An empirical model of firm entry with endogenous product-type choices
Seim(2006), RAND Journal of Economics

Wooyong Lee

31 Jan 2013
Introduction

- Before: entry model, identical products
  - In this paper: entry with simultaneous location choice

- Before or some other papers: complete info game
  - In this paper: incomplete info game
  - (greatly reduces computational burden)

- Results: video retail industry location data
  - Firms differentiate to shield profit from competition
  - More scope for product differentiation → more market power (more entry)
Outline

1. Summary of the paper
   - Model
   - Estimation
   - Data and Result

2. Contribution and weakness of the paper
   - Contribution: Endogenized product-quality choice
   - Contribution: Incomplete information on profits
   - Weakness of the paper

3. Suggestions for further research
Section 1

Summary of the paper
Model Setup

- market $m = 1, 2, \ldots, M$

let’s focus on one market $m$:
(script $m$ removed for ease of exposition)

- location $l = 0, 1, \ldots, L$
- firm $f = 1, 2, \ldots, F$
- decision $d_f = (d_{fl})_{l=0}^{L} \in \{0, 1\}^{L+1}$
Model Setup

- payoff $\Pi_{fl} = X_l \beta + \xi + h(\Gamma_l, n) + \varepsilon_{fl}$

$X$ : observable market characteristics
$\xi$ : unobservable market characteristics, known to firms
$h()$ : effect of competition on profits; specified later
$\Gamma_l$ : interaction between $l$ and 1 through $L$
$n = (n_l)_{l=1}^L$ : number of firms at location $l$
$\varepsilon_{fl}$ : idiosyncratic profitability of firm $f$ at location $l$
$\Pi_{f0}$ normalized to 0
Assumptions

Assumption (Independent symmetric private values)
\[ \varepsilon_1, \ldots, \varepsilon_F \text{ are private information to the players and are independently and identically distributed draws from the distribution } G(\cdot) \]

Assumption (Additively separable competitors’ effects)
\[ h(\Gamma, l, n) = \sum_{k=1}^{L} \gamma_{kl} n_k \]

Assumption (Distance bands)
Let \( d_{kl} \) be the distance between location \( k \) and \( l \).
\[ \gamma_{kl} = \gamma_{k'l} = \gamma_b \text{ if } D_b \leq \gamma_{kl}, \gamma_{k'l} < D_{b+1}, \text{ where } D_b \text{ and } D_{b+1} \text{ denote cutoffs that define a distance band.} \]
Under distance bands $b = 0, 1, \ldots, B$, with corresponding cutoff points $D_b$ and $D_{b+1}$ where $D_0 = 0$, the resulting payoff function is given by

$$\Pi_{fl} = X_l \beta + \xi + \sum_b \gamma_b N_{bl} + \varepsilon_{fl}$$

$\gamma_b$ : competitive effect in distance band $b$

$N_{bl} = \sum_k I_{kl}^b n_k$ : number of firms in distance band $b$

$I_{kl}^b := \mathbb{I}(d_{kl} \in [D_b, D_{b+1}])$, $\mathbb{I}(\cdot)$ : indicator function

$n_k$ : number of firms at location $k$

$\mathcal{E} := \sum_b N_{bl}$ : total number of firms in the market
Conjectures and equilibrium

FIGURE 1
IMPACT ON PROFITS OF COMPETITORS’ LOCATIONS: ILLUSTRATION
Conjectures and equilibrium

Given the expected competitor distribution across market locations, firms calculate its expected profit at each location $l$:

$$ \mathbb{E}(\Pi_{fl}) = X_l \beta + \xi + \sum_b \gamma_b \mathbb{E}(N_{bl}) + \varepsilon_{fl} $$

$$ := \mathbb{E}(\bar{\Pi}_{fl}) + \varepsilon_{fl} $$

- note that $\mathbb{E}(N_{bl}) = \sum_k I^{b}_{kl} \mathbb{E}(n_k)$

Structure of the game:

1. firms decide whether or not to enter $\rightarrow \mathcal{E}$ determined
2. Given number of firms $\mathcal{E}$, firms decide locations

$\rightarrow$ solve backward
Due to the symmetry of rival’s types, firm $f$’s perception of firm $g$’s location strategy is the same for all competitors.

Let $\theta_1 = (\beta, \gamma)$. Given $\xi, X, \mathcal{E}, \theta_1$, the probability that competitor $g$ chooses location $l$ is given by

$$p_{gl}(\xi, X, \mathcal{E}, \theta_1) = \mathbb{P}(\mathbb{E}(\bar{\Pi}_{gl}) + \varepsilon_{gl} > \mathbb{E}(\bar{\Pi}_{gk}) + \varepsilon_{gk} \quad \forall k \neq l, \forall g \neq f)$$

Expected number of competitors at location $k$ is

- $\mathbb{E}(n_k) = (\mathcal{E} - 1)p_{gk}$
- $\mathbb{E}(N_{bl}) = \sum_k \mathbb{I}_{kl} \mathbb{E}(n_k) = \sum_k \mathbb{I}_{kl}(\mathcal{E} - 1)p_{gk} + \mathbb{I}(b = 0)$
Conjectures and equilibrium

**Assumption**

ε are i.i.d. draws from a type 1 extreme-value distribution with scale parameter normalized to one.

By this assumption, there exists a closed-form formula for $p_{gl}$:

$$p_{gl} = \frac{\exp\{\mathbb{E}(\bar{\Pi}_gl)\}}{\sum_{k=1}^{\mathcal{L}} \exp\{\mathbb{E}(\bar{\Pi}_{gk})\}}$$

Since types are symmetric, every firm has the same equilibrium conjecture of its competitors’ location choices:

- $p_g := (p_{gl})_{l=1}^{\mathcal{L}} = p_f = p^* := (p^*_l)_{l=1}^{\mathcal{L}}$
Conjectures and equilibrium

A firm’s vector of equilibrium conjectures over all locations \( l \) is defined by

\[
p_l^* = \frac{\exp\{\mathbb{E}(\bar{\Pi}_l(X, p^*, \mathcal{E}, \theta_1))\}}{\sum_{k=1}^{\mathcal{L}} \exp\{\mathbb{E}(\bar{\Pi}_k(X, p^*, \mathcal{E}, \theta_1))\}}
\]

\[
= \frac{\exp\{X_l \beta + \gamma_0 + (\mathcal{E} - 1) \sum_b \gamma_b \sum_j \Pi_{jl}^b p_j^*\}}{\sum_{k=1}^{\mathcal{L}} \exp\{X_k \beta + \gamma_0 + (\mathcal{E} - 1) \sum_b \gamma_b \sum_j \Pi_{jk}^b p_j^*\}}
\]

for all \( l = 1, \ldots, \mathcal{L} \)
Conjectures and equilibrium

let’s analyze $\mathcal{E}$:

- In equilibrium, each entrant earns nonnegative profits in expectation, while any additional entrant would suffer losses.
- The probability of entry by a firm into a market involves a comparison of a weighted average of payoffs across locations to the normalized payoff of not entering.
Conjectures and equilibrium

- Given the assumption of i.i.d. type 1 extreme-value profitability types,

\[ P(entry) = \frac{\exp(\xi) \sum_{i=1}^{L} \exp\{E(\bar{\Pi}_l(X, p^*, \mathcal{E}, \theta_1))\}}{1 + \exp(\xi) \sum_{i=1}^{L} \exp\{E(\bar{\Pi}_l(X, p^*, \mathcal{E}, \theta_1))\}} \]

- Since the probability of entry is identical across competitors,

\[ \mathcal{E} = \mathcal{F} \times P(entry) \]

- Ways to choose \( \mathcal{F} \): 50% enter the market, \( \mathcal{F} = 50 \)
Assumption

*The expected number of entrants predicted by the model exactly equals the number of entrants observed in the data.*

With $\mathcal{E}$ being the observed number of entrants, by the entry probability equation,

$$
\xi = \ln \mathcal{E} - \ln(\mathcal{F} - \mathcal{E}) - \ln \left( \sum_{l=1}^{\mathcal{L}} \exp\{\mathbb{E}(\prod_l (X, p^*, \mathcal{E}, \theta_1))\} \right)
$$

Assumption

$\xi^m$ are i.i.d. random draws from a Normal distribution with parameter $\theta_2 = (\mu, \sigma)$ and independent of $\varepsilon$. 
Data from market \( m = 1, \ldots, M \)

Each market is treated as an independent \( \mathcal{F}^m \)-player location game

\( d^m = (d^m_1, \ldots, d^m_{\mathcal{F}}) \) : vector of actions taken by \( \mathcal{F}^m \) players in market \( m \)

The likelihood function is given by

\[
L(\theta_1, \theta_2) = \prod_{m=1}^{M} p_{\theta_1}(d^m|\xi^m, X^m, \mathcal{E}^m)g_{\theta_2}(\xi^m|X^m, \mathcal{E}^m, \mathcal{F}^m)
\]
Estimation Procedure

1. Initialize \((\theta_1, \theta_2)\)

2. Given \((\theta_1, \theta_2), (X^m)_{m=1}^M, F, (E^m)_{m=1}^M\), solve equilibrium conjecture equation for \(p^m\) for each \(m\) separately.

3. Calculate \(\xi^m\) with \(p^m, (\theta_1, \theta_2), (X^m)_{m=1}^M, F, (E^m)_{m=1}^M\) for each \(m\).

4. Estimate \((\theta_1, \theta_2)\) by maximizing likelihood function.
Estimation Issues

- initial value: a grid search over the parameter space
- likelihood maximization: Nelder-Mead optimization algorithm
- problematic when there are multiple equilibria: appendix proves uniqueness for two distance bands with four locations, simulation evidence suggests further result
Data description

- video retail industry
- videotapes are standardized; main differentiation by store location
- isolated markets; no large city nearby, no metropolitan area
- exclude tourist regions
- locations by Census tracts
### TABLE 1  Descriptive Statistics: Markets and Locations

<table>
<thead>
<tr>
<th></th>
<th>151 Sample Markets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Market Level</td>
<td></td>
</tr>
<tr>
<td>Population, market</td>
<td>74,367</td>
</tr>
<tr>
<td>Population, main city</td>
<td>59,428</td>
</tr>
<tr>
<td>Population, all tracts in market</td>
<td>92,563</td>
</tr>
<tr>
<td>Largest incorporated place within 10 miles</td>
<td>2,618</td>
</tr>
<tr>
<td>Largest incorporated place within 20 miles</td>
<td>7,916</td>
</tr>
<tr>
<td>Tract Level</td>
<td></td>
</tr>
<tr>
<td>Number of tracts</td>
<td>21.13</td>
</tr>
<tr>
<td>Number of store locations</td>
<td>18.72</td>
</tr>
<tr>
<td>Tract population</td>
<td>4,380</td>
</tr>
<tr>
<td>Area (square miles)</td>
<td>10.10</td>
</tr>
<tr>
<td>Average distance (miles) to other locations in market</td>
<td>3.49</td>
</tr>
</tbody>
</table>

Note: The largest incorporated place within 10 and 20 miles is relative to the centroid of the market’s main city. The distance between locations within a market is computed as the distance between the tracts’ population-weighted centroids. Demographic data are as of 1999.
### TABLE 2  Tract-Level Demographic Characteristics

<table>
<thead>
<tr>
<th>Demographic Characteristics</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>4,417</td>
<td>247</td>
<td>20,163</td>
</tr>
<tr>
<td>Population, within .5 miles of tract</td>
<td>4,952</td>
<td>247</td>
<td>23,676</td>
</tr>
<tr>
<td>Population, .5–3 miles of tract</td>
<td>42,281</td>
<td>0</td>
<td>145,499</td>
</tr>
<tr>
<td>Population, 3–10 miles of tract</td>
<td>54,817</td>
<td>0</td>
<td>169,271</td>
</tr>
<tr>
<td>Per capita income, within .5 miles of tract</td>
<td>17,807</td>
<td>3,484</td>
<td>60,347</td>
</tr>
<tr>
<td>Per capita income, .5–3 miles of tract</td>
<td>17,413</td>
<td>0</td>
<td>38,934</td>
</tr>
<tr>
<td>Per capita income, 3–10 miles of tract</td>
<td>19,417</td>
<td>0</td>
<td>38,452</td>
</tr>
<tr>
<td>Business Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establishment density per square mile</td>
<td>177.86</td>
<td>.15</td>
<td>5,239.48</td>
</tr>
</tbody>
</table>

Note: The tract’s total population is placed at the population-weighted centroid. Population within different distance bands to the tract under consideration is computed as the sum of the population in tracts for which the distance to the considered tract’s centroid falls within the specified range. Demographic data are as of 1999.
$d^0 = 0.5$ miles, $d^1 = 3$ miles
### Data description

**TABLE 3** Store Location Patterns, Sample Markets

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firms, market</td>
<td>13.68</td>
<td>4.00</td>
<td>33.00</td>
</tr>
<tr>
<td>Store Clustering</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firms, tract</td>
<td>.73</td>
<td>.00</td>
<td>9.00</td>
</tr>
<tr>
<td>Firms, within .5 miles of tract</td>
<td>.80</td>
<td>.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Firms, within .5–3 miles of tract</td>
<td>6.12</td>
<td>.00</td>
<td>27.00</td>
</tr>
<tr>
<td>Firms, within 3–10 miles of tract</td>
<td>7.94</td>
<td>.00</td>
<td>33.00</td>
</tr>
<tr>
<td>Location Patterns within City’s Area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to city center (miles)(^a)</td>
<td>3.02</td>
<td>.02</td>
<td>14.96</td>
</tr>
</tbody>
</table>

Note: All stores are placed at the tract’s population-weighted centroid. Competitors within different distance bands to a firm’s location are computed as the number of firms in tracts for which the distance to the firm’s tract falls in the specified range.

\(^a\) The city center is taken to be the population-weighted centroid of the market’s main city.
### TABLE 4  Parameter Estimates, Entry and Location-Choice Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Potential Entrant Pool</th>
<th>2 × Total Entrants</th>
<th>50 Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (Standard Error)</td>
<td>Marginal Effect</td>
<td>Coefficient (Standard Error)</td>
</tr>
<tr>
<td>Population_{0} (000)</td>
<td>1.8191 (.1534)</td>
<td>.0333</td>
<td>2.1258 (.1764)</td>
</tr>
<tr>
<td>Population_{1} (000)</td>
<td>1.3109 (.1200)</td>
<td>.0236</td>
<td>1.7349 (.1498)</td>
</tr>
<tr>
<td>Population_{2} (000)</td>
<td>.6070 (.1192)</td>
<td>.0121</td>
<td>1.1348 (.1486)</td>
</tr>
<tr>
<td>Business Density</td>
<td>-.8077 (.1458)</td>
<td>-.0155</td>
<td>-.8889 (.1477)</td>
</tr>
<tr>
<td>Average Per Capita Income_{0} (0000)</td>
<td>.9309 (.1136)</td>
<td>.0180</td>
<td>1.0380 (.1233)</td>
</tr>
<tr>
<td>Average Per Capita Income_{1} (0000)</td>
<td>1.0081 (.2081)</td>
<td>.0193</td>
<td>.9188 (.2043)</td>
</tr>
<tr>
<td>Average Per Capita Income_{2} (0000)</td>
<td>.4851 (.2512)</td>
<td>.0092</td>
<td>.4884 (.2601)</td>
</tr>
</tbody>
</table>
### TABLE 4  Parameter Estimates, Entry and Location-Choice Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Potential Entrant Pool</th>
<th>2 × Total Entrants</th>
<th>50 Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (Standard Error)</td>
<td>Marginal Effect</td>
<td>Coefficient (Standard Error)</td>
</tr>
<tr>
<td>( \nu_0 )</td>
<td>-3.4520 (0.3111)</td>
<td>-3.3853 (0.3266)</td>
<td></td>
</tr>
<tr>
<td>( \nu_1 )</td>
<td>-1.0103 (0.0745)</td>
<td>-1.0087 (0.0923)</td>
<td></td>
</tr>
<tr>
<td>( \nu_2 )</td>
<td>-0.3448 (0.0738)</td>
<td>-0.4870 (0.0934)</td>
<td></td>
</tr>
<tr>
<td>( \sigma )</td>
<td>3.5829 (0.3110)</td>
<td>4.6760 (0.4316)</td>
<td></td>
</tr>
<tr>
<td>( \mu )</td>
<td>-2.8764 (1.3425)</td>
<td>-7.0364 (1.5801)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Results based on 1999 demographic and firm data. Subscript 0 denotes the immediately adjacent locations to the chosen tract, within .5 miles in distance; subscript 1 denotes tracts at .5 to 3 miles in distance from the chosen tract; and subscript 2 denotes tracts at more than 3 miles distance from the chosen tract. Tract-level business density is defined as the number of establishments (0000) per square mile. \( \gamma \) denotes competitive effects, and \( \sigma \) and \( \mu \) are the estimates of the parameters of the distribution of \( \xi \).
Simulation Result

A: Isolating the effect of competition on entry

- Predicted entrants, growing city with fixed layout
- Predicted entrants, sample markets

Jamestown, NY

Predicted number of firms per 10,000 people

Population
Simulation Result

B: Tradeoff between access to market population and competition

Predicted number of firms per 10,000 people

Jamestown, NY

Population
Section 2

Contribution and weakness of the paper
Previous Works

- focused on tradeoff between demand and intensity of competition

- Early works on product quality differentiation: Hotelling (1929), Lancaster (1966, 1979) (competition on characteristic space)
Model for Endogenized Product Choice

- Empirically tractable model of competition on characteristic space (tractable for large $\mathcal{L}$)
- Unisolated market: locations affect each other via distance
Empirical Results

- Firms differentiate to shield from competition
- More scope for product differentiation $\rightarrow$ more market power (more entry)
- More disperse population $\rightarrow$ less gain from differentiation (slightly more entry in total)
- The two offset each other in video retail industry data
Previous Works

• Complete information game → computational burden
  → Limited number of locations

• Mazzeo(2002) : entry with product-quality choice, complete information game
  → uses motel industry data, computation is burdensome with 3 locations
Incomplete Information Game

- Introduce private information: idiosyncratic profitability sources: cost, intangible assets, customer service, inventory maintenance etc.
- As Rust (1994) notes: Bayesian Nash equilibrium conjectures are easy to derive.
- Other works under information asymmetry:
Cost of Assumptions

Distance band approach :

- All locations with same distance band have the same impact
- two at the north ≻ one at the north and one at the south

ε being i.i.d. type 1 extreme value distribution :

- idiosyncratic profitability uncorrelated across firms at a location, across location for a firm
- no spatial correlation
When there are multiple equilibria:

- matching location probabilities to observed location choices is problematic
- uniqueness for general setting is only verified by simulation
Section 3

Suggestions for further research
Further research so far

- entry game with multistore/multiproduct firms (Jia 2008)
- dynamic entry game
- timing game (Sweeting 2009)
- estimation with multiple equilibria (Aguirregabiria 2012: identical product case)
- estimation in case of location heterogeneity (Aguirregabiria 2012)
Further research so far

- effect of other factors on entry decisions: regulation (Gowrisankaran and Krainer 2011), organizational structure (Takechi and Higashida 2012), managerial ability (Goldfarb and Xiao 2011)
Suggestions

- investigation of conditions for unique equilibrium
- nonlinear effect of additional entrant
- relaxing distance band assumption
- relaxing distance band assumption
- investigation of conditions for unique equilibrium

Summary

Model

Estimation

Result

Contribution and Weakness

Product

Information

Weakness