ECO 2901
EMPIRICAL INDUSTRIAL ORGANIZATION

Lecture 2: Market Entry: Incomplete Information

Victor Aguirregabiria (University of Toronto)

January 16, 2020

Lecture 2: Empirical entry models of incomplete information

Outline

1. Model
2. Identification
3. Estimation
4. Empirical application: Supermarkets' choice of pricing strategy
   Ellickson & Misra (2008)

1. Model

Entry models with incomplete information

- A market with N potential entrants. If firm i is active in the market ($a_{im} = 1$), its profit is:

$$\Pi_{im} = x_{im}\ \beta_i + \omega_{im} + \varepsilon_{im} + \sum_{j \neq i} \delta_{ij}\ a_{jm}$$

- $\mathbf{x}_m = (x_{1m}, x_{2m}, ..., x_{Nm})$ is **common knowledge** to firms and observable to the researcher.
- $\omega_m = (\omega_{1m}, \omega_{2m}, ..., \omega_{Nm})$ is is **common knowledge** to firms but unobservable to the researcher.
- ε_{im} is **private information** of firm i, independent across firms, independent of $(\mathbf{x}_m, \omega_m)$, and unobservable to the researcher. For concreteness, $\varepsilon_{im} \sim iid\ N(0, 1)$.

Bayesian Nash Equilibrium (BNE)

- The information of firm i is $(\mathbf{x}_m, \omega_m, \varepsilon_{im})$.
- A player's strategy depends on the variables in his information set.
- Let $\alpha_i(\mathbf{x}_m, \omega_m, \varepsilon_{im})$ be a strategy function for firm i such that $\alpha_i : X \times \Omega \times \mathbb{R} \rightarrow \{0, 1\}$.
- A **Bayesian Nash Equilibrium (BNE)** is an N-tuple of strategy functions $\{\alpha_i(.) : i = 1, 2, ..., N\}$ such that $\alpha_i(\mathbf{x}_m, \omega_m, \varepsilon_{im}) = 1$ iff:

$$x_{im}\ \beta_i + \omega_{im} + \varepsilon_{im} + \sum_{j \neq i} \delta_{ij} \left[\int \alpha_j(\mathbf{x}_m, \omega_m, \varepsilon_{jm})\ d\Phi_j(\varepsilon_{jm}) \right] \geq 0$$

Conditional choice probabilities (CCPs)

- It is convenient to represent players's strategies – and a BNE – in terms of **Choice Probabilities.**
- Given a strategy function $\alpha_i(\mathbf{x}_m, \omega_m, \varepsilon_{im})$, the associated choice probability is the result of integration this strategy function over the distribution of the player's private information

$$P_i(\mathbf{x}_m, \omega_m) \equiv \int \alpha_i(\mathbf{x}_m, \omega_m, \varepsilon_{im}) \; d\Phi_i(\varepsilon_{im})$$

- It represents the expected behavior of player i from the point of view of the other players who do not know the private information ε_{im}.

BNE in terms of CCPs

- Firm i's expected profit is:

$$\Pi^e_{im} = x_{im}\ \beta_i + \omega_{im} + \varepsilon_{im} + \sum_{j \neq i} \delta_{ij}\ P_j(\mathbf{x}_m, \omega_m)$$

- Firm i's best response is:

$$\{a_{im} = 1\} \Leftrightarrow \left\{ \varepsilon_{im} < x_{im}\ \beta_i + \omega_{im} + \sum_{j \neq i} \delta_{ij}\ P_j(\mathbf{x}_m, \omega_m) \right\}$$

- And firm i's **best response probability function** is:

$$P_i\left(\mathbf{x}_m, \omega_m\right) \quad = \quad \Phi\left(x_{im}\ \beta_i + \omega_m + \sum_{j \neq i} \delta_{ij}\ P_j(\mathbf{x}_m, \omega_m) \right)$$

BNE in terms of CCPs [2]

- Given $(\mathbf{x}_m, \omega_m)$, a **Bayesian Nash equilibrium (BNE)** is a vector of probabilities $\mathbf{P}(\mathbf{x}_m, \omega_m) \equiv \{P_i(\mathbf{x}_m, \omega_m) : i = 1, 2, ..., N\}$ that solves the fixed point problem:

$$P_i(\mathbf{x}_m, \omega_m) \quad = \quad \Phi\left(x_{im}\ \beta_i + \omega_m + \sum_{j \neq i} \delta_{ij}\ P_j(\mathbf{x}_m, \omega_m) \right)$$

- In a BNE, firms' beliefs about their opponents' entry probabilities are the opponents' best responses to their own beliefs.
- By Brower FP Theorem, the model has at least one BNE.
- The equilibrium may not be unique.

Comment

- The first applications of entry models with incomplete information assumed that the only unobservables for the researcher where the private information variables ε_{im}. That is, they assume that $\omega_m = 0$.
- This restriction simplifies very substantially the identification and estimation of this type of models.
- However, it is quite unrealistic and it can be easily rejected by the data. This restriction implies that:

$$\Pr(a_{1m}, a_{2m}, ..., a_{Nm} \mid x_m) = \prod_{i=1}^{N} \Pr(a_{im} \mid x_m)$$

- Ignoring ω_m can induce substantial biases in the estimation of the parameters δ that measure players' strategic interactions.

2. Identification

Identification: Assumptions

- Suppose that we have a random sample of markets and we observe:

$$\{\mathbf{x}_m, a_{im} : m = 1, 2, ..., M;\ i = 1, 2, ..., N\}$$

- **Assumption 1:** ω_m is independent of $\mathbf{x}_m$ and it has a finite mixture distribution: $\omega_m \in \{c_1,\ c_2,\ ...,\ c_L\}$ with $\Pr(\omega_m = c_k) \equiv \lambda_k$.
- **Assumption 2:** $\{P_i^0(\mathbf{x}_m, \omega_m)\}$ is such that two markets, m and m', with the same common knowledge variables $(\mathbf{x}_m, \omega_m)$ select the same type of equilibrium.
- Under these assumptions, and standard rank conditions, we can identify the model parameters θ.

Identification: Step 1

- The proof of identification proceeds in two steps.
- First, we show that the probabilities $P_i^0(\mathbf{x}, \omega)$ are nonparametrically identified.
- This is obvious in the model with $\omega_m = \omega$ – no common knowledge unobserved heterogeneity – because:

$$P_i^0(\mathbf{x}) = \mathbb{E}(a_{im} \mid \mathbf{x}_m = \mathbf{x})$$

- In the model with $\omega_m = \omega$, the nonparametric identification of $P_i^0(\mathbf{x}, \omega)$ is based on the identification of **nonparametric finite mixture model**.

Identification: Step 1 (Nonparametric finite mixture)

- With common knowledge unobs, ω_m, the estimation of choice probabilities is more complicated. But there are many recent results (Hall & Zhou, 2003, Kasahara & Shimotsu, 2009, 2013).
- The model is:

$$\Pr(\mathbf{a}_m = a \mid \mathbf{x}_m = x) = \sum_{k=1}^{L} \lambda_k \left[\prod_{i=1}^{N} P_i^0(x, c_k) \right]$$

- Different results show the NP identification of $\lambda_k' s$ and $P_i^0(x, c_k)$'s.
- The key identification assumption is the independence of players' a_{im} conditional on $(\mathbf{x}_m, \omega_m)$.

Identification: Step 2

- Given $P_i^0(\mathbf{x}_m, \omega)$ for every market m and type ω, we can represent our model as a linear regression-like model:

$$\Phi^{-1}\left(P_i^0(\mathbf{x}_m)\right) = x_{im}\ \beta_i + \sum_{j \neq i} \delta_{ij}\ P_j^0(\mathbf{x}_m)$$

- Define $Y_{im} \equiv \Phi^{-1}\left(P_i^0(\mathbf{x}_m)\right)$; $Z_{im} \equiv (x_{im}, P_j^0(\mathbf{x}_m) : j \neq i)$; and $\theta_i \equiv (\beta_i, \delta_{ij} : j \neq i)$. Then,

$$Y_{im} = Z_{im}\ \theta_i + e_{im}$$

- θ_i is identified iff $\mathbb{E}\left(Z_{im}' Z_{im}\right)$ has full column rank. For this, we need exclusion restrictions, i.e., player specific variables in x_{im}. [or functional form identification].

3. Estimation

Maximum likelihood estimation (1)

- Suppose that the **only unobservables for the researcher are the private information variables** ε_{im}.
- If the model had unique equilibrium, then we could estimate θ by MLE:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \sum_{m=1}^{M} \sum_{i=1}^{N} a_{im} \ \ln P_i(\mathbf{x}_m, \theta) + (1 - a_{im}) \ln(1 - P_i(\mathbf{x}_m, \theta))$$

where $P_i(\mathbf{x}_m, \theta)$ is the unique equilibrium probability of player i given $(\mathbf{x}_m, \theta)$.

- However, when the model has multiple equilibria, the (standard) likelihood is not a function but a correspondence.

Maximum likelihood estimation (2)

- We still can define the MLE in a model with multiple equilibria.
- For any (θ, P), define the **extended likelihood function** is:

$$\begin{aligned} Q(\theta, P) &= \sum_{m=1}^{M} \sum_{i=1}^{N} a_{im} \ \ln \Phi \left(x_{im} \ \beta_i + P_{-i}(\mathbf{x}_m) \ \delta_i \right) \\ &+ \ (1 - a_{im}) \ \ln \Phi \left(-x_{im} \ \beta_i - P_{-i}(\mathbf{x}_m) \ \delta_i \right) \end{aligned}$$

where $P_{-i}(\mathbf{x}_m) = \{P_j(\mathbf{x}_m) : j \neq i\}$ and $\delta_i = \{\delta_{ij} : j \neq i\}$.

- This is a well-defined **function** for any values of (θ, P).

Maximum likelihood estimation (3)

- The MLE is defined as:

$$(\widehat{\theta}_{MLE}, \widehat{P}_{MLE}) = \arg\max_{\theta,P} \begin{cases} \max\limits_{P} \; Q(\theta, P) \\ \text{subject to:} \\ P_i(x_m) = \Phi\left(x_{im}\ \beta_i + P_{-i}(\mathbf{x}_m)\ \delta_i\right) \ \text{ for ever} \end{cases}$$

- This estimator has all the good properties of MLE under standard regularity conditions.
- However, it can be very difficult to implement in practice.
- It requires optimization with respect to P which is a high dimensional vector. Many local maxima.
- Judd and Su (2012). MPEC method.

Two-step Pseudo ML estimation (1)

- Let $\mathbf{P}^0$ be the vector of choice probabilities (for each i and x_m) in the population.
- It is possible to show that the true θ^0 uniquely maximizes $Q_\infty(\theta, \mathbf{P}^0)$.
- The two-step Pseudo ML estimator of θ^0 is defined as the sample counterpart of θ^0.
- That is:

$$\hat{\theta} = \arg\max Q_M(\theta, \widehat{\mathbf{P}^0})$$

where $\widehat{\mathbf{P}^0}$ is a consistent nonparametric estimator of P^0.

Two-step Pseudo ML estimation (2)

- The first-step can be just a Nadaraya-Watson Kernel estimator of the choice probabilities: $\widehat{P}_i(\mathbf{x})$.
- The second step is just a standard Probit model with likelihood:

$$\sum_{m=1}^{M} \sum_{i=1}^{N} a_{im} \ln \Phi \left(x_{im}\beta_i + \sum_{j \neq i} \delta_{ij} \ \widehat{P}_j(x_m) \right)$$
$$+(1 - a_{im}) \ln \Phi \left(-x_{im}\beta_i - \sum_{j \neq i} \delta_{ij} \ \widehat{P}_j(x_m) \right)$$

- It can be generalized to deal with unobserved heterogeneity ω_m.

K-Step Estimator

- The first-step nonparametric estimator can have large variance and finite sample bias because the curse of dimensionality in NP estimation.
- This translates into the two-step estimator of θ that can have also large variance and finite sample bias.
- The K-step estimator is a solution to this problem.
- Let $\hat{\theta}_i^{(1)}$ be the two-step estimator.
- Given $\hat{\theta}_i^{(1)}$ and $\widehat{P^{(0)}}$, we can construct new choice probabilities, $\widehat{P^{(1)}}$, that now are parametric and exploit part of the structure of the model:

$$\widehat{P^{(1)}}(\mathbf{x}_m) \quad = \quad \Phi\left(x_i \; \hat{\beta}_i^{(1)} + \sum_{j \neq i} \hat{\delta}_{ij}^{(1)} \; \widehat{P^{(0)}}(\mathbf{x}_m)\right)$$

- Under some regularity conditions (Kasahara & Shimotsu, 2009), $\widehat{P^{(1)}}$ has smaller variance and finite sample bias than $\widehat{P^{(0)}}$.

K-Step Estimator [2]

- Given the new estimator $\widehat{P^{(1)}}$, we can obtain a new estimator of θ:

$$\widehat{\theta}^{(2)} = \arg\max_{\theta} Q_M(\theta, \widehat{P^{(1)}})$$

with $Q_M(\theta, \widehat{P^{(1)}}) = \sum_{m=1}^{M} \sum_{i=1}^{N} a_{im} \ln \Phi \left(x_{im}\beta_i + \sum_{j \neq i} \delta_{ij} \ \widehat{P}_j^{(1)}(x_m) \right) + (1 - a_{im}) \ln \Phi \left(-x_{im}\beta_i - \sum_{j \neq i} \delta_{ij} \ \widehat{P}_j^{(1)}(x_m) \right)$

- We can also apply this procedure recursively to define a $K - step$ estimator.

- Under some regularity conditions (Kasahara & Shimotsu, 2009), $\widehat{\theta^{(K)}}$ with $K > 1$ has smaller variance and finite sample bias than $\widehat{\theta^{(1)}}$.

4. Empirical Application: Supermarkets Pricing Strategies

Empirical Application

- Ellickson & Misra (Marketing Science, 2008): **Supermarkets competition in market strategies**.
- Supermarket firms position themselves as **either everyday low prices** (EDLP) or **temporary sales promotions** (PROMO).
- PROMO can be more attractive to customers with:
  - low shopping/travel costs who visit stores more frequently;
  - low storage costs, such that they can buy for inventory when prices are low.
- A firm's optimal price strategy depends in a market:
  - The composition of consumers in that market.
  - The pricing strategy of competitors.

Empirical Questions

- [1] Do supermarket chains tailor their pricing strategies to local market conditions?

- [2] Are chains heterogeneous in their propensity to choose a particular pricing strategy?

-

- [3] **How do firms react to the expected pricing strategies of their rivals?**

  * Is there strategic complementarity of substitutability in the choice of pricing strategy?

  * If choice of pricing as entry model: substitutability.
  * If choice of pricing as choice of price level: complementarity.

Summary of Empirical Results

- [1] Consumer demographics play a significant role in the choice of local pricing strategies

  * EDLP is favored in low income, racially diverse markets;
  * PROMO clearly targets the rich.

- [2] Larger stores and vertically integrated chains are significantly more likely to adopt EDLP.

- [3] They find evidence of **strategic complementarity** in pricing decisions.

This is kind of surprising. My interpretation is that it has to do with different average prices associated with the two pricing strategies (which is also consistent with result [1]).

Model

- N_m supermarket chains in market m.
- $a_{im} \in \{E, H, P\}$ is the pricing mode of i in market m: $E = EDLP$; $P = PROMO$; $H = HYBRID$.
- (Expected) Profit function of choosing $a \in \{E, P\}$ for chain i in market m:

$$\pi^e_{im}(a) = \mathbf{x}_m \; \beta_a + \delta_{aE} \; \rho^E_{im} + \delta_{aP} \; \rho^P_{im} + \omega^{(1)}_{am} + \omega^{(2)}_{ai} + \varepsilon_{ima}$$

where, for $a \in \{E, P\}$:

$$\rho^a_{im} = \frac{1}{N_m - 1} \sum_{j \neq i} \Pr(a_{jm} = a \mid \mathbf{x}_m, \boldsymbol{\omega}_m)$$

Data

- "Trade Dimensions" dataset.
- **Definition of local market:**
  - Statistical clustering method (K-means) based on latitude, longitude, and ZipCode information of the stores.
  - Result of market definition (store clusters): Larger than a typical ZipCode, but significantly smaller than the average county. More than 8,000 markets over the whole US.

Estimates

	EDLP			HYBRID		
	Estimate	Std. err	T-stat	Estimate	Std. err	T-stat
Effect						
Intercept	−1.5483	0.2426	−6.3821	2.1344	0.2192	9.7372
Strategy variables						
$\hat{\rho}^{\mathrm{EDLP}}_{-i_c^{l_m}}$	4.4279	0.1646	26.9010	−2.0924	0.1595	−13.1185
$\hat{\rho}^{\mathrm{PROMO}}_{-i_c^{l_m}}$	−3.7733	0.1501	−25.1386	−6.3518	0.1351	−47.0155
MSA characteristics						
Size ('000 sq. miles)	0.0394	0.0848	0.4645	−0.0566	0.0804	−0.7039
Density (pop 10,000 per sq. mile)	−0.0001	0.0002	−0.4587	0.0006	0.0002	2.9552
Avg. food expenditure ($ '000)	−0.0375	0.0155	−2.4225	−0.0013	0.0141	−0.0904
Market variables						
Median household size	0.5566	0.1989	2.7983	0.2150	0.0900	2.3889
Median HH income	−0.0067	0.0019	−3.5385	0.0056	0.0017	3.2309
Proportion Black	0.6833	0.1528	4.4719	0.0139	0.1443	0.0963
Proportion Hispanic	0.5666	0.2184	2.5943	−0.0754	0.2033	−0.3708
Median vehicles in HH	−0.1610	0.0840	−1.9167	0.2263	0.1173	1.9292
Store characteristics						
Store size (sqft '000)	0.0109	0.0015	7.2485	0.0123	0.0014	8.8512
Vertically integrated	0.1528	0.0614	2.4898	0.0239	0.0550	0.4343
Chain characteristics						
Number of stores in chain	−0.0002	0.0001	−2.7692	0.0002	0.0001	3.5000
Chain effect	1.7278	0.0998	17.3176	2.8169	0.0820	34.3531
Chain/MSA effect	0.7992	0.0363	22.0408	0.9968	0.0278	35.8046

Estimates [2]

- **Coefficients on consumer demographics**

  - Consumer demographics play a strong role in determining pricing strategy.

  - EDLP is preferred in markets with poorer households.

- **Firm and store characteristics**

  - Competition from a higher share of EDLP increases the probability of choosing EDLP.

  - Strategic effects are quantitatively important: explain about 20% of the variation in EDLP profits.

- If pricing strategy is a form of product differentiation, we should not expect strategic complementarity. Alternative explanations: a plausible one is that EDLP and PROMO involve different average price levels (EDLP lower average prices, and PROMO higher).