Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices

Hiro Kasahara\(^1\)  
Katsumi Shimotsu\(^2\)

\(^1\)Department of Economics  
University of Western Ontario

\(^2\)Department of Economics  
Queen’s University
Finite Mixture Models of Dynamic Discrete Choices

\[ P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^{M} \pi^m P^m(\{a_t, x_t\}_{t=1}^T) \]

where \( a_t \in A = \{0, 1\} \) and \( x_t \in X = \{1, 2, \ldots, |X|\} \).

The Goal of this Paper:

- Identify \( \{\pi^m, P^m\} \)'s from \( P(\{a_t, x_t\}_{t=1}^T) \)
  without parametric assumptions

- Nonparametric Assumptions?
  – Markov property, Stationarity etc.
Why Nonparametric Identification?

- Identification of parametric/semiparametric finite mixture models.
- An initial nonparametric consistent estimate of type-specific component distributions
  ⇒ Applying two-step estimators for structural dynamic models to models with unobserved heterogeneity:
Our point of departure: Hall and Zhou (2003, AS)

“very little is known of the potential for consistent nonparametric inference in mixtures”

Two-type mixtures with independent marginals:

\[
F(y) = \pi \prod_{t=1}^{T} F^1_t(y_t) + (1 - \pi) \prod_{t=1}^{T} F^2_t(y_t).
\]

Hall and Zhou show:

- \( T \geq 3 \) is necessary and sufficient.
What is yet unknown

- Three or more types?
- Covariates?
- Relaxing independence of $y_t$ and $y_{t-1}$?

$\rightarrow$ We consider finite mixture models with covariates and state dependence.
Overview of Our Contribution

Three elements for nonparametric identification:

- The time-dimension of panel data
  - $T \geq 3$ in a simple case
  - $T \geq 6$ under first-order Markov process

- The number of the values the covariates can take $\approx$ the maximum number of types.

- The changes in the covariates must induce sufficiently heterogeneous variations in the choice probabilities across types.
Overview of Our Limitation

The key assumption is a Markov process:

\[ f(s_t|s_{t-1}, s_{t-2}, \ldots) = f(s_t|s_{t-1}). \]

In addition, either one of the following two assumptions must hold:

- **Stationarity**
  
  \[ P_t(choice|x, type) = P(choice|x, type) \]

- **Conditional Independence + Type-Invariant Transition**
Setup: Panel Dynamic Discrete Choice Model

\[ P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^M \pi^m p^m(x_1, a_1) \prod_{t=2}^T f_t^m(x_t|x_{t-1}, a_{t-1}) P_t^m(a_t|x_t, a_{t-1}). \]

where \( a_t \in A = \{0, 1\} \) and \( x_t \in X = \{1, 2, \ldots, |X|\} \).

- \( p^m(x_1, a_1) \): initial distribution for type \( m \)
- \( f_t^m(x_t|x_{t-1}, a_{t-1}) \): transition function for type \( m \)
- \( P_t^m(a_t|x_t, a_{t-1}) \): choice probability for type \( m \)
The Baseline Model

\[ P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^{M} \pi^m p^{*m}(x_1, a_1) \prod_{t=2}^{T} f(x_t|x_{t-1}, a_{t-1}) P^m(a_t|x_t). \]

- (a) Stationarity and Conditional Independence:
  \[ P_t^m(a_t|x_t, a_{t-1}) = P^m(a_t|x_t) \]

- (b) Type-Invariant Transition Functions:
  \[ f^m(x_{t+1}|x_t, a_t) = f(x_{t+1}|x_t, a_t) \]
The Baseline Model

Define

\[
\tilde{P}(\{a_t, x_t\}_{t=1}^T) = \frac{P(\{a_t, x_t\}_{t=1}^T)}{\prod_{t=2}^{T} f(x_t|x_{t-1},a_{t-1})}.
\]

\[
\tilde{P}(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^{M} \pi^m p^m(x_1, a_1) \prod_{t=2}^{T} P^m(a_t|x_t).
\]

- The number of restrictions \(\sim |X|^T\)
  \(>>\) The number of unknowns \(\sim M|X|\).
Nonparametric identification of the baseline model

Consider $T = 3$. Fix $a_t = 1$ for all $t$. Pick $x_1 = k$.

$$
\tilde{P}(\{1, x_t\}_{t=1}^3) = \sum_{m=1}^M \pi^m p^*^m(k, 1) P^m(a_2 = 1|x_2) P^m(a_3 = 1|x_3).
$$

Integrating out different elements gives the “marginals”:

$$
\tilde{P}(\{1, x_t\}_{t=1}^2) = \sum_{m=1}^M \pi^m p^*^m(k, 1) P^m(a_2 = 1|x_2),
$$

$$
\tilde{P}(\{1, x_1, 1, x_3\}) = \sum_{m=1}^M \pi^m p^*^m(k, 1) P^m(a_3 = 1|x_3),
$$

$$
\tilde{P}(\{1, x_1\}) = \sum_{m=1}^M \pi^m p^*^m(k, 1), \quad \text{etc.}
$$

Source of Identification: a variation of $x_2$ and $x_3$. 

---

Hiro Kasahara, Katsumi Shimotsu
Nonparametric identification of the baseline model

Notations

- **Unknowns:**
  \[
  \lambda_k^m = p^m((a_1, x_1) = (1, k)), \quad \lambda_\xi^m = P^m(a = 1|x = \xi),
  \]

- **Observables:**
  \[
  F_{x_1, x_2, x_3}^* = \tilde{P}(\{1, x_t\}_{t=1}^3), \quad F_{x_1, x_2}^* = \tilde{P}(\{1, x_t\}_{t=1}^2),
  \]
  \[
  F_{x_1, x_3}^* = \tilde{P}(\{1, x_1, 1, x_3\}), \quad F_{x_2, x_3} = \tilde{P}(\{1, x_t\}_{t=2}^3),
  \]
  \[
  F_{x_1}^* = \tilde{P}(\{1, x_1\}), \quad F_{x_2} = \tilde{P}(\{1, x_2\}), \quad F_{x_3} = \tilde{P}(\{1, x_3\}).
  \]
Consider $x_2$ and $x_3$ in $\{\xi_1, \xi_2, \ldots, \xi_{M-1}\}$.

- Unknowns:

\[
L_{(M \times M)} = \begin{bmatrix}
1 & \lambda^1_{\xi_1} & \cdots & \lambda^1_{\xi_{M-1}} \\
\vdots & \vdots & \ddots & \vdots \\
1 & \lambda^M_{\xi_1} & \cdots & \lambda^M_{\xi_{M-1}}
\end{bmatrix},
\]

$D_k = \text{diag}(\lambda^1_k, \ldots, \lambda^M_k)$, and $V = \text{diag}(\pi^1, \ldots, \pi^M)$. 
Nonparametric identification of the baseline model

- Observables:

\[
P = \begin{bmatrix}
1 & F_{\xi_1} & \cdots & F_{\xi_{M-1}} \\
F_{\xi_1} & F_{\xi_1,\xi_1} & \cdots & F_{\xi_1,\xi_{M-1}} \\
\vdots & \vdots & \ddots & \vdots \\
F_{\xi_{M-1}} & F_{\xi_{M-1},\xi_1} & \cdots & F_{\xi_{M-1},\xi_{M-1}}
\end{bmatrix},
\]

\[
P_k = \begin{bmatrix}
F^*_k & F^*_{k,\xi_1} & \cdots & F^*_{k,\xi_{M-1}} \\
F^*_{k,\xi_1} & F^*_{k,\xi_1,\xi_1} & \cdots & F^*_{k,\xi_1,\xi_{M-1}} \\
\vdots & \vdots & \ddots & \vdots \\
F^*_{k,\xi_{M-1}} & F^*_{k,\xi_{M-1},\xi_1} & \cdots & F^*_{k,\xi_{M-1},\xi_{M-1}}
\end{bmatrix}.
\]
Nonparametric identification of the baseline model

Factorization Equations:

\[ P = L'VL \] (\( M \times M \) equations)

\[ P_k = L'D_k VL \] (\( M \times M \) equations)

The \((M, M)\)-th equation of \( P_k = L'D_k VL \):

\[ \tilde{P}( (x_1, x_2, x_3) = (k, \xi_{M-1}, \xi_{M-1}) ) = \sum_{m=1}^{M} \pi^m p^*(k, 1) P^m(a_2 = 1|\xi_{M-1}) P^m(a_3 = 1|\xi_{M-1}) \]
“Dividing” $P_k$ by $P$:

\[
\text{eigenvalue}(P^{-1}P_k) = D_k,
\]

\[
\text{eigenvector}(P^{-1}P_k) = L,
\]

and

\[
V = (L')^{-1}PL^{-1}.
\]
Proposition 1

Assume $T \geq 3$. Let $\xi_j \in X$ for $j = 1, \ldots, M - 1$.

Suppose

1. $L$ is nonsingular for some $\{\xi_1, \ldots, \xi_{M-1}\}$,
2. $\lambda_\cdot^m \neq \lambda_\cdot^n$ for any $m \neq n$.

Then,

$$\{\pi^m, \lambda_\cdot^m, \lambda_\cdot^m\}_{m=1}^{M}$$ is identified from $\tilde{P}(\{1, x_t\}_t^{3})$'s.
A Simple Example

Consider $X = \{\xi_1, \xi_2\}$. The maximum number of types is $M = 3$.

$$
L = \begin{bmatrix}
1 & P_1^1(a = 1|x = \xi_1) & P_1^1(a = 1|x = \xi_2) \\
1 & P_2^2(a = 1|x = \xi_1) & P_2^2(a = 1|x = \xi_2) \\
1 & P_3^3(a = 1|x = \xi_1) & P_3^3(a = 1|x = \xi_2)
\end{bmatrix}
$$

(1) $L$ is nonsingular $\Rightarrow$

The changes in $x$’s induce heterogeneous variations in the choice probabilities across types.

(2) $\lambda_k^m \neq \lambda_k^n \Rightarrow$ The initial distributions are different across types.

\[
\begin{align*}
p_1^1(k, 1) & \neq p_2^2(k, 1) \\
p_1^1(k, 1) & \neq p_3^3(k, 1) \\
p_2^2(k, 1) & \neq p_3^3(k, 1)
\end{align*}
\]
First-order Markov Process under Stationarity

\[
P(\{a_t, x_t\}_{t=1}^T) = \sum_{m=1}^{M} \pi^m p^*(x_1, a_1) \prod_{t=2}^{T} f^m(x_t|x_{t-1}, a_{t-1}) P^m(a_t|x_t, a_{t-1})
\]

Define \( s_t = (a_t, x_t) \in S \).

\[
P(\{s_t\}_{t=1}^T) = \sum_{m=1}^{M} \pi^m q^*(s_1) \prod_{t=2}^{T} Q^m(s_t|s_{t-1}),
\]

where

\[
q^*(s_1) = p^*(x_1, a_1),
\]

\[
Q^m(s_t|s_{t-1}) = f^m(x_t|x_{t-1}, a_{t-1}) P^m(a_t|x_t, a_{t-1}).
\]
Problem:

$s_t$ appears both in $Q^m(s_t|s_{t-1})$ and $Q^m(s_{t+1}|s_t)$, and creates the dependence between these terms.

⇒ Cannot construct $P = L'VL$ and $P = L'D_kVL$.

Solution:

Look at every other period to break the dependence.
First-order Markov Process under Stationarity

Assume $T = 6$.

Fix $s_1 = s_3 = s_5 = \bar{s}$ and $s_6 = k$.

\[
P(\{s_t\}) = \sum_{m=1}^{M} \pi^m q^*(s_1) \prod_{t=2}^{6} Q^m(s_t|s_{t-1})
\]

\[
= \sum_{m=1}^{M} \pi^m q^*(\bar{s}) Q^m(s_2|\bar{s}) Q^m(\bar{s}|s_2) Q^m(s_4|\bar{s}) Q^m(\bar{s}|s_4) Q^m(s_6 = k|\bar{s}).
\]

Define

\[
\hat{\pi}^m_{\bar{s}} = \pi^m q^*(\bar{s}), \quad \lambda^m_{\bar{s}}(s) = Q^m(\bar{s}|s) Q^m(s|\bar{s})
\]

\[
\Rightarrow
\]

\[
P(\{s_t\}_{t=1}^{6}) = \sum_{m=1}^{M} \hat{\pi}^m_{\bar{s}} \lambda^m_{\bar{s}}(s_2) \lambda^m_{\bar{s}}(s_4) Q^m(k|\bar{s}).
\]
Proposition 6

Assume $T \geq 6$. Let $\xi_j \in S$ for $j = 1, \ldots, M - 1$.

Define

$$L_{\bar{s}} = \begin{bmatrix} 1 & \lambda_1^1(\xi_1) & \cdots & \lambda_1^1(\xi_{M-1}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_M^1(\xi_1) & \cdots & \lambda_M^1(\xi_{M-1}) \end{bmatrix}_{(M \times M)}.$$

Suppose

(1) $L_{\bar{s}}$ is nonsingular for some $\{\xi_1, \ldots, \xi_{M-1}\}$,

(2) $Q^m(k|\bar{s}) \neq Q^n(k|\bar{s})$ for any $m \neq n$.

Then,

$$\left\{ \tilde{\pi}_{\bar{s}}^m, \{ \lambda_{\bar{s}}^m(s), Q^m(s|\bar{s}) \}_{s \in S} \right\}_{m=1}^M$$

is identified from $P(\{s_t\}_{t=1}^T)$’s.
Identification of the number of components, $M$

How to choose $M$?

→ Important practical issue but economic theory usually does not provide guidance.

We may show two-periods of panel data may suffice for identifying the number of types
Proposition 3

Assume $T = 2$ and $X = \{1, 2, \ldots, |X|\}$.

$$P^* = \begin{bmatrix}
1 & \tilde{P}(x_2 = 1) & \cdots & \tilde{P}(x_2 = |X|) \\
\tilde{P}(x_1 = 1) & \tilde{P}((x_1, x_2) = (1, 1)) & \cdots & \tilde{P}((x_1, x_2) = (1, |X|)) \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{P}(x_1 = |X|) & \tilde{P}((x_1, x_2) = (|X|, 1)) & \cdots & \tilde{P}((x_1, x_2) = (|X|, |X|))
\end{bmatrix},$$

Then,

$$M \geq \text{rank}(P^*).$$
Note

\[ P^* = (L_1^*)' V L_2^* . \]

where

\[ L_1^* = \begin{bmatrix} 1 & \lambda_1^1 & \cdots & \lambda_{|X|}^1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_1^M & \cdots & \lambda_{|X|}^M \end{bmatrix}, \quad L_2^* = \begin{bmatrix} 1 & \lambda_1^1 & \cdots & \lambda_{|X|}^1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_1^M & \cdots & \lambda_{|X|}^M \end{bmatrix}. \]

If \( \text{rank}(L_1^*) = \text{rank}(L_2^*) = M \), then

\[ M = \text{rank}(P^*). \]
A Simple Example for Proposition 3

Suppose $M = 1$ and $X = \{1, 2\}$. Then,

$$L_1^* = [1, \lambda_1^*, \lambda_2^*], \quad L_2^* = [1, \lambda_1, \lambda_2]$$

and $V = 1$ while

$$P^* = (L_1^*)' VL_2^* = \begin{bmatrix}
1 & \lambda_1 & \lambda_2 \\
\lambda_1^* & \lambda_1^* \lambda_1 & \lambda_1^* \lambda_2 \\
\lambda_2^* & \lambda_2^* \lambda_1 & \lambda_2^* \lambda_2 
\end{bmatrix}.$$

Note that

$$\text{rank}(P^*) = M = 1.$$
Other Results

- Baseline Model + Nonstationarity
- Limited Transition Pattern: \( f(x'|x, a) = 0 \) for some \((x', x, a)\)
- Sieve Logit Estimator
Future Research

- Relaxing some of the assumptions
- Testing the number of components, $M$
- The degree of underidentification when $T = 2$

Applications:

- Measurement Errors
- Cross-sectional data with 3 independent variables