CCP Estimators

Robert A. Miller

University of Tokyo

October 2019

- Suppose the data comes from a long panel (either stationary or complete panel histories for finite lived agents).
- Also assume we know:
  1. the discount factor β
  2. the distribution of disturbances $G_t(\epsilon \,|x)$
  3. $u_{1t}(x)$ (or more generally one of the payoffs for each state and time).
  4. $u_{1t}(x) = 0$ (for notational convenience)
- Since the panel is long, $p_t(x)$ and hence $\psi_{jt}(x)$ are identified.
- There are, of course, alternative assumptions that deliver identification, and the methods described below are generic.

Unrestricted Maximum Likelihood Estimation

The likelihood

- To simplify the notation, consider a sample of N independently drawn observations on the whole history $t \in \{1, \ldots, T\}$ of individuals $n \in \{1, \ldots, N\}$, with data on their state variables decisions denoted by x_{nt}, and decisions denoted by d_{njt}.
- The joint probability distribution of the decisions and outcomes is:

$$\prod_{n=1}^{N} \prod_{t=1}^{T} \left(\sum_{j=1}^{J} \sum_{x'=1}^{X} d_{njt} I \left\{ x_{n,t+1} = x' \right\} p_{jt}(x) f_{jt}(x'|x) \right)$$

- Taking logs yields:

$$\sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} d_{njt} \left\{ \log \left[p_{jt}(x_{nt}) \right] + \sum_{x=1}^{X} I \left\{ x_{n,t+1} = x \right\} \log \left[f_{jt}(x|x_{nt}) \right] \right\}$$

Unrestricted Maximum Likelihood Estimation

The reduced form

- Note the choice probabilities are additively separable from the transition probabilities in the formula for the joint distribution of decisions and outcomes.
- Hence the estimation of the joint likelihood splits into one piece dealing with the choice probabilities conditional on the state, and another dealing with the transition conditional on the choice and state.
- Maximizing each additive piece separately with respect to $f_j(x'|x)$ and $p_t(x_{nt})$ we obtain the unrestricted ML estimators:

$$\widehat{f}_{jt}\left(x' \left| x\right.\right) = \frac{\sum_{n=1}^{N} I\left\{x_{nt} = x, d_{njt} = 1, x_{n,t+1} = x'\right\}}{\sum_{n=1}^{N} I\left\{x_{nt} = x, d_{njt} = 1\right\}}$$

and:

$$\widehat{p}_{jt}\left(x\right) = \frac{\sum_{n=1}^{N} I\left\{x_{nt} = x, d_{njt} = 1\right\}}{\sum_{n=1}^{N} I\left\{x_{nt} = x\right\}}$$

- Let $\kappa_{jt\tau}(x_{t+\tau+1}|x_t)$ denote the probability of reaching $x_{t+\tau+1}$ at $t+\tau+1$ from x_t by following action j at t and then always choosing the first action:

$$\kappa_{jt\tau}(x_{t+\tau+1}|x_t) \equiv \begin{cases} f_{jt}(x_{t+1}|x_t) & \tau = 0 \\ \sum_{x=1}^{X} f_{1,t+\tau}(x_{t+\tau+1}|x)\kappa_{jt,\tau-1}(x|x_t) & \tau = 1, \ldots \end{cases}$$

- Thus we can recursively estimate $\kappa_{jt\tau}(x_{t+\tau+1}|x_t)$ with:

$$\widehat{\kappa}_{jt\tau}(x_{t+\tau+1}|x_t) \equiv \begin{cases} \widehat{f}_{jt}(x_{t+1}|x_t) & \tau = 0 \\ \sum_{x=1}^{X} \widehat{f}_{1,t+\tau}(x_{t+\tau+1}|x)\widehat{\kappa}_{jt,\tau-1}(x|x_t) & \tau = t+1, \ldots \end{cases}$$

- Similarly we estimate $\psi_{jt}(x_t)$ with $\widehat{\psi}_{jt}(x_t)$ using the $\widehat{p}_{jt}(x)$ estimates of the CCPs.

Unrestricted Maximum Likelihood Estimation

Estimating the primitives

- From previous lectures:

$$
\begin{aligned}
u_{jt}(x_t) &= \psi_{1t}(x_t) - \psi_{jt}(x_t) \\
&\quad + \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^{\tau-t} \psi_{1,t+\tau}(x) \left[\kappa_{t1,\tau-1}(x|x_t) - \kappa_{tj,\tau-1}(x|x_t) \right]
\end{aligned}
$$

- Substituting $\widehat{\kappa}_{\tau-1}(x|x_t, j)$ for $\kappa_{\tau-1}(x|x_t, j)$ and $\psi_{jt}(x_t)$ with $\widehat{\psi}_{jt}(x_t)$ then yields:

$$
\begin{aligned}
\widehat{u}_{jt}(x_t) &\equiv \widehat{\psi}_{1t}(x_t) - \widehat{\psi}_{jt}(x_t) \\
&\quad + \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^{\tau-t} \widehat{\psi}_{1,t+\tau}(x) \left[\widehat{\kappa}_{t1,\tau-1}(x|x_t) - \widehat{\kappa}_{tj,\tau-1}(x|x_t) \right]
\end{aligned}
$$

- The stationary case is similar (and has the matrix representation we discussed in previous lectures).

- By the Law of Large Numbers $\widehat{f}_{jt}(x'|x)$ converges to $f_{jt}(x'|x)$ and $\widehat{p}_{jt}(x)$ converges to $p_{jt}(x)$, both almost surely.
- By the Central Limit Theorem both estimators converge at $\sqrt{N}$ and and have asymptotic normal distributions.
- Both $\widehat{f}_{jt}(x'|x)$ and $\widehat{p}_{jt}(x)$ are ML estimators for $f_{jt}(x'|x)$ and $p_{jt}(x)$ and obtain the Cramer-Rao lower bound asymptotically.
- Since and $u_{jt}(x)$ is exactly identified, it follows by the invariance principle that $\widehat{u}_{jt}(x)$ is consistent and asymptotically efficient for $u_{jt}(x_t)$, also attaining its Cramer Rao lower bound.
- The same properties apply to the stationary model.
- Note that greater efficiency can only be obtained by making functional form assumptions about $u_{jt}(x_t)$ and $f_{jt}(x'|x)$.
- False restrictions, such as adopting convenient functional forms for the payoffs, typically create misspecifications.

- In practice applications further restrict the parameter space.
- For example assume $\theta \equiv \left(\theta^{(1)}, \theta^{(2)}\right) \in \Theta$ is a closed convex subspace of Euclidean space, and:
  - $u_{jt}(x) \equiv u_j(x, \theta^{(1)})$
  - $f_{jt}(x|x_{nt}) \equiv f_{jt}(x|x_{nt}, \theta^{(2)})$
- We now define the model by (T, β, θ, g).
- Assume the DGP comes from (T, β, θ_0, g) where $\theta_0 \in \Theta^{(interior)}$.
- The ML estimator, denoted by θ_{ML}, maximizes:

$$\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} d_{njt}\left\{\ln\left[p_{jt}(x_{nt}, \theta)\right] + \sum_{x=1}^{X} I\left\{x_{n,t+1} = x\right\}\ln\left[f_{jt}\left(x|x_{nt}, \theta^{(2)}\right)\right]\right\}$$

over $\theta \in \Theta$ where $p_t(x, \theta)$ are the CCPs for (T, β, θ, g).

- A common variation on the ML estimator is:
  1. estimate $f_{jt}(x|x_{nt},\theta^{(2)})$ from the state transitions.
  2. obtain a limited information ML estimator $\theta^{(2)}_{LIML}$.
  3. estimate $\theta^{(1)}$ by searching over $p_t(x,\theta^{(1)},\theta^{(2)}_{LIML})$.
- More precisely we define:

$$\theta^{(2)}_{LIML} \equiv \arg\max_{\theta_2} \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{x=1}^{X} I\left\{x_{n,t+1}=x\right\} d_{njt}\log\left[f_{jt}(x|x_{nt},\theta^{(2)})\right]$$

$$\widehat{\theta}^{(1)}_{ML} \equiv \arg\max_{\theta_1} \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} d_{njt}\left\{\log\left[p_{jt}(x_{nt},\theta^{(1)},\theta^{(2)}_{LIML})\right]\right\}$$

- Note that:
  - when $\theta^{(2)}_0$, that is $f_{jt}(x|x_{nt})$, is known, $\widehat{\theta}^{(1)}_{ML}=\theta^{(1)}_{ML}$;
  - otherwise $\widehat{\theta}^{(1)}_{ML}$ is less efficient but computationally simpler than $\theta^{(1)}_{ML}$;
  - nevertheless both estimators solve for the optimal rule many times.

Quasi Maximum Likelihood Estimation

The steps (Hotz and Miller, 1993)

- The essential difference between this estimator and ML is this estimator substitutes an estimator of the continuation value into the likelihood rather than computing it from the optimal policy function:
  1. Estimate the reduced form $\widehat{p}$ and $\widehat{f}$ (or $\theta^{(2)}_{LIML}$) as above;
  2. Apply the Representation Theorem to obtain expressions for $v_{jt}(x_t) - v_{kt}(x_t)$;
  3. Substitute the reduced form estimates into these differences to obtain $\widehat{v}_{jt}\left(x, \theta^{(1)}\right) - \widehat{v}_{kt}\left(x, \theta^{(1)}\right)$ for any given $\theta^{(1)}$;
  4. Replace $v_{jt}(x_t)$ with $\widehat{v}_{jt}\left(x, \theta^{(1)}\right)$ in the random utility model (RUM) to obtain an estimate $\widehat{p}_{jt}\left(x, \theta^{(1)}\right)$ for any given $\theta^{(1)}$;
  5. Maximize the quasi-likelihood with respect to $\theta^{(1)}$.
- In effect we estimate a static RUM where differences in current utilities $u_j\left(x, \theta^{(1)}\right) - u_k\left(x, \theta^{(1)}\right)$ are augmented by a *dynamic correction factor* estimated in the first stage off the reduced form.

- In the second step, appealing to the Representation theorem, and the slides above $\widehat{v}_{jt}\left(x, \theta^{(1)}\right) - \widehat{v}_{kt}\left(x, \theta^{(1)}\right) =$

$$u_j\left(x, \theta^{(1)}\right) - u_k\left(x, \theta^{(1)}\right) - \sum_{\tau=1}^{T-t}\sum_{x=1}^{X}\beta^{\tau}\widehat{\psi}_{1,t+\tau}(x)\left[\begin{array}{c}\widehat{\kappa}_{kt,\tau-1}(x|x_t) \\ -\widehat{\kappa}_{jt,\tau-1}(x|x_t)\end{array}\right]$$

- In the last two steps we define:

$$\widehat{p}_{jt}\left(x, \theta^{(1)}\right) \equiv \int_{\epsilon_t}\prod_{k=1}^{J} I\left\{\begin{array}{l}\epsilon_{kt} - \epsilon_{jt} \\ \leq \widehat{v}_{jt}\left(x, \theta^{(1)}\right) - \widehat{v}_{kt}\left(x, \theta^{(1)}\right)\end{array}\right\} dG_t\left(\epsilon_t \,|x_t\right)$$

and:

$$\theta^{(1)}_{QML} \equiv \arg\max_{\theta_1}\sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J} d_{njt}\left\{\log\left[\widehat{p}_{jt}(x_{nt}, \theta^{(1)}, \theta^{(2)}_{LIML})\right]\right\}$$

Quasi Maximum Likelihood Estimation

Adjusting the asymptotic covariance for pre-estimation (Hotz and Miller, 1993)

- Form $P\left(\theta^{(1)}, p, f\right)$, a mapping from $\Theta^{(1)} \times P \times F$ to P with:

$$\kappa_{jt\tau}(x_{t+\tau+1}|x_t) \equiv \begin{cases} f_{jt}(x_{t+1}|x_t) & \tau = 0 \\ \sum_{x=1}^{X} f_{1,t+\tau}(x_{t+\tau+1}|x)\kappa_{jt,\tau-1}(x|x_t) & \tau = t+1, \ldots \end{cases}$$

$$\begin{aligned} v_{jt}\left(x, \theta^{(1)}\right) - v_{kt}\left(x, \theta^{(1)}\right) &= u_j\left(x, \theta^{(1)}\right) - u_k\left(x, \theta^{(1)}\right) \\ &\quad - \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^\tau \psi_{1,t+\tau}(x) \left[\begin{array}{l} \kappa_{kt,\tau-1}(x|x_t) \\ -\kappa_{jt,\tau-1}(x|x_t) \end{array} \right] \end{aligned}$$

$$p_{jt}\left(x, \theta^{(1)}\right) \equiv \int_{\epsilon_t} \prod_{k=1}^{J} I \left\{ \begin{array}{l} \epsilon_{kt} - \epsilon_{jt} \\ \leq v_{jt}\left(x, \theta^{(1)}\right) - v_{kt}\left(x, \theta^{(1)}\right) \end{array} \right\} g_t\left(\epsilon_t \,|x_t\right) d\epsilon_t$$

- Let:

$$\pi_{1n}\left(\theta^{(1)}, p, f\right) = W_N\left\{z_n \otimes \left[d_n - P\left(\theta^{(1)}, p, f\right)\right]\right\}$$

where W_N is a weighting matrix and z_n are instruments.

- Define the CCP estimator for $\theta^{(1)}_{CCP}$ by solving:

$$\sum_{n=1}^{N} \pi_{1n}\left(\theta^{(1)}, \widehat{p}, \widehat{f}\right) = 0$$

- Write the cell estimators as the solution to:

$$0 = \sum_{n=1}^{N} \pi_{2n}\left(\widehat{p}, \widehat{f}\right) = \sum_{n=1}^{N} \begin{bmatrix} I_n^d \left(d_n - \widehat{p}\right) \\ I_n^f \left(f_n - \widehat{f}\right) \end{bmatrix}$$

  where:
  - I_n^d is a $(J-1)XT$ dimensional row vector indicator function matching the state variables of n to the relevant CCP component(s) in p;
  - I_n^f is a $(J-1)X^2T$ dimensional row vector indicator function matching state variables and decision(s) of n to f components;
  - f_n is the outcome from the n making a choice given her state variables.
- For $k \in \{1, 2\}$ and $k' \in \{1, 2\}$ define:

$$\Omega_{kk'} \equiv E\left[\pi_{kn}\pi'_{k'n}\right] \qquad \Gamma_{11} \equiv E\left[\frac{\partial \pi_{1n}}{\partial \theta^{(1)}}\right] \qquad \Gamma_{12} \equiv E\left[\frac{\partial \pi_{1n}}{\partial p}, \frac{\partial \pi_{1n}}{\partial f}\right]$$

- Then the asymptotic covariance matrix for $\theta_{CCP}^{(1)}$, denoted by Σ_1, is:

$$\Sigma_1 = \Gamma_{11}^{-1}\left[\Omega_{11} + \Gamma_{12}\left(\Omega_{22} - \Omega_{21} - \Omega_{12}\right)\Gamma'_{12}\right]\Gamma_{11}^{-1\prime}$$

- Another approach is to match up the parametrization of $u_{jt}(x_t)$, denoted by $u_{jt}(x_t, \theta^{(1)})$, to its representation as closely as possible:
  1. Form the vector function where $\Psi(p, f)$ by stacking:

$$\begin{aligned} \Psi_{jt}(x_t, p, f) &\equiv \psi_{1t}(x_t) - \psi_{jt}(x_t) \\ &+ \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^{\tau} \psi_{1,t+\tau}(x) \left[\begin{array}{l} \kappa_{kt,\tau-1}(x|x_t) \\ -\kappa_{jt,\tau-1}(x|x_t) \end{array} \right] \end{aligned}$$

  2. Estimate the reduced form $\widehat{p}$ and $\widehat{f}$.
  3. Minimize the quadratic form to obtain:

$$\begin{aligned} \theta^{(1)}_{MD} &= \underset{\theta^{(1)} \in \Theta^{(1)}}{\arg\min} \left[u(x, \theta^{(1)}) - \Psi\left(\widehat{p}, \widehat{f}\right) \right]' \widetilde{W} \left[u(x, \theta^{(1)}) - \Psi\left(\widehat{p}, \widehat{f}\right) \right] \\ &= \underset{\theta^{(1)} \in \Theta^{(1)}}{\arg\min} \left[u(x, \theta^{(1)})' \widetilde{W} u(x, \theta^{(1)}) - 2\Psi\left(\widehat{p}, \widehat{f}\right)' \widetilde{W} u(x, \theta^{(1)}) \right] \end{aligned}$$

where $\widetilde{W}$, is a square $(J-1)\,TX$ weighting matrix.

Minimum Distance Estimators

Notes on minimizing the difference between unrestricted and restricted payoffs

- From the Representation theorem $u_{jt}(x_t, \theta_0^{(1)}) = \Psi_{jt}(x_t, p, f_0)$ if p are the CCPs for (T, β, θ_0, g).
- Furthermore $u_{jt}(x)$ is exactly identified from $\Psi_{jt}(x, p, f_0)$ without imposing any additional restrictions.
- Therefore parameterizing u with $\theta_0^{(1)}$ imposes overidentifying restrictions so $\theta_{MD}^{(1)}$ is consistent if the restrictions are true.
- Note $\theta_{MD}^{(1)}$ has a closed form if $u(x; \theta_0^{(1)})$ is linear in $\theta_0^{(1)}$.

Simulated Moments Estimators

- We could form a Methods of Simulated Moments (MSM) estimator from:
  1. Simulate a lifetime path from x_{nt_n} onwards for each j, using $\widehat{f}$ and $\widehat{p}$.
  2. Obtain estimates of $\widehat{E}\left[\epsilon_{jt} \middle| d^o_{jt} = 1, x_t\right]$.
  3. Stitch together a simulated lifetime utility outcome from the j^{th} choice at t_n onwards for n, denoted $\widehat{v}_{nj} \equiv \widehat{v}_{jt_n}\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)$.
  4. Form the $J-1$ dimensional vector $h_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)$ from:

$$
\begin{aligned}
h_{nj}\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right) \quad \equiv \quad & \widehat{v}_{jt_n}\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right) - \widehat{v}_{Jt_n}\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right) \\
& + \widehat{\psi}_{jt}(x_{nt_n}) - \widehat{\psi}_{Jt}(x_{nt_n})
\end{aligned}
$$

  5. Given a weighting matrix W_S and an instrument vector z_n minimize:

$$
N^{-1}\left[\sum_{n=1}^{N} z_n h_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)\right]' W_S \left[\sum_{n=1}^{N} z_n h_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)\right]
$$

- In the first step, given the state simulate a choice using $\widehat{p}$, and simulate the next state using $\widehat{f}$. In this way generate $\widehat{x}_{ns}$ and $\widehat{d}_{ns} \equiv \left(\widehat{d}_{n1s}, \ldots, \widehat{d}_{nJs}\right)$ for all $s \in \{t_n + 1, \ldots, T\}$.
- Generating this path does not exploit knowledge of G, only the CCPs.
- In the second step $\widehat{E}\left[\epsilon_{jt} \middle| d^o_{jt} = 1, x_t\right] \equiv$

$$p_{jt}^{-1}(x_t) \int_{\epsilon_t} \prod_{k=1}^{J} I\left\{\widehat{\psi}_{jt}(x_t) - \widehat{\psi}_{kt}(x_t) \leq \epsilon_{jt} - \epsilon_{kt}\right\} \epsilon_{jt} g(\epsilon_t)\, d\epsilon_t$$

- In Step 4 $\widehat{v}_{jt}\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right)$ is stitched together as:

$$u_{jt}(x_{nt_n}, \theta^{(1)}) + \sum_{s=t+1}^{T} \sum_{j=1}^{J} \beta^{t-1} 1\left\{\widehat{d}_{njs} = 1\right\} \left\{ \begin{array}{l} u_{js}(\widehat{x}_{ns}, \theta^{(1)}) \\ +\widehat{E}\left[\epsilon_{js} \middle| \widehat{x}_{ns}, \widehat{d}_{njs} = 1\right] \end{array} \right\}$$

- The solution has a closed form if $u_{jt}(x, \theta^{(1)})$ is linear in $\theta^{(1)}$.

Simulated Moments Estimators

Another MSM estimator

- Indeed ϵ_t could be simulated as well:
  1. Draw a realization $\widehat{\epsilon}$ from $G(\epsilon)$ for each $s \in \{t_n, \ldots, T\}$ and n.
  2. Set:
  $$\widehat{d}_{njs} = \prod_{k=1}^{J} I\left\{\widehat{\psi}_{js}(\widehat{x}_{ns}) - \widehat{\psi}_{ks}(\widehat{x}_{ns}) \leq \widehat{\epsilon}_{njs} - \widehat{\epsilon}_{nks}\right\}$$
  and stitch together:
  $$u_{jt}(x_{nt_n}, \theta^{(1)}) + \sum_{s=t+1}^{T} \sum_{j=1}^{J} \beta^{t-1} 1\left\{\widehat{d}_{njs} = 1\right\}\left\{u_{js}(\widehat{x}_{ns}, \theta^{(1)}) + \widehat{\epsilon}_{js}\right\}$$
  3. Minimize an analogous quadratic form to obtain $\theta^{(1)}$.
- Bajari, Benkard and Levin (2007) estimate an approximate reduced form of the policy function without exploiting the CCPs (pages 1341-1342, 2007), but acknowledge: "Our method requires that one be able to consistently estimate each firm's policy function, so this may limit our ability to estimate certain models (page 1345, 2007)."

Large Sample or Asymptotic Properties

Adjusting the asymptotic covariance for simulation as well (Pakes and Pollard, 1989)

- Simulation adds an additional, independent source of variation to the sample moments and hence the estimated asymptotic standard errors.
- Following the definition given in Lecture 7 suppose $\widehat{\theta}^{(1)}$ minimizes:

$$N^{-1}\left[\sum_{n=1}^{N} z_n h_n\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right)\right]' W_S \left[\sum_{n=1}^{N} z_n h_n\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right)\right]$$

- Then the additional component to the covariance matrix for $\widehat{\theta}^{(1)}$ is:

$$\Sigma_1^S \equiv S^{-1}\left(\Upsilon' W_S \Upsilon\right)^{-1} \Upsilon' W_S E\left[z_n h_n h_n' z_n'\right] W_S \Upsilon \left(\Upsilon' W_S \Upsilon\right)^{-1}$$

  where S is the number of simulations (per observation):

$$\Upsilon = E\left[\frac{z_n \partial h_n\left(x_{nt_n}, \theta_0^{(1)}, f_0, p_0\right)}{\partial \theta^{(1)}}\right]$$

- Note that $\Sigma_1^S \rightarrow 0$ as $S \rightarrow \infty$.

- Recall that for any extremum estimator for a problem satisfying standard regularity conditions that:

$$\theta^{(i+1)} = \theta^{(i)} - \left[\partial^2 Q_N\left(\theta^{(i)}\right)/\partial\theta\partial\theta'\right]^{-1}\left[\partial Q_N\left(\theta^{(i)}\right)/\partial\theta\right]$$

  where N indicates the sample, θ is the parameter value, and $Q_N(\theta)$ is the criterion function associated with the extremum estimator
- This algorithm converges to the maximand if the criterion function is strictly concave, and/or if $\theta^{(i)}$ is close enough to the maximum.
- The algorithm is based on the quadratic approximation:

$$\begin{aligned} Q_N(\theta) \simeq\ & Q_N\left(\theta^{(i)}\right) + \left[\partial Q_N\left(\theta^{(i)}\right)/\partial\theta\right]'\left(\theta - \theta^{(i)}\right) \\ & + \frac{1}{2}\left(\theta - \theta^{(i)}\right)'\left[\partial^2 Q_N\left(\theta^{(i)}\right)/\partial\theta\partial\theta'\right]\left(\theta - \theta^{(i)}\right) \end{aligned}$$

Asymptotic Efficiency

Iterating one step (Amemiya, 1985, pp 137 - 139)

- Suppose $\theta^{(1)}$ is a $\sqrt{N}$ consistent estimator for the (interior) true value $\theta_0 \in \Theta$, the parameter space. Then it is well known that $\theta^{(1)}$ has the same asymptotic properties as $\theta^{(\infty)}$, the limit of the sequence, namely:

$$\sqrt{N}\left(\theta^{(2)} - \theta_0\right) \underset{a.d.}{\sim} \left[p\lim \frac{1}{N} \frac{\partial^2 Q_N\left(\theta^{(1)}\right)}{\partial\theta\partial\theta'}\right]^{-1} \left[p\lim \frac{1}{\sqrt{N}} \frac{\partial Q_N\left(\theta^{(1)}\right)}{\partial\theta}\right]$$

- Specializing $Q_N(\theta) = \log L_N(\theta)$, the log likelihood, $\theta^{(2)}$ is asymptotically efficient, and $\sqrt{N}\left(\theta^{(2)} - \theta_0\right)$ is asymptotically normal with mean zero and covariance:

$$-\left\{\lim E\left[\frac{1}{N}\frac{\partial^2 \log L_N(\theta_0)}{\partial\theta\partial\theta'}\right]\right\}^{-1}$$

- This general principle can be applied to dynamic discrete choice models:
    1. Estimate the (unrestricted) CCPs.
    2. Use a CCP estimator to obtain $\theta^{(1)}$, the parameters characterizing the primitives.
    3. Solve for the CCPs as a mapping of $\theta^{(1)}$ and the state variables.
    4. Take one Newton-Raphson step to obtain $\theta^{(2)}$.
- Note that this procedure asmptotically guarantees the global optimum is selected.
- Starting out with a trial guess of a $\theta^{(0)}$ and then updating, the traditional way of implementing ML, reaches a local (not global) optimum at best ("Overshooting" cannot be ruled out).

Revisiting the Renewal Problem (Rust,1987)

Bus engines

- Recall 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 $d_{1t} = 1$).
- Zurcher sequentially maximizes expected discounted sum of payoffs:

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

where the iid choice-specific shocks, ϵ_{jt} are Type 1 extreme value.

- 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:

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

- Letting $p_1(x, s)$ denote the conditional choice probability (CCP) of replacing the engine given x and s:

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

- From previous lectures 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:
$$\begin{aligned} v_2(x,s) &= \theta_1 x + \theta_2 s + \beta V(x, s+1) \\ &= \theta_1 x + \theta_2 s + \beta\left\{v_1\left(x+1, s\right) - p_1(x+1, s) + 0.57\ldots\right\} \end{aligned}$$
- Similarly:
$$v_1(x,s) = \beta V(1,s) = \beta\left\{v_1(1,s) - \ln\left[p_1(1,s)\right] + 0.57\right\}\ldots$$
- Because the miles on a bus engine is the only factor affecting the value of the bus:
$$v_1(x+1, s) = v_1(1, s)$$

Revisiting the Renewal Problem

Using CCPs to represent differences in continuation values

- Differencing the expressions:

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

- Therefore:

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

- Intuitively the CCP for current replacement is the CCP for a static model with an offset term, to account for differences in continuation values from their exante value functions, $V(x,s+1) - V(1,s)$.
- The renewal property is a simple example of finite dependence.
- In general, models with finite dependence have offset terms that only depend on CCPs a finite number of periods into the future.

- 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:

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

- 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:

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

- 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)

- Suppose there is a finite maximum number of firms in a market at any one time denoted by I.
- If a firm exits, the next period an opening occurs to a potential entrant, who may decide to exercise this one time option, or stay out.
- At the beginning of each period every incumbent firm has the option of quitting the market or staying one more period.
- Let $d_t^{(i)} \equiv \left(d_{1t}^{(i)}, d_{2t}^{(i)} \right)$, where $d_{1t}^{(i)} = 1$ means i exits or stays out of the market in period t, and $d_{2t}^{(i)} = 1$ means i enters or does not exit.
- If $d_{2t}^{(i)} = 1$ and $d_{1,t-1}^{(i)} = 1$ then the firm in spot i at time t is an entrant, and if $d_{2,t-1}^{(i)} = 1$ the spot i at time t is an incumbent.

Entry Exit Game
State Variables

- In this application there are three components to the state variables and $x_t = (x_1, x_{2t}, s_t)$.
- The first is a permanent market characteristic, denoted by x_1, and is common across firms in the market. Each market faces an equal probability of drawing any of the possible values of x_1 where $x_1 \in \{1, 2, \ldots, 10\}$.
- The second, x_{2t}, is whether or not each firm is an incumbent, $x_{2t} \equiv \{d_{2t-1}^{(1)}, \ldots, d_{2t-1}^{(I)}\}$. Entrants pay a start up cost, making it more likely that stayers choose to fill a slot than an entrant.
- A demand shock $s_t \in \{1, \ldots, 5\}$ follows a first order Markov chain.
- In particular, the probability that $s_{t+1} = s_t$ is fixed at $\pi \in (0, 1)$, and probability of any other state occurring is equally likely:

$$\Pr\{s_{t+1} | s_t\} = \begin{cases} \pi \text{ if } s_{t+1} = s_t \\ (1-\pi)/4 \text{ if } s_{t+1} \neq s_t \end{cases}$$

- Each active firm produces one unit so revenue, denoted by y_t, is just price.
- Price is determined by:
    1. the supply of active firms in the market, $\sum_{i=1}^{I} d_{2t}^{(i)}$
    2. a permanent market characteristic, x_1
    3. the Markov demand shock s_t
    4. another temporary shock, denoted by η_t, distributed *iid* standard normal distribution, revealed to each market after the entry and exit decisions are made.
- The price equation is:

$$y_t = \alpha_0 + \alpha_1 x_1 + \alpha_2 s_t + \alpha_3 \sum_{i=1}^{I} d_{2t}^{(i)} + \eta_t$$

- We assume costs comprise a choice specific disturbance $\epsilon_{jt}^{(i)}$ that is privately observed, plus a linear function of z_t.
- Net current profits for exiting incumbent firms, and potential entrants who do not enter, are $\epsilon_{1t}^{(i)}$. Thus $U_1^{(i)}\left(x_t^{(i)}, s_t^{(i)}, d_t^{(-i)}\right) \equiv 0$.
- Current profits from being active are the sum of $\left(\epsilon_{2t}^{(i)} + \eta_t\right)$ and:

$$U_2^{(i)}\left(x_t^{(i)}, s_t^{(i)}, d_t^{(-i)}\right) \equiv \theta_0 + \theta_1 x_1 + \theta_2 s_t + \theta_3 \sum_{\substack{i'=1 \\ i' \neq i}}^{I} d_{2t}^{(i')} + \theta_4 d_{1,t-1}^{(i)}$$

  where θ_4 is the startup cost that only entrants pay.
- In equilibrium $E\left(\eta_t\right) = 0$ so:

$$u_j^{(i)}\left(x_t, s_t\right) = \theta_0 + \theta_1 x_1 + \theta_2 s_t + \theta_3 \sum_{\substack{i'=1 \\ i' \neq i}}^{I} p_2^{(i')}\left(x_t, s_t\right) + \theta_4 d_{1,t-1}^{(i)}$$

- We assume $\epsilon_{jt}^{(i)}$ is distributed Type 1 extreme value.
- The exit payoff is normalized to zero.
- We can now show the conditional value function for being active is:

$$\begin{aligned} v_2^{(i)}(x_t, s_t) &= u_2^{(i)}(x_t, s_t) \\ &\quad -\beta \sum_{x \in X} \sum_{s \in S} \left(\ln \left[p_1^{(i)}(x, s) \right] \right) f_2^{(i)}(x, s | x_t, s_t) \end{aligned}$$

- Note the future value term only depends on one-period-ahead CCPs and the transition probabilities of the state variables.
- Exiting is a terminal choice, another example of finite dependence.

Entry Exit Game

Monte Carlo

- The number of firms in each market is set to six and we simulated data for 3,000 markets.
- The discount factor is set at $\beta = 0.9$.
- Starting at an initial date with six potential entrants in the market, we solved the model, ran the simulations forward for twenty periods, and used the last ten periods to estimate the model.
- The key difference between this Monte Carlo and the renewal Monte Carlo is that the conditional choice probabilities have an additional effect on both current utility and the transitions on the state variables due to the effect of the choices of the firm's competitors on profits.

Entry Exit Game

Extract from Table 2 of Arcidiacono and Miller (2011)

Summary

Factors to consider when selecting an esimator

- There is a trade off between efficiency and computational ease:
  1. Asymptotic efficiency
     - ML and the 2 step (CCP/Newton) estimators are the most efficient.
  2. Small sample properties
     - Simulation induces additional variation that may increase the number of observations required to approximate the asymptotic distribution.
  3. Computational ease
     - MD with linear utility in the parameters has a closed form.
     - ML is the most burdensome and typically requires numerical approximations.
- Finally there is an open question about how well these estimators perform when the model is misspecified.