Economic theory vs. econometric model: an example

- Economic theory: Permanent income hypothesis
- Econometric model: Measurement error model

\[ y = \beta x^* + e \]
\[ x = x^* + \nu \]

\[ \begin{cases} 
  y : & \text{observed consumption} \\
  x : & \text{observed income} \\
  x^* : & \text{latent permanent income} \\
  \nu : & \text{latent transitory income} \\
  \beta : & \text{marginal propensity to consume} 
\end{cases} \]

- 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

\[ x_t = x_t^* + v_t \]
\[ x_t^* = x_{t-1}^* + \eta_t \]
\[ v_t = \rho_t v_{t-1} + \lambda_t \epsilon_{t-1} + \epsilon_t \]

\[ \eta_t : \text{permanent income shock in period } t \]
\[ \epsilon_t : \text{transitory income shock} \]
\[ x_t^* : \text{latent permanent income} \]
\[ v_t : \text{latent transitory income} \]

- Can a sample of \( \{x_t\}_{t=1,...,T} \) uniquely determine distributions of latent variables \( \eta_t, \epsilon_t, x_t^*, \) and \( v_t \)?
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
   - Effort and type in contract models
   - Unemployment and labor market participation
   - Cognitive and noncognitive skill formation
   - Matching models with latent indices
   - Income dynamics
4. conclusion
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$^+$ |
| $x_1$—no college | 0.876 | 0.111 | 0.000 |
| $x_2$—some college | 0.112 | 0.772 | 0.020 |
| $x_3$—BA$^+$ | 0.012 | 0.117 | 0.980 |

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

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

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

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

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

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

Graphical illustration of zero-mode measurement error

$f(x|x^*)$
## Latent variables in microeconomic models

<table>
<thead>
<tr>
<th>Empirical Models</th>
<th>Unobservables</th>
<th>Observables</th>
</tr>
</thead>
<tbody>
<tr>
<td>measurement error</td>
<td>true earnings</td>
<td>self-reported earnings</td>
</tr>
<tr>
<td>consumption function</td>
<td>permanent income</td>
<td>observed income</td>
</tr>
<tr>
<td>production function</td>
<td>productivity</td>
<td>output, input</td>
</tr>
<tr>
<td>wage function</td>
<td>ability</td>
<td>test scores</td>
</tr>
<tr>
<td>learning model</td>
<td>belief</td>
<td>choices, proxy</td>
</tr>
<tr>
<td>auction model</td>
<td>unobserved heterogeneity</td>
<td>bids</td>
</tr>
<tr>
<td>contract model</td>
<td>effort, type</td>
<td>outcome, state var.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Yingyao Hu (JHU)  Econometrics of Unobservables  2019  10 / 80
Our definition of measurement

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

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

- 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 $X^*$

<table>
<thead>
<tr>
<th>$X$</th>
<th>$X^*$</th>
<th>$L \geq K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>discrete ${x_1, x_2, ..., x_L}$</td>
<td>discrete ${x_1^<em>, x_2^</em>, ..., x_K^*}$</td>
<td>$L \geq K$</td>
</tr>
<tr>
<td>continuous</td>
<td>discrete ${x_1^<em>, x_2^</em>, ..., x_K^*}$</td>
<td>$L \geq K$</td>
</tr>
<tr>
<td>continuous</td>
<td>continuous</td>
<td>$L \geq K$</td>
</tr>
</tbody>
</table>

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

- observed & unobserved variables

<table>
<thead>
<tr>
<th>$X$</th>
<th>measurement</th>
<th>observables</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^*$</td>
<td>latent true variable</td>
<td>unobservables</td>
</tr>
</tbody>
</table>

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

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

- $f_{X^*}$ : 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 $X \in \{x_1, x_2, \ldots, x_L\}$ and $X^* \in \mathcal{X}^* = \{x_1^*, x_2^*, \ldots, x_K^*\}$

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

- matrix expression

  \[
  \begin{align*}
  \vec{\rho}_X &= \begin{bmatrix} f_X(x_1), f_X(x_2), \ldots, f_X(x_L) \end{bmatrix}^T \\
  \vec{\rho}_{X^*} &= \begin{bmatrix} f_{X^*}(x_1^*), f_{X^*}(x_2^*), \ldots, f_{X^*}(x_K^*) \end{bmatrix}^T \\
  M_{X|X^*} &= \begin{bmatrix} f_{X|x^*}(x_l|x_k^*) \end{bmatrix}_{l=1,2,\ldots,L;k=1,2,\ldots,K} \\
  \end{align*}
  \]

  $$\vec{\rho}_X = M_{X|X^*} \vec{\rho}_{X^*}.$$ 

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

  $$\text{Rank} \left( M_{X|X^*} \right) = \text{Cardinality} \left( \mathcal{X}^* \right)$$
two possible marginal distributions $\overrightarrow{p}_X^{a}$ and $\overrightarrow{p}_X^{b}$ are observationally equivalent, i.e.,

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

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

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

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 \( f_{X^*} \)

\[
L_{1\text{bnd}}(X^*) = \left\{ h : \int_{X^*} |h(x^*)| \, dx^* < \infty \text{ and } \sup_{x^* \in X^*} |h(x^*)| < \infty \right\}
\]

- define a linear operator

\[
L_{X|X^*} : L_{1\text{bnd}}(X^*) \to L_{1\text{bnd}}(X)
\]

\[
(L_{X|X^*} h)(x) = \int_{X^*} f_{X|X^*}(x|x^*) h(x^*) \, dx^*
\]

- operator equation

\[
f_X = L_{X|X^*} f_{X^*}
\]

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

\[
L_{X|X^*} h = 0 \text{ implies } h = 0 \text{ for any } h \in L_{1\text{bnd}}(X^*)
\]
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
  \[
  f_{X,Z}(x, z) = \sum_{x^* \in \mathcal{X}^*} f_{X|x^*}(x \mid x^*) f_{Z|x^*}(z \mid x^*) f_{X^*}(x^*)
  \]

- matrix expression
  \[
  M_{X,Z} = \begin{bmatrix} f_{X,Z}(x_l, z_j) \end{bmatrix}_{l=1,2,\ldots,L; j=1,2,\ldots,J} \\
  M_{Z|X^*} = \begin{bmatrix} f_{Z|X^*}(z_j \mid x_k^*) \end{bmatrix}_{j=1,2,\ldots,J; k=1,2,\ldots,K} \\
  D_{X^*} = diag \{ f_{X^*}(x_1^*), f_{X^*}(x_2^*), \ldots, f_{X^*}(x_K^*) \}
  \]
  \[
  M_{X,Z} = M_{X|X^*} D_{X^*} M_{Z|X^*}^T
  \]

- 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

\[ Y = \beta X^* + \eta \]

\[(X, X^*) \perp \eta \]

\[ X, X^* \in \{0, 1\} \]

- measurement error \( X - X^* \) is correlated with \( X^* \) in general

- \( f(y|x) \) is a mixture of \( f_\eta(y) \) and \( f_\eta(y - \beta) \)

\[
f(y|x) = \sum_{x^*=0}^1 f(y|x^*) f_{X^*|X}(x^*|x)
\]

\[= f_\eta(y) f_{X^*|X}(0|x) + f_\eta(y - \beta) f_{X^*|X}(1|x)
\]

\[\equiv f_\eta(y) P_x + f_\eta(y - \beta)(1 - P_x)\]
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
  \[ \text{sign} \{ \beta \} = \text{sign} \{ E(y|x = 1) - E(y|x = 0) \} \]
measurement error causes attenuation
a discrete latent regressor

\[ y = \beta x^* + \eta \]

\[(X, X^*) \perp \eta\]

\[ X, X^* \in \{x_1^*, x_2^*, \ldots, x_K^*\} \]

Chen Hu & Lewbel (2009): point identification generally holds

general models without \((X, X^*) \perp \eta\): partial identification

see Bollinger (1996) and Molinari (2008)
2-measurement model: linear model with classical error

- a simple linear regression model with zero means

\[ Y = \beta X^* + \eta \]
\[ X = X^* + \epsilon \]
\[ X^* \perp \epsilon \perp \eta \]

- \( \beta \) 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: \( \beta = 1 \)

\[
\begin{align*}
Y &= X^* + \eta \\
X &= X^* + \varepsilon
\end{align*}
\]

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

\[
f_{X^*}(x^*) = \frac{1}{2\pi} \int e^{-ix^*t} \Phi_{X^*}(t) dt \\
\Phi_{X^*} = E \left[ e^{itX^*} \right]
\]
2-measurement model: Kotlarski’s identity

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

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

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

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

- Kotlarski’s identity (1966)

$$
\Phi_{X^*}(t) = \exp \left[ \int_0^t \frac{\varepsilon E \left[ Ye^{isX} \right]}{Ee^{isX}} ds \right]
$$
2-measurement model: Kotlarski’s identity

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

\[
\begin{align*}
Y &= X^* + \eta \\
X &= X^* + \varepsilon
\end{align*}
\]

- 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)

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

- 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)

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

- 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

\[ Y = g(X^*) + \eta \]
\[ X = X^* + \epsilon \]
\[ X^* \perp \epsilon \perp \eta \]

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

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

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

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

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

- a global nonparametric point identification (exact identification if $J = K = L$)
Let $x, x^* \in \{x_1, x_2, x_3\}$ and $z \in \{z_1, z_2, z_3\}$, e.g., education levels.

\[
M_{x|x^*} = \begin{pmatrix}
n_{x|x^*}(x_1|x_1) & n_{x|x^*}(x_1|x_2) & n_{x|x^*}(x_1|x_3) \\
n_{x|x^*}(x_2|x_1) & n_{x|x^*}(x_2|x_2) & n_{x|x^*}(x_2|x_3) \\
n_{x|x^*}(x_3|x_1) & n_{x|x^*}(x_3|x_2) & n_{x|x^*}(x_3|x_3)
\end{pmatrix} \quad \Leftarrow \text{error structure}
\]

\[
M_{x^*|z} = \begin{pmatrix}
n_{x^*|z}(x_1|z_1) & n_{x^*|z}(x_1|z_2) & n_{x^*|z}(x_1|z_3) \\
n_{x^*|z}(x_2|z_1) & n_{x^*|z}(x_2|z_2) & n_{x^*|z}(x_2|z_3) \\
n_{x^*|z}(x_3|z_1) & n_{x^*|z}(x_3|z_2) & n_{x^*|z}(x_3|z_3)
\end{pmatrix} \quad \Leftarrow \text{IV structure}
\]

\[
D_{y|x^*} = \begin{pmatrix}
n_{y|x^*}(y|x_1) & 0 & 0 \\
0 & n_{y|x^*}(y|x_2) & 0 \\
0 & 0 & n_{y|x^*}(y|x_3)
\end{pmatrix} \quad \Leftarrow \text{latent model}
\]

\[
M_{y;x|z} = \begin{pmatrix}
n_{y;x|z}(y,x_1|z_1) & n_{y;x|z}(y,x_1|z_2) & n_{y;x|z}(y,x_1|z_3) \\
n_{y;x|z}(y,x_2|z_1) & n_{y;x|z}(y,x_2|z_2) & n_{y;x|z}(y,x_2|z_3) \\
n_{y;x|z}(y,x_3|z_1) & n_{y;x|z}(y,x_3|z_2) & n_{y;x|z}(y,x_3|z_3)
\end{pmatrix} \quad \Leftarrow \text{observed info.}
\]

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

- The main equation for a given \( y \)

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

\[\Leftrightarrow\]

\[
M_{y;x|z} = M_{x|x^*} D_{y|x^*} M_{x^*|z}
\]
Matrix equivalence

- The main equation for a given $y$

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

\[\iff\]

\[
 M_{y;x|z} = M_{x|x^*} D_{y|x^*} M_{x^*|z}
\]

- Similarly,

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

\[\iff\]

\[
 M_{x|z} = M_{x|x^*} M_{x^*|z}
\]
Matrix equivalence

- The main equation for a given $y$

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

$$\iff$$

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

- Similarly,

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

$$\iff$$

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

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

$$M_{y;x|z} M_{x|x^*}^{-1} = (M_{x|x^*} D_{y|x^*} M_{x^*|z}) \times (M_{x^*|z}^{-1} M_{x|x^*}^{-1})$$

$$= M_{x|x^*} D_{y|x^*} M_{x|x^*}^{-1}.$$
An inherent matrix diagonalization

- An eigenvalue-eigenvector decomposition:

\[
M_{y|x|z}M_{x|z}^{-1} = M_{x|x^*}D_{y|x^*}M_{x|x^*}^{-1}
\]

\[
= \begin{pmatrix}
    f_{x|x^*}(x_1|x_1) & f_{x|x^*}(x_1|x_2) & f_{x|x^*}(x_1|x_3) \\
    f_{x|x^*}(x_2|x_1) & f_{x|x^*}(x_2|x_2) & f_{x|x^*}(x_2|x_3) \\
    f_{x|x^*}(x_3|x_1) & f_{x|x^*}(x_3|x_2) & f_{x|x^*}(x_3|x_3)
\end{pmatrix}
\times
\begin{pmatrix}
    f_{y|x^*}(y|x_1) & 0 & 0 \\
    0 & f_{y|x^*}(y|x_2) & 0 \\
    0 & 0 & f_{y|x^*}(y|x_3)
\end{pmatrix}
\times
\begin{pmatrix}
    f_{x|x^*}(x_1|x_1) & f_{x|x^*}(x_1|x_2) & f_{x|x^*}(x_1|x_3) \\
    f_{x|x^*}(x_2|x_1) & f_{x|x^*}(x_2|x_2) & f_{x|x^*}(x_2|x_3) \\
    f_{x|x^*}(x_3|x_1) & f_{x|x^*}(x_3|x_2) & f_{x|x^*}(x_3|x_3)
\end{pmatrix}^{-1}
\]

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

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

\[
\{\clubsuit, \heartsuit, \spadesuit\} \xrightarrow{1\text{-to-}1} \{x_1, x_2, x_3\}
\]

- Decompositions with different indexing are observationally equivalent,

\[
M_{y:x|z} M_{x|z}^{-1} = M_{x|x^*} D_{y|x^*} M_{x|x^*}^{-1}
\]

\[
= \begin{pmatrix}
f_{x|x^*}(x_1|\clubsuit) & f_{x|x^*}(x_1|\heartsuit) & f_{x|x^*}(x_1|\spadesuit) \\
f_{x|x^*}(x_2|\clubsuit) & f_{x|x^*}(x_2|\heartsuit) & f_{x|x^*}(x_2|\spadesuit) \\
f_{x|x^*}(x_3|\clubsuit) & f_{x|x^*}(x_3|\heartsuit) & f_{x|x^*}(x_3|\spadesuit)
\end{pmatrix}
\]

\[
\times \begin{pmatrix}
f_{y|x^*}(y|\clubsuit) & 0 & 0 \\
0 & f_{y|x^*}(y|\heartsuit) & 0 \\
0 & 0 & f_{y|x^*}(y|\spadesuit)
\end{pmatrix}
\]

\[
\times \begin{pmatrix}
f_{x|x^*}(x_1|\clubsuit) & f_{x|x^*}(x_1|\heartsuit) & f_{x|x^*}(x_1|\spadesuit) \\
f_{x|x^*}(x_2|\clubsuit) & f_{x|x^*}(x_2|\heartsuit) & f_{x|x^*}(x_2|\spadesuit) \\
f_{x|x^*}(x_3|\clubsuit) & f_{x|x^*}(x_3|\heartsuit) & f_{x|x^*}(x_3|\spadesuit)
\end{pmatrix}^{-1}
\]

- 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 ♠, ♥, ♣ are identified under zero-mode assumption.
  
  - For example, error distribution $f_{x|x^*}$ is the same as in Kane et al (1999).

\[
\begin{pmatrix}
\text{no clg. } - x_1 : \\
\text{some clg. } - x_2 : \\
\text{BA}^+ - x_3 :
\end{pmatrix}
\begin{pmatrix}
f_{x|x^*}(x_1|♣) \\
f_{x|x^*}(x_2|♣) \\
f_{x|x^*}(x_3|♣)
\end{pmatrix}
= 
\begin{pmatrix}
0.111 \\
0.772 \\
0.117
\end{pmatrix}
\]

$\arg\max_{x_i} f_{x|x^*}(x_i|♣) = ♣$

$\Rightarrow$ The model $f_{y|x^*}$ and the error structure $f_{x|x^*}$ are identified.

\[x_2 = \arg\max_{x_i} f_{x|x^*}(x_i|♣)
\]

"$x_2$ is the mode"

\[\Rightarrow ♣ = x_2 \text{ (some college)}\]
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.,
     \( f(y|x_1^*) \neq f(y|x_2^*) \) 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 \( \Pr(y > 0|x^*) \)
Eigendecomposition 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

- $X, Z, \text{ and } X^*$ 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 \( L_{x|x^*} \).
Identification: from matrix to integral operator

- 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).
The main equation

\[ L_{y|x|z} = L_{x|x^*} D_{y|x^*|x^*} L_{x^*|z}. \]

- for a function \( g \),

\[
\left[ L_{y|x|z} g \right](x) = \int f_{y,x|z} (y, x|z) g(z) \, dz \\
= \int \int f_{x|x^*} (x|x^*) f_{y|x^*} (y|x^*) f_{x^*|z} (x^*|z) \, dx^* g(z) \, dz \\
= \int f_{x|x^*} (x|x^*) f_{y|x^*} (y|x^*) \int f_{x^*|z} (x^*|z) g(z) \, dz \, dx^* \\
= \int f_{x|x^*} (x|x^*) f_{y|x^*} (y|x^*) \left[ L_{x^*|z} g \right](x^*) \, dx^* \\
= \int f_{x|x^*} (x|x^*) \left[ D_{y|x^*|x^*} L_{x^*|z} g \right](x^*) \, dx^* \\
= \left[ L_{x|x^*} D_{y|x^*|x^*} L_{x^*|z} g \right](x). 
\]

Similarly,

\[ L_{x|z} = L_{x|x^*} L_{x^*|z}. \]
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 $L^{-1}_{x|x^*}$ 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
  \[ \int f_{x* \mid z}(x^* \mid z) h(x^*) \, dx^* = 0 \text{ for all } z \implies h = 0 \]

- Implications:
  - It is equivalent to the injectivity of \( L_{x^* \mid z} \).
  - Inverse \( L_{x^* \mid z}^{-1} \) exists on its domain.
  - It is a necessary condition to achieve point identification using IV.
  - Implied by the bounded completeness of \( f_{x^* \mid z} \), e.g., exponential family.

- Since \( L_{x \mid z} = L_{x \mid x^*} L_{x^* \mid z} \) and \( L_{x \mid x^*} \) is injective, the injectivity of \( L_{x^* \mid z} \) is implied by:
  - **Assumption 2(ii):** \( L_{z \mid x} \) is injective.
An inherent spectral decomposition

- \( L^{-1}_{x|x^*} \) and \( L^{-1}_{x|z} \) exist
  \[ \implies \text{an inherent spectral decomposition} \]
  \[
  L_{y;x|z} L^{-1}_{x|z} = (L_{x|x^*} D_{y;x^*|x^*} L_{x^*|z}) \times (L_{x|x^*} L_{x^*|z})^{-1} = L_{x|x^*} D_{y;x^*|x^*} L^{-1}_{x|x^*}.
  \]

- 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**: \(\sup_{y \in Y} \sup_{x^* \in X^*} f_{y|x^*}(y|x^*) < \infty\).

  \[\implies\text{boundedness of } L_{x|z} L_{x|z}^{-1}, \text{the observed operator on the LHS}.\]

- **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”.

  \[\implies\text{We need to “freeze” these degrees of freedom to show that } L_{x|x^*} \text{ and } D_{y;x^*|x^*} \text{ are uniquely determined by } L_{y;x|z} L_{x|z}^{-1}.\]
A close look at weighted sum of projections

- **Discrete case:**

\[
L_{y|x|z} L_{x|z}^{-1} = L_{x|x^*} D_{y;x^*|x^*} L_{x|x^*}^{-1}
\]

\[
= f_{y|x^*}(y|x_1) \times L_{x|x^*} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} L_{x|x^*}^{-1}
\]

\[
+ f_{y|x^*}(y|x_2) \times L_{x|x^*} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} L_{x|x^*}^{-1}
\]

\[
+ f_{y|x^*}(y|x_3) \times L_{x|x^*} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} L_{x|x^*}^{-1}
\]

- **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.
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 \( x^* = M \left[ f_{x|x^*}(\cdot |x^*)\right] \) for all \( x^* \).

- Intuition: Consider another variable \( \tilde{x}^* \) related to \( x^* \) by \( \tilde{x}^* = R (x^*) \).

\[
\Rightarrow M \left[ f_{x|\tilde{x}^*}(\cdot |\tilde{x}^*)\right] = M \left[ f_{x|x^*}(\cdot |R(\tilde{x}^*))\right] = R(\tilde{x}^*) \neq \tilde{x}^*.
\]

\( \Rightarrow \) 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 $M \left[f_{X|X^*}(\cdot|\bar{x}^*)\right] = x^*$ for all $x^* \in \mathcal{X}^*$.

- then

  \[ f_{X,Y,Z} \text{ uniquely determines } f_{X,Y,Z,X^*} \]

  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 $Y$.
- 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 $x_2 = x^* + \epsilon_2$
  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

\[
\begin{array}{ccc}
X_{t-1} & \uparrow & X_t \\
\uparrow & \uparrow & \uparrow \\
\rightarrow & \rightarrow & \rightarrow \\
X_{t-1}^* & \rightarrow & X_t^* & \rightarrow & X_{t+1}^* \\
\end{array}
\]

- a 3-measurement model

\[ X_{t-1} \perp X_t \perp X_{t+1} \mid X_t^*, \]
\{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)

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

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

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
Estimation: discrete case

- Estimate the matrices directly
  \[ L_{y; x, z} = \left( \begin{array}{ccc} f_{y;x|z}(y, x_1, z_1) & f_{y;x|z}(y, x_1, z_2) & f_{y;x|z}(y, x_1, z_3) \\ f_{y;x|z}(y, x_2, z_1) & f_{y;x|z}(y, x_2, z_2) & f_{y;x|z}(y, x_2, z_3) \\ f_{y;x|z}(y, x_3, z_1) & f_{y;x|z}(y, x_3, z_2) & f_{y;x|z}(y, x_3, z_3) \end{array} \right) \]

- 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 $\omega(.)$

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

- Define

$$M_{X,\omega,Z} = [E[\omega(Y)|X=x_k,Z=z_l]f_{X,Z}(x_k,z_l)]_{k=1,2,...,K; l=1,2,...,K}$$

$$D_{\omega|X^*} = \text{diag} \{E[\omega(Y)|x_1^*], E[\omega(Y)|x_2^*], \ldots, E[\omega(Y)|x_K^*]\}$$

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

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

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

- Estimation mimics identification procedure
Estimation: discrete case

- May also use extremum estimator with restrictions

\[
\left( \hat{M}_X|X^*, D_\omega|X^* \right) = \arg \min_{M,D} \left\| \hat{M}_{X,\omega,Z} \left( \hat{M}_{X,Z} \right)^{-1} M - M \times D \right\|
\]

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

Closed-form estimator
- mimic identification procedure
- don’t need optimization or iteration
- less nuisance parameters than semiparametric estimators
- but may not be efficient
Closed-form estimators

- a 3-measurement model

\[
\begin{align*}
  x_1 &= g_1(x^*) + \epsilon_1 \\
  x_2 &= g_2(x^*) + \epsilon_2 \\
  x_3 &= g_3(x^*) + \epsilon_3
\end{align*}
\]

- 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

\[
\hat{\alpha} = \left( \hat{\beta}, \hat{\eta}, \hat{f}_1, \hat{f}_2 \right) \\
= \text{arg max } \frac{1}{n} \sum_{i=1}^{n} \ln \int f_{y|x^*}(y_i|x^*; \beta, \eta) f_1(x_i|x^*) f_2(x^*|z_i) \, dx^*
\]

- \( \beta \): parameter vector of interest
- \( \eta, f_1, f_2 \): \( \infty \)-dimensional nuisance parameters
- \( A_n \): space of series approximations
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
    $$\hat{\alpha} = (\hat{\beta}, \hat{\eta}, \hat{f}_1, \hat{f}_2)$$
    $$= \arg \max_{(\beta, \eta, f_1, f_2) \in A_n} \frac{1}{n} \sum_{i=1}^{n} \ln \int f_{y|x^*}(y_i|x^*; \beta, \eta) f_1(x_i|x^*) f_2(x^*|z_i) dx^*. $$
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
- Effort and type in contract models
- Unemployment and labor market participation
- Cognitive and noncognitive skill formation
- Dynamic discrete choice with unobserved state variables
- Matching models with latent indices
- 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):
  - IPV auction model:
    - \( \mathcal{N}^* \): \# of potential bidders
    - \( A \): \# of actual bidders
    - \( b \): observed bids
    - \( b(x_i; \mathcal{N}^*) = \begin{cases} 
      x_i - \frac{\int_r^{x_i} F_{\mathcal{N}^*}(s) \mathcal{N}^* - 1 \, ds}{F_{\mathcal{N}^*}(x_i) \mathcal{N}^* - 1} & \text{for } x_i \geq r \\
      0 & \text{for } x_i < r.
    \end{cases} \)
  - conditional independence

\[ f(A_t, b_{1t}, b_{2t} | b_{1t} > r, b_{2t} > r) = \sum_{\mathcal{N}^*} f(A_t | A_t \geq 2, \mathcal{N}^*) f(b_{1t} | b_{1t} > r, \mathcal{N}^*) f(b_{2t} | b_{2t} > r, \mathcal{N}^*) \times f(\mathcal{N}^* | b_{1t} > r, b_{2t} > r) \]
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

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

Auctions with heterogeneous beliefs

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

<table>
<thead>
<tr>
<th>Type</th>
<th>Belief about other bidders’ behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>all other bidders are type-(L0) (bid naively)</td>
</tr>
<tr>
<td>2</td>
<td>all other bidders are type-1</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>(k)</td>
<td>all other bidders are type-((k - 1))</td>
</tr>
</tbody>
</table>

- Specification of type-\(L0\) 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
  \[ a_1 \perp a_2 \perp \ldots \perp a_N \mid e^* \]
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

\[
\begin{align*}
Y_{t-1} & \quad \uparrow & \quad Y_t & \quad \uparrow & \quad Y_{t+1} \\
\quad \rightarrow & \quad X_{t-1}^* & \quad \rightarrow & \quad X_t^* & \quad \rightarrow & \quad X_{t+1}^*
\end{align*}
\]

\[
\begin{align*}
\downarrow & \quad Z_{t-1} & \quad \downarrow & \quad Z_t & \quad \downarrow & \quad Z_{t+1}
\end{align*}
\]

- 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

\[
\begin{align*}
\Pr (Z_{t+1}, Y_t, R_t, Z_t) &= \sum_{X_{t+1}^*} \sum_{X_t^*} \Pr (Z_{t+1}|X_{t+1}^*) \Pr (Z_t|X_t^*) \Pr (X_{t+1}^*, X_t^*, Y_t, R_t).
\end{align*}
\]
Online credit markets for peer-to-peer lending attract dispersed and anonymous borrowers and lenders, and often require no collateral.

The problems of asymmetric information are two-fold:

1. Borrowers differ in their inherent risks $\Rightarrow$ Adverse Selection;
2. Additional incentives are necessary to motivate borrowers to exert effort $\Rightarrow$ Moral Hazard.

Xin (2018, Job market paper) sets up a dynamic structural model to formalize

1. borrowers’ repayment decisions,
2. lenders’ investment strategies,
3. websites’ pricing schemes,

when both hidden information (adverse selection) and hidden actions (moral hazard) are present.

identification strategies to recover the dist. of borrowers’ private types and costs of effort, and utility primitives, and estimate the model using a large dataset from Prosper.com.
Let the index for two loans be \( t - 1 \) and \( t \).

Key elements in the model:
1. Outcomes of the loan (default, late payment): \( O_t, O_{t-1} \);
2. Observed characteristics (debt-to-income ratio, credit grade): \( X_t, X_{t-1} \);
3. Effort choices: \( e_t, e_{t-1} \);
4. Borrower’s type: \( c \).

Dynamic structure motivated by the model:
Step 1: Identify Type Distribution

Observables, $X_t = \{\text{Financial Status}(Z_t), \text{Credit Grade}(K_t)\}$.

Three pieces of information, independent conditional on type.

$$f(O_t, X_t, O_{t-1}, X_{t-1}) = \sum_c \underbrace{f(c, X_{t-1}, O_{t-1})}_{\text{Init. Char.}} \underbrace{f(X_t|X_{t-1}, O_{t-1}, c)}_{\text{Transition of States}} \underbrace{f(O_t|c, X_t)}_{\text{Outcome Realized}}$$

Type distribution $f(c|X_{t-1}, O_{t-1})$ is identified for borrowers with multiple loans. (Hu and Shum, 2012)
Step 2: Identify Effort Choice Probabilities

Loan outcomes include borrowers’ default and late payment performances, $O_t = \{D_t, L_t\}$.

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

1. Conditional on effort, default and late payment are independent.
2. Effort choice is related to borrower’s type.

Following Hu (2008), effort choice probabilities and outcome realization process are identified.
Feng & Hu (2013 AER): Let $X_t^*$ and $X_t$ denote the true and self-reported labor force status.

monthly CPS $\{X_{t+1}, X_t, X_{t-9}\}$

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

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

- Cunha Heckman & Schennach (2010 ECMA)
  \[ X_t^* = (X_{C,t}^*, X_{N,t}^*) \] cognitive and noncognitive skill
  \[ l_t = (l_{C,t}, l_{N,t}) \] parental investments
- for \( k \in \{C, N\} \), skills evolve as
  \[ X_{k,t+1}^* = f_{k,s}(X_t^*, l_t, X_P^*, \eta_{k,t}), \]
  where \( X_P^* = (X_{C,P}^*, X_{N,P}^*) \) are parental skills
- latent factors
  \[ X^* = \left( \{X_{C,t}^*\}_{t=1}^T, \{X_{N,t}^*\}_{t=1}^T, \{l_{C,t}\}_{t=1}^T, \{l_{N,t}\}_{t=1}^T, X_{C,P}^*, X_{N,P}^* \right) \]
- 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
Dynamic discrete choice with unobserved state variables

- 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 $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$
Diamond & Agarwal (2017): 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
  \[ v(X_i, \varepsilon_i) = h(X_i) + \varepsilon_i. \]  
  \[ (1) \]
- the workers’ preference for firm $j$ is described by
  \[ u(Z_j, \eta_j) = g(Z_j) + \eta_j. \]  
  \[ (2) \]
- 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.
Matching models with latent indices

- 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
Income dynamics

- Arellano Blundell & Bonhomme (2017): 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

\[
\begin{align*}
    x_t &= x_t^* + v_t \\
    x_t^* &= x_{t-1}^* + \eta_t \\
    v_t &= \rho_t v_{t-1} + \lambda_t \epsilon_{t-1} + \epsilon_t
\end{align*}
\]

\[\begin{align*}
    \eta_t &: \text{permanent income shock in period } t \\
    \epsilon_t &: \text{transitory income shock} \\
    x_t^* &: \text{latent permanent income} \\
    v_t &: \text{latent transitory income}
\end{align*}\]

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

$$x_t = \nu_t + x^*_t$$

$$\frac{\Delta x_{t+2}}{\rho_{t+2} - 1} - \Delta x_{t+1} = \nu_t + \frac{\lambda_{t+2} \epsilon_{t+1} + \epsilon_{t+2} + \eta_{t+2}}{\rho_{t+2} - 1} - \eta_{t+1}$$

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

*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 updated manuscript for details

*The Econometrics of Unobservables*

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

at [Yingyao Hu's webpage](mailto:Yingyao Hu's webpage)

Comments are welcome. Thank you for your interest.