Conditional Independence and the Inversion Theorem

Robert A. Miller

DSE at Lausanne

August 2023

- Several of this week's sessions are devoted to:
  - solving more complex models with more sophisticated and faster algorithms *(Simon, John S., Yucheng, Felix, Jesus F)*.
  - developing simply computed estimators, for example based on conditional choice probabilities (CCP), that essentially replace policy functions with relative frequencies and draw heavily on the strengths of larger data sets *(Bob and Whitney)*
- These lectures typically do one of:
  - integrate the model solution into the estimation routine with a nested fixed point algorithm, for example NFXP (*John R., Bertel, Fedor)*
  - exploit model data generating process (without solving it) to determine identification and obtain estimates (*this pair of lectures)*
  - use numerical values drawn from published empirical work to quantify model solution, sometimes called calibration (*Felix*)

- Each approach has its strengths and weaknesses.
- Incorporating the equilibrium solution within the estimation algorithm:
  - yields the maximum likelihood estimator.
  - is a way to achieve asymptotic efficiency.
  - and the fixed point algorithm doubles as the solution to counterfactuals.
- Separating inference from the model solution:
  - gives the identification conditions.
  - yields less efficient but much faster estimates.
  - requires the model solution to compute counterfactuals.
- Calibration methods:
  - can dispense with the estimation step altogether.
  - disconnect sample variation from population probabilities.
  - does not typically gives estimates of precision.

- Relevant factors for this debate might be:
  - the kind of data including how much
  - the complexity of the model
  - the sensitivity of the estimates to the underlying assumptions
  - what specifically, is the policy question being asked
- But let's not debate now which method should be applied when.
- These two lectures are dealt with more comprehensively in:
  - https://comlabgames.com/structuraleconometrics/
- Much of the material in this lecture comes from my papers:
  - "Conditional choice probabilities and the estimation of dynamic models," with V. Joseph Hotz, *The Review of Economic Studies*, 1993, pp. 497 - 531.
  - "Identifying dynamic discrete choice models off short panels," with Peter Arcidiacono, *Journal of Econometrics*, 2020, pp.473 - 485.

- Let $T \in \{1, 2, \ldots\}$ with $T \leq \infty$ denote the horizon of the optimization problem and $t \in \{1, \ldots, T\}$ denote the time period.
- Each period the individual chooses amongst J mutually exclusive actions.
- Let $d_t \equiv (d_{1t}, \ldots, d_{Jt})$ where $d_{jt} = 1$ if action $j \in \{1, \ldots, J\}$ is taken at time t and $d_{jt} = 0$ if action j is not taken at t.
- $x_t \in \{1, \ldots, X\}$ for some finite positive integer X for each t.
- $\epsilon_t \equiv (\epsilon_{1t}, \ldots, \epsilon_{Jt})$ where $\epsilon_{jt} \in \mathbb{R}$ for all (j, t).
- Assume the data comprises observations on (d_t, x_t).
- The joint mixed density function for the state in period $t + 1$ conditional on (x_t, ϵ_t), denoted by $g_{t,x,\epsilon}(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t)$, satisfies the *conditional independence assumption*:

$$g_{t,j,x,\epsilon}(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t) = g_{t+1}(\epsilon_{t+1} | x_{t+1}) f_{jt}(x_{t+1} | x_t)$$

where $g_t(\epsilon_t | x_t)$ is a conditional density for the disturbances, and $f_{jt}(x_{t+1} | x)$ is a transition probability for x conditional on (j, t).

Framework

Bounded additively separable preferences

- Denote the discount factor by $\beta \in (0,1)$ and the current payoff from taking action j at t given (x_t, ϵ_t) by $u_{jt}(x_t) + \epsilon_{jt}$.
- To ensure a transversality condition is satisfied, assume $\{u_{jt}(x)\}_{t=1}^{T}$ is a bounded sequence for each $(j,x) \in \{1,\ldots,J\} \times \{1,\ldots,X\}$, and so is:
$$\left\{ \int \max\left\{ |\epsilon_{1t}|, \ldots, |\epsilon_{Jt}| \right\} g_t\left(\epsilon_t | x_t\right) d\epsilon_t \right\}_{t=1}^{T}$$
- At the beginning of each period t the agent observes the realization (x_t, ϵ_t) chooses d_t to sequentially maximize:
$$E\left\{ \sum_{\tau=t}^{T} \sum_{j=1}^{J} \beta^{\tau-1} d_{j\tau}\left[u_{j\tau}(x_\tau) + \epsilon_{j\tau} \right] | x_t, \epsilon_t \right\} \tag{1}$$
where the expectation is taken over future realized values $x_{t+1}, \ldots, x_T$ and $\epsilon_{t+1}, \ldots, \epsilon_T$ conditional on (x_t, ϵ_t).

- 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 the *social surplus function* as:

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

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

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

- Integrating $d_{jt}^o(x_t, \epsilon)$ over $\epsilon \equiv (\epsilon_1, \ldots, \epsilon_J)$ define the *conditional choice probabilities* CCPs by:

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

Extension to Dynamic Markov Games

Players, choices and state variables

- Consider a dynamic game for I countable players:
  1. $d_t^{(i)} \equiv \left(d_{t1}^{(i)}, \ldots, d_{tJ}^{(i)} \right)$ choice of player i in period t.
  2. $d_t \equiv \left(d_t^{(1)}, \ldots, d_t^{(I)} \right)$ choices of all the players in period t.
  3. $d_t^{(-i)} \equiv \left(d_t^{(1)}, \ldots, d_t^{(i-1)}, d_t^{(i+1)}, \ldots, d_t^{(I)} \right)$ choices of all but i in t.
  4. x_t value of state variables of the game in period t.
  5. $F\left(x_{t+1} \left| x_t, d_t \right. \right)$ transition probability for x_{t+1} given (x_t, d_t).
  6. $F_j\left(x_{t+1} \left| x_t, d_t^{(-i)} \right. \right) \equiv F\left(x_{t+1} \left| x_t, d_t^{(-i)}, d_{jt}^{(i)} = 1 \right. \right)$ transition probability for x_{t+1} given x_t, i choosing j, and everyone else $d_t^{(-i)}$.

- The summed discounted payoff to i from playing the game is:

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

  where:

  1. $U_j^{(i)} \left(x_t, d_t^{(-i)} \right)$ depends on the choices of all the players.
  2. $\epsilon_t^{(i)} \equiv \left(\epsilon_{1t}^{(i)}, \ldots, \epsilon_{Jt}^{(i)} \right)$ is *iid* across i with density $g\left(\epsilon_t^{(i)} | x_t \right)$.
  3. neither $d_t^{(-i)}$ nor $\epsilon_t^{(-i)}$ are observed by i.

- Analogous to the single agent setup define:

  1. $p_j^{(i)}(x_t) = \int d_j^{(i)} \left(x_t, \epsilon_t^{(i)} \right) g\left(\epsilon_t^{(i)} \right) d\epsilon_t^{(i)}$ as the CCP for the i choosing j in period t.
  2. $P\left(d_t^{(-i)} | x_t \right) = \prod_{i'=1, i' \neq i}^{I} \left(\sum_{j=1}^{J} d_{jt}^{(i')} p_j^{(i')}(x_t) \right)$ as the CCP for all the other players choosing $d_t^{(-i)}$ in period t.

- Then $\left(p_1^{(i)}(x_t), \ldots, p_J^{(i)}(x_t)\right)$ is an equilibrium if $d_j^{(i)}\left(x_t, \epsilon_t^{(i)}\right)$ solves the individual optimization problem (1) for each $\left(i, x_t, \epsilon_t^{(i)}\right)$ when:

$$u_j^{(i)}(x_t) = \sum_{d_t^{(-i)}} P\left(d_t^{(-i)} \,|x_t\right) U_j^{(i)}\left(x_t, d_t^{(-i)}\right) \tag{2}$$

and:

$$f_j^{(i)}\left(x_{t+1} \,\middle|\, x_t^{(i)}\right) = \sum_{d_t^{(-i)}} P\left(d_t^{(-i)} \,\middle|\, x_t^{(i)}\right) F_j\left(x_{t+1} \,\middle|\, x_t, d_t^{(-i)}\right) \tag{3}$$

- To analyze dynamic games taking this form:
  1. interpret $u_j^{(i)}(x_t)$ with (2) and $f_j^{(i)}\left(x_{t+1} \,\middle|\, x_t^{(i)}\right)$ with (3)
  2. in estimation treat the *best reply function* as the solution to a dynamic discrete choice optimization problem within the equilibrium played out by the *data generating process* DGP.

Inversion

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

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

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

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

where $G_{jt}(\epsilon\,|x) \equiv \partial G_t(\epsilon\,|x) / \partial \epsilon_j$.

Inversion

CCPs are invertible in conditional valuation functions (Hotz and Miller, 1993)

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

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

- $Q_{jt}(\delta, x)$ is the probability choosing j in a static random utility model (RUM) with payoff $\delta_j + \epsilon_j$ and disturbance distribution $G_t(\epsilon \,|x)$.
- $Q_t(\delta, x) \equiv (Q_{1t}(\delta, x), \ldots Q_{J-1,t}(\delta, x))'$ is invertible in δ.
- This *inversion theorem* implies:

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

- Define the *conditional value function correction* as:

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

- 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 $\psi_{jt}(x)$.

- 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:

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

- Then:

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

Representation

Telescoping the conditional value function one period forward

- From its definition:

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

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

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

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

Representation

A representation of conditional value functions dispensing with maximization

- From Arcidiacono and Miller (2011, 2019):

$$\begin{aligned} v_{jt}(x_t) &= u_{jt}(x_t) \\ &+ \sum_{\tau=t+1}^{T}\sum_{k=1}^{J}\sum_{x=1}^{X}\beta^{\tau-t}\left\{ \begin{array}{c} (u_{k\tau}(x)+\psi_k[p_\tau(x)]) \\ \times\omega_{k\tau}(x,j)\kappa_{\tau-1}(x|x_t,j) \end{array} \right\} \end{aligned} \tag{4}$$

where the weights $\omega_{k\tau}(x_\tau, j)$ satisfy:

$$-\infty < \omega_{k\tau}(x_\tau, j) < \infty \text{ and } \sum_{k=1}^{J}\omega_{k\tau}(x_\tau, j) = 1$$

while the $\tau - \textit{period}$ state transitions $\kappa_\tau(x_{\tau+1}|x_t, j)$ are defined as:

$$\kappa_\tau(x_{\tau+1}|x_t,j) \equiv \left\{ \begin{array}{l} \kappa_t(x_{t+1}|x_t,j) \equiv f_{jt}(x_{t+1}|x_t) \\ \sum_{x_\tau=1}^{X}\sum_{k=1}^{J}\omega_{k\tau}\,(x_\tau,j)\, f_{k\tau}(x_{\tau+1}|x_\tau)\kappa_{\tau-1}(x_\tau|x_t,j) \end{array} \right.$$

- The optimization model is fully characterized by (T, β, f, g, u).
- The data comprise observations for a real or synthetic panel on the observed part of the state variable, x_t, and decision outcomes, d_t.
- From the definition of the optimal decision rule, and then appealing to the inversion theorem:

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

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

Identification

Identifying the conditional value function correction

- From their respective definitions:

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

- 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

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

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

- Assume (T, β, g) is known, and note f is identified (by inspection).
- We seek to identify u off the data generating process (x_t, d_t).
- The representation result for valuation functions implies:

$$
\begin{aligned}
u_{jt}(x) &= u_{1t}(x) + \psi_{1t}(x) - \psi_{jt}(x) \\
&+ \sum_{\tau=t+1}^{T} \sum_{x_\tau=1}^{X} \beta^{\tau-t} \left\{ \begin{array}{l} [u_{1\tau}(x_\tau) + \psi_{1t}(x_\tau)] \times \\ [\kappa_{\tau-1}(x_\tau|x,1) - \kappa_{\tau-1}(x_\tau|x,j)] \end{array} \right\}
\end{aligned} \tag{5}
$$

- If (T, β, f, g) is known, along a payoff, say the first, is also known for every state and time, then u is exactly point identified.

- In stationary models, let I denote the X dimensional identity matrix, and define:

$$u_j \equiv (u_j(1), \ldots, u_j(X))'$$

$$\Psi_j \equiv \left[\psi_j(1) \ldots \psi_j(X)\right]'$$

and:

$$F_j \equiv \begin{bmatrix} f_j(1|1) & \ldots & f_j(X|1) \\ \vdots & \ddots & \vdots \\ f_j(1|X) & \ldots & f_j(X|X) \end{bmatrix}$$

Then for all j:

$$u_j = \Psi_1 - \Psi_j - u_1 + \beta \left(F_1 - F_j\right) \left[I - \beta F_1\right]^{-1} \left(\Psi_1 + u_1\right)$$

Estimation

Unrestricted estimates from the identification equation

- Assume $u_{1t}(x) = 0$ and set:

$$\widehat{p}_{jt}(x) = \sum_{n=1}^{N} \mathbf{1}\{x_{nt} = x, d_{njt} = 1\} \Big/ \sum_{n=1}^{N} \mathbf{1}\{x_{nt} = x\}$$

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

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

to obtain $\widehat{\psi}_{jt}(x)$ and hence from (5):

$$\begin{aligned} \widehat{u}_{jt}(x_t) \equiv\ & \widehat{\psi}_{1t}(x_t) - \widehat{\psi}_{jt}(x_t) \\ & + \sum_{\tau=1}^{T-t} \sum_{x=1}^{X} \beta^{\tau-t} \widehat{\psi}_{1,t+\tau}(x) \left[\widehat{\kappa}_{t1,\tau-1}(x | x_t) - \widehat{\kappa}_{tj,\tau-1}(x | x_t) \right] \end{aligned} \quad (6)$$

- As above, there is an equivalent matrix form for the stationary case.

- In practice all applications further restrict the parameter space to increase precision (at the expense of potential bias).
- 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 can now define the model by (T, β, θ, g).
- Assume the DGP comes from (T, β, θ_0, g) where:

$$\theta_0 \equiv \left(\theta_0^{(1)}, \theta_0^{(2)}\right) \in \Theta^{(interior)}$$

- For example many applications assume:
  - β is known
  - $u_{jt}(x) \equiv x'\theta_j^{(1)}$ is linear in x and does not depend on t
  - $f_{jt}(x|x_{nt})$ is degenerate, x following a deterministic law of motion.

- One approach is to estimate:
  - $\theta^{(2)}$ with LIML off the transitions $f_{jt}(x|x_{nt}, \theta^{(2)})$
  - $\theta_0^{(1)}$ by minimizing the distance between the unrestricted estimates $\widehat{u}_{jt}(x_t)$ given in (6) and its parameterization $u_{jt}(x_t, \theta^{(1)})$:

$$\theta_{MD}^{(1)} = \underset{\theta^{(1)} \in \Theta^{(1)}}{\arg\min} \left[u(x, \theta^{(1)}) - \widehat{u}(x_t) \right]' W \left[u(x, \theta^{(1)}) - \widehat{u}(x_t) \right]$$

    where $u(x, \theta^{(1)})$ and $\widehat{u}(x_t)$ are stacked vectors of $u_{jt}(x_t, \theta^{(1)})$ and $\widehat{u}_{jt}(x_t)$, and W is a weight matrix (MD).
- Note:
  - $\theta_{MD}^{(1)}$ has a closed form if $u(x; \theta_0^{(1)})$ is linear in $\theta_0^{(1)}$.
  - the overidentifying restrictions can be tested.

- Alternatively to implement a QML estimator, first estimate $p_{jt}(x)$, $\theta_0^{(2)}$ and $\kappa_\tau(x|t, x_t, k, \theta_0^{(2)})$ and $\psi_{1t}(x)$ as above, and then:

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

where in T1EV applications:

$$\widehat{p}_{jt}(x, \theta^{(1)}, \widehat{h}) = \frac{\exp\left[u_{jt}(x, \theta^{(1)}) + \widehat{h}_{jt}(x)\right]}{\sum_{k=1}^{J} \exp\left[u_{kt}(x, \theta^{(1)}) + \widehat{h}_{kt}(x)\right]}$$

and $\widehat{h}_{kt}(x)$ is a numeric *dynamic correction factor* defined:

$$\widehat{h}_{jt}(x) \equiv \sum_{\tau=t+1}^{T} \sum_{x_\tau=1}^{X} \beta^{\tau-t} \widehat{\psi}_{1\tau}(x_\tau) \kappa_{\tau-1}(x_\tau | t, x, j, \theta_{LIML}^{(2)})$$

- Similarly, to form a MSM estimator first:
  1. Estimate $p_{jt}(x)$, $\theta_0^{(2)}$ and $\kappa_\tau(x|t, x_t, k, \theta_0^{(2)})$ and $\psi_{kt}(x)$ for all $k \in \{1, \ldots, K\}$ as above.
  2. Simulate a lifetime path from x_{nt_n} onwards for each j, using $\widehat{f}$ and $\widehat{p}$. This generates $\widehat{x}_{ns}$ and $\widehat{d}_{ns} \equiv \left(\widehat{d}_{n1s}, \ldots, \widehat{d}_{nJs}\right)$ for all $s \in \{t_n + 1, \ldots, T\}$.
  3. Obtain estimates of:

$$\widehat{E}\left[\epsilon_{jt} \middle| d_{jt}^o = 1, x_t\right] \equiv$$

$$p_{jt}^{-1}(x_t) \int_{\epsilon_t} \prod_{k=1}^{J} \mathbf{1}\left\{ \begin{array}{r} \widehat{\psi}_{jt}(x_t) - \widehat{\psi}_{kt}(x_t) \\ \leq \epsilon_{jt} - \epsilon_{kt} \end{array} \right\} \epsilon_{jt} dG(\epsilon_t)$$

  or simulate it from the selected population $\widehat{\epsilon}_{jt}$.

Simulated Moments Estimators

The last three steps for an MSM estimator

1. Stitch together a simulated lifetime utility outcome for each n from the j^{th} choice at t_n onwards: $\widehat{v}_{jt_n}\left(x_{nt_n};\theta^{(1)},\widehat{f},\widehat{p}\right)\equiv$

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

2. Form the $J-1$ dimensional vector $h_n\left(x_{nt_n};\theta^{(1)},\widehat{f},\widehat{p}\right)$ from:

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

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

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