

# The Econometrics of Unobservables: Identification, Estimation, and Empirical Applications

Yingyao Hu

Department of Economics  
Johns Hopkins University

September 29, 2017

# Economic theory vs. econometric model: an example

- economic theory: Permanent income hypothesis
- econometric model: Measurement error model

$$y = \beta x^* + e$$

$$x = x^* + v$$

$$\left\{ \begin{array}{l} y : \text{observed consumption} \\ x : \text{observed income} \\ x^* : \text{latent permanent income} \\ v : \text{latent transitory income} \\ \beta : \text{marginal propensity to consume} \end{array} \right.$$

- maybe the most famous application of measurement error models

# A canonical model of income dynamics: an example

- permanent income: a random walk process
- transitory income: an ARMA process

$$x_t = x_t^* + v_t$$

$$x_t^* = x_{t-1}^* + \eta_t$$

$$v_t = \rho_t v_{t-1} + \lambda_t \epsilon_{t-1} + \epsilon_t$$

$$\left\{ \begin{array}{l} \eta_t : \text{ permanent income shock in period } t \\ \epsilon_t : \text{ transitory income shock} \\ x_t^* : \text{ latent permanent income} \\ v_t : \text{ latent transitory income} \end{array} \right.$$

- Can a sample of  $\{x_t\}_{t=1, \dots, T}$  uniquely determine distributions of latent variables  $\eta_t$ ,  $\epsilon_t$ ,  $x_t^*$ , and  $v_t$ ?

- ① example: permanent income hypothesis vs measurement error model
- ① empirical evidences on measurement error
- ② measurement models: observables vs unobservables
  - definition of measurement and general framework
  - 2-measurement model
  - 2.1-measurement model
  - 3-measurement model
  - dynamic measurement model
  - estimation (closed-form, extremum, semiparametric)
- ③ empirical applications with latent variables
  - auctions with unobserved heterogeneity
  - multiple equilibria in incomplete information games
  - dynamic learning models
  - unemployment and labor market participation
  - cognitive and noncognitive skill formation
  - two-sided matching
  - income dynamics
- ④ conclusion

# Measurement error: empirical evidences and assumptions

- Kane, Rouse, and Staiger (1999): Self-reported education  $x$  conditional on true education  $x^*$ . (Data source: National Longitudinal Class of 1972 and Transcript data)

$f_{x x^*}(x_i x_j)$	$x^*$ — true education level		
$x$ — self-reported education	$x_1$ —no college	$x_2$ —some college	$x_3$ —BA <sup>+</sup>
$x_1$ —no college	0.876	0.111	0.000
$x_2$ —some college	0.112	0.772	0.020
$x_3$ —BA <sup>+</sup>	0.012	0.117	0.980

- Finding I: more likely to tell the truth than any other possible values

$$f_{x|x^*}(x^*|x^*) > f_{x|x^*}(x_i|x^*) \text{ for } x_i \neq x^*.$$

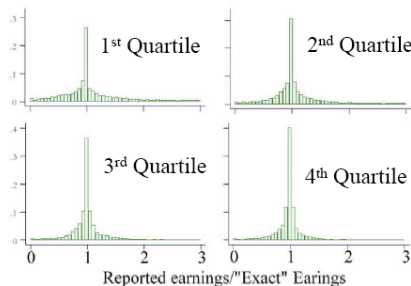
⇒ error equals zero at the mode of  $f_{x|x^*}(\cdot|x^*)$ .

- Finding II: more likely to tell the truth than to lie.  $f_{x|x^*}(x^*|x^*) > 0.5$ .

⇒ invertibility of the matrix  $[f_{x|x^*}(x_i|x_j)]_{i,j}$  in the table above.

# Measurement error: empirical evidences and assumptions

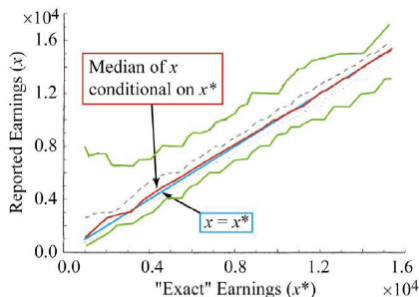
- Chen, Hong & Tarozzi (2005): ratio of self-reported earnings  $x$  vs. true earnings  $x^*$  by quartiles of true earnings. (Data source: 1978 CPS/SS Exact Match File)



- Finding I: distribution of measurement error depends on  $x^*$ .
- Finding II: distribution of measurement error has a zero mode.

# Measurement error: empirical evidences and assumptions

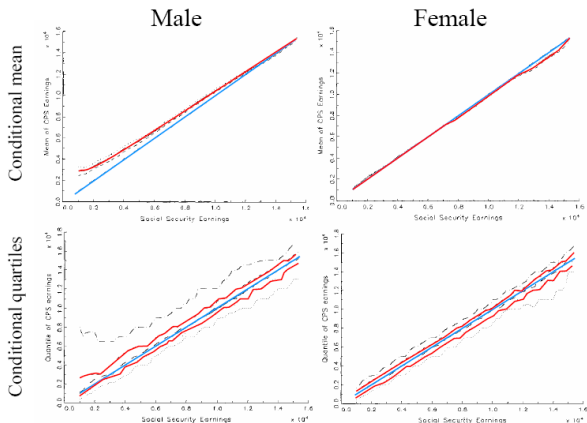
- Bollinger (1998, page 591): percentiles of self-reported earnings  $x$  given true earnings  $x^*$  for males. (Data source: 1978 CPS/SS Exact Match File)



- Finding I: distribution of measurement error depends on  $x^*$ .
- Finding II: distribution of measurement error has a zero median.

# Measurement error: empirical evidences and assumptions

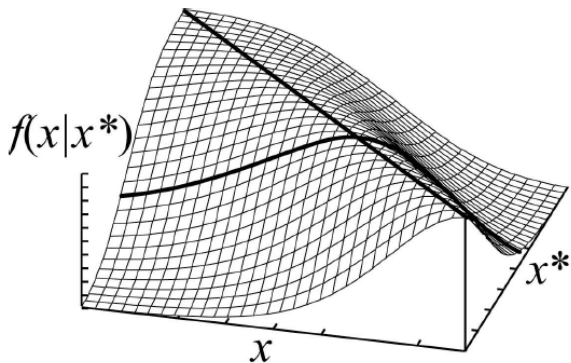
- Self-reporting errors by gender



Source: Bollinger (1998) with data from Bound & Krueger (1991)



# Graphical illustration of zero-mode measurement error



# Latent variables in microeconomic models

empirical models	unobservables	observables
measurement error	true earnings	self-reported earnings
consumption function	permanent income	observed income
production function	productivity	output, input
wage function	ability	test scores
learning model	belief	choices, proxy
auction	unobserved heterogeneity	bids
...	...	...

# Our definition of measurement

- $X$  is defined as a measurement of  $X^*$  if

cardinality of support( $X$ )  $\geq$  cardinality of support( $X^*$ ).

- there exists an injective function from support( $X^*$ ) into support( $X$ ).
- equality holds if there exists a bijective function between two supports.
- number of possible values of  $X$  is not smaller than that of  $X^*$

$X$	$X^*$	
discrete $\{x_1, x_2, \dots, x_L\}$	discrete $\{x_1^*, x_2^*, \dots, x_K^*\}$	$L \geq K$
continuous	discrete $\{x_1^*, x_2^*, \dots, x_K^*\}$	
continuous	continuous	

- $X - X^*$ : measurement error (classical if independent of  $X^*$ )

# A general framework

- observed & unobserved variables

$X$	measurement	observables
$X^*$	latent true variable	unobservables

- economic models described by distribution function  $f_{X^*}$

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

$f_{X^*}$  : latent distribution

$f_X$  : observed distribution

$f_{X|X^*}$  : relationship between observables & unobservables

- identification: Does observed distribution  $f_X$  uniquely determine model of interest  $f_{X^*}$ ?

# Relationship between observables and unobservables

- discrete  $X \in \{x_1, x_2, \dots, x_L\}$  and  $X^* \in \mathcal{X}^* = \{x_1^*, x_2^*, \dots, x_K^*\}$

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

- matrix expression

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

- given  $M_{X|X^*}$ , observed distribution  $f_X$  uniquely determine  $f_{X^*}$  if

$$\text{Rank}(M_{X|X^*}) = \text{Cardinality}(\mathcal{X}^*)$$

# Identification and observational equivalence

- two possible marginal distributions  $\vec{p}_{X^*}^a$  and  $\vec{p}_{X^*}^b$  are observationally equivalent, i.e.,

$$\vec{p}_X = M_{X|X^*} \vec{p}_{X^*}^a = M_{X|X^*} \vec{p}_{X^*}^b$$

- that is, different unobserved distributions lead to the same observed distribution

$$M_{X|X^*} h = 0 \text{ with } h := \vec{p}_{X^*}^a - \vec{p}_{X^*}^b$$

- identification of  $f_{X^*}$  requires

$$M_{X|X^*} h = 0 \text{ implies } h = 0$$

that is, two observationally equivalent distributions are the same.  
This condition can be generalized to the continuous case.

# Identification in the continuous case

- define a set of bounded and integrable functions containing  $f_{X^*}$

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

- define a linear operator

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

- operator equation

$$f_X = L_{X|X^*} f_{X^*}$$

- identification requires injectivity of  $L_{X|X^*}$ , i.e.,

$$L_{X|X^*} h = 0 \text{ implies } h = 0 \text{ for any } h \in \mathcal{L}_{bnd}^1(\mathcal{X}^*)$$

## A 2-measurement model

- definition: two measurements  $X$  and  $Z$  satisfy

$$X \perp Z \mid X^*$$

- two measurements are independent conditional on the latent variable

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

- matrix expression

$$M_{X,Z} = [f_{X,Z}(x_l, z_j)]_{l=1,2,\dots,L; j=1,2,\dots,J}$$

$$M_{Z|X^*} = [f_{Z|X^*}(z_j|x_k^*)]_{j=1,2,\dots,J; k=1,2,\dots,K}$$

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

$$M_{X,Z} = M_{X|X^*} D_{X^*} M_{Z|X^*}^T$$

- suppose that matrices  $M_{X|X^*}$  and  $M_{Z|X^*}$  have a full rank, then

$$\text{Rank}(M_{X,Z}) = \text{Cardinality}(\mathcal{X}^*)$$



## 2-measurement model: binary case

- a binary latent regressor

$$\begin{aligned} Y &= \beta X^* + \eta \\ (X, X^*) &\perp \eta \\ X, X^* &\in \{0, 1\} \end{aligned}$$

- measurement error  $X - X^*$  is correlated with  $X^*$  in general
- $f(y|x)$  is a mixture of  $f_\eta(y)$  and  $f_\eta(y - \beta)$

$$\begin{aligned} f(y|x) &= \sum_{x^*=0}^1 f(y|x^*)f_{X^*|X}(x^*|x) \\ &= f_\eta(y)f_{X^*|X}(0|x) + f_\eta(y - \beta)f_{X^*|X}(1|x) \\ &\equiv f_\eta(y)P_x + f_\eta(y - \beta)(1 - P_x) \end{aligned}$$

## 2-measurement model: binary case

- observed distributions  $f(y|x = 1)$  and  $f(y|x = 0)$  are mixtures of  $f(y|x^* = 1)$  and  $f(y|x^* = 0)$  with different weights  $P_1$  and  $P_2$



$$f(y|x = 1) - f(y|x = 0) = [f_{\eta}(y - \beta) - f_{\eta}(y)](P_0 - P_1)$$

- if  $|P_0 - P_1| \leq 1$ , then

$$|f(y|x = 1) - f(y|x = 0)| \leq |f(y|x^* = 1) - f(y|x^* = 0)|$$

- leads to partial identification

## 2-measurement model: binary case

- parameter of interest

$$\beta = E(y|x^* = 1) - E(y|x^* = 0)$$

- bounds

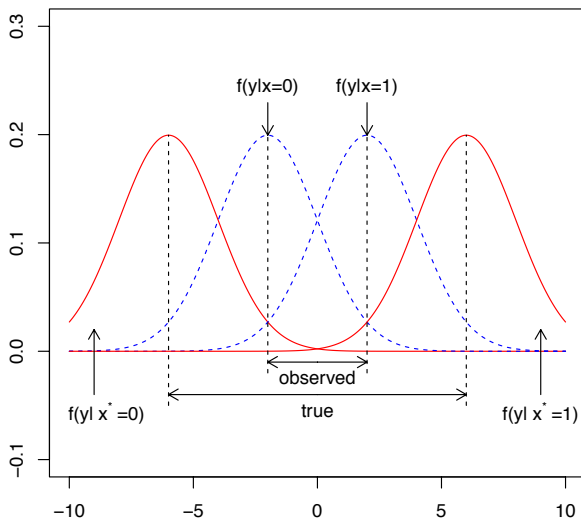
$$|\beta| \geq |E(y|x = 1) - E(y|x = 0)|$$

- If  $\Pr(x^* = 0|x = 0) > \Pr(x^* = 0|x = 1)$ , i.e.,  $P_0 - P_1 > 0$ , then

$$\text{sign}\{\beta\} = \text{sign}\{E(y|x = 1) - E(y|x = 0)\}$$

## 2-measurement model: binary case

- measurement error causes attenuation



- a discrete latent regressor

$$\begin{aligned}y &= \beta x^* + \eta \\(X, X^*) &\perp \eta \\X, X^* &\in \{x_1^*, x_2^*, \dots, x_K^*\}\end{aligned}$$

- Chen Hu & Lewbel (2009): point identification generally holds
- general models without  $(X, X^*) \perp \eta$  : partial identification  
see Bollinger (1996) and Molinari (2008)

- a simple linear regression model with zero means

$$Y = \beta X^* + \eta$$

$$X = X^* + \varepsilon$$

$$X^* \perp \varepsilon \perp \eta$$

- $\beta$  is generally identified (from observed  $f_{Y,X}$ )  
except when  $X^*$  is normal (Reiersol 1950)

## 2-measurement model: Kotlarski's identity

- a useful special case:  $\beta = 1$

$$Y = X^* + \eta$$

$$X = X^* + \varepsilon$$

## 2-measurement model: Kotlarski's identity

- a useful special case:  $\beta = 1$

$$Y = X^* + \eta$$

$$X = X^* + \varepsilon$$

- distribution function & characteristic function of  $X^*$  ( $i = \sqrt{-1}$ )

$$f_{X^*}(x^*) = \frac{1}{2\pi} \int e^{-ix^*t} \Phi_{X^*}(t) dt \quad \Phi_{X^*} = E \left[ e^{itX^*} \right]$$



## 2-measurement model: Kotlarski's identity

- a useful special case:  $\beta = 1$

$$Y = X^* + \eta$$

$$X = X^* + \varepsilon$$

- distribution function & characteristic function of  $X^*$  ( $i = \sqrt{-1}$ )

$$f_{X^*}(x^*) = \frac{1}{2\pi} \int e^{-ix^*t} \Phi_{X^*}(t) dt \quad \Phi_{X^*} = E \left[ e^{itX^*} \right]$$

- Kotlarski's identity (1966)

$$\Phi_{X^*}(t) = \exp \left[ \int_0^t \frac{iE[Ye^{isX}]}{Ee^{isX}} ds \right]$$

## 2-measurement model: Kotlarski's identity

- a useful special case:  $\beta = 1$

$$Y = X^* + \eta$$

$$X = X^* + \varepsilon$$

- distribution function & characteristic function of  $X^*$  ( $i = \sqrt{-1}$ )

$$f_{X^*}(x^*) = \frac{1}{2\pi} \int e^{-ix^*t} \Phi_{X^*}(t) dt \quad \Phi_{X^*} = E \left[ e^{itX^*} \right]$$

- Kotlarski's identity (1966)

$$\Phi_{X^*}(t) = \exp \left[ \int_0^t \frac{iE[Ye^{isX}]}{Ee^{isX}} ds \right]$$

- latent distribution  $f_{X^*}$  is uniquely determined by observed distribution  $f_{Y,X}$  with a closed form

## 2-measurement model: Kotlarski's identity

- Kotlarski's identity (1966)

$$\Phi_{X^*}(t) = \exp \left[ \int_0^t \frac{iE [Ye^{isX}]}{Ee^{isX}} ds \right]$$

- intuition:

$$\text{Var}(X^*) = \text{Cov}(Y, X)$$

- All the moments of  $X^*$  may be written as a function of joint moments of  $Y$  and  $X$  with a closed form
- first introduced to econometrics by Li and Vuong (1998). Li (2002, JoE) first used the result to consistently estimate regression models with classical measurement errors.

## 2-measurement model: nonlinear model with classical error

- a nonparametric regression model

$$Y = g(X^*) + \eta$$

$$X = X^* + \varepsilon$$

$$X^* \perp \varepsilon \perp \eta$$

- Schennach & Hu (2013 JASA):  $g(\cdot)$  is generally identified except some parametric cases of  $g$  or  $f_{X^*}$
- a generalization of Reiersol (1950, ECMA)
- 2-measurement model needs strong specification assumptions for nonparametric identification: additivity, independence

## 2.1-measurement model

- “0.1 measurement” refers to a 0-1 dichotomous indicator  $Y$  of  $X^*$
- definition of 2.1-measurement model:  
two measurements  $X$  and  $Z$  and a 0-1 indicator  $Y$  satisfy

$$X \perp Y \perp Z \mid X^*$$

- for  $y \in \{0, 1\}$

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

- an important message: adding “0.1 measurement” in a 2-measurement model is enough for nonparametric identification, i.e., under mild conditions,

$f_{X,Y,Z}$  uniquely determines  $f_{X,Y,Z,X^*}$

$$f_{X,Y,Z,X^*} = f_{X|X^*} f_{Y|X^*} f_{Z|X^*} f_{X^*}$$

- a global nonparametric point identification  
(exact identification if  $J = K = L$ )

# Identification: discrete case (Hu, 2008, JE)

- Let  $x, x^* \in \{x_1, x_2, x_3\}$  and  $z \in \{z_1, z_2, z_3\}$ , e.g., education levels.

$$M_{x|x^*} = \begin{pmatrix} f_{x|x^*}(x_1|x_1) & f_{x|x^*}(x_1|x_2) & f_{x|x^*}(x_1|x_3) \\ f_{x|x^*}(x_2|x_1) & f_{x|x^*}(x_2|x_2) & f_{x|x^*}(x_2|x_3) \\ f_{x|x^*}(x_3|x_1) & f_{x|x^*}(x_3|x_2) & f_{x|x^*}(x_3|x_3) \end{pmatrix} \leftarrow \text{error structure}$$

$$M_{x^*|z} = \begin{pmatrix} f_{x^*|z}(x_1|z_1) & f_{x^*|z}(x_1|z_2) & f_{x^*|z}(x_1|z_3) \\ f_{x^*|z}(x_2|z_1) & f_{x^*|z}(x_2|z_2) & f_{x^*|z}(x_2|z_3) \\ f_{x^*|z}(x_3|z_1) & f_{x^*|z}(x_3|z_2) & f_{x^*|z}(x_3|z_3) \end{pmatrix} \leftarrow \text{IV structure}$$

$$D_{y|x^*} = \begin{pmatrix} f_{y|x^*}(y|x_1) & 0 & 0 \\ 0 & f_{y|x^*}(y|x_2) & 0 \\ 0 & 0 & f_{y|x^*}(y|x_3) \end{pmatrix} \leftarrow \text{latent model}$$

$$M_{y;x|z} = \begin{pmatrix} f_{y;x|z}(y, x_1|z_1) & f_{y;x|z}(y, x_1|z_2) & f_{y;x|z}(y, x_1|z_3) \\ f_{y;x|z}(y, x_2|z_1) & f_{y;x|z}(y, x_2|z_2) & f_{y;x|z}(y, x_2|z_3) \\ f_{y;x|z}(y, x_3|z_1) & f_{y;x|z}(y, x_3|z_2) & f_{y;x|z}(y, x_3|z_3) \end{pmatrix} \leftarrow \text{observed info.}$$

- $M_{y;x|z}$  contains the same information as  $f_{y,x|z}$ .

# Matrix equivalence

- The main equation for a given  $y$

$$f_{y,x|z}(y, x|z) = \sum_{x^*} f_{x|x^*}(x|x^*) f_{y|x^*}(y|x^*) f_{x^*|z}(x^*|z)$$



$$M_{y;x|z} = M_{x|x^*} D_{y|x^*} M_{x^*|z}$$

# Matrix equivalence

- The main equation for a given  $y$

$$f_{y,x|z}(y, x|z) = \sum_{x^*} f_{x|x^*}(x|x^*) f_{y|x^*}(y|x^*) f_{x^*|z}(x^*|z)$$



$$M_{y;x|z} = M_{x|x^*} D_{y|x^*} M_{x^*|z}$$

- Similarly,

$$f_{x|z}(x|z) = \sum_{x^*} f_{x|x^*}(x|x^*) f_{x^*|z}(x^*|z)$$



$$M_{x|z} = M_{x|x^*} M_{x^*|z}$$



# Matrix equivalence

- The main equation for a given  $y$

$$\boxed{f_{y,x|z}(y, x|z) = \sum_{x^*} f_{x|x^*}(x|x^*) f_{y|x^*}(y|x^*) f_{x^*|z}(x^*|z)}$$
$$\Updownarrow$$
$$\boxed{M_{y;x|z} = M_{x|x^*} D_{y|x^*} M_{x^*|z}}$$

- Similarly,

$$\boxed{f_{x|z}(x|z) = \sum_{x^*} f_{x|x^*}(x|x^*) f_{x^*|z}(x^*|z)}$$
$$\Updownarrow$$
$$\boxed{M_{x|z} = M_{x|x^*} M_{x^*|z}}$$

- Eliminate  $M_{x^*|z}$

$$\begin{aligned} M_{y;x|z} M_{x|z}^{-1} &= (M_{x|x^*} D_{y|x^*} M_{x^*|z}) \times (M_{x^*|z}^{-1} M_{x|x^*}^{-1}) \\ &= M_{x|x^*} D_{y|x^*} M_{x|x^*}^{-1}. \end{aligned}$$

# An inherent matrix diagonalization

- An eigenvalue-eigenvector decomposition:

$$\begin{aligned}M_{y|x|z} M_{x|z}^{-1} &= M_{x|x^*} D_{y|x^*} M_{x|x^*}^{-1} \\&= \begin{pmatrix} f_{x|x^*}(x_1|x_1) & f_{x|x^*}(x_1|x_2) & f_{x|x^*}(x_1|x_3) \\ f_{x|x^*}(x_2|x_1) & f_{x|x^*}(x_2|x_2) & f_{x|x^*}(x_2|x_3) \\ f_{x|x^*}(x_3|x_1) & f_{x|x^*}(x_3|x_2) & f_{x|x^*}(x_3|x_3) \end{pmatrix} \\&\quad \times \begin{pmatrix} f_{y|x^*}(y|x_1) & 0 & 0 \\ 0 & f_{y|x^*}(y|x_2) & 0 \\ 0 & 0 & f_{y|x^*}(y|x_3) \end{pmatrix} \\&\quad \times \begin{pmatrix} f_{x|x^*}(x_1|x_1) & f_{x|x^*}(x_1|x_2) & f_{x|x^*}(x_1|x_3) \\ f_{x|x^*}(x_2|x_1) & f_{x|x^*}(x_2|x_2) & f_{x|x^*}(x_2|x_3) \\ f_{x|x^*}(x_3|x_1) & f_{x|x^*}(x_3|x_2) & f_{x|x^*}(x_3|x_3) \end{pmatrix}^{-1}\end{aligned}$$

- For  $\clubsuit \in \{x_1, x_2, x_3\}$ , i.e., an index of eigenvalues and eigenvectors:
  - eigenvalues:  $f_{y|x^*}(y|\clubsuit)$
  - eigenvectors:  $[f_{x|x^*}(x_1|\clubsuit), f_{x|x^*}(x_2|\clubsuit), f_{x|x^*}(x_3|\clubsuit)]^T$

# Ambiguity Inside the decomposition

- Ambiguity in indexing eigenvalues and eigenvectors, i.e.,

$$\{\clubsuit, \heartsuit, \spadesuit\} \stackrel{1\text{-to-1}}{\longleftrightarrow} \{x_1, x_2, x_3\}$$

- Decompositions with different indexing are observationally equivalent,

$$\begin{aligned} M_{y|x|z} M_{x|z}^{-1} &= M_{x|x^*} D_{y|x^*} M_{x|x^*}^{-1} \\ &= \begin{pmatrix} f_{x|x^*}(x_1|\clubsuit) & f_{x|x^*}(x_1|\heartsuit) & f_{x|x^*}(x_1|\spadesuit) \\ f_{x|x^*}(x_2|\clubsuit) & f_{x|x^*}(x_2|\heartsuit) & f_{x|x^*}(x_2|\spadesuit) \\ f_{x|x^*}(x_3|\clubsuit) & f_{x|x^*}(x_3|\heartsuit) & f_{x|x^*}(x_3|\spadesuit) \end{pmatrix} \\ &\quad \times \begin{pmatrix} f_{y|x^*}(y|\clubsuit) & 0 & 0 \\ 0 & f_{y|x^*}(y|\heartsuit) & 0 \\ 0 & 0 & f_{y|x^*}(y|\spadesuit) \end{pmatrix} \\ &\quad \times \begin{pmatrix} f_{x|x^*}(x_1|\clubsuit) & f_{x|x^*}(x_1|\heartsuit) & f_{x|x^*}(x_1|\spadesuit) \\ f_{x|x^*}(x_2|\clubsuit) & f_{x|x^*}(x_2|\heartsuit) & f_{x|x^*}(x_2|\spadesuit) \\ f_{x|x^*}(x_3|\clubsuit) & f_{x|x^*}(x_3|\heartsuit) & f_{x|x^*}(x_3|\spadesuit) \end{pmatrix}^{-1} \end{aligned}$$

- Identification of  $f_{x|x^*}$  boils down to identification of symbols  $\clubsuit, \heartsuit, \spadesuit$ .

# Restrictions on eigenvalues and eigenvectors

- Eigenvalues are distinctive if  $x^*$  is relevant, i.e.,
    - $f_{y|x^*}(y|x_i) \neq f_{y|x^*}(y|x_j)$  with  $x_i \neq x_j$  for some  $y$ .
  - Symbols  $\clubsuit$ ,  $\heartsuit$ ,  $\spadesuit$  are identified under zero-mode assumption.
- For example, error distribution  $f_{x|x^*}$  is the same as in Kane et al (1999).

no clg. – $x_1$ :	$f_{x x^*}(x_1 \clubsuit)$	=	$\begin{pmatrix} 0.111 \\ 0.772 \\ 0.117 \end{pmatrix}$
some clg. – $x_2$ :	$f_{x x^*}(x_2 \clubsuit)$		
BA <sup>+</sup> – $x_3$ :	$f_{x x^*}(x_3 \clubsuit)$		

$x_2 = \arg \max_{x_i} f_{x|x^*}(x_i|\clubsuit)$   
“ $x_2$  is the mode”

zero-mode assumption
----------------------

$\arg \max_{x_i} f_{x|x^*}(x_i|\clubsuit) = \clubsuit$   
“truth at the mode”

$\clubsuit = x_2$  (some college)

- Similarly, we can identify  $\heartsuit$  and  $\spadesuit$ .  
 $\implies$  The model  $f_{y|x^*}$  and the error structure  $f_{x|x^*}$  are identified.

# Uniqueness of the eigen decomposition

- uniqueness of the eigenvalue-eigenvector decomposition (Hu 2008 JE)

1. distinctive eigenvalues:  $\exists$  a nontrivial set of  $y$ , s.t.,

$$f(y|x_1^*) \neq f(y|x_2^*) \text{ for any } x_1^* \neq x_2^*$$

2. eigenvectors are columns in  $M_{X|X^*}$ , i.e.,  $f_{X|X^*}(\cdot|x^*)$ . A natural normalization is  $\sum_x f_{X|X^*}(x|x^*) = 1$  for all  $x^*$

3. ordering of the eigenvalues or eigenvectors

That is to reveal the value of  $x^*$  for either  $f_{X|X^*}(\cdot|x^*)$  or  $f(y|x^*)$  from one of below

a.  $x^*$  is the mode of  $f_{X|X^*}(\cdot|x^*)$ : very intuitive, people are more likely to tell the truth; consistent with validation study

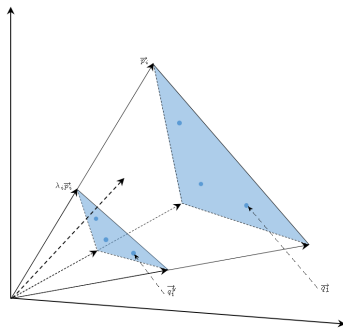
b.  $x^*$  is a quantile of  $f_{X|X^*}(\cdot|x^*)$ : useful in some applications

c.  $x^*$  is the mean of  $f_{X|X^*}(\cdot|x^*)$ : useful when  $x^*$  is continuous

d.  $E(g(y)|x^*)$  is increasing in  $x^*$  for a known  $g$ , say

$$\Pr(y > 0|x^*)$$

## 2.1-measurement model: geometric illustration



### Eigen-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  $Y = 1$ :  
 $\vec{q}_1^Y = \vec{p}_{Y_1, X|z_1} = [f_{Y, X|Z}(1, x_1|z_1), f_{Y, X|Z}(1, x_2|z_1), f_{Y, X|Z}(1, x_3|z_1)]^T$

# Discrete case without ordering conditions: finite mixture

- conditional independence with general discrete  $X$ ,  $Y$ ,  $Z$ , and  $X^*$  (Allman, Matias and Rhodes, 2009, Ann Stat)
- advantages:
  - 1 cardinality of  $X^*$  can be larger than that of  $X$  or  $Z$  or both
  - 2 a lower bound on the so-called Kruskal rank is sufficient for identification up to permutation. (but ordering is innocuous)
- disadvantages:
  - 1 Kruskal rank is hard to interpret in economic models, not testable as regular rank
  - 2 not clear how to extend to the continuous case
- cf. classic local parametric identification condition:  
Number of restrictions  $\geq$  Number of unknowns
- cf. 2.1 measurement model:
  - 1 reach the lower bound on the Kruskal rank:  $2\text{Cardinality}(X^*) + 2$
  - 2 directly extend to the continuous case
  - 3 values of  $X^*$  may have economic meaning

## 2.1-measurement model: continuous case

- $X$ ,  $Z$ , and  $X^*$  are continuous

$$f(y, x, z) = \int f(y|x^*)f(x|x^*)f(x^*, z)dx^*$$

- share the same idea as the discrete case in Hu (2008)
- from matrix to integral operator

diagonal matrix	$\Rightarrow$	“diagonal” operator (multiplication)
matrix diagonalization	$\Rightarrow$	spectral decomposition
eigenvector	$\Rightarrow$	eigenfunction

- nontrivial extension, highly technical
- Hu & Schennach (2008, ECMA)



# From conditional density to integral operator

- From 2-variable function to an integral operator

$$f_{x|x^*}(\cdot|\cdot)$$

$\Downarrow$

$$(L_{x|x^*}g)(x) = \int f_{x|x^*}(x|x^*)g(x^*)dx^* \quad \text{for any } g.$$

- Operator  $L_{x|x^*}$  transforms unobserved  $f_{x^*}$  to observed  $f_x$ , i.e.,  
 $f_x = L_{x|x^*}f_{x^*}$ .

$$\left( \begin{array}{c} f_{x^*}(x^*) \\ \text{distribution of } x^* \end{array} \right) \xrightarrow{L_{x|x^*}} \left( \begin{array}{c} f_x(x) \\ \text{distribution of } x \end{array} \right)$$

- $f_{x|x^*}(\cdot|\cdot)$  is called the *kernel* function of  $L_{x|x^*}$ .

- From matrix to integral operator

$$L_{y;x|z}g = \int f_{y,x|z}(y, \cdot | z) g(z) dz$$

$$L_{x|z}g = \int f_{x|z}(\cdot | z) g(z) dz$$

$$L_{x|x^*}g = \int f_{x|x^*}(\cdot | x^*) g(x^*) dx^*$$

$$L_{x^*|z}g = \int f_{x^*|z}(\cdot | z) g(z) dz$$

$$D_{y;x^*|x^*}g = f_{y|x^*}(y | \cdot) g(\cdot) .$$

- $L_{y;x|z}$  :  $y$  viewed as a fixed parameter.
- $D_{y;x^*|x^*}$  : “diagonal” operator (multiplication by a function).

# Identification: operator equivalence

- The main equation

$$L_{y;x|z} = L_{x|x^*} D_{y;x^*|x^*} L_{x^*|z}.$$

– for a function  $g$ ,

$$\begin{aligned} [L_{y;x|z}g](x) &= \int f_{y,x|z}(y, x|z) g(z) dz \\ &= \int \int f_{x|x^*}(x|x^*) f_{y|x^*}(y|x^*) f_{x^*|z}(x^*|z) dx^* g(z) dz \\ &= \int f_{x|x^*}(x|x^*) f_{y|x^*}(y|x^*) \int f_{x^*|z}(x^*|z) g(z) dz dx^* \\ &= \int f_{x|x^*}(x|x^*) f_{y|x^*}(y|x^*) [L_{x^*|z}g](x^*) dx^* \\ &= \int f_{x|x^*}(x|x^*) [D_{y;x^*|x^*} L_{x^*|z}g](x^*) dx^* \\ &= [L_{x|x^*} D_{y;x^*|x^*} L_{x^*|z}g](x). \end{aligned}$$

- Similarly,

$$L_{x|z} = L_{x|x^*} L_{x^*|z}.$$

# Identification: a necessary condition on error distribution

- Intuition: if  $f_{x|x^*}$  is known, we want  $f_{x^*}$  to be identifiable from  $f_x$ .
  - That is, if  $f_{x^*}$  and  $\tilde{f}_{x^*}$  are observationally equivalent as follows:

$$f_x(x) = \int f_{x|x^*}(x|x^*) f_{x^*}(x^*) dx^* = \int f_{x|x^*}(x|x^*) \tilde{f}_{x^*}(x^*) dx^*,$$

then  $f_{x^*} = \tilde{f}_{x^*}$ .

- In other words, let  $h = f_{x^*} - \tilde{f}_{x^*}$ , we want

$$\int f_{x|x^*}(x|x^*) h(x^*) dx^* = 0 \text{ for all } x \implies h = 0.$$

- An equivalent condition:
  - **Assumption 2(i):**  $L_{x|x^*}$  is injective.
- Implications:
  - Inverse  $L_{x|x^*}^{-1}$  exists on its domain.
  - Assumption 2(i) is implied by *bounded completeness* of  $f_{x|x^*}$ , e.g., exponential family.

# A necessary condition on instrumental variable

- Intuition: same as before

$$\int f_{x^*|z}(x^*|z)h(x^*) dx^* = 0 \text{ for all } z \implies h = 0$$

- Implications:
  - It is equivalent to the injectivity of  $L_{x^*|z}$ .
  - Inverse  $L_{x^*|z}^{-1}$  exists on its domain.
  - Used in Newey & Powell (2003) and Darolles, Florens & Renault (2005).
  - It is a necessary condition to achieve point identification using IV.
  - Implied by the bounded completeness of  $f_{x^*|z}$ , e.g., exponential family.
- Since  $L_{x|z} = L_{x|x^*}L_{x^*|z}$  and  $L_{x|x^*}$  is injective, the injectivity of  $L_{x^*|z}$  is implied by:
  - **Assumption 2(ii):**  $L_{x|z}$  is injective.

# An inherent spectral decomposition

- $L_{x|x^*}^{-1}$  and  $L_{x|z}^{-1}$  exist  
     $\implies$  an inherent spectral decomposition

$$\begin{aligned}L_{y;x|z} L_{x|z}^{-1} &= (L_{x|x^*} D_{y;x^*|x^*} L_{x^*|z}) \times (L_{x|x^*} L_{x^*|z})^{-1} \\ &= L_{x|x^*} D_{y;x^*|x^*} L_{x|x^*}^{-1}.\end{aligned}$$

- An eigenvalue-eigenfunction decomposition of an observed operator on LHS
  - Eigenvalues:  $f_{y|x^*}(y|x^*)$ , kernel of  $D_{y;x^*|x^*}$ .
  - Eigenfunctions:  $f_{x|x^*}(\cdot|x^*)$ , kernel of  $L_{x|x^*}$ .

# Identification: uniqueness of the decomposition

- **Assumption 3:**  $\sup_{y \in \mathcal{Y}} \sup_{x^* \in \mathcal{X}^*} f_{y|x^*}(y|x^*) < \infty$ .  
 $\implies$  boundedness of  $L_{y;x|z} L_{x|z}^{-1}$ , the observed operator on the LHS.
- Theorem XV.4.5 in Dunford & Schwartz (1971):  
*The representation of a bounded linear operator as a “weighted sum of projections” is unique.*
- Each “eigenvalue”  $\lambda = f_{y|x^*}(y|x^*)$  is the weight assigned to the projection onto a linear subspace  $S(\lambda)$  spanned by the corresponding “eigenfunction(s)”  $f_{x|x^*}(\cdot|x^*)$ .
- However, there are ambiguities inside “weighted sum of projections”.  
 $\implies$  We need to “freeze” these degrees of freedom to show that  $L_{x|x^*}$  and  $D_{y;x^*|x^*}$  are uniquely determined by  $L_{y;x|z} L_{x|z}^{-1}$ .

# A close look at weighted sum of projections

- Discrete case:

$$\begin{aligned}L_{y;x|z}L_{x|z}^{-1} &= L_{x|x^*}D_{y;x^*|x^*}L_{x|x^*}^{-1} \\&= f_{y|x^*}(y|x_1) \times L_{x|x^*} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} L_{x|x^*}^{-1} \\&+ f_{y|x^*}(y|x_2) \times L_{x|x^*} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} L_{x|x^*}^{-1} \\&+ f_{y|x^*}(y|x_3) \times L_{x|x^*} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} L_{x|x^*}^{-1}\end{aligned}$$

- Continuous case:

$$L_{y;x|z}L_{x|z}^{-1} = \int_{\sigma} \lambda P(d\lambda)$$



# Identification: uniqueness of the decomposition

- **Ambiguity I:** Eigenfunctions  $f_{x|x^*}(\cdot|x^*)$  are defined only up to a constant:
  - Solution: Constant determined by  $\int f_{x|x^*}(x|x^*) dx = 1$ .
  - Intuition: Eigenfunctions are conditional densities, therefore, are automatically normalized.
- **Ambiguity II:** If  $\lambda$  is a degenerate eigenvalue, more than one possible eigenfunctions.
  - Solution: **Assumption 4:** for all  $x_1^*, x_2^* \in \mathcal{X}^*$ , the set

$$\{y : f_{y|x^*}(y|x_1^*) \neq f_{y|x^*}(y|x_2^*)\}$$

*has positive probability whenever  $x_1^* \neq x_2^*$ .*

- Intuition: eigenvalues  $f_{y|x^*}(y_1|x^*)$  and  $f_{y|x^*}(y_2|x^*)$  share the same eigenfunction  $f_{x|x^*}(\cdot|x^*)$ . Therefore,  $y$  is helpful to distinguish eigenfunctions.
- Note: this assumption is weaker than (or implied by) the monotonicity assumptions typically made in the nonseparable error literature

# Identification: uniqueness of the decomposition

- **Ambiguity III:** Freedom in indexing eigenvalues: e.g., use  $x^*$  or  $(x^*)^3$ ?
  - Solution: the zero “location” assumption, i.e., **Assumption 5:** *there exists a known functional  $M$  such that  $x^* = M[f_{x|x^*}(\cdot|x^*)]$  for all  $x^*$ .*
  - Intuition: Consider another variable  $\tilde{x}^*$  related to  $x^*$  by  $\tilde{x}^* = R(x^*)$ .
    - $\implies M[f_{x|\tilde{x}^*}(\cdot|\tilde{x}^*)] = M[f_{x|x^*}(\cdot|R(\tilde{x}^*))] = R(\tilde{x}^*) \neq \tilde{x}^*$ .
    - $\implies$  Only one possible  $R$ : the identity function.
- Examples of  $M$ 
  - error has a zero mean:  $M[f] = \int xf(x)dx$  (thus, allow classical error)
  - error has a zero mode:  $M[f] = \arg \max_x f(x)$
  - error has a zero  $\tau$ -th quantile:  $M[f] = \inf \{x^* : \int 1(x \leq x^*) f(x)dx \geq \tau\}$
- Importance: this assumption is based on the findings from validation studies.

## 2.1-measurement model: continuous case

- key identification conditions:
  - 1) all densities are bounded
  - 2) the operators  $L_{X|X^*}$  and  $L_{Z|X}$  are injective.
  - 3) 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.
  - 4) there exists a known functional  $M$  such that  $M[f_{X|X^*}(\cdot|x^*)] = x^*$  for all  $x^* \in \mathcal{X}^*$ .

- then

$f_{X,Y,Z}$  uniquely determines  $f_{X,Y,Z,X^*}$

with

$$f_{X,Y,Z,X^*} = f_{X|X^*} f_{Y|X^*} f_{Z|X} f_{X^*}$$

- a global nonparametric point identification

## 3-measurement model

- definition: three measurements  $X$ ,  $Y$ , and  $Z$  satisfy

$$X \perp Y \perp Z \mid X^*$$

- can always be reduced to a 2.1-measurement model.  
all the identification conditions remain with a general  $\mathcal{Y}$ .
- doesn't matter which is called dependent variable, measurement, or instrument.

- examples:

Hausman Newey & Ichimura (1991)

add  $x^* = \gamma z + u$ ,  $z$  instrument,  $g(\cdot)$  is a polynomial

Schennach (2004): use a repeated measurement  $x_2 = x^* + \varepsilon_2$

general  $g(\cdot)$ , use ch.f. Kotlarski's identity

Schennach (2007): use IV:  $x^* = \gamma z + u$   $u \perp z$

general  $g(\cdot)$ , use ch.f. similar to Kotlarski's identity

# Hidden Markov model: a 3-measurement model

- an unobserved Markov process

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

- a measurement  $X_t$  of the latent  $X_t^*$  satisfying

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

- a hidden Markov model

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

- a 3-measurement model

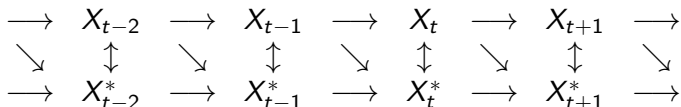
$$X_{t-1} \perp X_t \perp X_{t+1} \mid X_t^*,$$

# dynamic measurement model

- $\{X_t, X_t^*\}$  is a first-order Markov process satisfying

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

- Flow of chart



- Hu & Shum (2012, JE): nonparametric identification of the joint process
- Special case with  $X_t^* = X_{t-1}^*$  needs 4 periods of data.  
cf. 6 periods in Kasahara and Shimotsu (2009)

- Hu & Shum (2012): nonparametric identification of the joint process. (use Carroll Chen & Hu (2010, JNPS))
- key identification assumptions:
  - 1) for any  $x_{t-1} \in \mathcal{X}$ ,  $M_{X_t|X_{t-1}, X_{t-2}}$  is invertible.
  - 2) 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  $x_t^* \neq \bar{x}_t^*$  in  $\mathcal{X}^*$ 
$$\Delta_{x_t} \Delta_{x_{t-1}} \ln f_{X_t|X_t^*, X_{t-1}}(x_t^*) \neq \Delta_{x_t} \Delta_{x_{t-1}} \ln f_{X_t|X_t^*, X_{t-1}}(\bar{x}_t^*)$$
  - 3) for any  $x_t \in \mathcal{X}$ ,  $E[X_{t+1}|X_t = x_t, X_t^* = x_t^*]$  is increasing in  $x_t^*$ .
- joint distribution of five periods of data  $f_{X_{t+1}, X_t, X_{t-1}, X_{t-2}, X_{t-3}}$  uniquely determines Markov transition kernel  $f_{X_t, X_t^*|X_{t-1}, X_{t-1}^*}$

## Other approaches: use a secondary sample

- $\{Y, X\}, \{X^*\}$  (administrative sample) Hu & Ridder (2012)
- $\{Y, X\}, \{X, X^*\}$  (validation sample) Chen, Hong & Tamer (2005) among many other papers in econometrics & statistics
- $\{Y, X, W\}, \{Y_a, X_a, W_a\}$  (auxiliary survey sample) Carroll, Chen & Hu (2010) with model of interest  $f(Y|X^*, W) = f(Y_a|X_a^*, W_a)$
- also related to literature on missing data, where  $X^*$  can be considered as missing



- Estimate the matrices directly

$$L_{y;x,z} = \begin{pmatrix} f_{y;x|z}(y, x_1, z_1) & f_{y;x|z}(y, x_1, z_2) & f_{y;x|z}(y, x_1, z_3) \\ f_{y;x|z}(y, x_2, z_1) & f_{y;x|z}(y, x_2, z_2) & f_{y;x|z}(y, x_2, z_3) \\ f_{y;x|z}(y, x_3, z_1) & f_{y;x|z}(y, x_3, z_2) & f_{y;x|z}(y, x_3, z_3) \end{pmatrix}$$

- Use sample proportion
- Use kernel density estimator with continuous covariates
- Identification is globe, nonparametric, and constructive
- Mimic identification procedure:
  - a unique mapping from  $f_{y,x,z}$  to  $f_{y|x^*}$ ,  $f_{x|x^*}$ , and  $f_{x^*,z}$
- Easy to compute without optimization or iteration
- May have problems with a small sample: estimated prob outside [0,1]

## Estimation: discrete case

- Eigen decomposition holds after averaging over  $Y$  with a known  $\omega(\cdot)$

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

- Define

$$M_{X,\omega,Z} = [E[\omega(Y) | X = x_k, Z = z_l] f_{X,Z}(x_k, z_l)]_{k=1,2,\dots,K; l=1,2,\dots,K}$$
$$D_{\omega|X^*} = \text{diag}\{E[\omega(Y) | x_1^*], E[\omega(Y) | x_2^*], \dots, E[\omega(Y) | x_K^*]\}$$

- 

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

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

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

- Estimation mimics identification procedure

- May also use extremum estimator with restrictions

$$\left( \widehat{M}_{X|X^*}, \widehat{D}_{\omega|X^*} \right) = \arg \min_{M,D} \left\| \widehat{M}_{X,\omega,Z} \left( \widehat{M}_{X,Z} \right)^{-1} M - M \times D \right\|$$

such that

- 1) each entry in  $M$  is in  $[0, 1]$
  - 2) each column sum of  $M$  equals 1
  - 3)  $D$  is diagonal
  - 4) entries in  $M$  satisfies the ordering Assumption
- See Bonhomme et al. (2015, 2016) for more extremum estimators

- Global nonparametric identification elements of interest can be written as a function of observed distributions
  - continuous case: Kotlarski's identity
  - nonparametric regression with measurement error: Schennach (2004b, 2007), Hu and Sasaki (2015)
  - discrete case: eigen-decomposition in Hu (2008)
- Closed-form estimator
  - mimic identification procedure
  - don't need optimization or iteration
  - less nuisance parameters than semiparametric estimators
  - but may not be efficient

- a 3-measurement model

$$x_1 = g_1(x^*) + \epsilon_1$$

$$x_2 = g_2(x^*) + \epsilon_2$$

$$x_3 = g_3(x^*) + \epsilon_3$$

- normalization:  $g_3(x^*) = x^*$
- Schennach (2004b):  $g_2(x^*) = x^*$
- Hu and Sasaki (2015):  $g_2$  is a polynomial
- Hu and Schennach (2008):  $g_1$  and  $g_2$  are nonparametrically identified
- Open question: Do closed-form estimators for  $g_1$  and  $g_2$  exist?

# Estimation: a sieve semiparametric MLE

- Based on :

$$f_{y,x|z}(y, x|z) = \int f_{y|x^*}(y|x^*) f_{x|x^*}(x|x^*) f_{x^*|z}(x^*|z) dx^*$$

- Approximate  $\infty$ -dimensional parameters, e.g.,  $f_{x|x^*}$ , by truncated series

$$\hat{f}_1(x|x^*) = \sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \hat{\gamma}_{ij} p_i(x) p_j(x^*),$$

– where  $p_k(\cdot)$  are a sequence of known univariate basis functions.

- Sieve Semiparametric MLE

$$\begin{aligned} \hat{\alpha} &= (\hat{\beta}, \hat{\eta}, \hat{f}_1, \hat{f}_2) \\ &= \arg \max_{(\beta, \eta, f_1, f_2) \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^n \ln \int f_{y|x^*}(y_i|x^*; \beta, \eta) f_1(x_i|x^*) f_2(x^*|z_i) dx^* \end{aligned}$$

$$\left\{ \begin{array}{ll} \beta : & \text{parameter vector of interest} \\ \eta, f_1, f_2 : & \infty\text{-dimensional nuisance parameters} \\ \mathcal{A}_n : & \text{space of series approximations} \end{array} \right.$$

# Estimation: handling moment conditions

- Use  $\eta$  to handle moment conditions:
  - For parametric likelihoods: omit  $\eta$ .
  - For moment condition models: need  $\eta$ .
- Model defined by:

$$E [m (y, x^*, \beta) | x^*] = 0.$$

- Method:
  - Define a family of densities  $f_{y|x^*} (y|x^*, \beta, \eta)$  such that

$$\int m (y, x^*, \beta) f_{y|x^*} (y|x^*, \beta, \eta) dx^* = 0, \quad \forall x^*, \beta, \eta.$$

- Use sieve MLE

$$\begin{aligned} \hat{\alpha} &= (\hat{\beta}, \hat{\eta}, \hat{f}_1, \hat{f}_2) \\ &= \arg \max_{(\beta, \eta, f_1, f_2) \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^n \ln \int f_{y|x^*} (y_i|x^*; \beta, \eta) f_1(x_i|x^*) f_2(x^*|z_i) dx^*. \end{aligned}$$

# Estimation: consistency and normality

- Consistency of  $\hat{\alpha}$ 
  - Conditions: too technical to show here.
  - **Theorem (consistency)**: *Under sufficient conditions, we have*

$$\|\hat{\alpha} - \alpha_0\|_s = o_p(1).$$

- Proof: use Theorem 4.1 in Newey and Powell (2003).

- Asymptotic normality of parameters of interest  $\hat{\beta}$ .
  - Conditions: even more technical.
  - **Theorem (normality)**: *Under sufficient conditions, we have*

$$\sqrt{n} \left( \hat{\beta} - \beta_0 \right) \xrightarrow{d} N \left( 0, J^{-1} \right).$$

- Proof: use Theorem 1 in Shen (1997) and Chen and Shen (1998).



# Empirical applications with latent variables

- auctions with unknown number of bidders
- auctions with unobserved heterogeneity
- auctions with heterogeneous beliefs
- multiple equilibria in incomplete information games
- dynamic learning models
- unemployment and labor market participation
- cognitive and noncognitive skill formation
- dynamic discrete choice with unobserved state variables
- two-sided matching
- income dynamics

# First-price sealed-bid auctions

- Bidder  $i$  forms her own valuation of the object:  $x_i$ 
  - Bidders' values are private and independent
  - Common knowledge: value distribution  $F$ , number of bidders  $N^*$
- Bidder  $i$  chooses bid  $b_i$  to maximize her expected utility function

$$U_i = (x_i - b_i) \Pr(\max_{j \neq i} b_j < b_i)$$

- Winning probability  $\Pr(\max_{j \neq i} b_j < b_i)$  depends on bidder  $i$ 's belief about her opponents' bidding behavior
- Perfectly correct beliefs about opponents' bidding behavior  
→ Nash equilibrium

# Auctions with unknown number of bidders

- An Hu & Shum (2010, JE):

$$\text{IPV auction model: } \begin{cases} N^*: \# \text{ of potential bidders} \\ A: \# \text{ of actual bidders} \\ b: \text{ observed bids} \end{cases}$$

- bid function

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

- conditional independence

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

# Auctions with unobserved heterogeneity

- $s_t^*$  is an auction-specific state or unobserved heterogeneity

$$b_{it} = s_t^* \times a_i(x_i)$$

- 2-measurement model

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

and

$$\ln b_{1t} = \ln s_t^* + \ln a_1$$

$$\ln b_{2t} = \ln s_t^* + \ln a_2$$

- in general

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

- Li Perrigne & Vuong (2000), Krasnokutskaya (2011), Hu McAdams & Shum (2013 JE)

# Auctions with heterogeneous beliefs

- An (2016): empirical analysis on Level- $k$  belief in auctions
- Bidders have different levels of sophistication  $\Rightarrow$  Heterogeneous (possibly incorrect) beliefs about others' behavior
- Beliefs (types) have a hierarchical structure

Type	Belief about other bidders' behavior
1	all other bidders are type- $L_0$ (bid naïvely)
2	all other bidders are type-1
$\vdots$	$\vdots$
$k$	all other bidders are type- $(k - 1)$

- Specification of type- $L_0$  is crucial, assumed by the researchers
- Help explain overbidding and non-equilibrium behavior
- Observe joint distribution of a bidder's bids in three auctions, assuming bidder's belief level doesn't change across auctions
- three bids are independent conditional on belief level

# Multiple equilibria in incomplete information games

- Xiao (2014): a static simultaneous move game
- utility function

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

- expected payoff of player  $i$  from choosing action  $a_i$

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

- Bayesian Nash Equilibrium is defined as a set of choice probabilities  $\Pr(a_i)$  s.t.

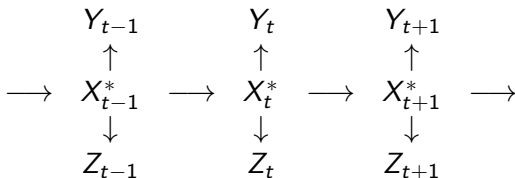
$$\Pr(a_i = k) = \Pr\left(\left\{\Pi_i(k) + \epsilon_i(k) > \max_{j \neq k} \Pi_i(j) + \epsilon_i(j)\right\}\right)$$

- let  $e^*$  denote the index of equilibria

$$a_1 \perp a_2 \perp \dots \perp a_N \mid e^*$$

# Dynamic learning models

- Hu Kayaba & Shum (2013 GEB): observe choices  $Y_t$ , rewards  $R_t$ , proxy  $Z_t$  for the agent's belief  $X_t^*$
- $Z_t$ : eye movement



- a 3-measurement model

$$Z_t \perp Y_t \perp Z_{t-1} \mid X_t^*$$

- learning rule  $\Pr(X_{t+1}^* | X_t^*, Y_t, R_t)$  can be identified from

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

# Unemployment and labor market participation

- Feng & Hu (2013 AER): Let  $X_t^*$  and  $X_t$  denote the true and self-reported labor force status.
- monthly CPS  $\{X_{t+1}, X_t, X_{t-9}\}_i$ ;
- local independence

$$\Pr(X_{t+1}, X_t, X_{t-9}) = \sum_{X_{t+1}^*} \sum_{X_t^*} \sum_{X_{t-9}^*} \Pr(X_{t+1}|X_{t+1}^*) \times \\ \times \Pr(X_t|X_t^*) \Pr(X_{t-9}|X_{t-9}^*) \Pr(X_{t+1}^*, X_t^*, X_{t-9}^*).$$

- assume

$$\Pr(X_{t+1}^*|X_t^*, X_{t-9}^*) = \Pr(X_{t+1}^*|X_t^*)$$

- a 3-measurement model

$$\Pr(X_{t+1}, X_t, X_{t-9}) \\ = \sum_{X_t^*} \Pr(X_{t+1}|X_t^*) \Pr(X_t|X_t^*) \Pr(X_{t-9}^*, X_t^*),$$



# Cognitive and noncognitive skill formation

- Cunha Heckman & Schennach (2010 ECMA)  
 $X_t^* = (X_{C,t}^*, X_{N,t}^*)$  cognitive and noncognitive skill  
 $I_t = (I_{C,t}, I_{N,t})$  parental investments
- for  $k \in \{C, N\}$ , skills evolve as

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

where  $X_P^* = (X_{C,P}^*, X_{N,P}^*)$  are parental skills

- latent factors

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

- measurements of these factors

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

- key identification assumption

$$X_1 \perp X_2 \perp X_3 \mid X^*$$

- a 3-measurement model

- Hu & Shum (2012 JE)
- $W_t = (Y_t, M_t)$ 
  - $Y_t$  agent's choice in period  $t$
  - $M_t$  observed state variable
  - $X_t^*$  unobserved state variable
- for Markovian dynamic optimization models

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

$f_{Y_t | M_t, X_t^*}$  conditional choice probability for the agent's optimal  
 $f_{M_t, X_t^* | Y_{t-1}, M_{t-1}, X_{t-1}^*}$  joint law of motion of state variables

- $f_{W_{t+1}, W_t, W_{t-1}, W_{t-2}}$  uniquely determines  $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$

# Two-sided matching model

- Agarwal & Diamond (2013): an economy containing  $n$  workers with characteristics  $(X_i, \varepsilon_i)$  and  $n$  firms described by  $(Z_j, \eta_j)$
- researchers observe  $X_i$  and  $Z_j$
- a firm ranks workers by a human capital index as

$$v(X_i, \varepsilon_i) = h(X_i) + \varepsilon_i. \quad (1)$$

- the workers' preference for firm  $j$  is described by

$$u(Z_j, \eta_j) = g(Z_j) + \eta_j. \quad (2)$$

- 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 pairwise stable equilibrium, where no two agents on opposite sides of the market prefer each other over their matched partners.

# Two-sided matching model

- when the numbers of firms and workers are both large, The joint distribution of  $(X, Z)$  from observed pairs then satisfies

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

$$f(X|q) = f_\varepsilon(F_V^{-1}(q) - h(X))$$

$$f(Z|q) = f_\eta(F_U^{-1}(q) - g(Z))$$

a 2-measurement model

- $h$  and  $g$  may be identified up to a monotone transformation.  
intuition:  $f_{Z|X}(z|x_1) = f_{Z|X}(z|x_2)$  for all  $z$  implies  $h(x_1) = h(x_2)$
- in many-to-one matching

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

a 3-measurement model

- Arellano Blundell & Bonhomme (2014): nonlinear aspect of income dynamics
- pre-tax labor income  $y_{it}$  of household  $i$  at age  $t$

$$y_{it} = \eta_{it} + \varepsilon_{it}$$

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

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

- transitory component  $\varepsilon_{it}$  is independent over time
- $\{y_{it}, \eta_{it}\}$  is a hidden Markov process with

$$y_{i,t-1} \perp y_{it} \perp y_{i,t+1} \mid \eta_{it}$$

- a 3-measurement model

# A canonical model of income dynamics: a revisit

- Permanent income: a random walk process
- Transitory income: an ARMA process

$$x_t = x_t^* + v_t$$

$$x_t^* = x_{t-1}^* + \eta_t$$

$$v_t = \rho_t v_{t-1} + \lambda_t \epsilon_{t-1} + \epsilon_t$$

$$\left\{ \begin{array}{l} \eta_t : \text{ permanent income shock in period } t \\ \epsilon_t : \text{ transitory income shock} \\ x_t^* : \text{ latent permanent income} \\ v_t : \text{ latent transitory income} \end{array} \right.$$

- Can a sample of  $\{x_t\}_{t=1, \dots, T}$  uniquely determine distributions of latent variables  $\eta_t$ ,  $\epsilon_t$ ,  $x_t^*$ , and  $v_t$ ?

# A canonical model of income dynamics: a revisit

- Define

$$\Delta x_{t+1} = x_{t+1} - x_t$$

- Estimate AR coefficient

$$\rho_{t+1} \frac{1 - \rho_{t+2}}{1 - \rho_{t+1}} = \frac{\text{COV}(\Delta x_{t+2}, x_{t-1})}{\text{COV}(\Delta x_{t+1}, x_{t-1})}$$

- Use Kotlarski's identity

$$x_t = v_t + x_t^*$$
$$\frac{\Delta x_{t+2}}{\rho_{t+2} - 1} - \Delta x_{t+1} = v_t + \frac{\lambda_{t+2}\epsilon_{t+1} + \epsilon_{t+2} + \eta_{t+2}}{\rho_{t+2} - 1} - \eta_{t+1}$$

- Joint distribution of  $\{x_t\}_{t=1, \dots, T} \geq 3$  uniquely determines distributions of latent variables  $\eta_t$ ,  $\epsilon_t$ ,  $x_t^*$ , and  $v_t$ . (Hu, Moffitt, and Sasaki, 2016)

## *The Econometrics of Unobservables*

- a solution to the endogeneity problem
- integration of microeconomic theory and econometric methodology
- economic theory motivates our intuitive assumptions
- global nonparametric point identification and estimation
- flexible nonparametrics applies to large range of economic models
- latent variable approach allows researchers to go beyond observables



See my review paper (Hu, 2017) for details

“The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics”

at [▶ Yingyao Hu's webpage](#)

Comments are welcome. Thank you for your interest.