Conditional Choice Probabilities
Second of two lectures for:
DSE 2021 Summer School in Bonn

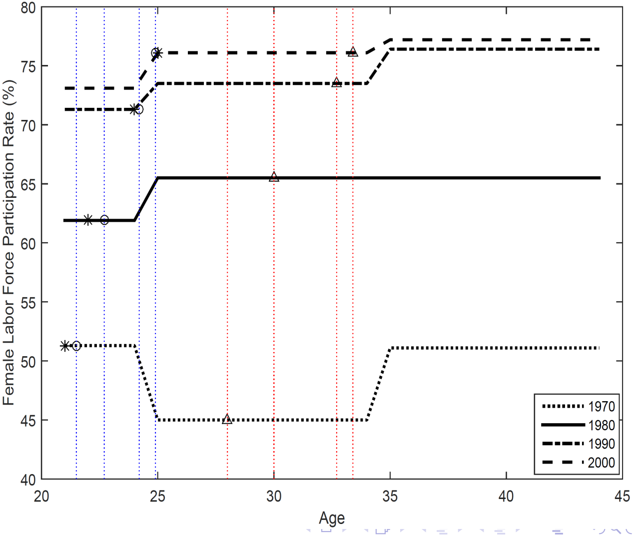
Robert A. Miller

Carnegie Mellon University

August 2021

Census tracks participation, marriage, first birth and home purchase by age

Labor force participation rate by age for 1970 - 2000. "Star" denotes median age at first marriage, "circle" denotes average age at first birth, "triangle" denotes average age at first homeownership. Age at first marriage is taken from the US Census Bureau, age at first birth is taken from the National Vital Statistical Reports (Mathews and Hamilton, 2002), age at first homeownership is computed from the PSID, whereas labor force participation rates are taken from publications of the US Bureau of Labor Statistics (Toossi, 2002, 2012).



Female Labor Supply, Fertility and Home Ownership

Figures 3 and 4 from Khorunzhina and Miller

Figure 3: Timing of children and first homeownership.

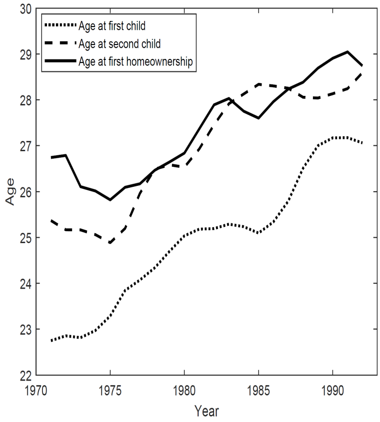
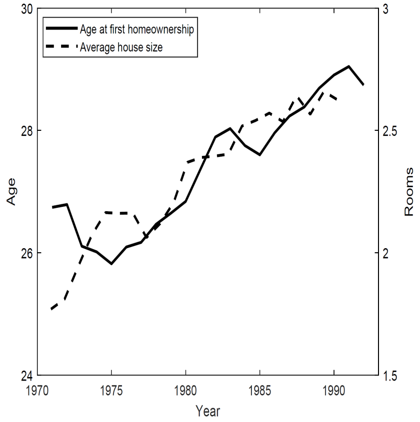
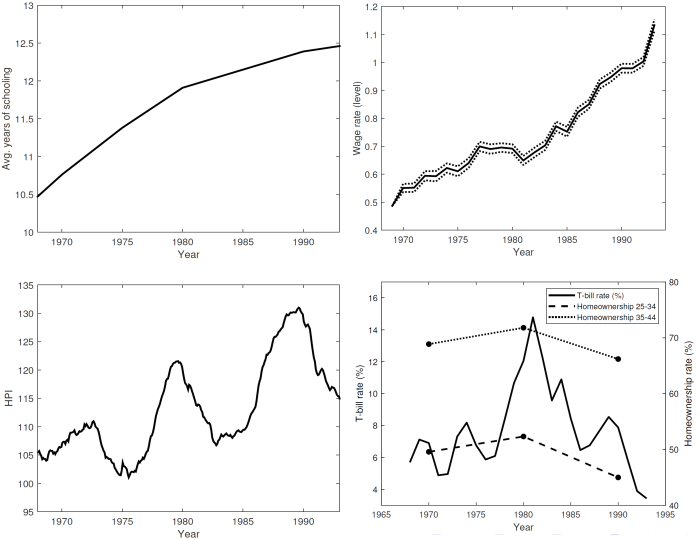


Figure 4: First homeownership and average home size adjusted by family size.



Contributing factors are schooling, wages, housing prices, interest rate

Educational attainment for female population 15 years old and over measured as the average years of total schooling, constructed based on data from Barro and Lee (2013). Wage rate is computed by the authors based on the PSID data sample. US National Home Price Index is based on Shiller (2015), whereas one-year Treasury constant maturity rate (GS1) is retrieved from FRED, Federal Reserve Bank of St. Louis. Decennial Censuses (Ruggles et al., 2019) are used for construction of 1970 – 1990 homeownership rates.



- We develop a nonstationary dynamic model of household choices for:
    1. fertility
    2. female labor supply
    3. first home purchase decisions.
- The secular (nonstationary) drivers in this model include:
    1. educational attainment
    2. wages
    3. interest rates
    4. housing prices
- We use the PSID to estimate the preference parameters of the model.
- We conduct counterfactual simulations on steady state economies with the estimated preferences to decompose what happens when each of the driving factors is changed.

- $b_t \in \{0,1\}$, where $b_t = 1$ if a child is born at time t.
- $f_t \in \{0,1\}$, where $f_t = 1$ means female works at time t.
- $a_t \in \{0,1\}$, where $a_t = 0$ means continuing to rent at t and $a_t = 1$ means first home is purchased.
- Consolidating the choices let $d_{kt} \in \{0,1\}$ where $\sum_{k=0}^{7} d_{kt} = 1$ where $(a_t, b_t, f_t) = (0,0,0)$ by $d_{0t} = 1$ and otherwise set $d_{kt} = 1$ for:

$$\begin{aligned} k \equiv \; & (1-a_t)\, b_t \,(1-f_t) + 2\,(1-a_t)\,(1-b_t)\, w_t + 3\,(1-a_t)\, b_t f_t \\ & + 4a_t\,(1-b_t)\,(1-f_t) + 5a_t b_t\,(1-f_t) + 6a_t\,(1-b_t)\, f_t + \dots \end{aligned}$$

- Purchasing the first house is a once-in-a-lifetime decision.
- If $a_t = 1$ then $a_s = 0$ for $s \in \{t+1, \dots, T\}$, and $\sum_{k=0}^{3} d_{kt} = 1$.
- Hence the model restricts homeowners to four discrete choices, tenants to eight.
- $c_t \in \mathcal{R}$ denotes nonhousing consumption, a continuous choice.

- The household's lifetime utility is modeled as:

$$-\sum_{\tau=t}^{\infty}\sum_{j=0}^{7}\beta^{\tau-t}d_{jt}\exp\left(u_{j\tau}^{h}+u_{j\tau}^{b}+u_{j\tau}^{l}-\rho c_{\tau}-\varepsilon_{j\tau}\right)$$

where:

  - β denotes the subjective discount factor.
  - $u_{j\tau}^{h}$ indexes current utility payoff from housing.
  - $u_{j\tau}^{b}$ indexes the expected lifetime utility of giving birth and raising a child.
  - $u_{j\tau}^{l}$ indexes the current utility of current leisure.
  - ρ is the constant absolute risk aversion parameter.
  - $\epsilon_{\tau} \equiv (\epsilon_{1\tau}, \ldots, \epsilon_{J\tau})$ is revealed at the beginning of the period t, has continuous support and is *IID* with density function $g(\epsilon_{t})$.

- We parameterize the index functions as:

$$\begin{aligned} u_{jt}^{h} &\equiv a_t \left(x_t' \theta_0 + b_t \theta_1 + f_t \theta_2 + b_t f_t \theta_3 \right) \\ &\quad + a_t s_t \left(x_t' \phi_0 + s_t \phi_1 + l_t \phi_2 \right) \\ u_{jt}^{b} &\equiv b_t \left(x_t' \gamma_0 + f_t \gamma_1 + s_t \gamma_2 \right) \\ u_{jt}^{l} &\equiv f_t x_t' \delta_0 + \delta_1 x_t' l_t + \delta_{20} l_t^2 + \delta_{21} l_t l_{t-1} \end{aligned}$$

where:

- x_t is a set of fixed or time varying attributes that characterize the decision maker (including age, education and marital status) along with previous fertility, housing and labor market outcomes.
- s_t measures house size in period t.
- $l_t \in [0, 1]$ is female labor supply in t where $l_t \in (0, 1]$ implies $f_t = 1$.
- s_{t-1} is included in u_{jt}^{h} to all for adjustments associated with resale and future housing purchases.
- $x_t' \delta_0$ is the fixed cost of working, and lagged labor supply affects the marginal utility of current leisure.

- Assume prices, interest rates and hence aggregate fluctuations are known in advance.
- Denote by:
  - W_t household financial wealth at the beginning of period t.
  - y_t income from real wages paid to the female for work in period t
  - $\widetilde{y}_t$ other income in period t
  - i_t the period t interest rate.
  - $R(s_t, q_t)$ rent by tenants.
  - $H(s_t, q_t)$ the house price, which depends on house size, quality and aggregate factors.
- The law of motion for disposable household wealth is:

$$
\begin{aligned}
(1 + i_t)^{-1} W_{t+1} \quad \leq \quad & W_t + y_t + \widetilde{y}_t - c_t \\
& - R(s_t, q_t) \prod_{\tau=1}^{t} (1 - a_\tau) - a_t H(s_t, q_t)
\end{aligned}
$$

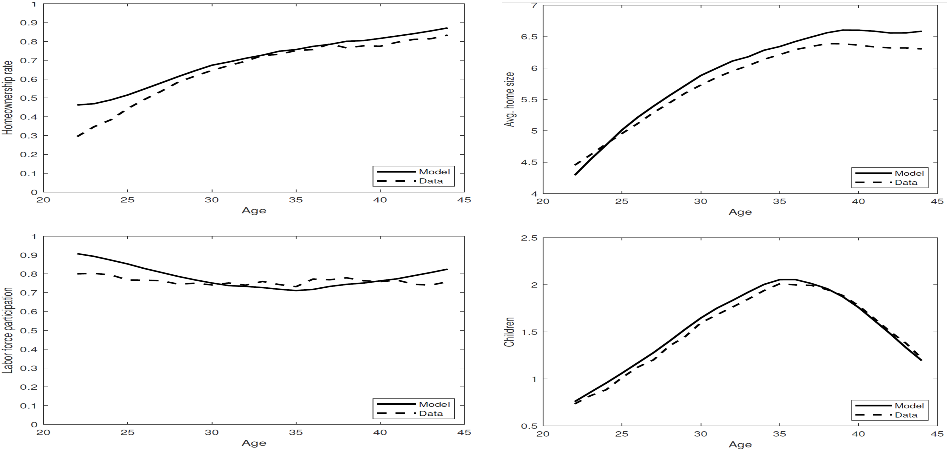
Female Labor Supply, Fertility and Home Ownership

State variables

- The household directly controls some state variables, including:
  - W_t current financial wealth.
  - $(b_t, \ldots, b_{t-18})$ family composition.
  - $l_{t\ 1} \in [0, 1]$ female lagged labor supply.
  - A_t home ownership.
- The other state variables include:
  - s_t house size, a Markov process with transition density $f(s_t | s_{t-1})$ when the household rents, and when household owns, $s_t = s_{t-1}$.
  - q_t aggregate variables for housing prices.
  - Y_t wage rates.
  - B_t current price of a bond in t paying one consumption unit each period into perpetuity.
  - D fixed demographics for each woman including age and education.
- Summarizing the state variables:

$$(D, B_t, q_t, Y_t, b_t, \ldots, b_{t-18}, A_t, s_t, l_{t-1}, W_t)$$

Within sample one period cross sectional forecasting of home ownership, house size, participation and children

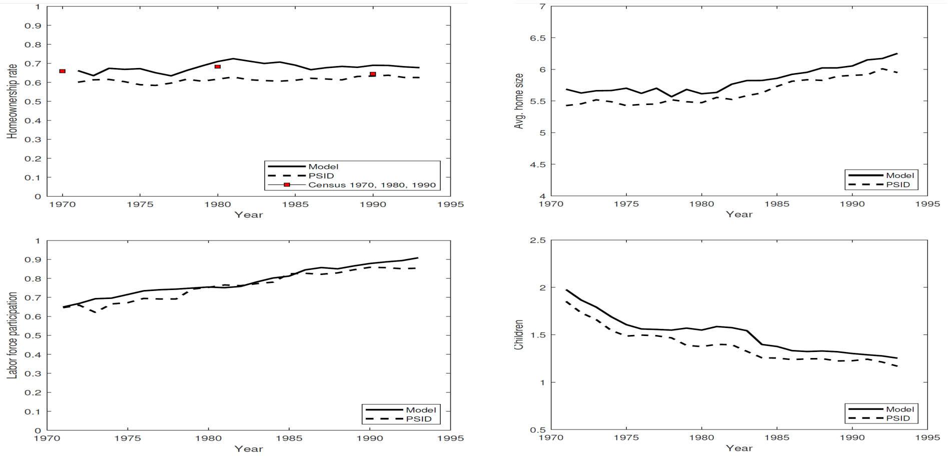


One-period in-sample model prediction versus data: life-cycle dimension.

Female Labor Supply, Fertility and Home Ownership

Figure 6 from Khorunzhina and Miller

Within sample one period time series forecasting home ownership, house size, participation and children



One-period in-sample model prediction versus data: time dimension

Female Labor Supply, Fertility and Home Ownership

Changing wages, education, home ownership prices and interest rates

- In a nonstationary we generally cannot infer from the data the time series processes that generate future paths.
- In principle we can, however, solve for hypothetical economies in the unknown nonstationary processes generating the data are replaced with processes we specify.
- This study compares two steady state economies (but also compute the transitions).
- The benchmark stationary economy somewhat resembles 1971:
  - we generate a stationary (overlapping generations) population from the 23 year olds approximating the 1971 population distribution.
  - the wage premium from education, housing prices, and the interest rate are fixed at their 1971 values.
  - the preference parameters are our estimates.

Unrestricted CCP Estimation

Data and outside knowledge

- We compare this benchmark with a stationary economy where we increase:
  1. *wages to the 1991 level (almost double their 1971 level):* $\Longrightarrow$first home purchase postponed, labor force participation increases, births fall.
  2. *education attainment by 1.5 years:* $\Longrightarrow$first home purchase brought forwards slightly, labor force participation increases, births fall.
  3. *housing prices by 15 percent:* $\Longrightarrow$first home purchase postponed, labor force participation slightly increases, births barely affected.
  4. *the interest rate from 4.88% (1971) to 5.87% (1991):* $\Longrightarrow$first home purchase brought forwards, reduces labor force participation falls, births increase.

Unrestricted CCP Estimation

Data and outside knowledge

- Stepping back from the two examples and the application, how do we apply CCP in estimation?
- Suppose the data generating process is from:
  - a stationary distribution and/or
  - complete histories for finite lived agents in a steady state economy.
- This is called a *long* panel (Arcidiacono and Miller, 2020).
- 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, we can identify:
  - the CCPs $p_t(x)$ and hence
  - the correction factors $\psi_{jt}(x) \equiv V_{jt}(x) - v_{jt}(x)$.

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

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

- Taking logs yields:

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

- 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 estimators:

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

and:

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

Unrestricted CCP Estimation

Estimating an intermediate probability distribution

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

$$\kappa_\tau(x_{\tau+1}|t, x_t, j) \equiv \begin{cases} f_{jt}(x_{t+1}|x_t) & \tau = t \\ \sum_{x=1}^{X} f_{1\tau}(x_{\tau+1}|x)\kappa_{\tau-1}(x|t, x_t, j) & \tau = t+1, \ldots \end{cases} \tag{2}$$

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

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

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

Unrestricted CCP Estimation

Utility parameter estimates

- The invertibility and representation theorems imply:

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

- 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 a matrix representation).

Unrestricted CCP Estimation

Asymptotic efficiency

- 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.
- Thus the unrestricted ML and CCP estimators are identical.
- Greater efficiency can only be obtained by making functional form assumptions on $u_{jt}(x_t)$ and $f_{jt}(x'|x)$.

Restricting the Parameter Space

Parameterizing the primitives

- In practice applications further restrict the parameter space.
- For example partition $\theta \equiv \left(\theta^{(1)}, \theta^{(2)}\right) \in \Theta$ and assume:
  - $u_{jt}(x) \equiv u_j(x, \theta^{(1)})$
  - $f_{jt}(x|x_{nt}) \equiv f_{jt}(x|x_{nt}, \theta^{(2)})$
  - Θ is a closed convex subspace of Euclidean space
- The model is now defined by (T, β, θ, f, g).
- Assume the DGP comes from $(T, \beta, \theta_0, f_0, g)$ where:

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

- For example many applications assume:
  - β and $f_{jt}(x|x_{nt})$ are known
  - $u_{jt}(x) \equiv x'\theta_j^{(1)}$ is linear in x and does not depend on t
  - the vector of unobserved variables $\epsilon \equiv (\epsilon_1 \dots \epsilon_J)$ is IID.

- Let $p_{tj}(x, \theta)$ denote the CCP when θ induces the DGP.
- The log likelihood for this model is then:

$$\sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} d_{njt} \left\{ \ln \left[p_{jt}(x_{nt}, \theta) \right] + \sum_{x=1}^{X} \mathbf{1} \left\{ x_{n,t+1} = x \right\} \ln \left[f_{jt}(x | x_{nt}, \theta^{(2)}) \right] \right\} \tag{3}$$

where $p_{jt}(x, \theta) \equiv$

$$\left[\begin{array}{l} \int_{-\infty}^{\epsilon_j + v_{tj}(x,\theta) - v_{t1}(x,\theta)} \cdots \int_{-\infty}^{\epsilon_j + v_{tj}(x,\theta) - v_{t,j-1}(x,\theta)} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\epsilon_j + v_{tj}(x,\theta) - v_{t,j+1}(x,\theta)} \cdots \int_{-\infty}^{\epsilon_j + v_{tj}(x,\theta) - v_{tJ}(x,\theta)} \end{array} \right] g_t(\epsilon | x) \, d\epsilon \tag{4}$$

and $v_{jt}(x, \theta)$ is the conditional valuation function for the j^{th} choice given θ.

- A Quasi Maximum Likelihood (QML) estimates:
  1. $\theta_0^{(2)}$ with $\theta_{LIML}^{(2)}$ from the data on $f_{jt}\left(x|x_t,\theta^{(2)}\right)$.
  2. $\kappa_\tau(x|t,x_t,k,\theta_0^{(2)})$ with $\kappa_\tau(x|t,x_t,k,\theta_{LIML}^{(2)})$ using $f_{jt}\left(x|x_t,\theta_{LIML}^{(2)}\right)$.
  3. $\psi_{1t}(x)$ with $\widehat{\psi}_{1t}(x)$ by substituting cell estimators $\widehat{p}_{jt}(x)$ for $p_{jt}(x)$.
  4. $v_{jt}\left(x,\theta^{(1)},\theta_0^{(2)}\right)$ with $\widehat{v}_{jt}\left(x,\theta^{(1)},\theta_{LIML}^{(2)}\right)$ for any given $\theta^{(1)}$, where "$\widehat{\ }$" indicates $\widehat{\psi}_{1t}(x)$ replaces $\psi_{1t}(x)$.
  5. $p_{jt}\left(x,\theta^{(1)},\theta_0^{(2)}\right)$ with $\widehat{p}_{jt}\left(x,\theta^{(1)},\theta_{LIML}^{(2)}\right)$ by substituting $\widehat{v}_{jt}\left(x,\theta^{(1)},\theta_{LIML}^{(2)}\right)$ for $v_{jt}\left(x,\theta^{(1)},\theta_0^{(2)}\right)$ in (4).
  6. $\theta_0^{(1)}$ with $\theta_{QML}^{(1)}$ by maximizing (3) after replacing $p_{jt}\left(x,\theta^{(1)},\theta^{(2)}\right)$ with $\widehat{p}_{jt}\left(x,\theta^{(1)},\theta_{LIML}^{(2)}\right)$.

Quasi Maximum Likelihood Estimation

Elaborating the first three steps in QML estimation

- Working through each step:
  1. As previously noted, this step is quite common whenever $f_{jt}\left(x|x_{nt},\theta^{(2)}\right)$ must be estimated, regardless of whether or not a CCP estimation is adopted:

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

  2. Here (2) is replaced with:

$$\begin{aligned} & \kappa_{\tau}(x_{\tau+1}|t,x_t,j,\theta^{(2)}_{LIML}) \\ \equiv & \begin{cases} f_{jt}(x_{t+1}|x_t,\theta^{(2)}_{LIML}) & \tau = t \\ \sum_{x=1}^{X} f_{1\tau}(x_{\tau+1}|x,\theta^{(2)}_{LIML})\kappa_{\tau-1}(x|t,x_t,j,\theta^{(2)}_{LIML}) & \tau = t+1,\ldots \end{cases} \end{aligned}$$

  3. For example if ϵ_t is T1EV, then $\widehat{\psi}_{1t}(x) = 0.57\ldots - \ln\left[p_{1t}(x)\right]$.

Quasi Maximum Likelihood Estimation

Elaborating the last three steps in QML estimation

- With respect to the last three steps:
  4. Appealing to the Representation theorem:

$$\widehat{v}_{jt}\left(x,\theta^{(1)},\theta^{(2)}_{LIML}\right)=u_{jt}(x,\theta^{(1)})+\widehat{h}_{jt}\left(x\right)$$

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

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

  5. In T1EV applications:

$$\widehat{p}_{jt}(x,\theta^{(1)},\theta^{(2)}_{LIML})=\frac{\exp\left[u_{jt}(x,\theta^{(1)})+\widehat{h}_{jt}\left(x\right)\right]}{\sum_{k=1}^{J}\exp\left[u_{kt}(x,\theta^{(1)})+\widehat{h}_{kt}\left(x\right)\right]}$$

  6. The name QML derives from the ML look-alike formula:

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

Minimum Distance Estimators (Altug and Miller, 1998)

Minimizing the difference between unrestricted and restricted current payoffs

- 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:

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

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

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

  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 $u_{jt}(x_t, \theta_0^{(1)}) = \Psi_{jt}\left(x_t, p, f_0\right)$ if p are the CCPs for $\left(T, \beta, \theta_0, g\right)$.
- Furthermore $u_{jt}(x)$ is exactly identified from $\Psi_{jt}\left(x, p, f_0\right)$ 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 $\theta_0^{(1)}$.
- One could form a simulated methods of moments estimator (Hotz, Miller, Sanders, Smith, 1994) based on this representation of the orthogonality conditions.
- Both the QML and MD estimators are less efficient than ML.

Unobserved Heterogeneity

A trade off in identification between functional form restrictions

- We have already shown that the model is exactly identified up to a normalization if the distribution of unobserved variables is known.
- Even in that case he model is underidentified for counterfactuals on transitions.
- Assumptions on preferences and transitions can help: for example nonstationary transitions and stable preferences (an exclusion restriction).
- What if we want to relax assumption that the distribution of unobserved variables is known?
- Then we must place assumptions on the way systematic payoffs are parameterized: note these are identifying assumptions.

Unobserved Heterogeneity

Rust's (1987) bus engine revisited

- To illustrate the approach which applies to quite generally to dynamic discrete choice optimization problems and dynamic games (Arcidiacono and Miller, 2011), let us revisit Mr. Zurcher's garage.
- Recall he 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 T1EV.
- Zurcher sequentially maximizes expected discounted sum of payoffs:

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

- Now suppose s is unobserved.

Unobserved Heterogeneity

ML estimation when CCPs are known (infeasible)

- To show how the EM algorithm helps, consider the infeasible case where $s \in \{1, \dots, S\}$ is unobserved but $p(x, s)$ is known.
- Let π_s denote population probability of being in unobserved state s.
- Supposing β is known the ML estimator for this "easier" problem is:

$$\{\hat{\theta}, \hat{\pi}\} = \arg\max_{\theta, \pi} \sum_{n=1}^{N} \ln \left[\sum_{s=1}^{S} \pi_s \prod_{t=1}^{T} l(d_{nt} | x_{nt}, s, p, \theta) \right]$$

where $p \equiv p(x, s)$ is a string of probabilities assigned/estimated for each (x, s) and $l(d_{nt} | x_{nt}, s_n, p, \theta)$ is derived from our representation of the conditional valuation functions and takes the form:

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

- Maximizing over the sum of a log of summed products is computationally burdensome.

Unobserved Heterogeneity

Why EM is attractive (when CCPs are known)

- The EM algorithm is a computationally attractive alternative to directly maximizing the likelihood.
- Denote by $d_n \equiv (d_{n1}, \ldots, d_{nT})$ and $x_n \equiv (x_{n1}, \ldots, x_{nT})$ the full sequence of choices and mileages observed in the data for bus n.
- At the m^{th} iteration:

$$
\begin{aligned}
q_{ns}^{(m+1)} &= \Pr\left\{ s \,\middle|\, d_n, x_{n,} \theta^{(m)}, \pi_s^{(m)}, p \right\} \\
&= \frac{\pi_s^{(m)} \prod_{t=1}^{T} l(d_{nt} | x_{nt}, s, p, \theta^{(m)})}{\sum_{s'=1}^{S} \pi_{s'}^{(m)} \prod_{t=1}^{T} l(d_{nt} | x_{nt}, s', p, \theta^{(m)})}
\end{aligned}
$$

$$
\pi_s^{(m+1)} = N^{-1} \sum_{n=1}^{N} q_{ns}^{(m+1)}
$$

$$
\theta^{(m+1)} = \arg\max_{\theta} \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{t=1}^{T} q_{ns}^{(m+1)} \ln[l(d_{nt} | x_{nt}, s, p, \theta)]
$$

Our algorithm begins by setting initial values for $\theta^{(1)}$, $\pi^{(1)}$, and $p^{(1)}(\cdot)$:

Step 1 Compute $q_{ns}^{(m+1)}$ as:

$$q_{ns}^{(m+1)} = \frac{\pi_s^{(m)} \prod_{t=1}^{T} l\left[d_{nt}|x_{nt}, s, p^{(m)}, \theta^{(m)}\right]}{\sum_{s'=1}^{S} \pi_s^{(m)} \prod_{t=1}^{T} l\left(d_{nt}|x_{nt}, s', p^{(m)}, \theta^{(m)}\right)}$$

Step 2 Compute $\pi_s^{(m+1)}$ according to:

$$\pi_s^{(m+1)} = \frac{\sum_{n=1}^{N} q_{ns}^{(m+1)}}{N}$$

Step 3 Update $p^{(m+1)}(x, s)$ using one of two rules below

Step 4 Obtain $\theta^{(m+1)}$ from:

$$\theta^{(m+1)} = \arg\max_{\theta} \sum_{n=1}^{N} \sum_{s=1}^{S} \sum_{t=1}^{T} q_{ns}^{(m+1)} \ln\left[l\left(d_{nt}|x_{nt}, s_n, p^{(m+1)}, \theta\right)\right.$$

Unobserved Heterogeneity

Updating the CCPs

- Take a weighted average of decisions to replace engine, conditional on x, where weights are the conditional probabilities of being in unobserved state s.

Step 3A Update CCPs with:

$$p^{(m+1)}(x,s) = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T} d_{1nt} q_{ns}^{(m+1)} I(x_{nt}=x)}{\sum_{n=1}^{N}\sum_{t=1}^{T} q_{ns}^{(m+1)} I(x_{nt}=x)}$$

- Alternatively use the identity from model that likelihood returns CCP of replacing the engine:

Step 3B Update CCPs with:

$$p^{(m+1)}(x_{nt},s_n) = l(d_{nt1}=1|x_{nt},s_n,p^{(m)},\theta^{(m)})$$

First Monte Carlo

Table 1 of Arcidiacono and Miller (2011)

		s Observed			s Unobserved		Time Effects	
	DGP	FIML	CCP	Ignoring s CCP	FIML	CCP	s Observed CCP	s Unobserved CCP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
θ_0 (intercept)	2	2.0100	1.9911	2.4330	2.0186	2.0280		
		(0.0405)	(0.0399)	(0.0363)	(0.1185)	(0.1374)		
θ_1 (mileage)	−0.15	−0.1488	−0.1441	−0.1339	−0.1504	−0.1484	−0.1440	−0.1514
		(0.0074)	(0.0098)	(0.0102)	(0.0091)	(0.0111)	(0.0121)	(0.0136)
θ_2 (unobs. state)	1	0.9945	0.9726		1.0073	0.9953	0.9683	1.0067
		(0.0611)	(0.0668)		(0.0919)	(0.0985)	(0.0636)	(0.1417)
β (discount factor)	0.9	0.9102	0.9099	0.9115	0.9004	0.8979	0.9172	0.8870
		(0.0411)	(0.0554)	(0.0591)	(0.0473)	(0.0585)	(0.0639)	(0.0752)
Time (minutes)		130.29	0.078	0.033	275.01	6.59	0.079	11.31
		(19.73)	(0.0041)	(0.0020)	(15.23)	(2.52)	(0.0047)	(5.71)

[a] Mean and standard deviations for 50 simulations. For columns 1–6, the observed data consist of 1000 buses for 20 periods. For columns 7 and 8, the intercept (θ0) is allowed to vary over time and the data consist of 2000 buses for 10 periods. See the text and the Supplemental Material for additional details.