The Identification Power of the Markov Assumption in Dynamic Discrete Choice Models

John Lazarev

New York University

March 26, 2015

Motivation

Motivation

- Suppose we have an important economic question
- To answer this question, we use data on the dynamic behavior of economic agents
- Dynamic discrete choice models are tools to analyze these data
- The Markov assumption is a key part of these models
- Many important implementations
- Many important papers about the properties of these tools

1. What testable restrictions does the Markov assumption place on the data?
2. What happens if the agent knows more than the researcher?
3. What can the data tell us about the unobserved information?

Roadmap of Today's Talk

1. Dynamic Discrete Choice Framework with Unobserved Heterogeneity
2. The Observed Markov Property
   - Observable Implications of the Markov Assumption
   - Unobserved Heterogeneity that Preserves the Markov Property
   - Unobserved Heterogeneity that Violates the Markov Property
   - Identification of the Unobserved State
3. Testing the Markov Property
4. Special Cases
   - Persistent Types
   - Stationary Initial State
   - Time Dependent Policy Function
5. Monte Carlo Evidence and Applications

Data

"Short" Panel:

- Agents/markets: $i = 1, \ldots, n$
- Time periods: $t = 1, \ldots, T$
- Actions: a_{it}
- Environment: x_{it}
- Asymptotics: $n \to \infty$, T fixed

Can directly estimate: $P\left\{\left(a_t, x_t\right), t = 1, \ldots, T\right\}$

- Discrete time index: $t = 1, 2, \dots$
- State space: $s_t \in \mathcal{S}$, cardinality is known: $|\mathcal{S}| = S$
- Probability distribution over the initial state: $P(s_1)$
- Action set: $a_t \in \mathcal{A}$, cardinality is known: $|\mathcal{A}| = A$
- State transition probability function: $P_t(s_{t+1}|s_1, a_1, \dots, s_t, a_t)$
- Utility functional: $V(s_t) = \mathbb{E}_t \left[\sum_{\tau=t}^{+\infty} \beta^{\tau-t} v(s_\tau, a_\tau) + \epsilon_\tau(a_\tau)\right]$, that consists of:
    - the single period utility function $v(s_t, a_t) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$
    - the agent's discount factor $\beta \in (0, 1)$
    - the idiosyncratic action-specific error term $\epsilon_\tau = (\epsilon_\tau(1), \dots, \epsilon_\tau(A))$ drawn independently across time from a distribution defined by a p.d.f: $f(\epsilon_t|s_t) : \mathbb{R}^A \times \mathcal{S} \to \mathbb{R}$

Assumptions

1. The state transition $P_t(s_{t+1}|s_t, a_t)$ has the Markov property: $P_t(s_{t+1}|s_1, a_1, \dots, s_t, a_t) = P_t(s_{t+1}|s_t, a_t)$ for any $t > 1$.
2. The state transition is time homogenous: $P_t(s_{t+1}|s_t, a_t) = P(s_{t+1}|s_t, a_t)$ for any t.

Proposition There exists an optimal decision rule that has the Markov property: $d^*(s_t, \epsilon_t) = a_t$

Conditional choice probabilities: $P(a_t|s_t) = \mathbb{E}[\mathcal{I}(d^*(s_t, \epsilon_t) = a_t)]$.

The Unobserved Heterogeneity

Unobserved Heterogeneity = the researcher has only partial information about the state of the world

Formally,

- $s_t = (x_t, u_t)$
- the agent observes s_t
- the researcher observes x_t but not u_t
- define a coarsening function $U : \mathcal{S} \to \mathcal{X}$
- denote $X = |U(\mathcal{S})|$
- unobserved heterogeneity is present when $X < S$

The Observed Markov Property

Data

- The date are defined by $P\left\{(a_t, x_t), t = 1, \ldots, T\right\}$
- Its dimensionality: $(AX)^T - 1$

Can these data be rationalized by a dynamic discrete choice model that has the Markov property?

- without unobserved heterogeneity
- with unobserved heterogeneity

No Unobserved Heterogeneity

Assume $S = X$

Model

- Elements of the Structure:
  - initial state: $P(s_1)$
  - state transition: $P(s_{t+1}|s_t, a_t)$
  - optimal policy: $P(a_t|s_t)$
- Its dimensionality: $(S-1)+(S-1)AS+(A-1)S = AS^2-1$

Thus, the Markov assumption reduces the dimensionality of the object from $(AS)^T - 1$ to $AS^2 - 1$. The number of imposed restrictions grows exponentially with T.

Example

- $S = 3$, $A = 2$, $T = 5$
- dimensionality of the generic process: $6^5 - 1 = 46,656$
- dimensionality under the Markov assumption: $2 \cdot 3^1 - 1 = 7$
- Probability of a randomly chosen process to violate the Markov property: 99.985%

Unobserved Heterogeneity

Suppose $X < S$. What can we infer from the data about the unobserved heterogeneity?

Four reasons for **lack of identification**:

1. relabeling
2. irrelevance (zero-probability states)
3. ovefitting
4. collinearity

- Dimensionality of the data: $(AX)^T - 1$
- Dimensionality of the model: $AS^2 - 1$

Proposition 2: If $T < \frac{\log(A)+2\log(S)}{\log(A)+\log(X)}$, then the primitives of the model are not identified.
Proposition 3: If $X < \frac{S^{2/T}}{A^{\frac{T-1}{T}}}$, then the primitives of the model are not identified.
Proposition 4: Generically, the dimensionality of the data does not equal to the dimensionality of the model.
Thus,

- The panel cannot be too short
- The unobserved heterogeneity cannot be too rich
- The unobserved heterogeneity is either under- or overidentified

Definition Two states s and s' are called collinear if:

- $U(s) = U(s')$
- $P(a|s) = P(a|s')$ for all $a \in \mathcal{A}$
- $\sum_{\tilde{s} \text{ s.t. } U(\tilde{s})=const} P(\tilde{s}|a,s) = \sum_{\tilde{s} \text{ s.t. } U(\tilde{s})=const} P(\tilde{s}|a,s')$

Proposition 5 The primitives of the model are identified for some T if and only if the state space $\mathcal{S}$ does not have collinear states.

Proposition 6 Suppose the data are Markov. Then either there is no unobserved heterogeneity or the state space has collinear states.

Should we be worried about unobserved heterogeneity?

- If the data are Markov, then
  - either there is no unobserved heterogeneity
  - or we cannot tell anything about the unobserved state
- If the data are not Markov, then
  - we can identify unobserved heterogeneity as long as it is not too rich
  - if we can identify unobserved heterogeneity, we will have a set of over identifying restrictions that we can test

Testing the Markov Property

Standard testing framework:

- H_0 = the Markov model
- H_1 = unrestricted process

Three tests are available:

- Estimate the Markov model and run a LM test
- Estimate the unrestricted model and run a Wald test
- Estimate both models and run a LR test

Special Cases

Persistent Types

- J persistent types: the unobserved part u_t doesn't change over time
- dimensionality of the model:
  - initial state: $(X-1)J$
  - state transition: $(X-1)XJ$
  - policy function: $(A-1)XJ$
  - distribution of types: $(J-1)$
  - **overall**: $X^2J + AXJ - 1$
- dimensionality of the Data: $(AX)^T - 1$
- Threshold $T = \frac{\log(X)+\log(J)+\log(A+X)}{\log(A)+\log(X)}$
- Threshold $J = \frac{A^T X^{T-1}}{A+X}$
- No collinear states if different types either act or affect state transition differently

Special Cases

Stationary Initial State

- $P(S_1)$ is the stationary distribution of the Markov process
- Dimensionality of the model reduces to $(S-1)AS + (A-1)S = AS^2 - S$
- Dimensionality of the data: $(AX)^T - 1$
- Threshold $T = \frac{\log(AS^2-S+1)}{\log(A)+\log(X)}$
- Threshold $X = \frac{(AS^2-S+1)^{1/T}}{A}$

- Suppose the optimal policy is time-dependent (e.g. finite-horizon)
- Dimensionality of the model:
  - initial state: $(S-1)$
  - state transition: $(S-1)SA$
  - policy function: $(A-1)ST$
  - **overall**: $AS^2 + (T-1)S(A-1) - 1$
- Dimensionality of the data: $(AX)^T - 1$
- Dimensionality of the model grows linearly in T
- Dimensionality of the data grows exponentially in T
- Therefore, the data are still informative

Main Takeaways from the Paper

1. Unlike many econometric models, the Markov assumption imposes a number of *testable* restrictions on the data.
2. If the data do not reject the Markov property, we cannot say much about unobserved heterogeneity.
3. If the data reject the Markov property, then unobserved heterogeneity can be potentially recovered from the data.

To Be Investigated

1. Monte Carlo properties of different tests
2. What can we say about unobserved heterogeneity in the well-known applications (e.g. Rust (1987), Ryan (2012))?
3. What exactly will we estimate if a part of the state is unobserved but the Markov property still holds?
4. What exactly will we estimate if a part of the state is unobserved and the Markov property fails?