Unobserved Heterogeneity Revisited

Robert A. Miller

Structural Econometrics

October 2017

Distributional Assumptions about the Unobserved Variables

A trade off

- We have already shown that the model is exactly identified up to a normalization if the distribution of unobserved variables is known.
- The model is underidentifed 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.

Motivating Example

Rust's (1987) bus engine revisited

- 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 $d_{1t} = 1$).
- Transitory iid choice-specific shocks, ϵ_{jt} are Type 1 Extreme value.
- Zurcher sequentially maximizes expected discounted sum of payoffs:

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

Motivating Example

ML Estimation when CCP's are known (infeasible)

- To show how the EM algorithm helps, consider the infeasible case where $s \in \{1, \ldots, S\}$ 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.

Motivating Example

Why EM is attractive (but also infeasible when CCP's 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:

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

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

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

Motivating Example

Steps in our algorithm when is s is unobserved and CCP's are unknown

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:

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

Motivating Example

Updating the CCP's

- 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 CCP's with:

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

- Or in a stationary infinite horizon model use identity from model that likelihood returns CCP of replacing the engine:

  Step 3B Update CCP's with:

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

First Monte Carlo

Finite horizon renewal problem

- Suppose $s \in \{0, 1\}$ equally weighted.
- There are two observed 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:

$$f(\Delta_t | x_2) = \exp\left[-x_2(\Delta_t - 25)\right] - \exp\left[-x_2(\Delta_t - 24.875)\right]$$

and x_2 is a multiple 0.01 drawn from a discrete equi-probability distribution between 0.25 and 1.25.

- Let θ_{0t} be an aggregate shock (denoting cost fluctuations say).
- 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 the observed state variables by $x_t \equiv (x_{1t}, x_2)$, this translates to:

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

First Monte Carlo

Table 1 of Arcidiacono and Miller (2011)

		s Observed			s Unobserved		Time Effects	
				Ignoring s			s Observed	s Unobserved
	DGP	FIML	CCP	CCP	FIML	CCP	CCP	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 (θ_0) 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.

The Estimators

Maximization

- We parameterize $u_{jt}(z_t)$ and $G(\epsilon_t)$ by θ, $f_{jt}(z_{t+1}|z_t)$ with α, and following our motivating example, we define two estimators.
- Given any conditional choice probability mapping $\widehat{p}$, both maximize the joint log likelihood:

$$(\widehat{\theta}, \widehat{\pi}, \widehat{\alpha}) = \underset{(\theta,\pi,\alpha,)}{\arg\max} \sum_{n=1}^{\mathcal{N}} \sum_{s=1}^{\mathcal{S}} \pi_s \log L\left(d_n, x_n \left| x_{n1}, s ; \theta, \widehat{p}\right.\right)$$

where $L\left(d_n, x_n \left| x_{n1}, s ; \theta, \widehat{p}\right.\right)$ is the likelihood of the (panel length) sequence (d_n, x_n):

$$L\left(d_n, x_n \left| x_{n1}, s ; \theta, \pi, p\right.\right) = \prod_{t=1}^{\mathcal{T}} \mathcal{L}_t\left(d_{nt}, x_{nt+1} \left| x_{nt}, s ; \theta, \pi, p\right.\right)$$

and $\mathcal{L}_t\left(d_{nt}, x_{nt+1} \left| x_{nt}, s_{nt} ; \theta, \pi, p\right.\right)$ is:

$$\sum_{j=1}^{J} d_{jnt} l_{jt}(x_{nt}, s_{nt}, \theta, \pi, p) f_{jt}\left(x_{n,t+1} \left| x_{nt}, s_{nt}, \theta\right.\right)$$

The Estimators

Using the Likelihood to obtain the CCP's

- The difference between the estimators arises from how $\widehat{p}$ is defined.
- The first estimator is based on the fact that $l_{jt}(x_{nt}, s_n, \theta, \alpha, \pi, p)$ is the likelihood of observing individual n make choice j at time t given s_n.
- Accordingly define $\widehat{p}\,(x, s)$ to solve:

$$\widehat{p}_{jt}\,(x, s) = l_{jt}(x, s; \widehat{\theta}, \widehat{\alpha}, \widehat{\pi}, \widehat{p})$$

- The large sample properties are standard.

- Let $\widehat{L}_n(s_n = s)$ denote the joint likelihood of the data for n and being in unobserved state s evaluated at $\left(\widehat{\theta}, \widehat{\alpha}, \widehat{\pi}, \widehat{p}\right)$.

$$\widehat{L}_n(s_n = s) \equiv \widehat{\pi}_s L\left(d_n, x_n \left| x_{n1}, s ; \widehat{\theta}, \widehat{p}\right.\right)$$

- Also denote by $\widehat{L}_n$ the likelihood of observing (d_n, x_n) given parameter values $\left(\widehat{\pi}, \widehat{\theta}, \widehat{p}\right)$:

$$\begin{aligned} \widehat{L}_n &\equiv \sum_{s=1}^{\mathcal{S}} \widehat{\pi}_s L\left(d_n, x_n \left| x_{n1}, s ; \widehat{\theta}, \widehat{p}\right.\right) \\ &= \sum_{s=1}^{\mathcal{S}} \widehat{L}_n(s_n = s) \end{aligned}$$

- As an estimated sample approximation, $N^{-1} \sum_{n=1}^{N} \left[\widehat{L}_n(s_n = s) / \widehat{L}_n\right]$ is the fraction of the population in s.

The Estimators

Another CCP "fixed point"

- Similarly:
  1. $N^{-1} \sum_{n=1}^{N} \left[I(x_{nt} = x) \widehat{L}_n(s_n = s) / \widehat{L}_n \right]$ is the estimated fraction of the population in s with x at t.
  2. $N^{-1} \sum_{n=1}^{N} \left[d_{jnt} I(x_{nt} = x) \widehat{L}_n(s_n = s) / \widehat{L}_n \right]$ is the estimated fraction choosing j at t as well.
- We define:

$$\widehat{p}_{jt}(x, s) = \left[\sum_{n=1}^{\mathcal{N}} d_{jnt} I(x_{nt} = x) \frac{\widehat{L}_n(s_n = s)}{\widehat{L}_n} \right] \Bigg/ \left[\sum_{n=1}^{\mathcal{N}} I(x_{nt} = x) \frac{\widehat{L}_n(s_n = s)}{\widehat{L}_n} \right]$$

- Compared to the first one this estimator has similar properties but imposes less structure.

Unobserved Markov Chain

Extending the estimation framework

- Up until now we have been assuming that the unobserved component is time invariant.
- Now suppose s_t is a Markov chain where $\pi\left(s_{t+1} | s_t\right)$ is an exogenous probability transition where:

$$f_{jt}\left(x_{t+1}, s_{t+1} \left| x_t, s_t\right.\right) = \pi\left(s_{t+1} | s_t\right) f_{jt}\left(x_{t+1} \left| x_t, s_t\right.\right)$$

  and $\pi_1\left(s_1 \left| x_1\right.\right)$ is the probability of being in (unobserved) state s_1 conditional on (observed) state x_1 in the first period.
- The intuition for the simpler case follows through in this generalization.

- We obtain the CCP's, $p^{(m)}(x, s)$, at the m^{th} step using one of the two ways we described above.
- The EM algorithm is used in obtaining $\theta^{(m)}$ in the same way as before.
- So we only have to determine:
  1. $q_{nst}^{(m)}$, the probability of n being in unobserved state s at time t,
  2. $\pi_1^{(m)}(s_1 | x_1)$, the probability distribution over the initial unobserved states conditional on the initial observed states,
  3. $\pi^{(m)}(s' | s)$, the transition probabilities of the unobserved states.

Unobserved Markov Chain

The remaining pieces of the algorithm

1. $q_{nst}^{(m+1)}$ follows from Bayes rule:

$$q_{nst}^{(m+1)} = \frac{L_n^{(m)}(s_{nt} = s)}{L_n^{(m)}}$$

2. Averaging over $q_{ns1}^{(m+1)}$:

$$\pi_1^{(m+1)}(s\,|x) = \frac{\sum_{n=1}^{\mathcal{N}} q_{ns1}^{(m+1)} I(x_{n1} = x)}{\sum_{n=1}^{\mathcal{N}} I(x_{n1} = x)}$$

3. Let $q_{ns't|s}$ denote the probability of n being in unobserved state s' at time t conditional on the data and also on being in unobserved state s at time $t-1$. We base $\pi^{(m+1)}(s'|s)$ on sample analogs of the identity:

$$\pi(s'|s) = \frac{E_n\left[q_{ns't|s}q_{nst-1}\right]}{E_n\left[q_{nst-1}\right]}$$

where E_n is the expectation taken over the population.

A Third Two Stage Estimator

An unrestricted first stage estimator

- Alternatively, subject to identification, we could estimate unrestricted CCPs in a first stage, and then plug them into the structural part of the econometric model.
  1. Estimate $f_{jt}(x_{n,t+1}|x_{nt}, s_n, \alpha)$ that is α and $p_{jt}(x, s)$ from an unrestricted likelihood formed from:
  $$\prod_{j=1}^{J}\left[p_{jt}\left(x_{nt}, s_n\right) f_{jt}(x_{n,t+1}|x_{nt}, s_n, \alpha)\right]^{d_{jnt}}$$
  2. Estimate θ using the conditional choice probabilities and the unobserved transitions on the unobserved variables obtained in the first step.
- Apart from not estimating θ, the essential difference in the first stage of this alternative estimator and the previous one, is that the likelihood components of this one come from $p_{jt}(x, s)$ not $L_j\left(d_t \middle| x, s; \widehat{\theta}, \widehat{\alpha}, \widehat{\pi}, \widehat{p}\right)$.

Second Monte Carlo

Structure

- Entrants pay startup cost to compete in the market, but not incumbents.
- Paying startup cost now transforms entrant into incumbent next period.
- Declining to compete in any given period is tantamount to exit.
- When a firm exits another firm potentially enters next period.

Second Monte Carlo

Dynamics

- There are two sources of dynamics in this model.
- An entrant depreciates startup cost over its anticipated lifetime.
- Since it is more costly for an entrant to start operations, than for an incumbent to continue, the number of incumbents signals how much competition the firm faces in the current period, and consequently affects its own decision whether to exit the industry or not.

- Each market has a permanent market characteristic, denoted by x_1, common to each player within the market and constant over time, but differing independently across markets, with equal probabilities on support $\{1, \ldots, 10\}$.
- The number of firm exits in the previous period is also common knowledge to the market, and this variable is indicated by:

$$x_{2t} \equiv \sum_{h=1}^{I} d_{1,t-1}^{(h)}$$

- This variable is a useful predictor for the number of firms that will compete in the current period.
- Intuitively, the more players paying entry costs, the lower the expected number of competitors.

- The unobserved state variable $s_t \in \{1, \ldots, 5\}$ follows a first order Markov chain.
- We assume that the probability of the unobserved variable remaining unchanged in successive periods is fixed at some $\pi \in (0, 1)$, and that if the state does change, any other state is equally likely to occur with probability $(1 - \pi)/4$.
- We generated also price data on each market, denoted by w_t, with the equation:

$$w_t = \alpha_0 + \alpha_1 x + \alpha_2 s_t + \alpha_3 \sum_{h=1}^{I} d_{1t}^{(h)} + \eta_t$$

where η_t is distributed as a standard normal disturbance independently across markets and periods, revealed to each market after the entry and exit decisions are made.

Second Monte Carlo

Utility and number of firms and markets

- The flow payoff of an active firm i in period t, net of private information $\epsilon_{2t}^{(i)}$ is modeled as:

$$U_2\left(x_t^{(i)}, s_t^{(i)}, d_t^{(-i)}\right) = \theta_0 + \theta_1 x + \theta_2 s_t + \theta_3 \sum_{h=1}^{I} d_{1t}^{(h)} + \theta_4 d_{1,t-1}^{(i)}$$

- We normalize exit utility as $U_1\left(x_t^{(i)}, s_t^{(i)}, d_t^{(-i)}\right) = 0$
- We assume $\epsilon_{jt}^{(i)}$ is distributed as Type 1 Extreme Value.
- The number of firms in each market in our experiment is 6.
- We simulated data for $3,000$ markets, and set $\beta = 0.9$.
- Starting at an initial date with 6 entrants in the market, we ran the simulations forward for twenty periods.

Second Monte Carlo

Table 2 of Arcidiacono and Miller (2011)

	DGP (1)	s_t Observed (2)	Ignore s_t (3)	CCP Model (4)	CCP Data (5)	Two-Stage (6)	No Prices (7)
Profit parameters							
θ_0 (intercept)	0	0.0207	−0.8627	0.0073	0.0126	−0.0251	−0.0086
		(0.0779)	(0.0511)	(0.0812)	(0.0997)	(0.1013)	(0.1083)
θ_1 (obs. state)	0.05	−0.0505	−0.0118	−0.0500	−0.0502	−0.0487	−0.0495
		(0.0028)	(0.0014)	(0.0029)	(0.0041)	(0.0039)	(0.0038)
θ_2 (unobs. state)	0.25	0.2529		0.2502	0.2503	0.2456	0.2477
		(0.0080)		(0.0123)	(0.0148)	(0.0148)	(0.0158)
θ_3 (no. of competitors)	−0.2	−0.2061	0.1081	−0.2019	−0.2029	−0.1926	−0.1971
		(0.0207)	(0.0115)	(0.0218)	(0.0278)	(0.0270)	(0.0294)
θ_4 (entry cost)	−1.5	−1.4992	−1.5715	−1.5014	−1.4992	−1.4995	−1.5007
		(0.0131)	(0.0133)	(0.0116)	(0.0133)	(0.0133)	(0.0139)
Price parameters							
α_0 (intercept)	7	6.9973	6.6571	6.9991	6.9952	6.9946	
		(0.0296)	(0.0281)	(0.0369)	(0.0333)	(0.0335)	
α_1 (obs. state)	−0.1	−0.0998	−0.0754	−0.0995	−0.0996	−0.0996	
		(0.0023)	(0.0025)	(0.0028)	(0.0028)	(0.0028)	
α_2 (unobs. state)	0.3	0.2996		0.2982	0.2993	0.2987	
		(0.0045)		(0.0119)	(0.0117)	(0.0116)	
α_3 (no. of competitors)	−0.4	−0.3995	−0.2211	−0.3994	−0.3989	−0.3984	
		(0.0061)	(0.0051)	(0.0087)	(0.0088)	(0.0089)	
π (persistence of unobs. state)	0.7			0.7002	0.7030	0.7032	0.7007
				(0.0122)	(0.0146)	(0.0146)	(0.0184)
Time (minutes)		0.1354	0.1078	21.54	27.30	15.37	16.92
		(0.0047)	(0.0010)	(1.5278)	(1.9160)	(0.8003)	(1.6467)

[a]Mean and standard deviations for 100 simulations. Observed data consist of 3000 markets for 10 periods with 6 firms in each market. In column 7, the CCP's are updated with the model. See the text and the Supplemental Material for additional details.