Conditional Independence

Robert A. Miller

Econometrics 2

October 2018

Recapitulation

A dynamic discrete choice model

- Each period $t \in \{1, 2, \ldots, T\}$ for $T \leq \infty$, an individual chooses among J mutually exclusive actions.
- Let d_{jt} equal one if action $j \in \{1, \ldots, J\}$ is taken at time t and zero otherwise:

$$d_{jt} \in \{0, 1\}$$

$$\sum_{j=1}^{J} d_{jt} = 1$$

- Suppose that actions taken at time t can potentially depend on the state $z_t \in Z$.
- The current period payoff at time t from taking action j is $u_{jt}(z_t)$.
- Given choices $(d_{1t}, \ldots, d_{Jt})$ in each period $t \in \{1, 2, \ldots, T\}$ the individual's expected utility is:

$$E\left\{\sum_{t=1}^{T}\sum_{j=1}^{J}\beta^{t-1} d_{jt} u_{jt}(z_t)\right\}$$

Recapitulation

Value function and optimization

- Writing the optimal decision rule as $d_t^o(z) \equiv (d_{1t}^o(z_t), \ldots, d_{Jt}^o(z_t))$, and denoting the value function by $V_t(z_t)$, we obtained:

$$
\begin{aligned}
V_t(z_t) &= \sum_{t=1}^{T} \sum_{j=1}^{J} d_{jt}^o u_{jt}(z_t) \\
&= \sum_{j=1}^{J} d_{jt}^o \left[u_{jt}(z_t) + \beta \sum_{z_{t+1}=1}^{Z} V_{t+1}(z_{t+1}) f_{jt}(z_{t+1} | z_t) \right]
\end{aligned}
$$

- Let $v_{jt}(z_t)$ denote the flow payoff of action j plus the expected future utility of behaving optimally from period $t+1$ on:

$$
v_{jt}(z_t) \equiv u_{jt}(z_t) + \beta \sum_{z_{t+1}=1}^{Z} V_{t+1}(z_{t+1}) f_{jt}(z_{t+1} | z_t)
$$

- Bellman's principle implies:

$$
d_{jt}^o(z_t) \equiv \prod_{k=1}^{K} I\{v_{jt}(z_t) \geq v_{kt}(z_t)\}
$$

- Partitioning the states $z_t \equiv (x_t, \epsilon_t)$ into those which are observed, x_t, and those that are not, ϵ_t, indexing a given specification of $u_{jt}(z_t)$, $f_{jt}\left(z_{t+1}|z_t\right)$ and β by $\theta \in \Theta$, we showed the maximum likelihood estimator, $\theta_{ML} \in \Theta$ selects θ to maximize the joint probability of the observed occurrences:

$$\prod_{n=1}^{N} \int_{\epsilon_T} \ldots \int_{\epsilon_1} \left[\begin{array}{l} \sum_{j=1}^{J} I\left\{d_{njT} = 1\right\} d_{jT}^{o}\left(x_{nT}, \epsilon_T\right) \times \\ \prod\limits_{t=1}^{T-1} H_{nt}\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t\right.\right) g\left(\epsilon_1 \left| x_{n1}\right.\right) \end{array} \right] d\epsilon_1 \ldots d\epsilon_T$$

where:

$$H_{nt}\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t\right.\right) \equiv$$
$$\sum_{j=1}^{J} I\left\{d_{njt} = 1\right\} d_{jt}^{o}\left(x_{nt}, \epsilon_t\right) f_{jt}\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t\right.\right)$$

is the probability density of the pair $\left(x_{n,t+1}, \epsilon_{t+1}\right)$ conditional on $\left(x_{nt}, \epsilon_t\right)$ when choices are optimal for θ, and $d_{njt} = 1$.

Recapitulation

A computational challenge

- What are the computational challenges to enlarging the state space?
  1. Computing the value function;
  2. Solving for equilibrium in a multiplayer setting;
  3. Integrating over unobserved heterogeneity.
- These challenges have led researchers to compromises on several dimensions:
  1. Keep the dimension of the state space small;
  2. Assume all choices and outcomes are observed;
  3. Model unobserved states as a matter of computational convenience;
  4. Consider only one side of market to finesse equilibrium issues;
  5. Adopt parameterizations based on convenient functional forms.

Separable Transitions in the Observed Variables

A simplification

- We could assume that for all (j, t, x_t, ϵ_t) the transition of the observed variables does not depend on the unobserved variables:

$$f_{jt}\left(x_{t+1}\left|x_t, \epsilon_t\right.\right) = f_{jt}\left(x_{t+1}\left|x_t\right.\right)$$

- Since x_{t+1} conveys all the information of x_t for the purposes of forming probability distributions at $t+1$:

$$\begin{aligned} f_{jt}\left(x_{t+1}, \epsilon_{t+1}\left|x_t, \epsilon_t\right.\right) &\equiv g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}, x_t, \epsilon_t\right.\right) f_{jt}\left(x_{t+1}\left|x_t, \epsilon_t\right.\right) \\ &\equiv g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}, \epsilon_t\right.\right) f_{jt}\left(x_{t+1}\left|x_t\right.\right) \end{aligned}$$

- The ML estimator maximizes the same criterion function but $H_{nt}\left(x_{n,t+1}, \epsilon_{t+1}\left|x_{nt}, \epsilon_t\right.\right)$ simplifies to:

$$\begin{aligned} &H_{nt}\left(x_{n,t+1}, \epsilon_{t+1}\left|x_{nt}, \epsilon_t\right.\right) = \\ &\sum_{j=1}^{J} I\left\{d_{njt} = 1\right\} d_{jt}^{o}\left(x_{nt}, \epsilon_t\right) g_{t+1}\left(\epsilon_{t+1}\left|x_{n,t+1}, \epsilon_t\right.\right) f_{jt}\left(x_{n,t+1}\left|x_{nt}\right.\right) \end{aligned}$$

Separable Transitions in the Observed Variables

Exploiting separability in estimation

- Note $f_{jt}(x_{t+1} | x_t)$ is identified for each (j,t) from the transitions.
- Instead of jointly estimating the parameters, we could use a two stage estimator to reduce computation costs:

1. Estimate $f_{jt}(x_{t+1} | x_t)$ with a cell estimator (for x finite), a nonparametric estimator, or a parametric function;
2. Define:

$$\widehat{H}_{nt}(x_{n,t+1}, \epsilon_{t+1} | x_{nt}, \epsilon_t; \theta) \equiv \sum_{j=1}^{J} \left[\begin{array}{l} I\left\{d_{njt}=1\right\} d_{jt}^{o}(x_{nt}, \epsilon_t; \theta) \\ \times g_{t+1}(\epsilon_{t+1} | x_{n,t+1}, \epsilon_t; \theta) \widehat{f}_{jt}(x_{n,t+1} | x_{nt}) \end{array} \right]$$

3. Select the remaining (preference) parameters to maximize:

$$\prod_{n=1}^{N} \int_{\epsilon} \prod_{t=1}^{T} \widehat{H}_{nt}(x_{n,t+1}, \epsilon_{t+1} | x_{nt}, \epsilon_t ; \theta) g_1(\epsilon_1 | x_{n1} ; \theta) d\epsilon$$

4. Correct standard errors from the first stage estimator to account for the loss in asymptotic efficiency.

Conditional Independence

Conditional independence defined

- Separable transitions do not, however, free us from:
  1. the curse of multiple integration;
  2. numerically optimization to obtain the value function.
- Suppose in addition, that conditional on x_{t+1} the unobserved variable ϵ_{t+1} is independent of x_t (often the case in practice) and ϵ_t (a strong assumption).
- Conditional independence embodies both assumptions:

$$\begin{aligned} f_{jt}\left(x_{t+1}\left|x_t,\epsilon_t\right.\right) &= f_{jt}\left(x_{t+1}\left|x_t\right.\right) \\ g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1},x_t,\epsilon_t\right.\right) &= g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.\right) \end{aligned}$$

  It implies:

$$f_{jt}\left(x_{t+1},\epsilon_{t+1}\left|x_t,\epsilon_t\right.\right) = f_{jt}\left(x_{t+1}\left|x_t\right.\right) g_{t+1}\left(\epsilon_{t+1}\left|x_{t+1}\right.\right)$$

- Note that the model in Assignment 1 does not satisfy conditional independence, because posterior beliefs are unobserved state variables governed by a controlled Markov process.

Conditional Independence

Simplifying expressions within the likelihood

- Conditional independence simplifies $H_{nt}\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t \right.\right)$ to:

$$H_{nt}\left(x_{n,t+1}, \epsilon_{t+1} \left| x_{nt}, \epsilon_t \right.\right) =$$
$$\sum_{j=1}^{J} I\left\{d_{njt} = 1\right\} d_{jt}^{o}\left(x_{nt}, \epsilon_t\right) g_{t+1}\left(\epsilon_{t+1} \left| x_{n,t+1} \right.\right) f_{jt}\left(x_{n,t+1} \left| x_{nt} \right.\right)$$

- Also note that:

$$\begin{aligned}
& \prod_{t=1}^{T} \left\{ \sum_{j=1}^{J} I\left\{d_{njt} = 1\right\} d_{jt}^{o}\left(x_{nt}, \epsilon_t\right) f_{jt}\left(x_{n,t+1} \left| x_{nt} \right.\right) \right\} \\
= \; & \prod_{t=1}^{T} \left\{ \sum_{j=1}^{J} I\left\{d_{njt} = 1\right\} f_{jt}\left(x_{n,t+1} \left| x_{nt} \right.\right) \right\} \\
& \times \prod_{t=1}^{T} \left\{ \sum_{j=1}^{J} I\left\{d_{njt} = 1\right\} d_{jt}^{o}\left(x_{nt}, \epsilon_t\right) \right\}
\end{aligned}$$

- Hence the contribution of $n \in \{1, \ldots, N\}$ to the likelihood is the product of:

$$\prod_{t=1}^{T-1} \sum_{j=1}^{J} I\{d_{njt} = 1\} f_{jt}(x_{n,t+1} | x_{nt})$$

and:

$$\int_{\epsilon_T} \ldots \int_{\epsilon_1} \prod_{t=1}^{T-1} \sum_{j=1}^{J} I\{d_{njt} = 1\} d_{jt}^{o}(x_{nt}, \epsilon_t) g_{t+1}(\epsilon_{t+1} | x_{n,t+1}) g_1(\epsilon_1 | x_{n1}) d\epsilon_1 \ldots$$

- The second expression simplifies to:

$$\prod_{t=1}^{T} \left[\sum_{j=1}^{J} I\{d_{njt} = 1\} \int_{\epsilon_t} d_{jt}^{o}(x_{nt}, \epsilon_t) g_t(\epsilon_t | x_{nt}) d\epsilon_t \right]$$

Conditional Independence

Conditional choice probabilities defined

- Under conditional independence, we define for each (t, x_t) the conditional choice probability (CCP) for action j as:

$$\begin{aligned} p_{jt}(x_t) &\equiv \int_{\epsilon_t} d_{jt}^o(x_{nt}, \epsilon_t)\, g_t(\epsilon_t | x_{nt})\, d\epsilon_t \\ &= E\left[d_{jt}^o(x_t, \epsilon_t) | x_t\right] \\ &= \int_{\epsilon_t} \prod_{k=1}^{J} I\{v_{kt}(x_{nt}, \epsilon_t) \leq v_{jt}(x_{nt}, \epsilon_t)\}\, g_t(\epsilon_t | x_{nt})\, d\epsilon_t \end{aligned}$$

- Using this notation, the likelihood can now be compactly expressed as:

$$\begin{aligned} &\sum_{n=1}^{N} \sum_{t=1}^{T-1} \sum_{j=1}^{J} I\{d_{njt} = 1\} \ln\left[f_{jt}(x_{n,t+1} | x_{nt})\right] \\ &+ \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} I\{d_{njt} = 1\} \ln p_{jt}(x_t) \end{aligned}$$

- Conditional independence implies that $v_{jt}(x_t, \epsilon_t)$ only depends on ϵ_t through $u_{jt}(x_t, \epsilon_t)$ because:

$$v_{jt}(x_t, \epsilon_t) \equiv$$
$$u_{jt}(x_t, \epsilon_t) + \beta \int_{\epsilon_t+1} \sum_{x_{t+1}=1}^{X} V_{t+1}(x_{t+1}, \epsilon_{t+1}) f_{jt}\left(x_{t+1} \left| x_t\right.\right) g_{t+1}\left(\epsilon_{t+1} \left| x_{t+1}\right.\right)$$

- Without further loss of generality we now define:

$$u_{jt}(x_t, \epsilon_t) \equiv E\left[u_{jt}(x_t, \epsilon_t) \left| x_t\right.\right] + \epsilon^*_{jt} \equiv u^*_{jt}(x_t) + \epsilon^*_{jt}$$

- In this way we redefine the primitives by the preferences $u^*_{jt}(x_t)$, the observed variables transitions $f_{jt}\left(x_{t+1} \middle| x_t\right)$, and the distribution of unobserved variables $g^*_t\left(\epsilon^*_t \middle| x_t\right)$ where $\epsilon^*_t \equiv \left(\epsilon^*_{1t}, \ldots, \epsilon^*_{Jt}\right)$.

Conditional Independence

Conditional value functions defined

- Given conditional independence, define the conditional valuation function as:

$$v^*_{jt}(x_t) \equiv u^*_{jt}(x_t) + \beta \int_{\epsilon_t+1} \sum_{x_{t+1}=1}^{X} V^*_{t+1}(x_{t+1}, \epsilon^*_{t+1}) f_{jt}(x_{t+1} | x_t) g^*_{t+1}(\epsilon^*_{t+1} | x_{t+1})$$

- Thus $p_{jt}(x)$ is found by integrating over $(\epsilon_{1t}, \ldots, \epsilon_{Jt})$ in the regions:

$$\epsilon^*_{kt} - \epsilon^*_{jt} \leq v^*_{jt}(x_t) - v^*_{kt}(x_t)$$

hold for all $k \in \{1, \ldots, J\}$. That is $p_{jt}(x_t)$ can be rewritten:

$$\int_{\epsilon_t} \prod_{k=1}^{J} I \{ v_{kt}(x_{nt}, \epsilon_t) \leq v_{jt}(x_{nt}, \epsilon_t) \} g_t(\epsilon_t | x_t) d\epsilon_t$$

$$= \int_{\epsilon_t} \prod_{k=1}^{J} I \{ \epsilon^*_{kt} - \epsilon^*_{jt} \leq v^*_{jt}(x_{nt}) - v^*_{kt}(x_{nt}) \} g^*_t(\epsilon^*_t | x_t) d\epsilon^*_t$$

- Suppose we only had data on the last period T, and wished to estimate the preferences determining choices in T.
- By definition this is a static problem in which $v_{jT}(z_T) \equiv u_{jT}(z_T)$.
- For example to the probability of observing the J^{th} choice is:

$$p_{JT}(z_T) \equiv \int_{-\infty}^{\epsilon_{JT}+u_{JT}(z_T)-u_{1T}(z_T)} \cdots \int_{-\infty}^{\epsilon_{JT}+u_{JT}(z_T)-u_{J-1,T}(z_T)} \int_{-\infty}^{\infty} g_T(\epsilon_T | x_T) \, d\epsilon_T$$

- The only essential difference between a estimating a static discrete choice model using ML and a estimating a dynamic model satisfying conditional independence using ML is that parametrizations of $v_{jt}(x_t)$ based on $u_{jt}(x_t)$ do not have a closed form, but must be computed numerically.

Another Renewal Problem (Rust,1987)

Bus engines

- The structural model estimated in my JPE paper is a renewal problem: if there is only one occupation to work in and an infinite number of jobs within that occupation, then no job is revisited after a spell in it ends, and leaving aside the life cycle trend, every new job match restarts life.
- Replacing engines can be modeled that way too. Mr. Zurcher decides whether to replace the existing engine ($d_{1t} = 1$), or keep it for at least one more period ($d_{2t} = 1$).
- Bus mileage advances 1 unit ($x_{t+1} = x_t + 1$) if Zurcher keeps the engine ($d_{2t} = 1$) and is set to zero otherwise ($x_{t+1} = 0$ if $d_{1t} = 1$).
- Transitory iid choice-specific shocks, ϵ_{jt} are Type 1 extreme value.
- Zurcher sequentially maximizes expected discounted sum of payoffs:

$$E\left\{\sum_{t=1}^{\infty}\beta^{t-1}\left[d_{2t}(\theta_1 x_t + \theta_2 s + \epsilon_{2t}) + d_{1t}\epsilon_{1t}\right]\right\}$$

- Let $V(x_t, s)$ denote the ex-ante value function at the beginning of period t, the discounted sum of current and future payoffs just before ϵ_t is realized and before the decision at t is made.
- We also define the conditional value function for each choice as:

$$v_j(x, s) = \begin{cases} \beta V(0, s) & \text{if } j = 1 \\ \theta_1 x + \theta_2 s + \beta V(x+1, s) & \text{if } j = 2 \end{cases}$$

- Let $p_1(x, s)$ denote the conditional choice probability (CCP) of replacing the engine given x and s.
- The parametric assumptions about the transitory shocks imply:

$$p_1(x, s) = \frac{1}{1 + \exp\left[v_2(x, s) - v_1(x, s)\right]}$$

- An ML estimator could be formed off this equation following the steps described above.

- One can show (next mini) that when ϵ_{jt} is Type 1 extreme value, then for all $(x, s,)$:

$$V(x, s) = v_j(x, s) - \beta \log\left[p_j(x, s)\right] + 0.57\ldots$$

Therefore the conditional valuation function of not replacing is:

$$\begin{aligned} v_2(x, s) &= \theta_1 x + \theta_2 s + \beta V(x, s+1) \\ &= \theta_1 x + \theta_2 s + \beta\left\{v\left({}_1 x + 1, s\right) - p_1(x+1, s) + 0.57\ldots\right\} \end{aligned}$$

- Similarly:

$$v_1(x, s) = \beta V(0, s) = \beta\left\{v_1(0, s) - \ln\left[p_1(0, s)\right] + 0.57\right\}\ldots$$

- Because the miles on a bus engine is the only factor affecting the value of the bus:

$$v_1(x+1, s) = v_1(0, s)$$

Another Renewal Problem

Using CCPs to represent differences in continuation values

- Hence:

$$v_2(x,s) - v_1(x,s) = \theta_1 x + \theta_2 s + \beta \ln[p_1(0,s)] - \beta \ln[p_1(x+1,s)]$$

- Therefore:

$$\begin{aligned} p_1(x,s) &= \frac{1}{1+\exp[v_2(x,s)-v_1(x,s)]} \\ &= \frac{1}{1+\exp\left\{\theta_1 x + \theta_2 s + \beta \ln\left[\frac{p_1(0,s)}{p_1(x+1,s)}\right]\right\}} \end{aligned}$$

- Intuitively the CCP for current replacement is the CCP for a static model with an offset term.
- The offset term accounts for differences in continuation values using future CCPs that characterize optimal future replacements.

- Consider the following CCP estimator.
- Form first stage estimate for $p_1(x, s)$, called $\hat{p}_1(x, s)$ from the relative frequencies:

$$\hat{p}_1(x, s) = \frac{\sum_{n=1}^{N} d_{1nt} I(x_{nt} = x) I(s_n = s)}{\sum_{n=1}^{N} I(x_{nt} = x) I(s_n = s)}$$

- In second stage substitute $\hat{p}_1(x, s)$ into the likelihood as incidental parameters and estimate θ_1 and θ_2 with a logit:

$$\frac{d_{1nt} + d_{2nt} \exp(\theta_1 x_{nt} + \theta_2 s_n + \beta \ln [\hat{p}_1(0, s_n)] - \beta \ln [\hat{p}_1(x_{nt} + 1, s_n)]}{1 + \exp(\theta_1 x_{nt} + \theta_2 s_n + \beta \ln [\hat{p}_1(0, s_n)] - \beta \ln [\hat{p}_1(x_{nt} + 1, s_n)])}$$

Monte Carlo Study (Arcidiacono and Miller, 2011)

Modifying the bus engine problem

- Suppose bus type $s \in \{0, 1\}$ is equally weighted.
- There are two other state variables
  1. total accumulated mileage:
  $$x_{1t+1} = \begin{cases} \Delta_t \text{ if } d_{1t} = 1 \\ x_{1t} + \Delta_t \text{ if } d_{2t} = 1 \end{cases}$$
  2. permanent route characteristic for the bus, x_2, that systematically affects miles added each period.
- We assume $\Delta_t \in \{0, 0.125, \ldots, 24.875, 25\}$ is drawn from a truncated exponential distribution:

$$f(\Delta_t | x_2) = \exp\left[-x_2(\Delta_t - 25)\right] - \exp\left[-x_2(\Delta_t - 24.875)\right]$$

and x_2 is a multiple 0.01 drawn from a discrete equi-probability distribution between 0.25 and 1.25.

Monte Carlo Study

Including aggregate shocks in panel estimation

- Let θ_{0t} denote an aggregate shock (denoting fully anticipated cost fluctuations). Then the difference in current payoff from retaining versus replacing the engine is:

$$u_{2t}(x_{1t}, s) - u_{1t}(x_{1t}, s) \equiv \theta_{0t} + \theta_1 \min\{x_{1t}, 25\} + \theta_2 s$$

- Denoting $x_t \equiv (x_{1t}, x_2)$, this implies:

$$\begin{aligned} v_{2t}(x_t, s) - v_{1t}(x_t, s) &= \theta_{0t} + \theta_1 \min\{x_{1t}, 25\} + \theta_2 s \\ &\quad + \beta \sum_{\Delta_t \in \Lambda} \left\{ \ln\left[\frac{p_{1t}(0, s)}{p_{1t}(x_{1t} + \Delta_t, s)}\right]\right\} f(\Delta_t | x_2) \end{aligned}$$

- In the first three columns of the next table each sample is on 1000 buses for 20 periods, while in the fourth column we assume 2000 buses are observed for 10 periods.
- The mean and standard deviations are compiled from 50 simulations.

Monte Carlo Study

Extract from Table 1 of Arcidiacono and Miller (2011)

Monte Carlo for Optimal Stopping Problem[+]

	DGP (1)	FIML (2)	CCP (3)	Time effects CCP (4)
θ_0 (Intercept)	2	2.0100	1.9911	
		(0.0405)	(0.0399)	
θ_1 (Mileage)	-0.15	-0.1488	-0.1441	-0.1440
		(0.0074)	(0.0098)	(0.0121)
θ_2 (Type)	1	0.9945	0.9726	0.9683
		(0.0611)	(0.0668)	(0.0636)
β (Discount Factor)	0.9	0.9102	0.9099	0.9172
		(0.0411)	(0.0554)	(0.0639)
Time (Minutes)		130.29	0.078	0.079
		(19.73)	(0.0041)	(0.0047)

[+] Mean and standard deviations for fifty simulations. For columns (2) and (3), the observed data consist of 1000 buses for 20 periods. For column (4), the intercept (θ_0) is allowed to vary over time and the data consist of 2000 buses for 10 periods.