# Bayesian Analysis in Partially Identified Models

Yuan Liao

Presented at:

CentER for Research in Economics and Business
Tilburg University

March 4, 2010

# Outline

# Examples of Loss of Identification

**Missing Data Problem.**

$Y \in \{0, 1\}$, whether the treatment is successful.

$Z \in \{0, 1\}$, indicator of missing data. $Y$ is observed iff $Z = 1$.

$\theta_0 = P(Y = 1)$: the parameter of interest.

$$\theta_0 = P(Y = 1|Z = 1)P(Z = 1) + \boxed{P(Y = 1|Z = 0)} P(Z = 0)$$

- “Missing at random” assumes that

  $P(Y = 1|Z = 0) = P(Y = 1|Z = 1)$.

- Partial identification approach: $P(Y = 1|Z = 0) \in [0, 1]$.

  Therefore $\theta_0$ is not point identified, but satisfies

$$P(Y = 1|Z = 1)P(Z = 1) \leq \theta_0$$

$$\leq P(Y = 1|Z = 1)P(Z = 1) + P(Z = 0)$$

**Interval censored data**

Let $Y \in [Y_1, Y_2]$. $Y_1$ and $Y_2$ are observed, but not $Y$. Then $EY_1 \leq EY \leq EY_2$. If $\theta_0 = EY$, moment inequalities:

$$E(Y_2 - \theta_0) \geq 0, E(\theta_0 - Y_1) \geq 0$$

**Interval regression** $Y = X^T\theta_0 + \epsilon$, $E(Z\epsilon) = 0$. $Y \in [Y_1, Y_2]$

$Z$ is instrumental variable. $Z_s \geq 0 \Rightarrow$

$$E(Z_s Y_1) \leq E(Z_s X)^T\theta_0 \leq E(Z_s Y_2)$$

Without further assumptions, the parameter is not identified.

$\Rightarrow$ **Moment Inequalities**:

$$E_X m(X, \theta_0) \geq 0$$

- $m(.,.) : \mathcal{X} \times \Theta \to \mathbb{R}^p$, is a known function of $(x, \theta)$.
- $\theta_0 \in \Theta \subset \mathbb{R}^d$.
- $X$ is observable random variable.

- Recall the interval censored example:

$E(Y_1) \leq \theta_0 \leq E(Y_2)$. Data$= (Y_1, Y_2)$

$$E_X m(X, \theta_0) = E \begin{pmatrix} \theta_0 - Y_1 \\ Y_2 - \theta_0 \end{pmatrix} \geq 0$$

- Interval regression example:

$E(ZY_1) \leq E(ZX)^T \theta_0 \leq E(ZY_2)$, $Data = (Z, X, Y_{12})$

$$E_X m(Data, \theta_0) = E \begin{pmatrix} ZX^T\theta_0 - ZY_1 \\ ZY_2 - ZX^T\theta_0 \end{pmatrix} \geq 0$$

- Missing data problem:

$$E_X m(X, \theta_0) = E \begin{pmatrix} \theta_0 - I(Y = 1, Z = 1) \\ I(Y = 1, Z = 1) + I(Z = 0) - \theta_0 \end{pmatrix} \geq 0$$

- Loss of identification due to:

  if define

$$\Omega = \{\theta \in \Theta : E_X m(X, \theta) \geq 0\}.$$

  $\Omega \neq \{\theta_0\}$

- $\theta_0$: *partially identified* on $\Omega$;

  $\Omega$: *identified region*.

# Literature

**Frequentist**

- Consistent set estimator of $\Omega$:

  Horowitz and Manski (2000), Chernozhukov, Hong and Tamer (2007), Beresteanu and Molinari (2008).

- Confidence region of $\theta_0$: Imbens and Manski (2004), Manski and Tamer (2002), Rosen (2008), etc.

- Hypothesis tests: Bugni (2008), Canay (2008), etc.

- Model selection: Andrews and Soares (2007), Shi (2010).

**Bayesian approach**

Moon and Schorfheid (2009)

- Prior placed on nuisance parameter, related to $\theta_0$
- assume the true likelihood function

**Our approach:**

- Study general $E_X m(X, \theta) \geq 0$, prior on $\theta_0$ directly
- Assume less: derive likelihood from moment inequalities
- Make inference on the moments and model

# In This Paper:

1. Assume less on incomplete data: Identification

   $\Leftarrow$ Bayesian Moment **Inequalities**

2. Assume less on distribution: likelihood

   $\Leftarrow$ Bayesian **Moment** inequalities

3. Incorporate prior information

   $\Leftarrow$ **Bayesian** moment inequalities

# Proposed Bayesian Approach

**Key Questions:**

1. How to derive the posterior of $\theta_0$ from $E_X m(X, \theta_0) \geq 0$?
2. Is the derived posterior reliable in some sense?
3. What can we say about $\theta_0$ and the ID region?
4. What if some moment inequalities are mis-specified?

   How likely?

   (Moment and Model Selection)

**Question 1: How to derived the posterior?**

- $p(\theta|Data) \propto p(\theta) \times$likelihood.

  Construct the likelihood first.

- In practice,

  What is known: moment inequalities.

  What is not known: the true likelihood function

- Derive from $E_X m(X, \theta) \geq 0$.

- Define $\lambda_0 = E_X m(X, \theta_0)$, and $\bar{m}(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(X_i, \theta)$.
- CLT $\Rightarrow \sqrt{n}(\bar{m}(\theta_0) - \lambda_0) \rightarrow^d N(0, V)$

$$L(\theta, \lambda) \propto \exp\left(-\frac{n}{2}(\bar{m}(\theta) - \lambda)^T V^{-1}(\bar{m}(\theta) - \lambda)\right)$$

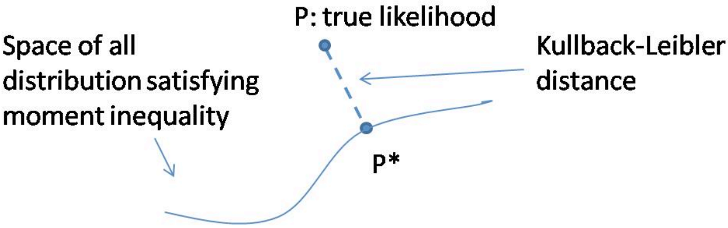
- QUESTION: Does it have a likelihood interpretation?

  Likelihood$=$ $p(Data|\theta)$.

  ANSWER: Yes We Can! **Best approximation to the true likelihood** subject to moment conditions (Kim 2002).

  “limited information likelihood”

# Limited Information Likelihood




$$L(\theta, \lambda) \to^{p} P^{*}$$

$L(\theta, \lambda)$ is called the “limited information likelihood”.

- $\lambda = E_X m(X, \theta)$.

  Put prior $p(\lambda) = \psi e^{-\lambda^T \psi}$, $\lambda \geq 0$.

- Integrate out $\lambda$:

$$p(\theta | Data) \quad \propto \quad p(\theta) \int_{[0,\infty)^p} L(\theta, \lambda) p(\lambda) d\lambda$$

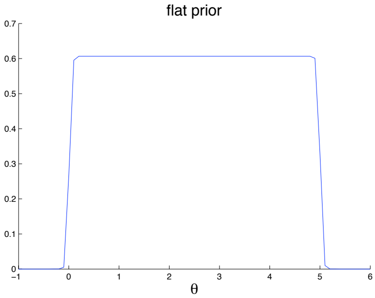
## Question 2: Is the derived posterior reliable?

Answer: show posterior consistency.

## What do I mean by "Posterior Consistency"?

ID region: $\Omega = \{\theta \in \Theta : E_X m(X, \theta) \geq 0\}$.

For example, when $\Omega = [EY_1, EY_2] = [0, 5]$,



# Technical Assumptions

1. The parameter space is compact.
2. The identified region has nonempty interior $int(\Omega)$.
3. $Em_j(X,.) : \Theta \rightarrow \mathbb{R}$ is continuous for each component $j$.
4. The prior $p(\theta)$ is continuous and bounded away from zero on $\Omega$.

# Posterior Consistency

## Theorem 1 (Posterior Consistency)

*Under regularity conditions:*

1. *For any small neighborhood of* $\Omega$*, for some* $a > 0$*,*

$$P(\theta \notin \Omega^{\delta} | Data) = o_p(e^{-an})$$

2. $\forall$ *open set* $A \subset int(\Omega)$*,*

$$\liminf_{n\to\infty} P(\theta \in A | Data) > 0$$

Heuristic Proof:

$$p(\theta|D) \propto p(\theta) \int_{\lambda \geq 0} \exp(-\frac{n}{2}||\bar{m}(\theta) - \lambda||_V^2) p(\lambda) d\lambda$$

where

$$||\bar{m}(\theta) - \lambda||_V^2 = (\bar{m}(\theta) - \lambda)^T V^{-1} (\bar{m}(\theta) - \lambda)$$

$V > 0$.

ID region: $\Omega = \{\theta \in \Theta : E_X m(X, \theta) \geq 0\}$

## Question 3: What can we say about $\theta_0$ and the ID region?

Numerical Example: Missing Data.

- $Y = I(employment)$
- $Z = I(Nonmissing)$
- $P(Z = 0) = 0.33$,
- $P(Y = 1|missing) = 0.1$, $P(Y = 1|Nonmissing) = 0.7$.
- $\boxed{\theta = P(Y = 1) = 0.5}$
- Simulated 5000 data, where 1622 were missing.

$$\hat{P}(Y = 1|Nonmissing) = \frac{2355}{5000 - 1622} = 0.69$$

- "Missing at random" estimates $\hat{\theta} = 0.69$

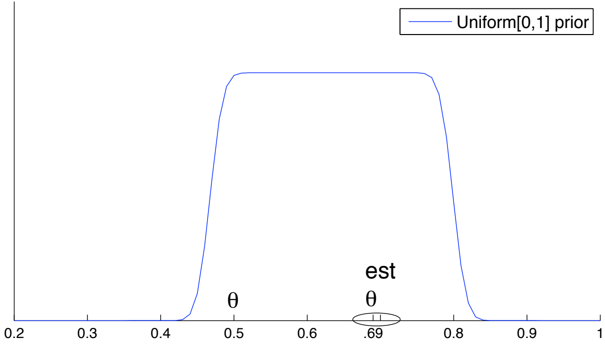
  95% confidence interval $[0.68, 0.71]$.

## Moment inequality approach

$$lower = P(Y = 1|Z = 1)P(Z = 1)$$

$$upper = P(Y = 1|Z = 1)P(Z = 1) + P(Z = 0)$$

True ID region: $[0.47, 0.80]$. If $p(\theta) \sim Uniform[0, 1]$



Previous study (with additional effort or by a different method):

25 Y's are observed both for people with $Z = 1$ and $Z = 0$.

$\hat{p} = 0.45$, $se = \sqrt{\frac{\hat{p}(1-\hat{p})}{25}} = .1$

prior of $\theta$: $N(.45, 0.1^2)$.

Figure: posterior density

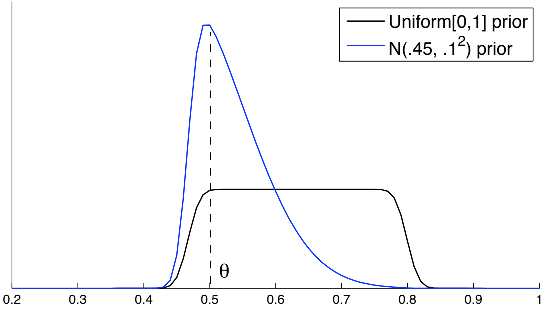


Table: 95% Confidence v.s. 95% Credible Interval

| Method | C.I. | Length |
|---|---|---|
| Freq. Confidence<br>Imbens & Manski (2004) | $[0.465, 0.801]$ | 0.336 |
| Bayes. Credible<br>Uniform$[0,1]$ prior | $[0.477, 0.772]$ | 0.295 |
| Bayes. Credible<br>$N(.45, 0.1^2)$ prior | $[0.465, 0.679]$ | 0.214 |

**Key Questions:**

1. How to derive the posterior of $\theta_0$ from $E_X m(X, \theta_0) \geq 0$?
2. Is the derived posterior reliable in some sense?
3. What can we say about $\theta_0$ and the ID region?
4. What if some moment inequalities are mis-specified?

   How likely?

   (Moment and Model Selection)

# Moment and Model Selection

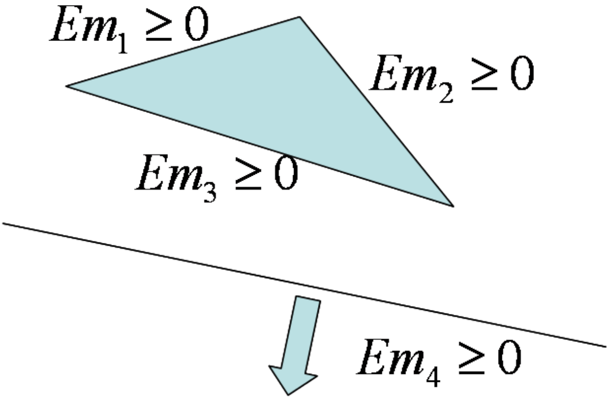
- Suppose we have $p$ moment inequalities:

$$Em_i(X, \theta) \geq 0, \quad i = 1, ..., p$$

where $\theta = (\theta_1, ..., \theta_k) \in \Theta_1 \times \Theta_k$

- If incorrectly specified,

$$\Omega = \{\theta \in \Theta : Em_i(X, \theta) \geq 0, i = 1, ..., p\} = \emptyset$$


$Em_1 \geq 0$
$Em_2 \geq 0$
$Em_3 \geq 0$
$Em_4 \geq 0$

# Example

**Linear regression model**: $Y = X^T\theta + \epsilon$.

$Z$: candidate instrumental variables.

- Fix $Z_s$ such that $E(Z_s\epsilon) = 0$, then

$$E(Z_s(Y - X^T\theta)) = 0$$

  **Model selection**: select covariates in $X$ to model $Y$.

  Set $\Theta = \Theta_1 \times \{0\} \times \{0\}, ...$

- Fix $\Theta$, if $E(Z_s\epsilon) \neq 0$, it is possible that

$$\{\theta \in \Theta : E(Z_s(Y - X^T\theta)) = 0\} = \emptyset$$

  **Moment selection**: select valid IV.

**Interval regression model**: $Y = X^T\theta + \epsilon$, $Y$ is censored in $[Y_1, Y_2]$. $Z$ : candidate instrumental variables.

- Fix $Z_s \geq 0$ such that $E(Z_s\epsilon) = 0$, then
$$EZ_s \begin{pmatrix} Y_2 - X^T\theta \\ X^T\theta - Y_1 \end{pmatrix} \geq 0 \tag{3.1}$$
  **Model selection**: Select parameter space, e.g., $\Theta = \Theta_1 \times \{0\} \times \{0\}, \ldots$
- Fix $\Theta$: if $E(Z_s\epsilon) \neq 0$, then it is possible that (**??**) does not hold for any $\theta \in \Theta$.
  **Moment selection**: Select valid IV.

Problem: choose suitable combination $(Z_s, \Theta)$, such that ID region is not empty.

- $C_s(M_s, \Theta_s)$: selected combination
- A combination $C_s$ is called compatible, if

$$\Omega = \{\theta_s \in \Theta_s : EM_s(X, \theta_s) \geq 0\}$$

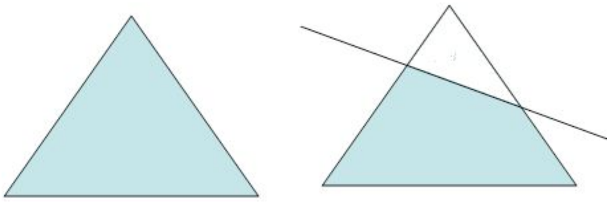
is not empty.

## Goals of Moment/Model Selection

1. **Select compatible combinations**: define non-empty identified regions.
2. Among compatible, **select the optimal one**:

   (i) Simpler models are favored.

   (ii) Contains as many moment inequalities as possible.



- Let $\lambda = EM(X, \theta_s)$, $M = (M_s, M_s^c)^T$.
- Limited Information Likelihood:

$$L(\theta_s, \lambda_s, \lambda_s^c) = \exp(-\frac{n}{2} G(\theta_s, \lambda_s, \lambda_s^c))$$

where

$$G(\theta_s, \lambda_s, \lambda_s^c) = \left\{ \begin{pmatrix} \bar{M}_s(\theta_s) - \lambda_s \\ \bar{M}_s^c(\theta_s) - \lambda_s^c \end{pmatrix}^T V^{-1} \begin{pmatrix} \bar{M}_s(\theta_s) - \lambda_s \\ \bar{M}_s^c(\theta_s) - \lambda_s^c \end{pmatrix} \right\}$$

- marginal posterior of $C_s$.

$P(C_s|Data) \propto \int L(\theta_s, \lambda, C_s) p(\theta_s|C_s) p(\lambda_s|C_s) p(\lambda_s^c|C_s)$

$d\theta_s d\lambda_s d\lambda_s^c \times p(C_s)$

# Goal 1: Selecting Compatible $C_s$

## Theorem 2

*Under some regularity conditions,*

1. *If $C_s$ is compatible ($\Omega \neq \emptyset$),*

$$\liminf_{n\to\infty} p(C_s|Data) > 0$$

2. *If $C_s$ is not compatible ($\Omega = \emptyset$), then for some $\alpha > 0$,*

$$p(C_s|Data) = o_p(e^{-\alpha n})p(C_s)$$

# Goal 2: Selecting Optimal $C_s$

## Theorem 3

*Put equal priors on all $C_s$, $p(\lambda_s) = \psi \exp(-\psi^T \lambda_s)$.*

$$p(\lambda_s^c | C_s) \sim MVN(0, \sigma_n^2 I), \quad p(\theta_s | C_s) \sim MVN(0, n\sigma_n^2 I).$$

*$\sigma_n^2 \to \infty$, but slower than exponential rate. Define*

$$C^* = \arg\max_{C_s} p(C_s | Data)$$

***Asymptotically, $C^*$ is compatible, and with the largest***

$$\dim(M_s) - \dim(\Theta_s).$$

# Conclusions

- Loss of Identification exists commonly in statistical models incomplete data: missing, censored.
- Our approach
  1. Assume less on incomplete data: vs Identification
     $\Leftarrow$ Bayesian Moment **Inequalities**
  2. Assume less on distribution: vs Moon & Schorfheid (2009)
     $\Leftarrow$ Bayesian **Moment** Inequalities
  3. Incorporate prior information: vs Frequentist
     $\Leftarrow$ **Bayesian** Moment Inequalities

Can select moment and models with posterior probability.

Max posterior:

- compatible model
- smallest ID
- simplest model

# Future Works

**Survival Analysis**

We observe i.i.d. observations of :

- $Y^* = \min(Y, C)$: censored survival time.
- $d = I(Y < C)$: indicator of censoring.

Model: $Y = X^T\theta + \epsilon$. Med$(\epsilon|X) = 0$.

Traditional app. Assume $\boxed{C \perp Y|X}$

Moment Ineq app. Khan and Tamer (2009):

$$\text{For any } h(X) \geq 0,$$
$$E[\frac{1}{2}h(X) - I(Y^* \geq X^T\theta)h(X)] \geq 0$$
$$E[dI(Y^* \leq X^T\theta)h(X) - \frac{1}{2}h(X)] \geq 0$$

## Theoretical Directions

- Limited Information Likelihood: derived directly from

$$E_X m(X, \theta) \geq 0$$

Other moment condition-based likelihood: **empirical likelihood** (EL, Owen 1990), **generalized empirical likelihood** (GEL, Newey and Smith 2001).

| | LIL | EL | GEL |
|---|---|---|---|
| point identi. | Kim (2002) | Lazar(2003) | CH(2003) |
| partial identi. | this talk | working paper | not yet |

- High dimensional variable selection
- Nonparametric IV regression

$$Y = g(X) + \epsilon$$

$E(\epsilon|W) = 0 \Rightarrow E(Y|W) = E(g(X)|W)$.

$g = \sum_{i=1}^{\infty} b_i \phi_i$, $g_b = \sum_{i=1}^{q} b_i \phi_i$.