Representation

Robert A. Miller

Structural Econometrics

September 2017

Reviewing Conditional Independence

State variables and transitions

- Each period t until T, for $T \leq \infty$, an individual chooses among J mutually exclusive actions.
- For each action $j \in \{1, \ldots, J\}$ and period $t \in \{1, \ldots, T\}$ define:

$$d_{jt} = \begin{cases} 1 \text{ if action } j \text{ taken at period } t \\ 0 \text{ if not} \end{cases}$$

- If action j is taken at time t, the probability of x_{t+1} occurring in period $t+1$ is denoted by $f_{jt}(x_{t+1}|x_t)$.

- Suppose $\epsilon_t \equiv (\epsilon_{1t}, \ldots, \epsilon_{Jt})$ is revealed at the beginning of the period t, has continuous support and is *iid* with density function $g_t(\epsilon \,|x_t)$.
- The individual's current period payoff from choosing j at time t is:

$$u_{jt}(x_t) + \epsilon_{jt}$$

- The individual chooses $d_t \equiv (d_{1t}, \ldots, d_{Jt})$ to sequentially maximize:

$$E\left\{\sum_{t=1}^{T}\sum_{j=1}^{J}\beta^{t-1} d_{jt}\left[u_{jt}(x_t) + \epsilon_{jt}\right]\right\}$$

  where at each period t the expectation is taken over the future values of $x_{t+1}, \ldots, x_T$ and $\epsilon_{t+1}, \ldots, \epsilon_T$.

Reviewing Conditional Independence

Optimization

- Denote the optimal decision rule at t as $d_t^o(x_t, \epsilon_t)$, with j^{th} element $d_{jt}^o(x_t, \epsilon_t)$ and define:

$$V_t(x_t) \equiv E\left\{\sum_{\tau=t}^{T}\sum_{j=1}^{J}\beta^{\tau-t-1}d_{j\tau}^o\left(x_\tau, \epsilon_\tau\right)\left(u_{j\tau}(x_\tau) + \epsilon_{j\tau}\right)\right\}$$

- The conditional value function, $v_{jt}(x_t)$, is defined as:

$$v_{jt}(x_t) = u_{jt}(x_t) + \beta\sum_{x=1}^{X}V_{t+1}(x)f_{jt}\left(x|x_t\right)$$

- Integrating $d_{jt}^o(x_t, \epsilon)$ over $\epsilon \equiv (\epsilon_1, \ldots, \epsilon_J)$:

$$p_{jt}(x_t) \equiv E\left[d_{jt}^o\left(x_t, \epsilon\right)|x_t\right] = \int d_{jt}^o\left(x_t, \epsilon\right)g_t\left(\epsilon\,|x_t\right)d\epsilon$$

- The starting point for our analysis is to define differences in the conditional valuation functions as:

$$\Delta v_{jkt}\left(x\right) \equiv v_{jt}\left(x\right) - v_{kt}\left(x\right)$$

- Although there are $J\left(J-1\right)$ differences all but $\left(J-1\right)$ are linear combinations of the $\left(J-1\right)$ basis functions.
- For example setting the basis functions as:

$$\Delta v_{jt}\left(x\right) \equiv v_{jt}\left(x\right) - v_{Jt}\left(x\right)$$

  then clearly:

$$\Delta v_{jkt}\left(x\right) = \Delta v_{jt}\left(x\right) - \Delta v_{kt}\left(x\right)$$

- Without loss of generality we focus on this particular basis function.

Inversion

Each CCP is a mapping of differences in the conditional valuation functions

- Using the definition of $\Delta v_{jt}(x)$:

$$\begin{aligned} p_{jt}(x) &\equiv \int d_{jt}^{o}(x,\epsilon)\, g_t(\epsilon\,|x)\, d\epsilon \\ &= \int I\{\epsilon_k \leq \epsilon_j + \Delta v_{jt}(x) - \Delta v_{kt}(x) \forall\ k \neq j\}\, g_t(\epsilon\,|x)\, d\epsilon \\ &= \int_{-\infty}^{\epsilon_j+\Delta v_{jt}(x)-\Delta v_{1t}(x)} \cdots \int_{-\infty}^{\epsilon_j+\Delta v_{jt}(x)-\Delta v_{J-1,t}(x)} \int_{-\infty}^{\epsilon_j+\Delta v_{jt}(x)} g_t(\epsilon\,|x)\, d\epsilon \end{aligned}$$

- Noting $g_t(\epsilon\,|x) \equiv \partial^J G_t(\epsilon\,|x)/\partial\epsilon_1, \ldots, \partial\epsilon_J$, integrate over $(\epsilon_1, \ldots, , \epsilon_{j-1}, \epsilon_{j+1} \ldots, \epsilon_J)$.
- Denoting $G_{jt}(\epsilon\,|x) \equiv \partial G_t(\epsilon\,|x)/\partial\epsilon_j$, yields:

$$p_{jt}(x) = \int_{-\infty}^{\infty} G_{jt}\left(\begin{array}{c} \epsilon_j + \Delta v_{jt}(x) - \Delta v_{1t}(x), \ldots \\ \ldots, \epsilon_j, \ldots, \epsilon_j + \Delta v_{jt}(x) \end{array} |x \right) d\epsilon_j$$

Inversion

There are as many CCPs as there are conditional valuation functions

- For any vector $J-1$ dimensional vector $\delta \equiv (\delta_1, \ldots, \delta_{J-1})$ define:

$$Q_{jt}(\delta, x) \equiv \int_{-\infty}^{\infty} G_{jt}(\epsilon_j + \delta_j - \delta_1, \ldots, \epsilon_j, \ldots, \epsilon_j + \delta_j | x) \, d\epsilon_j$$

- We interpret $Q_{jt}(\delta, x)$ as the probability taking action j in a static random utility model (RUM) where the payoffs are $\delta_j + \epsilon_j$ and the probability distribution of disturbances is given by $G_t(\epsilon | x)$.
- It follows from the definition of $Q_{jt}(\delta, x)$ that:

$$0 \leq Q_{jt}(\delta, x) \leq 1 \text{ for all } (j, t, \delta, x) \text{ and } \sum_{j=1}^{J-1} Q_{jt}(\delta, x) \leq 1$$

- In particular the previous slide implies that for any given (j, t, x):

$$p_{jt}(x) = \int_{-\infty}^{\infty} G_{jt}\left(\begin{array}{c} \epsilon_j + \Delta v_{jt}(x) - \Delta v_{1t}(x), \\ \ldots, \epsilon_j, \ldots, \epsilon_j + \Delta v_{jt}(x) \end{array} |x \right) d\epsilon_j \equiv Q_{jt}(\Delta v_t(x), x)$$

Theorem (Inversion)

For each (t, δ, x) *define:*

$$Q_t(\delta, x) \equiv (Q_{1t}(\delta, x), \ldots Q_{J-1,t}(\delta, x))'$$

Then the vector function $Q_t(\delta, x)$ *is invertible in* δ *for each* (t, x).

- Note that $p_{Jt}(x) = Q_{Jt}(\Delta v_t, x)$ is a linear combination of the other equations in the system because $\sum_{k=1}^{J} p_k = 1$.
- Let $p \equiv (p_1, \ldots, p_{J-1})$ where $0 \leq p_j \leq 1$ for all $j \in \{1, \ldots, J-1\}$ and $\sum_{j=1}^{J-1} p_j \leq 1$. Denote the inverse of $Q_{jt}(\Delta v_t, x)$ by $Q_{jt}^{-1}(p, x)$.
- The inversion theorem implies:

$$\begin{bmatrix} \Delta v_{1t}(x) \\ \vdots \\ \Delta v_{J-1,t}(x) \end{bmatrix} = \begin{bmatrix} Q_{1t}^{-1}[p_t(x), x] \\ \vdots \\ Q_{J-1,t}^{-1}[p_t(x), x] \end{bmatrix}$$

Inversion

Using the inversion theorem

- In finessing optimization and integration by exploiting conditional independence, how far can the three applications described in the previous two lectures be extended?
- We use the Inversion Theorem to:
  1. provide empirically tractable representations of the conditional value functions.
  2. analyze identification in dynamic discrete choice models.
  3. provide convenient parametric forms for the density of ϵ_t that generalize the Type 1 Extreme Value distribution.
  4. generalize the renewal and terminal state properties exploited in the first two examples, by obtaining restrictions on the state variable transitions used to implement CCP estimators.
  5. introduce new methods for incorporating unobserved state variables.

- From the definition of the optimal decision rule, and then appealing to the inversion theorem:

$$
\begin{aligned}
d_{jt}^{o}\left(x_t, \epsilon_t\right) &= \prod\nolimits_{k=1}^{J} 1\left\{\epsilon_{kt} - \epsilon_{jt} \leq v_{jt}(x) - v_{kt}(x)\right\} \\
&= \prod\nolimits_{k=1}^{J} 1\left\{\epsilon_{kt} - \epsilon_{jt} \leq \begin{array}{l} v_{jt}(x) - v_{Jt}(x_t) \\ -\left[v_{kt}(x) - v_{Jt}(x_t)\right] \end{array}\right\} \\
&= \prod\nolimits_{k=1}^{J} 1\left\{\epsilon_{kt} - \epsilon_{jt} \leq \Delta v_{jt}(x) - \Delta v_{kt}(x)\right\} \\
&= \prod\nolimits_{k=1}^{J} 1\left\{\epsilon_{kt} - \epsilon_{jt} \leq Q_{jt}^{-1}\left[p_t(x), x\right] - Q_{kt}^{-1}\left[p_t(x), x\right]\right\}
\end{aligned}
$$

- If $G_t\left(\epsilon \left| x\right.\right)$ is known and the data generating process (DGP$)$ is (x_t, d_t), then $p_t(x)$ and hence $d_t^{o}\left(x_t, \epsilon_t\right)$ are identified.

Corollaries of the Inversion Theorem

Definition of the conditional value function correction

- Define:

$$\psi_{jt}(x) \equiv V_t(x) - v_{jt}(x)$$

  is the conditional value function correction. In stationary settings, we drop the t subscript and write:

$$\psi_j(x) \equiv V(x) - v_j(x)$$

- Suppose that instead of taking the optimal action she committed to taking action j instead. Then the expected lifetime utility would be:

$$v_{jt}(x_t) + E_t\left[\epsilon_{jt} \left| x_t\right.\right]$$

  so committing to j before ϵ_t is revealed entails a loss of:

$$V_t(x_t) - v_{jt}(x_t) - E_t\left[\epsilon_{jt} \left| x_t\right.\right] = \psi_{jt}(x) - E_t\left[\epsilon_{jt} \left| x_t\right.\right]$$

- For example if $E_t\left[\epsilon_t \left| x_t\right.\right] = 0$, the loss simplifies to $\psi_{jt}(x)$.

Corollaries of the Inversion Theorem

Identifying the conditional value function correction

- From their respective definitions:

$$\begin{aligned} & V_t(x) - v_{it}(x) \\ = \; & \sum_{j=1}^{J} \left\{ p_{jt}(x) \left[v_{jt}(x) - v_{it}(x) \right] + \int \epsilon_{jt} d_{jt}^{o} \left(x_t, \epsilon_t \right) g_t \left(\epsilon_t \left| x \right. \right) d\epsilon_t \right\} \end{aligned}$$

- But:

$$v_{jt}(x) - v_{it}(x) = Q_{jt}^{-1} \left[p_t(x), x \right] - Q_{it}^{-1} \left[p_t(x), x \right]$$

and

$$\begin{aligned} & \int \epsilon_{jt} d_{jt}^{o} \left(x, \epsilon_t \right) g \left(\epsilon_t \left| x \right. \right) d\epsilon_t \\ = \; & \int \prod_{k=1}^{J} 1 \left\{ \begin{array}{l} \epsilon_{kt} - \epsilon_{jt} \\ \leq Q_{jt}^{-1} \left[p_t(x), x \right] - Q_{kt}^{-1} \left[p_t(x), x \right] \end{array} \right\} \epsilon_{jt} g_t \left(\epsilon_t \left| x \right. \right) d\epsilon_t \end{aligned}$$

Therefore $\psi_{it}(x) \equiv V_t(x) - v_{it}(x)$ is identified if $G_t \left(\epsilon \left| x \right. \right)$ is known and (x_t, d_t) is the DGP.

Conditional Valuation Function Representation

Telescoping one period forward

- From its definition:

$$v_{jt}(x_t) = u_{jt}(x_t) + \beta \sum_{x=1}^{X} V_{t+1}(x) f_{jt}(x_{t+1}|x_t)$$

- Substituting for $V_{t+1}(x_{t+1})$ using conditional value function correction we obtain for any k:

$$v_{jt}(x_t) = u_{jt}(x_t) + \beta \sum_{x=1}^{X} \left[v_{k,t+1}(x) + \psi_{k,t+1}(x) \right] f_{jt}(x|x_t)$$

- We could repeat this procedure ad infinitum, substituting in for $v_{k,t+1}(x)$ by using the definition for $\psi_{kt}(x)$.

Conditional Valuation Function Representation

Recursively defining the distribution of future state variables

- To formalize this idea, consider a random sequence of weights from t to T which begins with $\omega_{jt}(x_t, j) = 1$.
- For periods $\tau \in \{t+1, \ldots, T\}$, the choice sequence maps x_τ and the initial choice j into

$$\omega_\tau(x_\tau, j) \equiv \{\omega_{1\tau}(x_\tau, j), \ldots, \omega_{J\tau}(x_\tau, j)\}$$

where $\omega_{k\tau}(x_\tau, j)$ may be negative or exceed one but:

$$\sum_{k=1}^{J} \omega_{k\tau}(x_\tau, j) = 1$$

.

- The weight of state $x_{\tau+1}$ conditional on following the choices in the sequence is recursively defined by $\kappa_t(x_{t+1}|x_t, j) \equiv f_{jt}(x_{t+1}|x_t)$ and for $\tau = t+1, \ldots, T$:

$$\kappa_\tau(x_{\tau+1}|x_t, j) \equiv \sum_{x_\tau=1}^{X} \sum_{k=1}^{J} \omega_{k\tau}(x_\tau, j) f_{k\tau}(x_{\tau+1}|x_\tau)\kappa_{\tau-1}(x_\tau|x_t, j)$$

Theorem (Representation)

For any state $x_t \in \{1, \ldots, X\}$*, choice* $j \in \{1, \ldots, J\}$ *and weights* $\omega_\tau(x_\tau, j)$ *defined for periods* $\tau \in \{t, \ldots, T\}$*:*

$$\begin{aligned} v_{jt}(x_t) &= u_{jt}(x_t) \\ &+ \sum_{\tau=t+1}^{T} \sum_{k=1}^{J} \sum_{x=1}^{X} \beta^{\tau-t} \left[u_{k\tau}(x) + \psi_k[p_\tau(x)] \right] \omega_{k\tau}(x, j) \kappa_{\tau-1}(x | x_t, j \end{aligned}$$

- The theorem yields an alternative expression for $v_{jt}(x_t)$ that dispenses with recursive maximization.
- Intuitively, the individuals have already solved their optimization problem, so their decisions, as reflected in their CCPs, are informative of their value functions.

- Both theorems apply to this multiagent setting with two critical differences, and both are relevant for studying identification:
  1. $u_{jt}(x_t)$ is a primitive in single agent optimization problems, but $u_{jt}^{(i)}(x_t)$ is a reduced form parameter found by integrating $U_{jt}^{(i)}\left(x_t, d_t^{(\sim i)}\right)$ over the joint probability distribution $P_t\left(d_t^{(\sim i)} \left| x_t\right.\right)$.
  2. $f_{jt}\left(x_{t+1} \left| x_t\right.\right)$ is a primitive in single agent optimization problems, but $f_{jt}^{(i)}\left(x_{t+1} \left| x_t\right.\right)$ depends on CCPs of the other players, $P_t\left(d_t^{(\sim i)} \left| x_t\right.\right)$, as well as the primitive $F_{jt}\left(x_{t+1} \left| x_t, d_t^{(\sim i)}\right.\right)$. It is easy to interpret restrictions placed directly on $f_{jt}\left(x_{t+1} \left| x_t\right.\right)$ but placing restrictions on $F_{jt}\left(x_{t+1} \left| x_t, d_t^{(\sim i)}\right.\right)$ complicates matters in dynamic games because of the endogenous effects arising from $P_t\left(d_t^{(\sim i)} \left| x_t\right.\right)$ on $f_{jt}^{(i)}\left(x_{t+1} \left| x_t\right.\right)$.

- Are there tractable distributions $G_t(\epsilon | x)$ aside from the Type 1 Extreme Value?
- To keep the approach operational we have to compute $\psi_k(p)$ for at least some k.
- Suppose ϵ is drawn from the GEV distribution function:

$$G(\epsilon_1, \epsilon_2, \ldots, \epsilon_J) \equiv \exp\left[-\mathcal{H}(\exp[-\epsilon_1], \exp[-\epsilon_2], \ldots, \exp[-\epsilon_J])\right]$$

where $\mathcal{H}(Y_1, Y_2, \ldots, Y_J)$ satisfies the following properties:

1. $\mathcal{H}(Y_1, Y_2, \ldots, Y_J)$ is nonnegative, real valued, and homogeneous of degree one;
2. $\lim \mathcal{H}(Y_1, Y_2, \ldots, Y_J) \to \infty$ as $Y_j \to \infty$ for all $j \in \{1, \ldots, J\}$;
3. for any distinct $(i_1, i_2, \ldots, i_r)$ the cross derivative $\partial \mathcal{H}(Y_1, Y_2, \ldots, Y_J) / \partial Y_{i_1}, Y_{i_2}, \ldots, Y_{i_r}$ is nonnegative for r odd and nonpositive for r even.

- Suppose $G(\epsilon)$ factors into two independent distributions, one a nested logit, and the other any GEV distribution.
- Let $\mathcal{J}$ denote the set of choices in the nest and denote the other distribution by $G_0(Y_1, Y_2, \ldots, Y_K)$ let K denote the number of choices that are outside the nest.
- Then:

$$G(\epsilon) \equiv G_0(\epsilon_1, \ldots, \epsilon_K) \exp\left[-\left(\sum_{j\in\mathcal{J}} \exp\left[-\epsilon_j/\sigma\right]\right)^{\sigma}\right]$$

- The correlation of the errors within the nest is given by $\sigma \in [0,1]$ and errors within the nest are uncorrelated with errors outside the nest. When $\sigma = 1$, the errors are uncorrelated within the nest, and when $\sigma = 0$ they are perfectly correlated.

- Define $\phi_j(Y)$ as a mapping into the unit interval where

$$\phi_j(Y) = Y_j \mathcal{H}_j \left(Y_1, \ldots, Y_J \right) / \mathcal{H} \left(Y_1, \ldots, Y_J \right)$$

- Since $\mathcal{H}_j \left(Y_1, \ldots, Y_J \right)$ and $\mathcal{H} \left(Y_1, \ldots, Y_J \right)$ are homogeneous of degree zero and one respectively, $\phi_j(Y)$ is a probability, because $\phi_j(Y) \geq 0$ and $\sum_{j=1}^{J} \phi_j(Y) = 1$.

Lemma (GEV correction factor)

When ϵ_t is drawn from a GEV distribution, the inverse function of $\phi(Y) \equiv (\phi_2(Y), \ldots \phi_J(Y))$ exists, which we now denote by $\phi^{-1}(p)$, and:

$$\psi_j(p) = \ln \mathcal{H} \left[1, \phi_2^{-1}(p), \ldots, \phi_J^{-1}(p) \right] - \ln \phi_j^{-1}(p) + \gamma$$

Generalized Extreme Values

Correction factor for extended nested logit

Lemma

For the nested logit $G(\epsilon_t)$ defined above:

$$\psi_j(p) = \gamma - \sigma \ln(p_j) - (1-\sigma) \ln\left(\sum_{k \in \mathcal{J}} p_k\right)$$

- Note that $\psi_j(p)$ only depends on the conditional choice probabilities for choices that are in the nest: the expression is the same no matter how many choices are outside the nest or how those choices are correlated.
- Hence, $\psi_j(p)$ will only depend on $p_{j'}$ if ϵ_{jt} and $\epsilon_{j't}$ are correlated. When $\sigma = 1$, ϵ_{jt} is independent of all other errors and $\psi_j(p)$ only depends on p_j.