Semiparametric Information Bound of Dynamic Discrete Choice Models

Moshe Buchinsky    Jinyong Hahn    Kyoo il Kim
UCLA
University of Minnesota

June 20, 2008
Motivation
Motivation

- Unobserved heterogeneity of economic agents is a key element of empirical micro
- Dynamic discrete choice programming (DDP) models over twenty years in IO and Labor
  - Patent renewal, Bus engine replacement, Life cycle Job search, Fertility and female labor decision
- Unobserved heterogeneity in DDP?
  - Ignored
  - Fixed Effects or Mixtures
Identification critically depends on the structure of heterogeneity

- Not easy to check due to the complexity of DDP
- Even though the model is identified, such an identification is often fragile
- For example, the identification at infinity (Taber, 2000)
- Typically, we assume the number of types is small (Keane and Wolpin, 1997)
- Statistical inference may be sensitive to the number of types chosen by researcher

- Simple structure is essential to the identification?
Instead of fragile identification argument, we consider the possibility of robust inference i.e., whether the structural parameter is root-\(n\) estimable

\[ \text{root-}n \text{ estimable } = \text{positive semiparametric information bound} \]

Convenient testing procedure to determine whether the semiparametric information bound of DDP is zero

- ‘Theory by computer’: A simulation based approach
- Our test can be used to panel models or dynamic games
What we do in this paper

- Negative message
  - More pessimistic than Rust (1994) or Magnac & Thesmar (2002)

- MC study
  - A simplified version of Keane and Wolpin's (1997) model
  - We show that the semiparametric information is zero for this model
Information Bound of DDP Model
DDP as a gigantic discrete choice model with individual effects

- Set of alternatives is the collection of all possible lifetime sequences of choices
- For example, college attendance decision for $T = 3$, $\Omega = \{111, 110, 101, 100, 011, 010, 001, 000\}$
- e.g., a DDP model with $T$ periods and $J$ alternatives is viewed as a discrete choice model with $K = J^T$ choices

$$p(x, \theta_0, \alpha) = \begin{pmatrix} \Pr(d = \omega_1|x, \alpha) \\ \vdots \\ \Pr(d = \omega_{K-1}|x, \alpha) \end{pmatrix} = \begin{pmatrix} p_1(x, \theta_0, \alpha) \\ \vdots \\ p_{K-1}(x, \theta_0, \alpha) \end{pmatrix}$$

where $\alpha$ is individual effect, $x$'s are covariates, $\theta_0$ is the structural parameter
The question is whether the semiparametric information for $\theta_0$ is positive where the conditional distribution of $\alpha$ given $x$ is nonparametrically specified.
Necessary condition for identification

- We draw on Johnson's theorems

**Theorem (Johnson (2004, Theorem 2.2))**

*If $\theta_0$ is identified, then for some $x$ there exist a nonzero vector $c \equiv (c_0, \ldots, c_{K-1})'$ such that*

$$\forall \alpha \in A, \sum_{j=1}^{K-1} c_j p_j(x, \theta_0, \alpha) = c_0. \tag{1}$$

- This means when the structural parameter is identified, the $K - 1$ choice probabilities should be linearly dependent
- But this theorem cannot rule out fragile identification
Theorem is strengthened.

Theorem (Johnson (2004, Theorem 2.3))

If (1) does not hold for any \( x \in \mathcal{X} \), then the semiparametric information bound for \( \theta_0 \) is zero.

- As long as \( p(x, \theta, \alpha) \) is smooth in the neighborhood of \( \theta_0 \) and the score is bounded, the condition (1) for some \( x \) becomes necessary condition for positive information.
Panel Logit is identified: $T = 2$,

\[
p_{10} = p(y_{i1} = 1, y_{i2} = 0|x_i, \alpha_i) \\
= \frac{\exp(\alpha_i + x_{i1}\beta)}{1 + \exp(\alpha_i + x_{i1}\beta)} \frac{1}{1 + \exp(\alpha_i + x_{i2}\beta)}
\]

\[
p_{01} = p(y_{i1} = 0, y_{i2} = 1|x_i, \alpha_i) \\
= \frac{1}{1 + \exp(\alpha_i + x_{i1}\beta)} \frac{\exp(\alpha_i + x_{i2}\beta)}{1 + \exp(\alpha_i + x_{i2}\beta)}
\]

We have $p_{10}/p_{01} = \exp((x_{i1} - x_{i2})\beta)$. Regardless of $\alpha$, $p_{10}$ and $p_{01}$ have linear relationship given values of $x_{i1}$ and $x_{i2}$.
Panel Probit has no information bound: \( T = 2 \),

\[
\begin{align*}
p_{10} &= p(y_{i1} = 1, y_{i2} = 0 | x_i, \alpha_i) \\
&= \Phi(\alpha_i + x_{i1}\beta) \left( 1 - \Phi(\alpha_i + x_{i2}\beta) \right) \\
p_{01} &= p(y_{i1} = 0, y_{i2} = 1 | x_i, \alpha_i) \\
&= (1 - \Phi(\alpha_i + x_{i1}\beta)) \Phi(\alpha_i + x_{i2}\beta)
\end{align*}
\]

Only when \( x_{i1} = x_{i2} \), we have \( p_{10} = p_{01} \). This may happen with zero probability.
Understanding Johnson’s theorem as it is

- As $\alpha$ varies (holding $x$ and $\theta$ fixed), the function $p(x, \theta, \alpha)$ defines a set of points in $\mathbb{R}^{K-1}$:

$$C_{x,\theta} = \{p(x, \theta, \alpha) : \alpha \in \mathcal{A}\}$$

- The reduced form choice probabilities are given by the expected value of $p(x, \theta, \alpha)$ over $\mathcal{A}$ conditional on $x$ (given $\theta$):

$$p^*(x; \theta) = \int_{\mathcal{A}} p(x, \theta, \alpha) dF(\alpha|x) d\alpha$$

- $p^*(x; \theta)$ is a weighted average of points in the set $C_{x,\theta}$, with weight $dF(\alpha|x)$
- Note that the set of values $p^*(x; \theta)$ can take is equal to the convex hull of $C_{x,\theta}$, denoted by $H_{x,\theta}$
• Identification is only possible if the set $H_{x,\theta}$ for different values of $\theta$ do not overlap. To be precise, if

$$\exists \theta' \neq \theta_0 \text{ such that } \forall x \in \mathcal{X}, \ H_{x,\theta'} \cap H_{x,\theta_0} \neq \emptyset$$

then $\theta_0$ is not identified relative to $\theta'$.

• A necessary condition for identification is that $C_{x,\theta}$ must lie in a at most $K-2$ dimensional hyperplane, for at least one value of $x \in \mathcal{X}$.

• Otherwise the convex hull contains some $K-1$ dimensional volume and if the boundaries of these sets move continuously in $\theta$, then they must overlap.
Define

\[
p(x, \theta, \alpha) = \begin{bmatrix} 1 \\ p(x, \theta, \alpha) \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 1(d = \omega_1) \\ \vdots \\ 1(d = \omega_{K-1}) \end{bmatrix}
\]

If (1) is satisfied, then we can find vector \( c(x, \theta) \) such that \( c(x, \theta)' p(x, \theta, \alpha) = 0 \) for all \( \alpha \)
Our understanding of Johnson’s theorem

- This condition naturally leads to the equality

\[ E \left[ c(x, \theta)' y \mid x, \alpha \right] = c(x, \theta)' E \left[ y \mid x, \alpha \right] = c(x, \theta)' p(x, \theta, \alpha) = 0 \]

- By iterating the expectation, we obtain the moment restriction

\[ E \left[ c(x, \theta)' y \mid x \right] = 0 \]

which can be a basis of GMM estimation

- Finding such a vector \( c(x, \theta) \approx \) positive information
Therefore, (1) can be understood to be a sufficient condition for nonzero information

Johnson shows (1) is a necessary condition for nonzero information and we show it is also sufficient!

Note that this is about whether the (whole) information matrix is zero or not

It is possible that even when the information matrix is not zero, partial information of a certain parameter of interest may be still zero

Example, dynamic logit model
Our Testing Strategy
Our strategy

Consider the following Statements

Statement \((S^c)\)
\[
\forall x \in \mathcal{X}, \forall \text{ nonzero vector } c, \exists \alpha \text{ such that the equality (1) is violated}
\]

Statement \((S)\)
\[
\exists x \in \mathcal{X}, \exists \text{ a nonzero vector } c, \text{ such that the equality (1) is satisfied } \forall \alpha
\]

Note that the model is root-\(n\) consistently estimable if and only if \(S\) holds

We, therefore, want to test

\[
H_0^*: S \text{ holds}
\]
Our strategy

- Rejection of the null hypothesis $H_0^*$ makes us conclude that $S^c$ holds
  - Then, semiparametric information bound equals to zero
  - This turns out to be a little inconvenient
- Alternatively we propose a weaker test of $H_0^*$ but easier to implement
Practical necessary condition

- Define a $K \times K$ matrix

$$q(x; \alpha_1, \ldots, \alpha_K) \equiv \begin{bmatrix} 1 & \ldots & 1 \\ p(x, \theta, \alpha_1) & \ldots & p(x, \theta, \alpha_K) \end{bmatrix}$$

$$\alpha^{(K)} \equiv (\alpha_1, \ldots, \alpha_K)'.$$

and consider the following two statements:

**Statement (T⁻)**

$$\forall x \in \mathcal{X}, \exists \alpha^{(K)} \in \mathcal{A}^K \text{ such that } \det(q(x; \alpha_1, \ldots, \alpha_K)) \neq 0$$

**Statement (T)**

$$\exists x \in \mathcal{X}, \det(q(x; \alpha_1, \ldots, \alpha_K)) = 0 \forall \alpha^{(K)} \in \mathcal{A}^K$$

- If $S$ is satisfied, then $T$ is automatically satisfied.
Practical necessary condition

- $T$ is a necessary condition for the root-$n$ estimability.
- We, therefore, propose to test $H_0 : T$ holds.

- Violation of $T$ implies $S^c$, but $S^c$ implies that the information is zero!!
Stochastic Information Bound Test
1. Generate a $\alpha^{(K)} \in A^K$ such that $\alpha^{(K)}$ contains $K$ different values

2. Compute $\det(q(x; \alpha_1, \ldots, \alpha_K))$ for all $x \in X$
   - If $\det(q(x; \alpha_1, \ldots, \alpha_K)) = 0$ for some $x \in X$, then go back to Step 1
   - If $\det(q(x; \alpha_1, \ldots, \alpha_K)) \neq 0$ for all $x \in X$, then stop and declare the zero information of the model. (Rejection of $T^c$)

- This is often infeasible in practice
  - We typically do not have an analytic form of $p(x, \theta, \alpha)$
  - On the other hand, we can often generate choices, $d$
Stochastic information bound test

1. Generate a \( \alpha^{(K)} \in \mathcal{A}^K \) such that \( \alpha^{(K)} \) contains \( K \) different values.

2. Generate a random sample \( d_{1,k}, \ldots, d_{N,k} \). Let
   \[ \hat{p}(x, \theta_0, \alpha_k) = \frac{1}{N} \sum_{i=1}^{N} d_{i,k}. \]
   Define \( \hat{q}(x; \alpha_1, \ldots, \alpha_K) \) similarly. Compute \( \det(\hat{q}(x; \alpha_1, \ldots, \alpha_K)) \) for all \( x \in \mathcal{X} \).

   1. If the hypothesis \( \det(\hat{q}(x; \alpha_1, \ldots, \alpha_K)) = 0 \) cannot be rejected for some \( x \in \mathcal{X} \), then go back to Step 1.
   2. If the hypothesis \( \det(\hat{q}(x; \alpha_1, \ldots, \alpha_K)) = 0 \) can be rejected for all \( x \in \mathcal{X} \), then stop and declare the zero information of the model.
Understanding Our Test
Intuition behind our test

- When we reject \( T \) and so

\[
q(x; \alpha_1, \ldots, \alpha_K) \text{ or } \begin{bmatrix}
p(x, \theta_0, \alpha_1) & \cdots & p(x, \theta_0, \alpha_K) \\ p_K(x, \theta_0, \alpha_1) & \cdots & p_K(x, \theta_0, \alpha_K)
\end{bmatrix}
\]

is nonsingular, we claim the information bound must be zero.

- Let’s see why this must be the case...

- Note that even though \( \theta_0 \) has no information bound, we can still estimate the mixing ratios with the parametric rate (Bajari, Fox, Kim, Ryan, 2008)
Numerical Illustration
Keane and Wolpin (1997)

- Finite horizon model of life cycle career choice model
- Dynamic discrete choice model of schooling, work, and occupational choice
- 11 years of observations starting at the age of 16
- Unobserved Heterogeneity due to the unobserved endowments at the age 16
- To deal with this heterogeneity, KW use mixtures
- KW assume that the number of types is three or four
- KW estimate structural parameters together with type frequency parameters
We consider a simplified version of Keane and Wolpin (1997) with $J = 3$ and $T = 3$

Three alternatives: Attend school, Work in a white-collar occupation, Work in a blue-collar occupation

Reward Structure

- **Schooling Alternative:**
  \[ R_{i1t} = \exp \{ \alpha_{i1} + \varepsilon_{i1t} \} - tc \cdot I(E_{it} \geq 12) \]

- **White-collar Alternative:**
  \[ R_{i2t} = \exp \{ \alpha_{i2} + \beta_1 E_{it} + \beta_2 x_{i2t} + \varepsilon_{i2t} \} \]

- **Blue-collar Alternative:**
  \[ R_{i3t} = \exp \{ \alpha_{i3} + \gamma_1 E_{it} + \gamma_2 x_{i3t} + \varepsilon_{i3t} \} \]
- $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3})'$ denotes the activity specific endowments of the individual $i$
- $\varepsilon_{it} = (\varepsilon_{i1t}, \varepsilon_{i2t}, \varepsilon_{i3t})'$ represents idiosyncratic shock
- $E_{it}$ denote years of education $\{12, 13, 14\}$, $tc$ represents the direct costs of education, $x_{i2t}$ and $x_{i3t}$ denote occupation specific job experiences
- We simulate choices based on these reward functions by solving a dynamic programming problem
\( \alpha \sim \mathcal{N} \left( \begin{pmatrix} 2.5 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 2.0 & 0.2 \\ 0.2 & 2.5 \end{pmatrix} \right) \)

\( \alpha \sim \mathcal{N} \left( \begin{pmatrix} 2.5 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2.0 & 0.3 & 0.3 \\ 0.3 & 2.0 & -0.1 \\ 0.3 & -0.1 & 2.0 \end{pmatrix} \right) \)
We do conclude the information bound is zero for the KW model!

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Two Alternatives</th>
<th>Three Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>0.9499</td>
<td>0.2948</td>
</tr>
<tr>
<td>10,000</td>
<td>0.7709</td>
<td>0.3032</td>
</tr>
<tr>
<td>100,000</td>
<td>2.6099</td>
<td>0.4053</td>
</tr>
<tr>
<td>1,000,000</td>
<td>7.0276</td>
<td>2.5509</td>
</tr>
<tr>
<td>10,000,000</td>
<td>25.2266</td>
<td>4.4695</td>
</tr>
</tbody>
</table>
Alternative MC design

\[ \alpha \sim N \left( \begin{pmatrix} 2.5 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2.0 & 1.2 & -1.2 \\ 1.2 & 2.0 & -0.1 \\ -1.2 & -0.1 & 2.0 \end{pmatrix} \right) \]
Again the information bound is zero for the KW model!

Although it is harder to reject

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>t-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000</td>
<td>0.3224</td>
</tr>
<tr>
<td>100,000</td>
<td>0.2896</td>
</tr>
<tr>
<td>1,000,000</td>
<td>0.8443</td>
</tr>
<tr>
<td>10,000,000</td>
<td>1.9457</td>
</tr>
<tr>
<td>50,000,000</td>
<td>3.5732</td>
</tr>
</tbody>
</table>
We do reject the null hypothesis of the zero determinant.

It means the semiparametric information bound of the KW model is ZERO!

$T$-statistic grows as the number of simulations gets large, confirming the consistency of our test.

The growth rate approaches $10^{-1/2}$ as we increase the simulation size by 10.

MC result suggests that the root $n$ consistency of KW estimator is due to a simple mixture structure assumption.

Dependence makes it harder to reject.
Conclusion
Identification of dynamic discrete choice programming models may be sensitive to the (assumed) structure of individual heterogeneity.

When the *true* heterogeneity dominates the model, the structural parameter may not be estimable with the parameteric rate.

We propose a convenient testing procedure to determine whether the semiparametric information bound of DDP is zero or not.

From our MC study, we conclude that our concern is not too much.

We illustrate that the structural parameters have no information bound in the KW model.
Thank You !!!