|X^*}$  have a full rank  $K$ . Then  $K = \text{rank}(M_{X,Z})$ .*

**Proof.** In the 2-measurement model, we have  $K \leq L$  and  $K \leq J$ . Therefore,  $\text{rank}(M_{X,Z}) \leq \min\{J, K, L\} = K$ . Since  $M_{X|X^*}$  and  $M_{Z|X^*}$  have a full rank  $K$  and  $D_{X^*}$  has rank  $K$  by definition of the support, we have  $\text{rank}(M_{X,Z}) = K$ . ■

Although point identification may not be feasible without further assumptions, we may still have some partial identification results. Consider a linear regression model with a

---

<sup>2</sup>If  $M_{X|X^*}$  and  $M_{Z|X^*}^T$  are lower and upper triangular matrices, respectively, point identification may be feasible through the so-called LU decomposition. In general, this is also related to the literature on non-negative matrix factorization, which focuses more on existence and approximation, instead of uniqueness.

discrete regressor  $X^*$  as follows:

$$\begin{aligned} Y &= X^* \beta + \eta \\ Y &\perp X \mid X^* \end{aligned} \tag{14}$$

where  $X^* \in \{0, 1\}$  and  $E[\eta \mid X^*] = 0$ . Here the dependent variable  $Y$  takes the place of  $Z$  as a measurement of  $X^*$ .<sup>3</sup> 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  $X^*$ , we may have

$$\begin{aligned} &|E[Y \mid X^* = 1] - E[Y \mid X^* = 0]| \\ &\geq |E[Y \mid X = 1] - E[Y \mid X = 0]|. \end{aligned} \tag{15}$$

That means the observed difference provides a lower bound on the parameter of interest  $|\beta|$ . More partial identification results may be found in Bollinger (1996) and Molinari (2008). Furthermore, the model may be point identified under the assumption that the regression error  $\eta$  is independent of the regressor  $X^*$ . (See Chen, Hu, and Lewbel (2009) for details.)

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

$$\begin{aligned} X &= X^* + \epsilon \\ Z &= X^* + \epsilon' \end{aligned} \tag{16}$$

where  $X^*$ ,  $\epsilon$ , and  $\epsilon'$  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  $X^*$  may be written as a closed-form function of the observed distribution  $f_{X,Z}$ . Define  $\phi_{X^*}(t) = E[e^{itX^*}]$  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

$$\begin{aligned} f_{X^*}(x^*) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ix^*t} \phi_{X^*}(t) dt \\ \phi_{X^*}(t) &= \exp \left[ \int_0^t \frac{E[X e^{isZ}]}{E[e^{isZ}]} ds \right]. \end{aligned} \tag{17}$$

This is the so-called Kotlarski's identity (See Kotlarski (1965) and Rao (1992)). This result has been used in many empirical and theoretical studies, such as Li and Vuong (1998), Li, Perrigne, and Vuong (2000), Krasnokutskaya (2011), Schennach (2004a), and Evdokimov (2010).

---

<sup>3</sup>We follow the routine to use  $Y$  to denote a dependent variable instead of  $Z$ .



The intuition of Kotlarski's identity is that the variance of  $X^*$  is revealed by the covariance of  $X$  and  $Z$ , i.e.,  $\text{var}(X^*) = \text{cov}(X, Z)$ . Therefore, the higher order moments between  $X$  and  $Z$  may 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 may also apply to an extended model as follows:

$$\begin{aligned} X &= X^*\beta + \epsilon \\ Z &= X^* + \epsilon' \end{aligned} \tag{18}$$

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

$$\frac{\text{cov}(X, Z)}{\text{var}(Z)} \leq \beta \leq \frac{\text{var}(X)}{\text{cov}(X, Z)}. \tag{19}$$

However, these bounds may not be tight because the joint independence of  $X^*$ ,  $\epsilon$ , and  $\epsilon'$  may lead to point identification of  $\beta$ . Reiersøl (1950) shows that such point identification is feasible when  $X^*$  is not normally distributed. A more general extension is to consider

$$\begin{aligned} X &= g(X^*) + \epsilon \\ Z &= X^* + \epsilon', \end{aligned} \tag{20}$$

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 model in equation (20) is very close to a nonparametric regression model with a classical measurement error, except that the regression error  $\epsilon$  needs to be independent of the regressor  $X^*$ .

## 2.3 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:<sup>4</sup>

---

<sup>4</sup>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.

**Definition 3** A 2.1-measurement model contains two measurements  $X \in \mathcal{X}$  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

$$X \perp Y \perp Z \mid X^*, \quad (21)$$

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

### 2.3.1 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 1** The two measurements  $X$  and  $Z$  and the latent variable  $X^*$  share the same support  $\mathcal{X}^* = \{x_1^*, x_2^*, \dots, x_K^*\}$ .

This condition is not restrictive because the number of possible values in  $\mathcal{X}^*$  may be identified, as shown in Lemma 1, and one can always transform a discrete variable into one with less possible values.

The conditional independence in equation (21) implies<sup>5</sup>

$$f_{X,Y,Z}(x, y, z) = \sum_{x^* \in \mathcal{X}^*} f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*). \quad (22)$$

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

$$\begin{aligned} M_{X,y,Z} &= [f_{X,Y,Z}(x_i, y, z_j)]_{i=1,2,\dots,K; j=1,2,\dots,K} \\ D_{y|X^*} &= \text{diag} \{f_{Y|X^*}(y|x_1^*), f_{Y|X^*}(y|x_2^*), \dots, f_{Y|X^*}(y|x_K^*)\}. \end{aligned} \quad (23)$$

Equation (22) is then equivalent to

$$M_{X,y,Z} = M_{X|X^*} D_{y|X^*} D_{X^*} M_{Z|X^*}^T. \quad (24)$$

Next, we assume

**Assumption 2** Matrix  $M_{X,Z}$  has a full rank  $K$ .

Equation (13) then implies  $M_{X|X^*}$ ,  $D_{X^*}$ , and  $M_{Z|X^*}$  all have a full rank. We may then eliminate  $D_{X^*} M_{Z|X^*}^T$  to obtain

$$M_{X,y,Z} M_{X,Z}^{-1} = M_{X|X^*} D_{y|X^*} M_{X|X^*}^{-1}. \quad (25)$$

---

<sup>5</sup>Hui and Walter (1980) first consider the case where  $X^*$  is binary and show that this identification problem can be reduced to solving a quadratic equation.

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

**Assumption 4** *One of the following conditions holds:*

- 1)  $f_{X|X^*}(x_1|x_j^*) > f_{X|X^*}(x_1|x_{j+1}^*)$ ;
- 2)  $f_{X|X^*}(x^*|x^*) > f_{X|X^*}(\tilde{x}^*|x^*)$  for any  $\tilde{x}^* \neq x^* \in \mathcal{X}^*$ ;
- 3) *There exists a function  $\omega(\cdot)$  such that  $E[\omega(Y)|X^* = x_j^*] > E[\omega(Y)|X^* = x_{j+1}^*]$ .*

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  $y_0$ . We summarize the results as follows:

**Theorem 1** (Hu (2008)) *Under assumptions 1, 2, 3, and 4, the 2.1-measurement model in Definition 3 is nonparametrically 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*

$$f_{X,Y,Z,X^*} = f_{X|X^*}f_{Y|X^*}f_{Z|X^*}f_{X^*}. \quad (26)$$

Theorem 1 provides an exact identification result in the sense that the number of unknown probabilities is equal to the number of observed probabilities in equation (22). Assumption 1 implies that there are  $2K^2 - 1$  observed probabilities in  $f_{X,Y,Z}(x, y, z)$  on the left hand side of equation (22). 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 may directly mimic the identification procedure.

When supports of measurements  $X$  and  $Z$  are larger than that of  $X^*$ , we may still achieve the identification with minor modification of the conditions. Suppose supports  $\mathcal{X}$  and  $\mathcal{Z}$  are larger than  $\mathcal{X}^*$ , i.e.,  $\mathcal{X} = \{x_1, x_2, \dots, x_L\}$ ,  $\mathcal{Z} = \{z_1, z_2, \dots, z_J\}$ , and  $\mathcal{X}^* = \{x_1^*, x_2^*, \dots, x_K^*\}$  with  $L > K$  and  $J > K$ . By combining some values in the supports of  $X$  and  $Z$ , we may first transform  $X$  and  $Z$  to  $\tilde{X}$  and  $\tilde{Z}$  so that they share the same support  $\mathcal{X}^*$  as  $X^*$ . We may then identify  $f_{\tilde{X}|X^*}$  and  $f_{\tilde{Z}|X^*}$  by Theorem 1 with those assumptions imposed on  $(\tilde{X}, Y, \tilde{Z}, X^*)$ . 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 may consider the joint distribution

$$f_{\tilde{X},Z} = \sum_{x^* \in \mathcal{X}^*} f_{\tilde{X}|X^*} f_{Z|X^*} f_{X^*}, \quad (27)$$

which is equivalent to

$$M_{\tilde{X},Z} = M_{\tilde{X}|X^*} D_{X^*} M_{Z|X^*}^T. \quad (28)$$

Since we have identified  $M_{\tilde{X}|X^*}$  and  $D_{X^*}$ , we may identify  $M_{Z|X^*}$ , i.e.,  $f_{Z|X^*}$ , by inverting  $M_{\tilde{X}|X^*}$ . Similar argument holds for identification of  $f_{X|X^*}$ . This discussion implies that Assumptions 1 is not necessary. We keep it in Theorem 1 in order to show the minimum data information needed for nonparametric identification of the 2.1-measurement model.

### 2.3.2 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

$$\begin{aligned} \vec{p}_{X|x_i^*} &= [f_{X|X^*}(x_1|x_i^*), f_{X|X^*}(x_2|x_i^*), f_{X|X^*}(x_3|x_i^*)]^T \\ \vec{p}_{X|z} &= [f_{X|Z}(x_1|z), f_{X|Z}(x_2|z), f_{X|Z}(x_3|z)]^T. \end{aligned} \quad (29)$$

We have for each  $z$

$$\vec{p}_{X|z} = \sum_{i=1}^3 w_i^z (\vec{p}_{X|x_i^*}) \quad (30)$$

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  $\vec{p}_{X|x_1^*}$ ,  $\vec{p}_{X|x_2^*}$ , and  $\vec{p}_{X|x_3^*}$ . Similarly, if we consider the subsample with  $Y = 1$ , we have

$$\vec{p}_{y_1, X|z} = \sum_{i=1}^3 w_i^z (\lambda_i \vec{p}_{X|x_i^*}) \quad (31)$$

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

$$\vec{p}_{y_1, X|z} = [f_{Y,X|Z}(1, x_1|z), f_{Y,X|Z}(1, x_2|z), f_{Y,X|Z}(1, x_3|z)]^T. \quad (32)$$

That means vector  $\vec{p}_{y_1, X|z}$  is a weighted average of  $(\lambda_i \vec{p}_{X|x_i^*})$  for  $i = 1, 2, 3$ , where weights  $w_i^z$  are the same as in equation (30) from the whole sample. Notice that the direction of basis vectors  $(\lambda_i \vec{p}_{X|x_i^*})$  corresponding to the subsample with  $Y = 1$  is the same as the direction of basis vectors  $\vec{p}_{X|x_i^*}$  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  $\vec{p}_{X|z}$  to one spanned by  $\vec{p}_{y_1, X|z}$ , the basis vectors don't 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.,  $\lambda_i$ . This mapping is in fact  $M_{X,y,Z}M_{X,Z}^{-1}$  on the left hand side of equation (25). The variation in variable  $Z$  guarantees that such a mapping exists. Figure 1 illustrates this framework.

### 2.3.3 The continuous case

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

**Assumption 5** *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 6** *The operators  $L_{X|X^*}$  and  $L_{Z|X}$  are injective.*<sup>6</sup>

**Assumption 7** *For all  $\bar{x}^* \neq \tilde{x}^*$  in  $\mathcal{X}^*$ , the set  $\{y : f_{Y|X^*}(y|\bar{x}^*) \neq f_{Y|X^*}(y|\tilde{x}^*)\}$  has positive probability.*

**Assumption 8** *There exists a known function  $M$  such that  $M[f_{X|X^*}(\cdot|x^*)] = x^*$  for all  $x^* \in \mathcal{X}^*$ .*

The functional  $M[\cdot]$  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 2** *(Hu and Schennach (2008)) Under assumptions 5, 6, 7, and 8, the 2.1-measurement model in Definition 3 with a continuous  $X^*$  is nonparametrically 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 (26).*

This result implies that if we observe an additional binary indicator of the latent variable together with two measurements, we may relax the additivity and the independence assumptions in equation (16) and achieve nonparametric identification of very general models. Comparing the model in equation (16) 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.

---

<sup>6</sup> $L_{Z|X}$  is defined in the same way as  $L_{X|X^*}$  in equation (7).

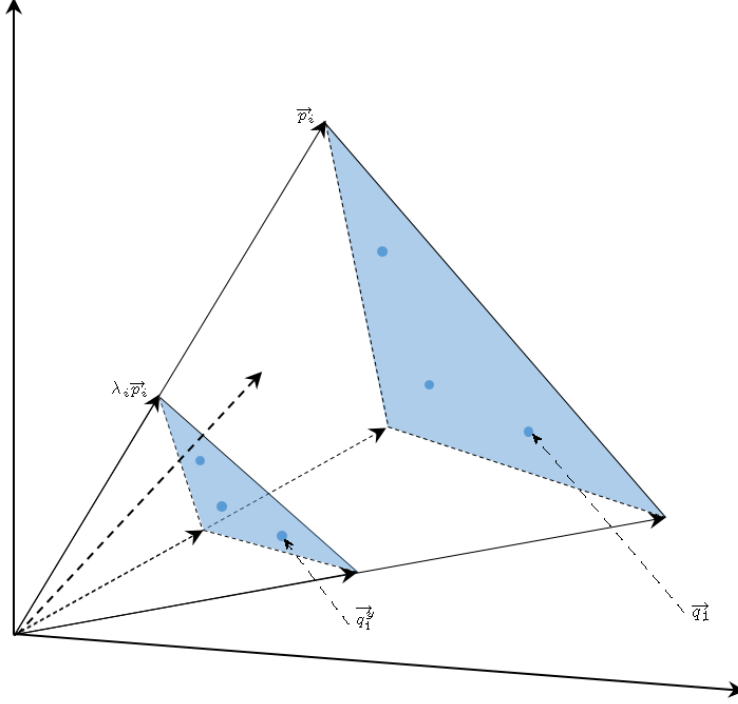


Figure 1: Eigenvalue-eigenvector decomposition in the 2.1-measurement model.

Eigenvalue:  $\lambda_i = f_{Y|X^*}(1|x_i^*)$ .

Eigenvector:  $\vec{p}_i = \vec{p}_{X|x_i^*} = [f_{X|X^*}(x_1|x_i^*), f_{X|X^*}(x_2|x_i^*), f_{X|X^*}(x_3|x_i^*)]^T$ .

Observed distribution in the whole sample:

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

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

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

### 2.3.4 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

$$\begin{aligned} \Pr(Y = 1|X^*) &\neq \Pr(Y = 1) \\ X &= X^*\gamma + \epsilon \\ Z &= X^*\gamma' + \epsilon' \end{aligned} \tag{33}$$

where  $(X^*, \epsilon, \epsilon')$  are mutually independent. We may interpret the error term  $\epsilon'$  as a performance shock in the SAT test. If coefficients  $\gamma$  and  $\gamma'$  are known, we may use  $X/\gamma$  and  $Z/\gamma'$  as the two measurements in equation (16) to identify the marginal distribution of ability without using the binary measurement  $Y$ . As shown in Hu and Sasaki (forthcoming), we may identify all the elements of interest in this model. Here we focus on the identification of the coefficients  $\gamma$  and  $\gamma'$  to illustrate the intuition of the identification results.

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

$$X = Z\gamma + (\epsilon - \epsilon'\gamma). \tag{34}$$

The OLS estimator corresponds to  $\frac{\text{cov}(X,Z)}{\text{var}(Z)} = \gamma \frac{\text{var}(X^*)}{\text{var}(X^*) + \text{var}(\epsilon')}$ , which is the well-known attenuation result with  $\left| \frac{\text{cov}(X,Z)}{\text{var}(Z)} \right| < |\gamma|$ . 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  $\epsilon'$  in the regression error  $(\epsilon - \epsilon'\gamma)$ . When an additional variable  $Y$  is available even if it is binary, however, we may use  $Y$  as an instrument to solve the endogeneity problem and identify  $\gamma$  as

$$\gamma = \frac{E[X|Y = 1] - E[X|Y = 0]}{E[Z|Y = 1] - E[Z|Y = 0]}. \tag{35}$$

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  $\gamma$ .

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

$$f_{X,Z|Y=y} = \int_{\mathcal{X}^*} f_{X|X^*} f_{Z|X^*} f_{X^*|Y=y} dx^*. \quad (36)$$

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 (34). 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  $\epsilon'$ , 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 (30) and (31) in Section 2.3.2.<sup>7</sup>

Furthermore, an arguably surprising result is that such identification of the 2.1 measurement model may still be 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 ( $f_{X|X^*}, f_{Y|X^*}, f_{Z|X^*}$ ), and the marginal distribution  $f_{X^*}$  as in equation (26). The joint distribution  $f_{X,Y,Z,X^*}$  is a four-dimensional function, while ( $f_{X|X^*}, f_{Y|X^*}, f_{Z|X^*}$ ) are three two-dimensional functions. Therefore, the number of unknowns are greatly reduced under the conditional independence assumption.

## 2.4 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 may 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 symmetric roles so that it doesn't 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  $X \in \mathcal{X}$ ,  $Y \in \mathcal{Y}$ , and  $Z \in \mathcal{Z}$  of the latent variable  $X^* \in \mathcal{X}^*$  satisfying*

$$X \perp Y \perp Z \mid X^*, \quad (37)$$

---

<sup>7</sup>Another way to look at this is that  $\gamma$  can also be expressed as

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



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 may always replace  $Y$  with a 0-1 binary indicator, e.g.,  $I(Y > EY)$ . In fact, we intentionally write the results in section 2.3 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 symmetric 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 doesn't matter which variable is called a dependent variable, a measurement, or an instrument variable in terms of identification. We summarize the results as follows:

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

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

$$X_{t+1}^* \perp \{X_s^*\}_{s \leq t-1} \mid X_t^*. \quad (38)$$

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

$$X_t \perp \{X_s, X_s^*\}_{s \neq t} \mid X_t^*. \quad (39)$$

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 may be shown in the flow chart as follows.

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

Consider a panel data set, where we observed three periods of data  $\{X_{t-1}, X_t, X_{t+1}\}$ . The conditions in equations (38) and (39) imply

$$X_{t-1} \perp X_t \perp X_{t+1} \mid X_t^*, \quad (40)$$

i.e.,  $(X_{t-1}, X_t, X_{t+1})$  are jointly independent conditional on  $X_t^*$ . Although the original model is dynamic, it may be reduced to a 3-measurement model as in equation (40). Given Corollary

1, we may nonparametrically identify  $f_{X_{t+1}|X_t^*}$ ,  $f_{X_t|X_t^*}$ ,  $f_{X_{t-1}|X_t^*}$ , and  $f_{X_t^*}$ . Under a stationarity assumption that  $f_{X_{t+1}|X_{t+1}^*} = f_{X_t|X_t^*}$ , we may then identify the Markov kernel  $f_{X_{t+1}^*|X_t^*}$  from

$$f_{X_{t+1}|X_t^*} = \int_{\mathcal{X}^*} f_{X_{t+1}|X_{t+1}^*} f_{X_{t+1}^*|X_t^*} dx_{t+1}^*, \quad (41)$$

by inverting the integral operator corresponding to  $f_{X_{t+1}|X_{t+1}^*}$ .<sup>8</sup> Therefore, it doesn't 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, Kayaba, and Shum (2013), Feng and Hu (2013), Wilhelm (2013), and Hu and Sasaki (2014), etc.

## 2.5 A dynamic measurement model

A natural extension to the hidden Markov model in equations (38)-(39) is to relax the local independence assumption in equation (39) when more periods of data are available. For example, we may allow direct serial correlation of observed measurement  $\{X_t\}$ . To this end, we assume the following:

**Assumption 9** *The joint process  $\{X_t, X_t^*\}$  is a first-order Markov process. Furthermore, the Markov kernel satisfies*

$$f_{X_t, X_t^* | X_{t-1}, X_{t-1}^*} = f_{X_t | X_t^*, X_{t-1}} f_{X_t^* | X_{t-1}, X_{t-1}^*}. \quad (42)$$

Equation (42) 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:

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

For simplicity, we focus on the discrete case and assume

**Assumption 10**  $X_t$  and  $X_t^*$  share the same support  $\mathcal{X}^* = \{x_1^*, x_2^*, \dots, x_K^*\}$ .

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

$$\begin{aligned} M_{X_{t+1}, x_t | x_{t-1}, X_{t-2}} &= \left[ f_{X_{t+1}, X_t | X_{t-1}, X_{t-2}}(x_i, x_t | x_{t-1}, x_j) \right]_{i=1,2,\dots,K; j=1,2,\dots,K} \\ M_{X_t | x_{t-1}, X_{t-2}} &= \left[ f_{X_t | X_{t-1}, X_{t-2}}(x_i | x_{t-1}, x_j) \right]_{i=1,2,\dots,K; j=1,2,\dots,K} \end{aligned} \quad (43)$$

---

<sup>8</sup>Without stationarity, one may 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  $(X_t, X_{t+1}, X_{t+2})$ .

**Assumption 11** (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  $(x_{t-1}, \bar{x}_{t-1}, \bar{x}_t)$  such that  $M_{X_{t+1}, x_t|x_{t-1}, X_{t-2}}$ ,  $M_{X_{t+1}, x_t|\bar{x}_{t-1}, X_{t-2}}$ ,  $M_{X_{t+1}, \bar{x}_t|x_{t-1}, X_{t-2}}$ , and  $M_{X_{t+1}, \bar{x}_t|\bar{x}_{t-1}, X_{t-2}}$  are invertible and that for all  $x_t^* \neq \tilde{x}_t^*$  in  $\mathcal{X}^*$

$$\Delta_{x_t} \Delta_{x_{t-1}} \ln f(x_t^*) \neq \Delta_{x_t} \Delta_{x_{t-1}} \ln f(\tilde{x}_t^*)$$

where  $\Delta_{x_t} \Delta_{x_{t-1}} \ln f(x_t^*)$  is defined as

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

**Assumption 12** For any  $x_t \in \mathcal{X}$ ,  $E[X_{t+1}|X_t = x_t, X_t^* = x_t^*]$  is increasing in  $x_t^*$ .

**Assumption 13** The Markov kernel is stationary, i.e.,

$$f_{X_t, X_t^*|X_{t-1}, X_{t-1}^*} = f_{X_2, X_2^*|X_1, X_1^*}. \quad (44)$$

We summarize the results as follows:

**Theorem 3** (Hu and Shum (2012)) Under assumptions 9, 10, 11, 12, and 13, 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 1, 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 (42) 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 to use 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. However, the 2.1-measurement model implies that the minimum data information for non-parametric identification is in fact "2.1 measurements" instead of "3 measurements". That is a reason why such a model, even with a time-varying unobserved state variable, can be identified using only four periods of data as in Hu and Shum (2012).

### 2.5.1 Illustrative Examples

In this section, we use a simple example to illustrate the identification strategy in Theorem 3, which is based on Carroll, Chen, and Hu (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 may be described as follows:

$$f_{Y,X|Z,S} = \int f_{Y|X^*,Z} f_{X|X^*,S} f_{X^*|Z,S} dx^*. \quad (45)$$

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, Chen, and Hu (2010) provide sufficient conditions for nonparametric identification of all the densities on the right hand side of equation (45). To illustrate the identification strategy, we consider the following parametric specification

$$\begin{aligned} Y &= X^* \beta + Z \gamma + \eta \\ X &= X^* + S \gamma' + \epsilon \\ X^* &= g(Z, S) + u, \end{aligned} \quad (46)$$

where  $(\beta, \gamma, \gamma')$  are unknown parameters. We use an unspecified function  $g$  to stress the correlation, instead of causality, between the income  $X^*$  and the family size  $Z$ .

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

$$\begin{aligned} Y &= (X - S \gamma' - \epsilon) \beta + Z \gamma + \eta \\ &= X \beta + Z \gamma - S \gamma' \beta + (\eta - \epsilon \beta). \end{aligned} \quad (47)$$

The regressor  $X$  is endogenous because it is correlated with the measurement error  $\epsilon$ . Note that the income  $X^*$  may vary with the family size  $Z$  and the sample indicator  $S$ , which are independent of  $\epsilon$ , the source of the endogeneity. Let  $(z_0, z_1)$  and  $(s_0, s_1)$  be possible values of  $Z$  and  $S$ , respectively. Since  $E[\eta|Z, S] = E[\epsilon|Z, S] = 0$ , we may have a difference-in-difference estimator for  $\beta$

$$\beta = \frac{[E(Y|z_1, s_1) - E(Y|z_0, s_1)] - [E(Y|z_1, s_0) - E(Y|z_0, s_0)]}{[E(X|z_1, s_1) - E(X|z_0, s_1)] - [E(X|z_1, s_0) - E(X|z_0, s_0)]} \quad (48)$$

The fact that there is no interaction term of  $Z$  and  $S$  on the right hand side of equation (47) is not due to our parametric specification but because of the conditional independence in equation (45).

In the dynamic model in Theorem 3, we have

$$f_{X_{t+1}, X_t, X_{t-1}, X_{t-2}} = \sum_{x^*} f_{X_{t+1}|X_t, X_t^*} f_{X_t|X_t^*, X_{t-1}} f_{X_t^*|X_{t-1}, X_{t-2}}. \quad (49)$$

To make it analogical to equation (45), we may re-write equation (49) as

$$f_{X_{t+1}, X_{t-2}|X_t, X_{t-1}} = \sum_{x^*} f_{X_{t+1}|X_t^*, X_t} f_{X_{t-2}|X_t^*, X_{t-1}} f_{X_t^*|X_t, X_{t-1}}. \quad (50)$$

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  $\eta$  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  $\epsilon$  represents an independent error term. Suppose we can replace  $X_t^*$  by  $G^{-1}(X_{t-2}, X_{t-1}, \epsilon)$  to obtain

$$X_{t+1} = H(G^{-1}(X_{t-2}, X_{t-1}, \epsilon), X_t, \eta), \quad (51)$$

where  $X_{t-2}$  is endogenous and correlated with  $\epsilon$ . The conditional independence in equation (50) implies that the variation in  $X_t$  and  $X_{t-1}$  may vary with  $X_t^*$ , but not with the error  $\epsilon$ . 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-difference 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  $(X_t, X_{t-1}) = (\text{high}, \text{high})$  and a subsample with  $(X_t, X_{t-1}) = (\text{high}, \text{low})$ , 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  $(X_t, X_{t-1}) = (\text{low}, \text{high})$  and  $(X_t, X_{t-1}) = (\text{low}, \text{low})$ . 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  $X_t^*$ .

## 2.6 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, we 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.

Consider the 2.1-measurement model in Theorem 2, where the observed density is associated with the unobserved ones as follows:

$$f_{X,Y,Z}(x, y, z) = \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*)f_{Y|X^*}(y|x^*)f_{Z|X^*}(z|x^*)f_{X^*}(x^*)dx^*. \quad (52)$$

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

$$f_{Y|X^*}(y|x^*) = f_{Y|X^*}(y|x^*; \theta). \quad (53)$$

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

$$\left(\hat{\theta}, \hat{f}_{X|X^*}, \hat{f}_{Z|X^*}, \hat{f}_{X^*}\right) = \arg \max_{(\theta, f_1, f_2, f_3) \in \mathcal{A}_N} \frac{1}{N} \sum_{i=1}^N \ln \int_{\mathcal{X}^*} f_1(X_i|x^*)f_{Y|X^*}(Y_i|x^*; \theta)f_2(Z_i|x^*)f_3(x^*)dx^*, \quad (54)$$

where  $\mathcal{A}_N$  is approximating sieve spaces which contain truncated series as parametric approximations to densities  $(f_{X|X^*}, f_{Z|X^*}, f_{X^*})$ . We may impose restrictions, such as Assumption 8, on the sieve spaces  $\mathcal{A}_N$ . The truncated series in the sieve spaces  $\mathcal{A}_N$  are usually linear combinations of known basis functions, such as polynomials or splines, in which the coefficients are treated as unknown parameters. The number of coefficients may increase 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 may converge at a fast rate, i.e.,  $\hat{\theta}$  may be  $\sqrt{n}$  consistent and asymptotically normally distributed under suitable assumptions. We refer to Hu and Schennach (2008) and its supplementary materials for more discussion on estimation in this framework.

Although the sieve MLE in (54) 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 may 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 (17) may be estimated using Kotlarski's identity with characteristic functions replaced by corresponding empirical characteristic functions. In general, one may consider a nonlinear regression model in the framework of the 3-measurement model as

$$\begin{aligned} Y &= g_1(X^*) + \eta, E[\eta|X^*] = 0 \\ X &= g_2(X^*) + \epsilon \\ Z &= g_3(X^*) + \epsilon', \end{aligned} \tag{55}$$

where  $\epsilon$  and  $\epsilon'$  are independent of  $X^*$  and  $\eta$ . 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.<sup>9</sup> Hu and Sasaki (forthcoming) 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.

In the case where  $X^*$  is discrete as Theorem 1, the identification strategy may also lead to a closed-form estimator for the unknown probabilities in the sense that one may mimic the identification procedure to solve for the unknowns. The eigenvector-eigenvalue decomposition in equation (25) may be considered as a procedure to minimize the absolute difference between the left hand side and the right hand side of equations (24) and (25), in fact, to zero. With a finite sample, certain estimated probabilities might be outside  $[0, 1]$ . One remedy is to minimize the absolute difference under the restrictions that all the probabilities are between 0 and 1. When the sample size becomes larger, the probability of using this remedy should be smaller when all the assumptions hold. This closed-form estimator performs well in empirical studies, such as An, Baye, Hu, Morgan, and Shum (2012), An, Hu, and Shum (2010), Feng and Hu (2013), and Hu, Kayaba, and Shum (2013).

Such closed-form estimators have their advantages that there are much fewer nuisance parameters than indirect estimators, such as the sieve MLE, and that their computation does 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.

---

<sup>9</sup>Schennach (2007) also provides a closed-form estimator for a similar nonparametric regression model using a generalized function approach.

### 3 Applications in microeconomic models with latent variables

A major breakthrough in the measurement error literature is the nonparametric identification of the 2.1-measurement model in section 2.3, 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.

#### 3.1 Auctions with unobserved heterogeneity

Unobserved heterogeneity has been a concern in the estimation of auction models for a long time. Li, Perrigne, and Vuong (2000) and Krasnokutskaya (2011) use the identification result of 2-measurement model in equation (16) to estimate auction models with separable unobserved heterogeneity. In a first-price auction indexed by  $t$  for  $t = 1, 2, \dots, T$  with zero reserve price, there are  $N$  symmetric risk-neutral bidders. For  $i = 1, 2, \dots, N$ , 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

$$b_{it} = s_t^* a_i, \tag{56}$$

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

$$b_{1t} \perp b_{2t} \mid s_t^*. \tag{57}$$

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

$$\begin{aligned} \ln b_{1t} &= \ln s_t^* + \ln a_1 \\ \ln b_{2t} &= \ln s_t^* + \ln a_2, \end{aligned} \tag{58}$$

where one may use Kotlarski's identity to achieve 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, Perrigne, and Vuong (2000).

Hu, McAdams, and Shum (2009) consider auction models with nonseparable unobserved heterogeneity. They assume 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.,

$$b_{1t} \perp b_{2t} \perp b_{3t} \mid s_t^*. \quad (59)$$

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.

### 3.2 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. When reserve prices are binding, those potential bidders whose values are less than the reserve price would not participate so that the observed number of bidders  $A$  is smaller than that of potential bidders  $N^*$ .

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  $[\underline{x}, \bar{x}]$ . 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:

$$b(x_i; N^*) = \begin{cases} x_i - \frac{\int_r^{x_i} F_{N^*}(s)^{N^*-1} ds}{F_{N^*}(x_i)^{N^*-1}} & \text{for } x_i \geq r \\ 0 & \text{for } x_i < r. \end{cases} \quad (60)$$

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

$$\begin{aligned} & f(A_t, b_{1t}, b_{2t} | b_{1t} > r, b_{2t} > r) \\ &= \sum_{N^*} f(A_t | A_t \geq 2, N^*) f(b_{1t} | b_{1t} > r, N^*) f(b_{2t} | b_{2t} > r, N^*) f(N^* | b_{1t} > r, b_{2t} > r). \end{aligned} \quad (61)$$

That means 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, Hu, and Shum (2010), the bid distribution, and therefore, the value distribution, may be nonparametrically identified. Furthermore, such identification is constructive and directly leads to an estimator.

### 3.3 Multiple equilibria in incomplete information games

Xiao (2013) considers a static simultaneous move game, in which player  $i$  for  $i = 1, 2, \dots, N$  chooses an action  $a_i$  from a choice set  $\{0, 1, \dots, K\}$ . Let  $a_{-i}$  denote actions of the other players, i.e.,  $a_{-i} = \{a_1, a_2, \dots, a_{i-1}, a_{i+1}, \dots, a_N\}$ . The player  $i$ 's payoff is specified as

$$u_i(a_i, a_{-i}, \epsilon_i) = \pi_i(a_i, a_{-i}) + \epsilon_i(a_i), \quad (62)$$

where  $\epsilon_i(k)$  for  $k \in \{0, 1, \dots, K\}$  is a choice-specific payoff shock for player  $i$ . 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

$$\sum_{a_{-i}} \pi_i(a_i, a_{-i}) \Pr(a_{-i}) + \epsilon_i(a_i) \equiv \Pi_i(a_i) + \epsilon_i(a_i) \quad (63)$$

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

$$\Pr(a_i = k) = \Pr\left(\left\{\Pi_i(k) + \epsilon_i(k) > \max_{j \neq k} \Pi_i(j) + \epsilon_i(j)\right\}\right). \quad (64)$$

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

However, multiple equilibria may exist for this problem, which means the observed choice probabilities may be 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.,

$$a_1 \perp a_2 \perp \dots \perp a_N \mid e^*. \quad (65)$$

Therefore, the observed correlation among the actions contains information on multiple equilibria. If the support of actions is larger than that of  $e^*$ , one may simply use three players' actions as three measurements for  $e^*$ . Otherwise, if there are enough players, one may partition the players into three groups and use the group actions as the three measurements. Xiao (2013) also extends this identification strategy to dynamic games.

### 3.4 Dynamic 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, Kayaba, and Shum (2013) use bandit experiments to nonparametrically 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 doesn't 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 is 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.,  $\Pr(X_{t+1}^* | X_t^*, Y_t, R_t)$ . They assume that the choice only depends on 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

$$Z_t \perp Y_t \perp Z_{t-1} \mid X_t^*. \quad (66)$$

Therefore, the proxy rule  $\Pr(Z_t | X_t^*)$  is nonparametrically identified. Under the local independence assumption, one may 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  $\Pr(X_{t+1}^* | X_t^*, Y_t, R_t)$  may be identified from the observed distribution  $\Pr(Z_{t+1}, Y_t, R_t, Z_t)$  through

$$\begin{aligned} & \Pr(Z_{t+1}, Y_t, R_t, Z_t) \\ &= \sum_{X_{t+1}^*} \sum_{X_t^*} \Pr(Z_{t+1} | X_{t+1}^*) \Pr(Z_t | X_t^*) \Pr(X_{t+1}^*, X_t^*, Y_t, R_t). \end{aligned} \quad (67)$$

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.

### 3.5 Unemployment and labor market 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  $i = 1, 2, \dots, N$ . 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,

$$\begin{aligned} & \Pr(X_{t+1}, X_t, X_{t-9}) \\ &= \sum_{X_{t+1}^*} \sum_{X_t^*} \sum_{X_{t-9}^*} \Pr(X_{t+1}|X_{t+1}^*) \Pr(X_t|X_t^*) \Pr(X_{t-9}|X_{t-9}^*) \Pr(X_{t+1}^*, X_t^*, X_{t-9}^*). \end{aligned} \quad (68)$$

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

$$\begin{aligned} & \Pr(X_{t+1}, X_t, X_{t-9}) \\ &= \sum_{X_t^*} \Pr(X_{t+1}|X_t^*) \Pr(X_t|X_t^*) \Pr(X_{t-9}^*, X_{t-9}), \end{aligned} \quad (69)$$

which implies a 3-measurement model. This model can be considered as an application of Theorem 1 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.

### 3.6 Dynamic discrete choice with unobserved state variables

Hu and Shum (2012) show that the transition kernel of a Markov process  $\{W_t, X_t^*\}$  may 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 may be decomposed as follows:

$$f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*} = f_{Y_t | M_t, X_t^*} f_{M_t, X_t^* | Y_{t-1}, M_{t-1}, X_{t-1}^*}. \quad (70)$$

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 may lead to a parametric GMM estimator. That implies an estimator for a dynamic discrete choice model with unobserved state variables, where one may 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 (2014) 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

$$X_t^* = \alpha^d + \beta^d X_{t-1}^* + \eta_t^d \text{ if } Y_{t-1} = d \in \{0, 1\}. \quad (71)$$

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

$$X_t = X_t^* + \epsilon_t. \quad (72)$$

Therefore, they obtain

$$X_{t+1} = \alpha^d + \beta^d X_t^* + \eta_{t+1}^d + \epsilon_{t+1} \quad (73)$$

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

### 3.7 Fixed effects in panel data models

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

$$Y_{it} = g(X_{it}, \alpha_i) + \xi_{it}, \quad (74)$$

where  $X_{it}$  is an explanatory variable,  $Y_{it}$  is the dependent variable,  $\xi_{it}$  is an independent error term, and  $\alpha_i$  represents fixed effects. In order to use Kotlarski's identity, Evdokimov (2010) considers the event where  $\{X_{i1} = X_{i2} = x\}$  for two periods of data to obtain

$$\begin{aligned} Y_{i1} &= g(x, \alpha_i) + \xi_{i1}, \\ Y_{i2} &= g(x, \alpha_i) + \xi_{i2}. \end{aligned} \quad (75)$$

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

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

$$Y_{it} = g(X_{it}, Y_{i,t-1}, U_{it}, \xi_{it}), \quad (76)$$

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

$$X_{i,t+1} \perp (Y_{it}, Y_{i,t-1}, X_{i,t-1}) \mid (X_{it}, U_{it}) \quad (77)$$

to obtain

$$f_{X_{i,t+1}, Y_{it}, X_{it}, Y_{i,t-1}, X_{i,t-1}} = \int f_{X_{i,t+1}|X_{it}, U_{it}} f_{Y_{it}|X_{it}, Y_{i,t-1}, U_{it}} f_{X_{it}, Y_{i,t-1}, X_{i,t-1}, U_{it}} dU_{it}. \quad (78)$$

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 (78) 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 (77) is not directly motivated by economic structure.

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

$$Y_{it} = g(X_{it}, \alpha_i' F_t + \xi_{it}), \quad (79)$$

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  $\xi_{it}$  for  $t = 1, 2, \dots, T$  are jointly independent conditional on  $\alpha_i$  and  $X_i = (X_{i1}, X_{i2}, \dots, X_{iT})$ , he obtains

$$Y_{i1} \perp Y_{i2} \perp \dots \perp Y_{iT} \mid (\alpha_i, X_i), \quad (80)$$

which may form 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  $\alpha_i$  than a general argument as in Theorem 2.

Sasaki (2013) considers a dynamic panel with unobserved heterogeneity  $\alpha_i$  and sample attrition as follows:

$$\begin{aligned} Y_{it} &= g(Y_{i,t-1}, \alpha_i, \xi_{it}) \\ D_t &= h(Y_{it}, \alpha_i, \eta_{it}) \\ Z_i &= \varsigma(\alpha_i, \epsilon_i) \end{aligned} \quad (81)$$

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

$$Y_{i3} \perp Z_i \perp Y_{i1} \mid (\alpha_i, Y_2 = y_2, D_2 = D_1 = 1). \quad (82)$$

In the case where  $\alpha_i$  is discrete, the model is identified using the results in Theorem 1. Sasaki (2013) also extends this identification result to more complicated settings.

### 3.8 Cognitive and noncognitive skill formation

Cunha, Heckman, and Schennach (2010) consider a model of cognitive and noncognitive skill formation, where for multiple periods of childhood  $t \in \{1, 2, \dots, T\}$ ,  $X_t^* = (X_{C,t}^*, X_{N,t}^*)$  stands for cognitive and noncognitive skill stocks in period  $t$ , respectively. The  $T$  childhood periods are divided into  $s \in \{1, 2, \dots, S\}$  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 noncognitive skills, respectively. For  $k \in \{C, N\}$ , they assume that skills evolve as follows:

$$X_{k,t+1}^* = f_{k,s}(X_t^*, I_t, X_P^*, \eta_{k,t}), \quad (83)$$

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

$$X^* = \left( \{X_{C,t}^*\}_{t=1}^T, \{X_{N,t}^*\}_{t=1}^T, \{I_{C,t}\}_{t=1}^T, \{I_{N,t}\}_{t=1}^T, X_{C,P}^*, X_{N,P}^* \right), \quad (84)$$

one may 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

$$X_j = g_j(X^*, \varepsilon_j) \quad (85)$$

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

$$X_1 \perp X_2 \perp X_3 \mid X^*, \quad (86)$$

this leads to a 3-measurement model and the distribution of  $X^*$  may 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.

### 3.9 Two-sided matching models

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

$$v(X_i, \varepsilon_i) = h(X_i) + \varepsilon_i. \quad (87)$$

The workers' preference for firm  $j$  is described by

$$u(Z_j, \eta_j) = g(Z_j) + \eta_j. \quad (88)$$

The preferences on both sides are public information in the market. Researchers are interested in the preferences, including functions  $h$ ,  $g$ , and distributions of  $\varepsilon_i$  and  $\eta_j$ .

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.,  $F_U(u(Z_j, \eta_j)) = q$  is matched with the worker with the  $q$ -th quantile position of the human capital index, i.e.,  $F_V(v(X_i, \varepsilon_i)) = q$ , 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

$$f(X, Z) = \int_0^1 f(X|q) f(Z|q) dq, \quad (89)$$

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

$$\begin{aligned} f(X|q) &= f_\varepsilon(F_V^{-1}(q) - h(X)) \\ f(Z|q) &= f_\eta(F_U^{-1}(q) - g(Z)), \end{aligned} \quad (90)$$

the functions  $h$  and  $g$  may 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.,  $f_{Z|X}(z|x_1) = f_{Z|X}(z|x_2)$  for all  $z$ , then the two workers must have the same observed part of the human capital index, i.e.,  $h(x_1) = h(x_2)$ . A similar argument also holds for function  $g$ . In order to further identify the model, Agarwal and Diamond (2013) considers many-to-one matching where one firm may have two or more identical slots for workers. In such a sample, they may 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

$$f(X_1, X_2, Z) = \int_0^1 f(X_1|q) f(X_2|q) f(Z|q) dq. \quad (91)$$

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

### 3.10 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, Blundell, and Bonhomme (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

$$y_{it} = \eta_{it} + \varepsilon_{it} \quad (92)$$

where  $\eta_{it}$  is the persistent component of income and  $\varepsilon_{it}$  is the transitory one. Furthermore, they assume that  $\varepsilon_{it}$  has a zero mean and is independent over time, and that the persistent

component  $\eta_{it}$  follows a first-order Markov process satisfying

$$\eta_{it} = Q_t(\eta_{i,t-1}, u_{it}) \quad (93)$$

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

The dynamic process  $\{y_{it}, \eta_{it}\}$  can be considered as a hidden Markov process as  $\{X_t, X_t^*\}$  in equations (38) and (39). 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

$$y_{i,t-1} \perp y_{it} \perp y_{i,t+1} \mid \eta_{it} \quad (94)$$

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

## 4 Summary

This paper 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.

## References

- AGARWAL, N., AND W. DIAMOND (2013): “Identification and Estimation in Two-Sided Matching Markets,” Discussion paper, Cowles foundation.
- AN, Y., M. R. BAYE, Y. HU, J. MORGAN, AND M. SHUM (2012): “Identification and Estimation of Online Price Competition with an Unknown Number of Firms,” Discussion paper, Indiana University, Kelley School of Business, Department of Business Economics and Public Policy.
- AN, Y., Y. HU, AND M. SHUM (2010): “Estimating First-Price Auctions with an Unknown Number of Bidders: A Misclassification Approach,” *Journal of Econometrics*, 157, 328–341.
- ARELLANO, M., R. BLUNDELL, AND S. BONHOMME (2014): “Household Earnings and Consumption: A Nonlinear Framework,” .

- BOLLINGER, C. (1996): “Bounding mean regressions when a binary regressor is mismeasured,” *Journal of Econometrics*, 73, 387–399.
- BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): “Measurement error in survey data,” *Handbook of econometrics*, 5, 3705–3843.
- CARROLL, R. J., X. CHEN, AND Y. HU (2010): “Identification and estimation of nonlinear models using two samples with nonclassical measurement errors,” *Journal of nonparametric statistics*, 22(4), 379–399.
- CARROLL, R. J., D. RUPPERT, L. A. STEFANSKI, AND C. M. CRAINICEANU (2012): *Measurement error in nonlinear models: a modern perspective*. CRC press.
- CHEN, X., H. HONG, AND D. NEKIPELOV (2011): “Nonlinear models of measurement errors,” *Journal of Economic Literature*, 49(4), 901–937.
- CHEN, X., H. HONG, AND E. TAMER (2005): “Measurement error models with auxiliary data,” *The Review of Economic Studies*, 72(2), 343–366.
- CHEN, X., Y. HU, AND A. LEWBEL (2009): “Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information,” *Statistica Sinica*, 19, 949–968.
- CHEN, X., AND X. SHEN (1998): “Sieve extremum estimates for weakly dependent data,” *Econometrica*, pp. 289–314.
- CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): “Estimating the technology of cognitive and noncognitive skill formation,” *Econometrica*, 78(3), 883–931.
- EVDOKIMOV, K. (2010): “Identification and estimation of a nonparametric panel data model with unobserved heterogeneity,” *Department of Economics, Princeton University*.
- FENG, S., AND Y. HU (2013): “Misclassification Errors and the Underestimation of the US Unemployment Rates,” *The American Economic Review*, 103(2), 1054–1070.
- FREYBERGER, J. (2012): “Nonparametric panel data models with interactive fixed effects,” Discussion paper, University of Wisconsin, Madison.
- FULLER, W. A. (2009): *Measurement error models*, vol. 305. John Wiley & Sons.
- GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): “Optimal Nonparametric Estimation of First-price Auctions,” *Econometrica*, 68(3), 525–574.
- HOTZ, J., AND R. MILLER (1993): “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *Review of Economic Studies*, 60, 497–529.
- HU, Y. (2008): “Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution,” *Journal of Econometrics*, 144, 27–61.
- HU, Y., Y. KAYABA, AND M. SHUM (2013): “Nonparametric learning rules from bandit experiments: the eyes have it!,” *Games and Economic Behavior*, 81, 215–231.

- HU, Y., D. McADAMS, AND M. SHUM (2009): “Nonparametric Identification of Auction Models with Non-Separable Unobserved Heterogeneity,” *working paper, Johns Hopkins University*.
- HU, Y., AND Y. SASAKI (2014): “Dynamic Discrete Choice Models with Proxies for Unobserved Technologies,” Discussion paper.
- (forthcoming): “Closed-Form Estimation of Nonparametric Models with Non-Classical Measurement Errors,” *Journal of Econometrics*.
- HU, Y., AND S. SCHENNACH (2008): “Instrumental Variable Treatment of Nonclassical Measurement Error Models,” *Econometrica*, 76, 195–216.
- HU, Y., AND M. SHUM (2012): “Nonparametric identification of dynamic models with unobserved state variables,” *Journal of Econometrics*, 171(1), 32–44.
- (2013): “Identifying dynamic games with serially-correlated unobservables,” *Structural Econometric Models*, 31, 97–113.
- HUI, S., AND S. WALTER (1980): “Estimating the error rates of diagnostic tests,” *Biometrics*, 36, 167–171.
- KASAHARA, H., AND K. SHIMOTSU (2009): “Nonparametric identification of finite mixture models of dynamic discrete choices,” *Econometrica*, 77(1), 135–175.
- KOTLARSKI, I. (1965): “On pairs of independent random variables whose product follows the gamma distribution,” *Biometrika*, pp. 289–294.
- KRASNOKUTSKAYA, E. (2011): “Identification and Estimation in Procurement Auctions under Unobserved Auction Heterogeneity,” *Review of Economic Studies*, 28, 293–327.
- LI, T., I. PERRIGNE, AND Q. VUONG (2000): “Conditionally Independent Private Information in OCS Wildcat Auctions,” *Journal of Econometrics*, 98, 129–161.
- LI, T., AND Q. VUONG (1998): “Nonparametric estimation of the measurement error model using multiple indicators,” *Journal of Multivariate Analysis*, 65(2), 139–165.
- MOLINARI, F. (2008): “Partial Identification of Probability Distributions with Misclassified Data,” *Journal of Econometrics*, 144, 81–117.
- RAO, B. (1992): *Identifiability in Stochastic Models*. Academic Press, Inc.
- REIERSØL, O. (1950): “Identifiability of a linear relation between variables which are subject to error,” *Econometrica*, pp. 375–389.
- SASAKI, Y. (2013): “Heterogeneity and selection in dynamic panel data,” Discussion paper, Citeseer.
- SCHENNACH, S. M. (2004a): “Estimation of nonlinear models with measurement error,” *Econometrica*, 72(1), 33–75.

- (2004b): “Nonparametric regression in the presence of measurement error,” *Econometric Theory*, 20(06), 1046–1093.
- (2007): “Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models,” *Econometrica*, 75(1), 201–239.
- (2012): “Measurement error in nonlinear models: A review,” Discussion paper, cemmap working paper, Centre for Microdata Methods and Practice.
- SCHENNACH, S. M., AND Y. HU (2013): “Nonparametric identification and semiparametric estimation of classical measurement error models without side information,” *Journal of the American Statistical Association*, 108(501), 177–186.
- SHEN, X. (1997): “On methods of sieves and penalization,” *The Annals of Statistics*, 25(6), 2555–2591.
- SHIU, J.-L., AND Y. HU (2013): “Identification and estimation of nonlinear dynamic panel data models with unobserved covariates,” *Journal of Econometrics*, 175(2), 116–131.
- WILHELM, D. (2013): “Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error,” Discussion paper, University College London.
- XIAO, R. (2013): “Identification and Estimation of Incomplete Information Games with Multiple Equilibria,” Discussion paper.