Identification and CCP Estimation

Robert A. Miller

DSE at UCL

June 2025

- Suppose that each period $t \in \{1, \ldots, T\}$, the agent observes the realization (x_t, ϵ_t), and 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:

- An integer $T \leq \infty$ denotes the horizon of the optimization problem.
- the individual chooses amongst J mutually exclusive actions.
- $d_t \equiv (d_{1t}, \ldots, d_{Jt})$ 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).
- conditional independence holds, meaning:

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

- 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:

$$\begin{aligned} V_t(x_t) &\equiv E\left[V_t^*(x_t, \epsilon_t)\,|x_t\right] \\ &= E\left\{\sum_{\tau=t}^{T}\sum_{j=1}^{J}\beta^{\tau-t-1}d_{j\tau}^o\left(x_\tau, \epsilon_\tau\right)\left(u_{j\tau}(x_\tau)+\epsilon_{j\tau}\right)|x_t\right\} \end{aligned}$$

- 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 $\epsilon \equiv (\epsilon_1, \ldots, \epsilon_J)$ the CCPs are defined by:

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

Inversion

Differences in conditional valuation functions

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

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

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

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

  then clearly:

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

- 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 $\Delta v_{jt}(x)$:

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

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

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

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:

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

- 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 $G_t(\epsilon \left| x \right.)$.
- 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):

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

Inversion

Proposition 1 of Hotz and Miller (1993)

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

Interpreting the inversion expression

- $Q_{jt}^{-1}(p, x)$ has an intuitive interpretation:
  - Given x and $p(x)$ the agent is indifferent between the j^{th} and J^{th} choices for values of ϵ'_{jt} and ϵ'_{Jt} satisfying:

$$\begin{aligned} v_{jt}(x) + \epsilon'_{jt} &= v_{Jt}(x) + \epsilon'_{Jt} \\ \Rightarrow \Delta v_{jt}(x) &= \epsilon'_j - \epsilon'_J \\ &= Q_{jt}^{-1}[p_t(x), x] \end{aligned}$$

  - Thus the value of $Q_{jt}^{-1}[p_t(x), x]$ is the difference between the j^{th} and J^{th} taste shocks that would make the agent indifferent between those two choices.
  - More generally the value of the vector mapping:

$$Q_t^{-1}[p_t(x), x] = \left(Q_{1t}^{-1}[p_t(x), x], \ldots, Q_{J-1t}^{-1}[p_t(x), x]\right)$$

  corresponds to the value of a vector $\epsilon'_t \equiv (\epsilon'_{1t}, \ldots, \epsilon'_{Jt})$ that renders the agent indifferent to all the choices.

- The inversion theorem exploits conditional independence to finesse optimization and integration.
- More specifically 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 *generalizing T1EV*.
  4. generalize the renewal and terminal state properties often used in empirical work to *finite dependence*, by obtaining restrictions on the state variable transitions used to implement CCP estimators.
  5. introduce new methods for *incorporating unobserved state variables*.

- 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:

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

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

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

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

Conditional Value Function Correction

Identifying the conditional value function correction

- From their respective definitions:

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

- 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

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

- 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 Value Function Correction

Correction factor for extended nested logit

Lemma

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

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

- 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.

Conditional Valuation Function Representation

Telescoping one period forward

- From its definition:

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

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

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 $\tau = t+1, \ldots, T$:

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

Conditional Valuation Function Representation

Theorem 1 of Arcidiacono and Miller (2011)

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.

Identifying the Primitives

Identifying assumptions and data generating process

- The optimization model is fully characterized by the time horizon, the utility flows, the discount factor, the transition matrix of the observed state variables, and the distribution of the unobserved variables, summarized with the notation (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.
- Following most of the empirical work in this area we consider identification when (T, β, f, g) are assumed to be known.
- Thus the goal is to identify u from (x_t, d_t) when (T, β, f, g) is known.

Identifying the Primitives

Observational Equivalence

- It is widely believed that u is only identified relative to one choice per period for each state.
- Can we say more than that?
- For each (x, t) let $I(x, t) \in \{1, \ldots, J\}$ denote any arbitrarily defined normalizing action and $c_t(x) \in \mathbb{R}$ its associated benchmark flow utility, meaning $u^*_{I(x,t),t}(x) \equiv c_t(x)$.
- Assume $\{c_t(x)\}_{t=1}^{T}$ is bounded for each $x \in \{1, \ldots, X\}$.
- Let $\kappa^*_\tau(x_{\tau+1}|x_t, j)$ denote the probability distribution of $x_{\tau+1}$, given a state of x_t taking action j at t, and then repeatedly taking the normalized action from period $t+1$ through to period τ.
- Thus $\kappa^*_t(x_{t+1}|x_t, j) \equiv f_{jt}(x_{t+1}|x_t)$ and for $\tau \in \{t+1, \ldots, T\}$:

$$\kappa^*_\tau(x_{\tau+1}|x_t, j) \equiv \sum\nolimits_{x=1}^{X} f_{I(x,\tau),\tau}(x_{\tau+1}|x)\kappa^*_{\tau-1}(x|x_t, j) \qquad (2)$$

Theorem (Observational Equivalence, Arcidiacono and Miller, 2020)

For each $R \in \{1, 2, \ldots\}$, define for all $x \in \{1, \ldots, X\}$, $j \in \{1, \ldots, J\}$ and $t \in \{1, \ldots, R\}$:

$$
\begin{aligned}
u_{jR}^{*}(x) &\equiv u_{jR}(x) + c_R(x) - u_{I(x,R),R}(x) &\quad (3)\\
u_{jt}^{*}(x) &\equiv u_{jt}(x) + c_t(x) - u_{I(x,t),t}(x) &\quad (4)\\
&+ \sum_{\tau=t+1}^{R} \sum_{x'=1}^{X} \beta^{\tau-t} \left\{ \begin{array}{l} \left[c_\tau(x') - u_{I(x,\tau),\tau}(x') \right] \times \\ \left[\kappa_{\tau-1}^{*}(x'|x_t, I(x,t)) - \kappa_{\tau-1}^{*}(x'|x_t, j) \right] \end{array} \right\}
\end{aligned}
$$

(T, β, f, g, u^{*}), is observationally equivalent to (T, β, f, g, u) in the limit of $R \rightarrow T$. Conversely suppose (T, β, f, g, u^{*}) is observationally equivalent to (T, β, f, g, u). For each date and state select any action $I(x,t) \in \{1, \ldots, J\}$ with payoff $u_{I(x,t),t}^{*}(x) \equiv c_t(x) \in \mathbb{R}$. Then (3) and (4) hold for all (t, x, j).

Identifying the Primitives

Identification off long panels (Arcidiacono and Miller,2020)

Theorem (Identification)

For all j, t, and x:

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

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

Identifying the Primitives

Asymptotic efficiency of the unresestricted estmator

- By the Law of Large Numbers the cell estimators $\widehat{f}_{jt}\left(x'\left|x\right.\right)$ and $\widehat{p}_{jt}\left(x\right)$ converge to their population analogues
- By the Central Limit Theorem both estimators converge at $\sqrt{N}$ and and have asymptotic normal distributions.
- Both $\widehat{f}_{jt}\left(x'\left|x\right.\right)$ and $\widehat{p}_{jt}\left(x\right)$ are ML estimators for $f_{jt}\left(x'\left|x\right.\right)$ 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.
- Greater efficiency can only be obtained by making functional form assumptions about $u_{jt}(x_t)$ and $f_{jt}\left(x'\left|x\right.\right)$.

Restricting the Parameter Space

Parameterizing the primitives

- 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:

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

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

- A Quasi Maximum Likelihood (QML) estimator can be obtained by estimating:
    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 $\widehat{\kappa}_\tau(x|t,x_t,k,\theta_{LIML}^{(2)})$ 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 $\theta^{(1)}$, given below.
    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 ML estimator.

- With respect to the last two 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 the numeric *dynamic correction factor* $\widehat{h}_{kt}\left(x\right)$ is defined:

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

  5. In T1EV applications:

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

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:

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

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

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

  where $\widetilde{W}$ is a square $(J-1)\,TX$ weighting matrix.
- Note $\theta_{MD}^{(1)}$ has a closed form if $u(x; \theta_0^{(1)})$ is linear in $\theta_0^{(1)}$.

Simulated Moments Estimators

A simulated moments estimator (Hotz, Miller, Sanders and Smith, 1994)

- 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 $\widehat{E}\left[\epsilon_{jt} \middle| d_{jt}^{o} = 1, x_t\right]$ from $\widehat{p}$.
  3. Stitch together a simulated lifetime utility outcome from the j^{th} choice at t_n onwards for n, to form $\widehat{v}_{jt}\left(x_{nt_n}, \theta^{(1)}, \widehat{f}, \widehat{p}\right)$.
  4. Form the $J-1$ dimensional vector $l_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)$ from:

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

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

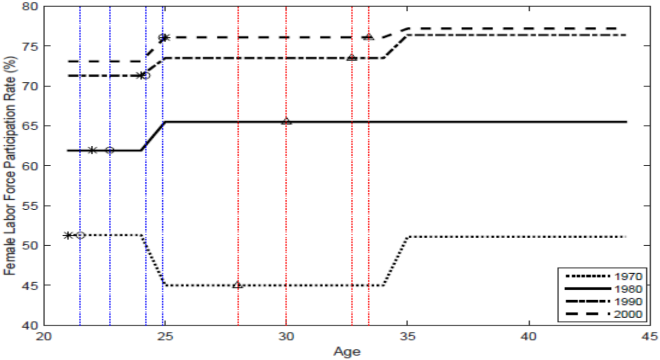
$$
N^{-1}\left[\sum_{n=1}^{N} z_n l_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)\right]' W_S \left[\sum_{n=1}^{N} z_n l_n\left(x_{nt_n}; \theta^{(1)}, \widehat{f}, \widehat{p}\right)\right]
$$

- The average age of a first-time home buyer was about 28 years old in the 1970s, about 30 in the 1990's, and is now about 32.5.
- This increase coincided with postponing marriage and fertility; the average age of mother at first birth rose from 22 forty years ago to 24 two decades ago, and is currently about 26.
- In contrast female labor-force participation rose from 48 percent in 1975, to 74 percent in 1995 and 76 percent in 2015, hours worked following a similar pattern.
- The median age of marriage and first birth practically coincide at each of the four census points (1970,...,2000):
  - but age at first home purchase is several years older
  - and the gap between first birth and first home purchase widened a little and then stabilized.

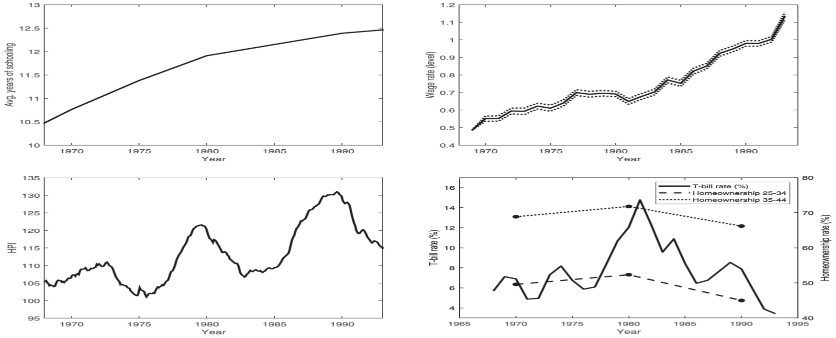
Home Ownership, Fertility and Labor Supply

Figure 1 from Khorunzhina and Miller (2021)

Figure 1: 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).



- Over these four decades:
    1. real wages rose.
    2. the real interest rate declined.
    3. housing prices rose and then fell.
    4. females became more educated.



A Dynamic Discrete Choice Model

Our approach

- We develop a nonstationary dynamic model of household choices for:
  1. fertility
  2. female labor supply
  3. first home purchase decisions.
- The nonstationary drivers in this model include:
  1. educational attainment
  2. female wages (conditional on education)
  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.

- Denote by:
  - $b_t \in \{0, 1\}$, where $b_t = 1$ if a child is born at time t.
  - $c_t \in \mathcal{R}$ denotes nonhousing consumption, a continuous choice.
  - $l_t \in \{0, 1\}$, where $l_t = 1$ means female works at time t.
  - $h_t \in \{0, 1\}$, where $h_t = 1$ means first home is purchased at t.
- If $h_t = 1$ then $h_\tau = 0$ for $\tau \in \{t + 1, \ldots, T\}$.
- Define homeownership by $h_t^* \equiv \sum_{\tau=1}^{t-1} h_\tau$. Then there are:
  - eight (b_t, l_t, h_t) discrete choices combinations if $h_t^* = 0$.
  - effectively four (b_t, l_t) combinations if $h_t^* = 1$.
- We label each possible choice permutation by $d_{jt} \in \{0, 1\}$ where:
  - $j \in \{0, \ldots, 7\}$ and if $h_t^* = 1$ then $j \in \{0, \ldots, 3\}$.
  - $\sum_{j=0}^{7} d_{j\tau} = 1$ and $\sum_{j=0}^{3} d_{j\tau} = 1$ if $h_t^* = 1$.

- We model household lifetime utility from t onwards as:

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

where j indexes the discrete choices at τ and:

- β denotes the subjective discount factor.
- u_{τ}^{h} indexes expected lifetime utility from purchasing first home.
- u_{τ}^{b} indexes net expected lifetime utility of raising a child.
- u_{τ}^{l} indexes the current utility of current leisure.
- ρ is the constant absolute risk aversion parameter.
- $\epsilon_{j\tau}$ is a period τ choice-specific disturbance with *iid* density $g\left(\epsilon_{j\tau}\right)$.

- We parameterize the index functions as:

$$
\begin{aligned}
u_t^h &\equiv \theta_0 + l_t\theta_1 + b_t\theta_2 + x_t l_t\theta_3 + l_{t-1}\theta_4 \\
&\quad + s_t\left(\theta_5 + x_t'\theta_6 + s_t^2\theta_7 + s_{t-1}\theta_8 + l_t^*\theta_9 + l_{t-1}^*\theta_{10}\right) \\
u_t^b &\equiv \gamma_0 + l_t\gamma_1 + x_t'\gamma_2 \\
&\quad + h_t^*\gamma_3 + (1-m_t)\,h_t^*\gamma_4 + s_t\gamma_5 \\
u_t^l &\equiv \delta_0 + x_t'\delta_1 + h_t^*\delta_2 + (1-m_t)\,h_t^*\delta_3 + s_t\delta_4 + l_{t-1}\delta_5 \\
&\quad + l_t^*\left[\delta_6 + x_t'\delta_7 + h_t^*\delta_8 + (1-m_t)\,h_t^*\delta_9 + l_t^*\delta_{10} + l_{t-1}^*\delta_{11}\right]
\end{aligned}
$$

where s_t measures house size in period t and:

- x_t are fixed or time varying attributes (including marital status and ages plus education of both spouses) along with previous fertility outcomes.
- $m_t \in \{0,1\}$ is marital status with $m_t = 1$ indicating married.
- $l_t^* \in [0,1]$ is female labor supply in t where $l_t^* \in (0,1]$ *iff* $l_t = 1$.

- Assume future spot prices and interest rates are known.
- 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.
- Defining gross flows before consumption as:

$$y_t^* \equiv y_t + \widetilde{y}_t - (1 - h_t^*) R(s_t, q_t) - h_t H(s_t, q_t)$$

the law of motion for disposable household wealth is:

$$(1 + i_t)^{-1} W_{t+1} \leq W_t + y_t^* - c_t$$

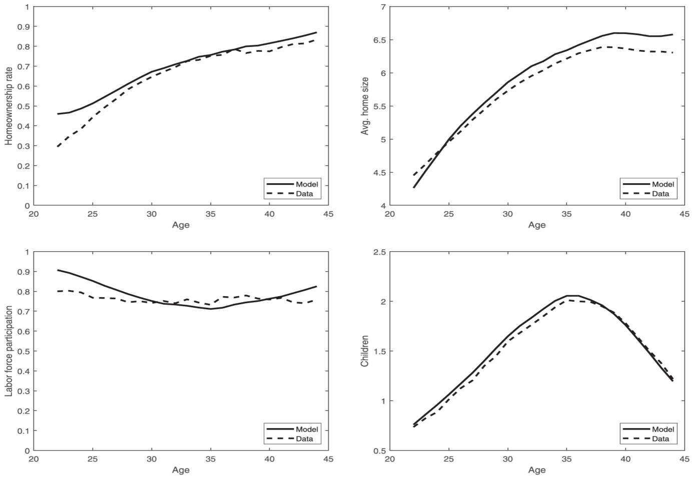
How Well does the Model fit the Data?

Panel data (PSID)

	Full sample	Owners	Renters
Age	32.4	33.9	29.7
Education	13.0	13.0	12.9
Married	0.82	0.92	0.64
Number of children	1.53	1.67	1.28
Home ownership rate	0.64		
House value for home owners		66,381	
Annual rent for renters			2,956
Move to owned house	0.087		
own-to-own**		0.062	
rent-to-own***		0.064	
Move to rental house	0.126		
rent-to-rent***			0.329
own-to-rent***			0.041
Number of rooms in dwelling	5.8	6.4	4.7
Labor force participation	0.753	0.736	0.783
Hours worked*	1497	1479	1527
Labor income*	11,070	11,504	10,341
Number of observations	43,504	27,871	15,633

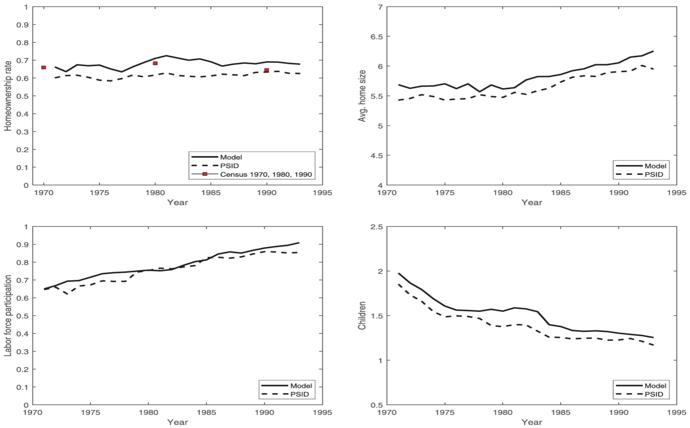
How Well does the Model fit the Data?

Over the life cycle: Figure 5 from Khorunzhina and Miller (2021)



How Well does the Model fit the Data?

Over the business cycle: Figure 6 from Khorunzhina and Miller (2021)



Counterfactuals

The problem of predicting the future

- If the time series process is stationary, the probability distribution characterizing the future is embodied in the past.
- Stationarity is not an attractive assumption when there is technological progress and secular demographic shifts.
- Except in special cases we cannot infer probability distributions characterizing the future from the data on nonstationary time series.
- We can nevertheless solve for hypothetical economies in which the unknown nonstationary processes generating the data are replaced with time series processes we specify.
- This study inoculates the counterfactual analysis against the unbalanced (PSID) sample and buffering from aggregate effects by comparing two steady state economies.
- Here we compare the long term differences of permanent shifts; we could also compute transitions from one steady state to the other.

Counterfactuals

Two steady state economies

- One benchmark stationary economy somewhat resembles the (PSID) economy in 1971. We:
  - generate an artificial population of 23 year olds that approximates the population distribution of that age group within the PSID.
  - set their preference parameters to our estimates.
  - fix the wage premium from education, housing prices, and the interest rate at the 1971 values.
  - successively apply the optimal rule for 25 years to attain a steady state economy (when supplemented by immigration).
- Coincidentally the aggregate statistics for this benchmark stationary economy are remarkably close to the corresponding analogues:
  - for the PSID in 1971,
  - and for the U.S. economy at large at that time.
- The other benchmark economy replaces the 1971 wage rate, housing prices, and interest rate with their 1991 values, but leaves the demographic composition unchanged.

Counterfactuals

Changing wages, education, home ownership prices and interest rates

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