# *Empirical Games of Market Entry and Spatial Competition in Retail Industries*

**Victor Aguirregabiria (University of Toronto)**

*Lecture at Boston College*
*April 29, 2015*

# Introduction

- **Distance** from a store to customers, wholesalers, and competitors has substantial effects on demand and costs.

- Study **determinants of when and where to open retail stores** is necessary to inform public policy and business debates: e.g.
    - Value of a merger between retail chains;
    - Spatial pre-emption;
    - Cannibalization between stores of the same chain;
    - Magnitude of economies of density.

- Empirical work on **structural estimation of these models** has been relatively recent and has followed the seminal work by Bresnahan and Reiss (1990, 1991a).

# Outline

1. **Models**
   **(a)** Static Models with Single-Store Firms
   **(b)** Static Models with Multi-Store Firms
   **(c)** Dynamic Models

2. **Data**

3. **Specification and Estimation**

4. **Some Econometric Issues**
   **(a)** Multiple equilibria
   **(b)** Unobserved market heterogeneity

# 1(a). Static Models - Single-Store Firms

- $N$ **retail firms**, indexed by $i \in \{1, 2, \ldots, N\}$, are potential entrants in a market.

- The **geographic market** is a compact set $\mathbb{C}$ in the Euclidean space $\mathbb{R}^2$, and it contains $L$ locations where firms can operate stores. Locations are indexed by $\ell \in \{1, 2, \ldots, L\}$.

- Firms play a **two-stage game**.
  - Stage 1: entry and store location decisions.
  - Stage 2: Compete in prices (or quantities) taking entry decisions as given.

## Static Models - Single-Store Firms: Two-stage Game

- A firm's **entry decision**:

$$\boldsymbol{a}_i \equiv \{a_{i\ell} : \ell = 1, 2, \dots, L\}$$

where $a_{i\ell} \in \{0,1\} = 1\{\text{firm } i \text{ has a store in location } \ell\}$

- **Consumers**. A consumer is characterized by his preference for the products that firms sell and by his geographical location or home address $h \in \{1, 2, \dots, H\}$.

- **Aggregate consumer demand** comes from a discrete choice model of differentiated products where both product characteristics and transportation costs affect demand. For

## Static Models - Single-Store Firms: Demand

- **Spatial logit model**. Demand for firm *i* with store at location $\ell$:

$$q_{i\ell} = \sum_{h=1}^{H} \left[ \frac{a_{i\ell}\, exp\{x_i\, \beta \, - \, \alpha\, p_{i\ell} - \, \tau(d_{h\ell})\}}{\sum_{j=1}^{N} \sum_{\ell'=1}^{L} a_{j\ell'}\, exp\{x_j\, \beta \, - \, \alpha\, p_{j\ell'} - \, \tau(d_{h\ell'})\}} \right] M(h)$$

- $d_{h\ell}$ represents the geographic distance between the home address *h* and the business location $\ell$.

- $\tau(.)$ is an increasing real-valued function that represents consumer transportation costs.

## Static Models - Single-Store Firms: Price Competition

- **Nash-Bertrand competition**. A firm chooses its price $p_{i\ell}$ to maximize its variable profit:

$$(p_{i\ell} - c_{i\ell})\, q_{i\ell} \quad \text{where } c_{i\ell} \text{ is the unit cost of store } (i, \ell)$$

- Equilibrium prices and quantities: $p_i^*(\ell, \boldsymbol{a}_{-i}, \mathbf{x})$ and $q_i^*(\ell, \boldsymbol{a}_{-i}, \mathbf{x})$ given that firm $i$ has a store at location $\ell$ and other firms' entry/location decisions are $\boldsymbol{a}_{-i} \equiv \{\boldsymbol{a}_j : j \neq i\}$,

- **Equilibrium (indirect) variable profit**:

$$VP_i^*(\ell, \boldsymbol{a}_{-i}, \mathbf{x}) = [p_i^*(\ell, \boldsymbol{a}_{-i}, \mathbf{x}) - c_{i\ell}]\, q_i^*(\ell, \boldsymbol{a}_{-i}, \mathbf{x})$$

## Static Models-Single-Store: Entry Complete Information

- **Entry game.** Total profit of firm $i$ in location $\ell$ is:

$$\pi_i(\ell, \boldsymbol{a}_{-i}, \mathbf{x}) = VP_i^*(\ell, \boldsymbol{a}_{-i}, \mathbf{x}) - EC_{i\ell}$$

$EC_{i\ell}$ is the entry cost of firm $i$ at location $\ell$. Profit of an inactive firm: $\pi_i(0, \boldsymbol{a}_{-i}, \mathbf{x}) = 0$.

- **Nash equilibrium.** ***Complete information game***
  N-tuple $\{\boldsymbol{a}_i^*: i = 1, 2, \dots, N\}$ where every firm follows its best response:

$$a_{i\ell}^*(\mathbf{x}) = 1\left\{ \pi_i(\ell, \boldsymbol{a}_{-i}^*, \mathbf{x}) \geq \pi_i\left(\ell', \boldsymbol{a}_{-i}^*, \mathbf{x}\right) \ for\ any\ \ell' \neq \ell \right\}$$

## Static Models-Single-Store: Entry Incomplete Information

- **Nash equilibrium.** ***Incomplete information game.*** Suppose that the entry cost of firm $i$ is:

$$EC_{i\ell} = ec_{i\ell} + \varepsilon_{i\ell}$$

where $ec_{i\ell}$ is public information and $\varepsilon_{i\ell}$ is private information.

- $\boldsymbol{\varepsilon}_i \equiv \{\varepsilon_{i\ell}\colon\ \ell = 1, 2, \dots, L\}$ is i.i.d. across firms with a distribution function $F_i$ continuously differentiable over $\mathbb{R}^L$.

- A firm's strategy is an *L-dimension* mapping:

$$\boldsymbol{\alpha}_i(\boldsymbol{\varepsilon}_i; \mathbf{x}) \equiv \{\ \alpha_{i\ell}(\boldsymbol{\varepsilon}_i; \mathbf{x}) :\ \ell = 1, 2, \dots, L\ \}$$

where $\alpha_{i\ell}(\boldsymbol{\varepsilon}_i; \mathbf{x})$ is a binary-valued function from.

## Static Models-Single-Store: Entry Incomplete Information

- Firms maximize expected profits.

$$\pi_i^e(\ell, \boldsymbol{\alpha}_{-i}\,,, \mathbf{x}) \equiv E_{\boldsymbol{\varepsilon}_{-i}}[\,\pi_i(\ell, \boldsymbol{\alpha}_{-i}(\boldsymbol{\varepsilon}_{-i}), \mathbf{x}\,)\,]$$

where $E_{\boldsymbol{\varepsilon}_{-i}}$ represents the expectation over the distribution of the private information of firms other than $i$.

- **Bayesian Nash equilibrium (BNE)**. N-tuple of strategy functions $\{\alpha_i^*: i = 1, 2, \ldots, N\}$ such that every firm maximizes its *expected profit*: for any $\boldsymbol{\varepsilon}_i$,

$$\alpha_{i\ell}^*(\boldsymbol{\varepsilon}_i; \mathbf{x}) = \; 1\{\,\pi_i^e(\ell, \boldsymbol{\alpha}_{-i}^*, \mathbf{x}) \geq \pi_i^e\left(\ell', \boldsymbol{\alpha}_{-i}^*\,, \mathbf{x}\right) \;\; for\; any\; \ell' \neq \ell\;\}$$

# 1(b). Static Models – Multi-Store Firms

- **Retail chains** are prominent in many retail industries: supermarkets, department stores, apparel, electronics, fast food restaurants, etc.

- **Cannibalization** (business stealing between stores of the same chain) and **economies of scope** (some operating costs are shared between stores of the same chain) are important factors in the entry and location decisions of a multi-store firm.

- **Economies of density**: economies of scope that increase when store locations are geographically closer to each other. Recent empirical literature on retail chains has emphasized the importance of these economies of density, i.e., Holmes (2011), Jia (2008), Ellickson, Houghton, and Timmins (2013), and Nishida (2015).

## Static Models - Multi-Store Firms

- Entry decision: $\boldsymbol{a}_i \equiv \{a_{i\ell} : \ell = 1, 2, \dots, L\}$, where now the vector $\boldsymbol{a}_i$ can take any value within the choice set $\{0,1\}^L$.

- Demand system still can be described using the same demand equation.

- Variable profit:

$$\sum_{\ell=1}^{L} a_{i\ell} \, (p_{i\ell} - c_{i\ell}) q_{i\ell}$$

- Firms compete in prices taking their store locations as given.
  - Uniforms pricing
  - Spatial price discrimination

## Static Models - Multi-Store: Price Competition

- First order conditions for price $p_{i\ell}$:

$$q_{i\ell} + \ (p_{i\ell} - c_{i\ell}) \frac{\partial q_{i\ell}}{\partial p_{i\ell}} + \sum_{\ell' \neq \ell} (p_{i\ell'} - c_{i\ell'}) \frac{\partial q_{i\ell'}}{\partial p_{i\ell}} = 0$$

- The third term captures how the pricing decision of the firm internalizes the cannibalization effect among its own stores.

- **Nash-Bertrand equilibrium** is a solution to the previous system.

- **Equilibrium (indirect) variable profit** of firm $i$ is:

$$VP_i^*(\boldsymbol{a}_i, \boldsymbol{a}_{-i}; \mathbf{x}) = \sum_{\ell=1}^{L} a_{i\ell} \left[p_i^*(\ell, \boldsymbol{a}_{-i}, \mathbf{x}) - c_{i\ell}\right] q_i^*(\ell, \boldsymbol{a}_{-i}, \mathbf{x}),$$

## Static Models - Multi-Store: Entry / Complete Information

- Total profit of the retail chain:

$$\pi_i(\boldsymbol{a}_i, \boldsymbol{a}_{-i}; \mathbf{x}) = VP_i^*(\boldsymbol{a}_i, \boldsymbol{a}_{-i}; \mathbf{x}) - EC_i(\boldsymbol{a}_i)$$

- **Entry costs** depend on the number of stores (i.e., (dis)economies of scale) and on the distance between the stores (e.g., economies of density). I'll provide examples of specifications.

- **Nash equilibrium**. N-tuple $\{\boldsymbol{a}_i^*: i = 1, 2, \dots, N\}$ such that:

$$\pi_i(\boldsymbol{a}_i^*, \boldsymbol{a}_{-i}^*; \mathbf{x}) \geq \pi_i(\boldsymbol{a}_i, \boldsymbol{a}_{-i}^*; \mathbf{x}) \quad for\ any\ \boldsymbol{a}_i \neq \boldsymbol{a}_i^*$$

## Static Models-Multi-Store: Entry / Incomplete Information

- Private information variables in entry costs.

- A **Bayesian Nash equilibrium** is an N-tuple of strategy functions $\{\boldsymbol{\alpha}_i^*: i = 1, 2, \dots, N\}$ such that every firm maximizes its *expected profit*:

$$\pi_i^e(\boldsymbol{\alpha}_i^*(\boldsymbol{\varepsilon}_i; \mathbf{x}), \boldsymbol{\alpha}_{-i}^*; \mathbf{x}) \geq \pi_i^e(\boldsymbol{a}_i, \boldsymbol{\alpha}_{-i}^*; \mathbf{x}) \quad for\ any\ \boldsymbol{a}_i \neq \boldsymbol{\alpha}_i^*(\boldsymbol{\varepsilon}_i; \mathbf{x})$$

# 1(c). Dynamic Models

- Opening a store is a forward-looking decision with significant non-recoverable entry costs due to firm and location-specific investments.

- The sunk cost of setting up new stores is a potentially important force behind the configuration of the spatial market structure.

- Time is discrete and indexed by $t \in \{\ldots, 0, 1, 2, \ldots\}$.

- At the beginning of period $t$ a firm's network of stores is represented by the vector $\boldsymbol{a}_{it} \equiv \{a_{i\ell t} : \ell = 1, 2, \ldots, L\}$,

- Market structure at period $t$: $\boldsymbol{a}_t \equiv \{\boldsymbol{a}_{it} : i = 1, 2, \ldots, N\}$

## Dynamic Models: Ericson-Pakes Framework

- Every period $t$ the model has two stages, similar to the ones described in the static game above.

- **Bertrand game:** Given firms' store networks $\boldsymbol{a}_t$, retail chains compete in prices in exactly the same way as in the Bertrand model described above. [Dynamic Pricing ...]

- This determines the indirect variable profit: $VP_i^*(\boldsymbol{a}_t; \boldsymbol{z}_t)$, where $\boldsymbol{z}_t$ is a vector of exogenous state variables in demand and costs.

- **Dynamic game:** every firm decides its network of stores for next period, $\boldsymbol{a}_{it+1}$, and pays at period $t$ the entry and exit costs.

## Dynamic Models: Markov Perfect Equilibrium

- The period profit of a firm is:

$$\pi_i(\boldsymbol{a}_{it+1}, \boldsymbol{a}_t, \boldsymbol{z}_t) = VP_i^*(\boldsymbol{a}_t; \boldsymbol{z}_t) - FC_i(\boldsymbol{a}_{it}; \boldsymbol{z}_t) - AC_i(\boldsymbol{a}_{it+1}, \boldsymbol{a}_{it})$$

  - $FC_i$ is the fixed cost of operating the network.
  - $AC_i$ is the cost of adjusting the network from $\boldsymbol{a}_{it}$ to $\boldsymbol{a}_{it+1}$.

- A ***Markov Perfect Equilibrium*** is an N-tuple of strategy functions $\{\alpha_i^*(\boldsymbol{a}_t, \boldsymbol{z}_t): i = 1, 2, \dots, N\}$ such that:

$$\alpha_i^*(\boldsymbol{a}_t, \boldsymbol{z}_t) = argmax_{\{\boldsymbol{a}_{it+1}\}} \begin{bmatrix} \pi_i(\boldsymbol{a}_{it+1}, \boldsymbol{a}_t, \boldsymbol{z}_t) \\ \\ +\delta\, E\left(V_i(\boldsymbol{a}_{it+1}, \alpha_{-i}^*(\boldsymbol{a}_t, \boldsymbol{z}_t), \boldsymbol{z}_{t+1}; \alpha_{-i}^*)\right) \end{bmatrix}$$

# 2. Data

- Sample of **geographic markets** with information on firms' entry decisions and consumer socio-economic characteristics over one or several periods of time.

- Number of firms and time periods is typically small and statistical inference (i.e., the construction of sample moments and the application of LLN and CLT) is based on a **"large" number of markets**.

- Some relevant features:
  (a) selection of geographic markets;
  (b) presence or not of within market spatial differentiation;
  (c) information on prices, quantities, or sales at the store level;

## 2(a). Data: Selection of Geographic Markets

- Following the seminal work by Bresnahan and Reiss (1990), most of the applications in the literature have followed the so called **"Isolated Small Markets" approach**.

- For instance, in BR paper: 149 small U.S. towns; town belongs to a county with fewer than 10,000 people; there is no other town with a population of over 1000 people within 25 miles of the central town; and there is no large city within 125 miles.

- Main motivation for using this sample selection is in the assumptions of spatial competition in the Bresnahan-Reiss model (See later): no spatial differentiation within a market; and no interactions across markets.

## 2(a). Data: Selection of Geographic Markets

- If BR model were estimated using a sample of large cities, we would spuriously find very small competition effects simply because there is negligible or no competition at all between stores located far away of each other within the city.

- The model also assumes that there is no competition between stores located in different markets. This assumption is plausible only if the market under study is not geographically close to other markets.

- Some important limitations of this approach:
  - Extrapolation of estimation results to urban markets.
  - Many interesting retail industries are predominantly urban.

## 2(b). Data: Within market spatial differentiation

- Empirical models of entry in retail markets that take into account the spatial locations and differentiation of stores within a city market.

- Seminal work by Seim (2006).

- A city is partitioned into many small locations or blocks, e.g., census tracts, or a uniform grid of square blocks.

- In contrast to the "isolated small towns" approach, these locations are not isolated, and the model allows for competition effects between stores at different locations.

- Information on the number of stores, consumer demographics, and input prices at the block level.

## 2(c). Data: Prices, quantities, or sales at store level

- Most applications of models of entry in retail markets use data with market entry information but without prices and quantities due to the lack of such data, e.g., Bresnahan and Reiss (1990), Mazzeo (2002), Seim (2006), or Jia (2008), among many others.

- These studies either do not try to separately identify variable profits from fixed costs, or they do it by assuming that the variable profit is proportional to an observable measure of market size.

- Data on prices and quantities at store level can substantially help the identification of these models. In particular, it is possible to consider a richer specification of the model that distinguishes between demand, variable cost, and fixed cost parameters, and includes unobservable variables into each of these components of the model.

## 2(c). Data: Prices, quantities, or sales at store level

- When price & quantity (or sales) data are available, a **sequential estimation approach** is quite convenient.
  - **First step**, data on prices and quantities at the store level can be used to estimate a spatial demand system, e.g., Davis (2006) for movie theatres or Houde (2012) for gas stations.
  - **Second step**, variable costs can be estimated using firms' best response functions in Bertrand or Cournot model.
  - **Third step**, we estimate fixed cost parameters using the entry game and information of firms' entry and store location decisions.

## 2(c). Data: Prices, quantities, or sales at store level

- In some applications, price and quantity are not available, but there is information on revenue at the store level, e.g., Ellickson and Misra, (2012), Aguirregabiria, Clark, and Wang (2013), Suzuki (2013).

- This information can be used to estimate a variable profit function in a first step, and then in a second step the structure of fixed costs is estimated.

# 3. Specification and Estimation

- The games of entry in retail markets that have been estimated in empirical applications have imposed different types of restrictions on the framework that I have presented above. Restrictions on:
  - firm and market heterogeneity;
  - firms' information;
  - spatial competition;
  - multi-store firms;
  - dynamics;
  - functional form of the structural functions.

- The motivations for these restrictions are diverse: Identification or precise enough estimates; researcher's limited information; computational simplicity.

# 3. Specification and Estimation

- Following approx. the chronological evolution of the literature, I will describe the specification and estimation of the following models.

  **(a) Homogeneous firms**

  **(b) Entry with endogenous product choice**

  **(c) Firm heterogeneity**

  **(d) Entry and spatial competition**

  **(e) Multi-store firms**

  **(f) Dynamic models**

## 3(a). Specification and Estimation: Homogeneous firms

- **Bresnahan and Reiss (1991a)**, study several retail and professional industries in US, i.e., pharmacies, tire dealers, doctors, and dentists.

- The main purpose of the paper is to estimate the "nature" or "degree" of competition for each of the industries: how fast variable profits decline when the number of firms in the market increases.

- For each industry, their dataset consists of a cross-section of *M* small "isolated markets". Assumptions:
    - Markets are assumed independent in terms of demand and competition.
    - A market consists of a single location, i.e., $L=1$, such that spatial competition is not explicitly incorporated in the model.

## 3(a). Specification and Estimation: Homogeneous firms

- For each local market, the researcher observes:
  - number of active firms ($n$).
  - measure of market size ($s$).
  - exogenous market characteristics in demand and/or costs ($\mathbf{x}$).

- Given this limited information, the researcher needs to restrict firm heterogeneity. All the potential entrants in a market are identical and have complete information on demand and costs.

- The profit of a store is:

$$\pi(n) = s * vp(\mathbf{x}, n) - EC(\mathbf{x}) - \varepsilon$$

## 3(a). Specification and Estimation: Homogeneous firms

- Implicitly, there is a two-stage game: stage 1, entry; stage 2, price or quantity competition between active stores that determines the variable profit $s * vp(\mathbf{x}, n)$.

- However, the form of competition between active firms is not explicitly modelled. Instead, the authors consider a flexible specification of the variable profit per capita $vp(\mathbf{x}, n)$ that is strictly decreasing but nonparametric in the number of active stores.

- This specification is consistent with a general model of competition (or collusion) between homogeneous firms, or even between symmetrically differentiated firms.

## 3(a). Specification and Estimation: Homogeneous firms

- **Nash-Equilibrium in the entry game:** number of firms $n^*$ satisfies:

  - Active firms want to be active: $\pi(n^*) \geq 0$
  - Inactive firms want to be out: $\pi(n^* + 1) < 0.$

- Since the profit function is strictly decreasing, the **equilibrium is unique** and it can be represented using the following expression:

$$\{n^* = n\} \Leftrightarrow \{\pi(n) \geq 0 \;\; and \;\; \pi(n+1) < 0\,\}$$

$$\Leftrightarrow \;\; \{s \; vp(\mathbf{x}, n+1) - EC(\mathbf{x}) < \varepsilon \leq s \; vp(\mathbf{x}, n) - EC(\mathbf{x})\}$$

## 3(a). Specification and Estimation: Homogeneous firms

- The equilibrium condition implies that the distribution of the number of active firms in a market is:

$$\Pr(n^* = n \mid s, \mathbf{x})$$
$$= F\big(s \;\; vp(\mathbf{x}, \mathrm{n}) - EC(\mathbf{x})\big) - F\big(s \;\; vp(\mathbf{x}, \mathrm{n}+1) - EC(\mathbf{x})\big)$$

where $F(.)$ is the CDF of $\varepsilon$.

- This is the structure of an ordered discrete choice model. The parameters of the model can be easily estimated by MLE.

## 3(b). Entry with Endogenous Product Choice

- **Mazzeo (2002)** studies market entry in the motel industry using local markets along U.S. interstate highways.

- A local market is defined as a narrow region around a highway exit.

- Mazzeo's model maintains most of the assumptions in Bresnahan-Reiss: no spatial competition, ex-ante homogeneous firms, complete information, no multi-store firms, and no dynamics.

- However, he extends Bresnahan-Reiss model by introducing endogenous product differentiation. Firms not only decide whether to enter in a market but they also choose the type of product: low quality hotel $E$, or high quality product $H$.

# 3(b). Entry with Endogenous Product Choice

- Product competition less intense and it can increase firms' profits. But firms have also an incentive to offer the type of product for which demand is stronger.

- The profit of an active hotel of type $T \in \{E, H\}$ is:

$$\pi_T(n_E, n_H) = s \;\; v_T(\mathbf{x}, n_E, n_H) - EC_T(\mathbf{x}) - \varepsilon_T$$

where $n_E$ and $n_H$ represent the number of active hotels with low and high quality, respectively, in the local market.

- $v_T(.)$ is the variable profit per capita and $EC_T(\mathbf{x}) + \varepsilon_T$ is the entry cost for type $T$ hotels, where $\varepsilon_T$ is unobservable to the researcher.

## 3(b). Entry with Endogenous Product Choice

- Mazzeo solves and estimates his model under two different equilibrium concepts: Stackelberg and what he terms a **"two-stage game."**

- **First stage:** total number of active hotels, $n \equiv n_E + n_H$, enter the market as long as there is some configuration $(n_E, n_H)$ where both low quality and high quality hotels make positive profits.

- F*irst-stage profit function* as:

$$\Pi(n) \equiv max_{\{n_E,n_H:\, n_E+n_H=n\}}\{min[\,\pi_E(n_E, n_H)\,,\ \pi_H(n_E, n_H)]\}$$

- The **equilibrium number of hotels** is the value $n^*$ that satisfies two conditions: $\Pi(n^*) \geq 0$ and $\Pi(n^* + 1) < 0$.

## 3(b). Entry with Endogenous Product Choice

- If the profit functions $\pi_E$ and $\pi_H$ are strictly decreasing functions in the number of active firms $(n_E, n_H)$, then $\Pi(n)$ is also a strictly decreasing function, and the equilibrium number of stores in the first stage, $n^*$, is unique.

- **Second stage**. Active hotels choose simultaneously their type or quality level. An equilibrium is a pair $(n_E^*, n_H^*)$ such that every firm chooses the type that maximizes its profit given the choices of the other firms:

$$\pi_E(n_E^*, n_H^*) \geq \pi_H(n_E^* - 1, n_H^* + 1)$$

$$\pi_H(n_E^*, n_H^*) \geq \pi_E(n_E^* + 1, n_H^* - 1)$$

# 3(b). Entry with Endogenous Product Choice

- Using these equilibrium conditions, it is possible to obtain a closed form expression for the (quadrangle) region in the space of the unobservables $(\varepsilon_E, \varepsilon_H)$ that generate a particular value of the equilibrium pair $(n_E^*, n_H^*)$.

- Let $R_\varepsilon(n_E, n_H; s, \mathbf{x})$ be the quadrangle region in $\mathbb{R}^2$ associated with the pair $(n_E, n_H)$ given exogenous market characteristics $(s, \mathbf{x})$, and let $F(\varepsilon_E, \varepsilon_H)$ be the CDF of the unobservable variables.

$$\Pr(n_E^* = n_E, n_H^* = n_H | s, \mathbf{x})$$
$$= \int 1\{(\varepsilon_E, \varepsilon_H) \in R_\varepsilon(n_E, n_H; s, \mathbf{x})\} \, dF(\varepsilon_E, \varepsilon_H)$$

## 3(b). Entry with Endogenous Product Choice

- **Mazzeo** finds that hotels have strong incentives to differentiate from their rivals to avoid nose-to-nose competition.

- **Ellickson and Misra** (2008) estimate a game for the US supermarket industry where supermarkets choose the type of "pricing strategy": 'Everyday Low Price' (EDLP) versus 'High-Low' pricing. They find evidence strategic complementarity between supermarkets.

- **Vitorino (2012)** estimates a game of store entry in shopping centers that allows for positive spillover effects among stores, and also unobserved market heterogeneity for the researcher. She finds that, after controlling for unobserved market heterogeneity, firms face business stealing effects but also significant incentives to collocate.

# 3(c). Firm Heterogeneity

- **Bresnahan and Reiss (1990)** estimate a static model of entry in the US automobile dealers industry that allows for firm heterogeneity.

- Each local market has two potential entrants, which we index with $i, j \in \{1,2\}$. The profit function of firm $i$ if active in the market is:

$$\pi_i(a_j) = s \ v(\mathbf{x}, a_j) - EC(\mathbf{x}) - \varepsilon_i$$

$\pi_i(0)$ is profit of firm $i$ under monopoly; $\pi_i(1)$ is profit under duopoly.

- **Nash equilibrium**: pair of actions $(a_i^*, a_j^*)$ such that:

$$a_i^* = 1\{\pi_i(a_j^*) \geq 0\} \quad and \quad a_j^* = 1\{\pi_j(a_i^*) \geq 0\}$$

## 3(c). Firm Heterogeneity

- Given this description of an equilibrium, we can derive the quadrangle regions in the space of the unobservables $(\varepsilon_i, \varepsilon_j)$ associated with the different equilibrium outcomes.

- Threshold for entry as a monopolist: $\Delta^M(\mathrm{s}, \mathbf{x}) \equiv s \;\; v(\mathbf{x}, 0) - EC(\mathbf{x})$; Threshold for entry as a duopolist: $\Delta^D(\mathrm{s}, \mathbf{x}) \equiv s \;\; v(\mathbf{x}, 1) - EC(\mathbf{x})$. Then,

$$
\begin{aligned}
(a_i^*, a_j^*) = (0,0) &\Leftrightarrow \{\varepsilon_i > \Delta^M(\mathrm{s}, \mathbf{x}) \;\; and \;\; \varepsilon_j > \Delta^M(\mathrm{s}, \mathbf{x})\} \\
(a_i^*, a_j^*) = (0,1) &\Leftrightarrow \{\varepsilon_i > \Delta^D(\mathrm{s}, \mathbf{x}) \;\; and \;\; \varepsilon_j \leq \Delta^M(\mathrm{s}, \mathbf{x})\} \\
(a_i^*, a_j^*) = (1,0) &\Leftrightarrow \{\varepsilon_i \leq \Delta^M(\mathrm{s}, \mathbf{x}) \;\; and \;\; \varepsilon_j > \Delta^D(\mathrm{s}, \mathbf{x})\} \\
(a_i^*, a_j^*) = (1,1) &\Leftrightarrow \{\varepsilon_i \leq \Delta^D(\mathrm{s}, \mathbf{x}) \;\; and \;\; \varepsilon_j \leq \Delta^D(\mathrm{s}, \mathbf{x})\}
\end{aligned}
$$

## 3(c). Firm Heterogeneity

- This model has multiple equilibria. For values of $(\varepsilon_i, \varepsilon_j)$ within the square region $[\Delta^D(\mathrm{s}, \mathbf{x}), \Delta^M(\mathrm{s}, \mathbf{x})] \times [\Delta^D(\mathrm{s}, \mathbf{x}), \Delta^M(\mathrm{s}, \mathbf{x})]$, outcomes $(0{,}1)$ and $(1{,}0)$ are both Nash equilibria.

- In this empirical application where all the exogenous state variables are common market characteristics and the two firms are identical in terms of observable characteristics and structural parameters, the model implies the unique distribution of the number of entrants.

$$\Pr(n^* = 2 | s, \mathbf{x}) = F(\Delta^D, \Delta^D)$$

$$\Pr(n^* = 1 | s, \mathbf{x}) = F(\Delta^M, +\infty) + F(+\infty, \Delta^M) - F(\Delta^M, \Delta^M) - F(\Delta^D, \Delta^D)$$

$$\Pr(n^* = 0 | s, \mathbf{x}) = 1 - F(\Delta^M, +\infty) - F(+\infty, \Delta^M) + F(\Delta^M, \Delta^M)$$

# 3(d). Entry and Spatial Competition

- How important is spatial differentiation to explain market power? **Seim (2006)** studies this question in the context of the video rental industry.

- A local market has $L$ business locations, indexed by $\ell$. For every business location point, Seim defines $B$ concentric rings around:
  - a first ring of radius $d_1$ (e.g., half a mile);
  - a second ring of radius $d_2$ (e.g., one mile);
  - and so on, where $d_1 < d_2 < \cdots < d_B$.

- The profit of a store in location $\ell$ depends on the number of other stores located within each of the B rings. Closer stores have stronger negative effects on profits.

## 3(d). Entry and Spatial Competition

- The specification of the profit function is:

$$\pi_{i\ell} = \mathbf{x}_\ell\,\beta + \sum_{\mathrm{b}=1}^{\mathrm{B}} \gamma_{\mathrm{b}}\, n_{\mathrm{b}\ell} + \xi_\ell + \varepsilon_{i\ell}$$

  - $\beta, \gamma_1, \gamma_2, \ldots, \gamma_{\mathrm{B}}$ are parameters;
  - $\mathbf{x}_\ell$ is a vector of observable exogenous characteristics;
  - $n_{\mathrm{b}\ell}$ is the number of stores in ring b around location $\ell$;
  - $\xi_\ell$ represents exogenous characteristics of location $\ell$ that are unobserved to the researcher but common and observable to firms;
  - $\varepsilon_{i\ell}$ is a component of the profit of firm $i$ in location $\ell$ that is private information to this firm, and are assumed i.i.d. over firms and locations with the Type 1 Extreme Value

# 3(d). Entry and Spatial Competition

- A firm does not know other firms' private information, and therefore, the number–of active stores at different ring-locations $\{n_{\mathrm{b}\ell}\}$ is unknown to this firm.

- Let $\{n^e_{\mathrm{b}\ell}\}$ be a firm's expectation about the number of stores active at ring-location $(\mathrm{b}, \ell)$. Best response function is:

$$a_{i\ell} = 1\left\{ \mathbf{x}_\ell \beta + \sum_{\mathrm{b}=1}^{\mathrm{B}} \gamma_{\mathrm{b}}\, n^e_{\mathrm{b}\ell} + \xi_\ell + \varepsilon_{i\ell} \right.$$

$$\left. \geq \mathbf{x}_{\ell'} \beta + \sum_{\mathrm{b}=1}^{\mathrm{B}} \gamma_{\mathrm{b}}\, n^e_{\mathrm{b}\ell'} + \xi_{\ell'} + \varepsilon_{i\ell'} \quad \forall \ell' \neq \ell \right\}$$

## 3(d). Entry and Spatial Competition

- Integrating this best response function over the distribution of the private information variables, we obtain the probability that the best response of a firm is to have a store in location $\ell$:

$$\mathrm{P}_{\ell} = \frac{exp\left\{\mathbf{x}_{\ell}\beta + \sum_{\mathrm{b}=1}^{\mathrm{B}} \gamma_{\mathrm{b}}\, n_{\mathrm{b}\ell}^{e} \; + \xi_{\ell}\right\}}{1 + \sum_{\ell'=1}^{L} exp\left\{\mathbf{x}_{\ell'}\beta + \sum_{\mathrm{b}=1}^{\mathrm{B}} \gamma_{\mathrm{b}}\, n_{\mathrm{b}\ell'}^{e} + \xi_{\ell'}\right\}}$$

- In **equilibrium**, firms' beliefs/expectations must be consistent:

$$n_{\mathrm{b}\ell}^{e} = N\left[\sum_{\ell'=1}^{L} D_{\ell\ell'}^{b}\; \mathrm{P}_{\ell'}\right]$$

where $D_{\ell\ell'}^{b}$ is a binary indicator of the event "$\ell'$belongs to ring b around $\ell$".

## 3(d). Entry and Spatial Competition

- System of $L$ equations with $L$ unknowns that define a fixed point mapping in the space of the vector of entry probabilities $\mathbf{P} = \{\mathrm{P}_\ell : \ell = 1, 2, \ldots, L\}$. An equilibrium of the model is a fixed point of this mapping. By Brower's Theorem an equilibrium exists.

- **Zhu and Singh (2009)** consider a similar model to study competition between big-box discount stores in US (i.e., Kmart, Target and Wal-Mart), Zhu and Singh (2009). They extend Seim's entry model by introducing firm heterogeneity.

- The model allows competition effects to be asymmetric across three different chains.

## 3(d). Entry and Spatial Competition

- The profit function of a store of chain $i$ at location $\ell$ is:

$$\pi_{i\ell} = \mathbf{x}_{\ell}\,\beta_i + \sum_{j\neq i}\sum_{b=1}^{B} \gamma_{bij}\, n_{b\ell j} + \xi_{\ell} + \varepsilon_{i\ell}$$

where $n_{b\ell j}$ represents the number of stores that chain $j$ has within the b-ring around location $\ell$. Despite the paper studies competition between retail chains, it still makes similar simplifying assumptions as in Seim's model that ignores important aspects of competition between retail chains. In particular, the model ignores economies of density, and firms' concerns on cannibalization between stores of the same chain. It assumes that the entry decisions of a retail chain are made independently at each location. Under these assumptions, the equilibrium of the model can be described as a vector of $N * L$ entry probabilities, one for each firm and location, that solves the following fixed point problem:

$$\mathrm{P}_{i\ell} = \frac{exp\left\{\mathbf{x}_{\ell}\,\beta_i + \sum_{j\neq i}\sum_{\mathrm{b}=1}^{\mathrm{B}}\gamma_{bij}\,N\left[\sum_{\ell'=1}^{L} D_{\ell\ell'}^{b}\ \mathrm{P}_{j\ell'}\right] + \xi_{\ell}\right\}}{1+\sum_{\ell'=1}^{L} exp\left\{\mathbf{x}_{\ell'}\,\beta_i + \sum_{j\neq i}\sum_{\mathrm{b}=1}^{\mathrm{B}}\gamma_{bij}\,N\left[\sum_{\ell''=1}^{L} D_{\ell'\ell''}^{b}\ \mathrm{P}_{j\ell''}\right] + \xi_{\ell'}\right\}}$$

The authors find substantial heterogeneity in the competition effects between these three big-box discount chains, and in the pattern of how these effects decline with distance. For instance, Wal-Mart's supercenters have a very substantial impact even at large distance.

Datta and Sudhir (2013) estimate an entry model of grocery stores that endogenizes both location and product type decisions. Their main interests are the consequence of zoning on market structure. Zoning often reduces firms' ability to avoid competition by locating remotely each other. Theory suggests that in such a market firms have a stronger incentive to differentiate their products. Their estimation results support this theoretical prediction. The authors also investigate different impacts of various types of zoning ("centralized zoning", "neighborhood zoning", "outskirt zoning") on equilibrium market structure.

# 3(e). Multi-Store Firms

As we have mentioned above, economies of density and cannibalization are potentially important factors in store location decisions of retail chains. A realistic model of competition between retail chains should incorporate this type of spillover effects. Taking into account these effects requires a model of competition between multi-store firms similar to the one in Section 2.1(b). The model takes into account the joint determination of a firm's entry decisions at different locations. A firm's entry decision is represented by the *L-dimension* vector $\boldsymbol{a}_i \equiv \{a_{i\ell} : \ell = 1, 2, \ldots, L\}$, with $a_{i\ell} \in \{0,1\}$, such that the set of possible actions contains $2^L$ elements. For instance, Jia (2008) studies competition between two chains (Wal-Mart and Kmart) over 2,065 locations (US counties). The number of possible decisions of a retail chain is $2^{2065}$, which is larger than $10^{621}$. It is obvious that, without further restrictions, computing firms' best responses is intractable.

Jia (2008) proposes and estimates a game of entry between Kmart and Wal-Mart over more than two thousand locations (counties). Her model imposes restrictions on the specification of firms' profits that imply the supermodularity of the game and facilitate substantially the computation of an equilibrium. Suppose that we index the two firms as $i$ and $j$. The profit function of a firm, say $i$, is $\Pi_i = VP_i(\boldsymbol{a}_i, \boldsymbol{a}_j) - EC_i(\boldsymbol{a}_i)$, where $VP_i$ is the variable profit function such that

$$VP_i(\boldsymbol{a}_i, \boldsymbol{a}_j) = \sum_{\ell=1}^{L} a_{i\ell} \left[\mathbf{x}_\ell \, \beta_i + \gamma_{ij} \, a_{j\ell}\right], \tag{29}$$

and $EC_i$ is the entry cost function such that

$$EC_i(\boldsymbol{a}_i) = \sum_{\ell=1}^{L} a_{i\ell} \left[ \theta_{i\ell}^{EC} - \frac{\theta^{ED}}{2} \sum_{\ell' \neq \ell} \frac{a_{i\ell'}}{d_{\ell\ell'}} \right]. \tag{30}$$

where $\mathbf{x}_\ell$ is a vector of market/location characteristics; $\gamma_{ij}$ is a parameter that represents the effect on the profit of firm $i$ of competition from a store of chain $j$; $\theta_{i\ell}^{EC}$ is the entry cost that firm $i$ would have in location $\ell$ in the absence of economies of density (i.e., if it were a single-store firm); $\theta^{ED}$ is a parameter that represents the magnitude of the economies of density and is assumed to be positive; and $d_{\ell\ell'}$ is the distance between locations $\ell$ and $\ell'$. Jia further assumes that the entry cost $\theta_{i\ell}^{EC}$ consists of three parts: $\theta_{i\ell}^{EC} = \theta_i^{EC} + (1-\rho)\xi_\ell + \varepsilon_{i\ell}$, where $\theta_i^{EC}$ is chain-fixed effects, $\rho$ is a scale parameter, $\xi_\ell$ is a location random effect, and $\varepsilon_{i\ell}$ is a firm-location error term. Both $\{\xi_\ell\}$ and $\{\varepsilon_{i\ell}\}$ are i.i.d. draws from the standard normal distribution and known to all the players when making decisions. To capture economies of density, the presence of the stores of the same firm at other locations is weighted by the inverse of the distance between locations, $1/d_{\ell\ell'}$. This term is multiplied by one-half to avoid double counting in the total entry cost of the retail chain.

The specification of the profit function in equations (29) and (30) imposes some important restrictions. Under this specification, locations are interdependent only through economies of density. In particular, there are no cannibalization effects between stores of the same chain at different locations. Similarly, there is no spatial competition between stores of different chains at different locations. In particular, this specification ignores the spatial competition effects between Kmart, Target, and Wal-Mart that Zhu and Singh (2009) find in their study. The specification also rules out cost savings that do not depend on store density such as lower wholesale prices due to strong bargaining power of chain stores. The main motivation for these restrictions is to have a supermodular game that facilitates very substantially the computation of an equilibrium, even when the model has a large number of locations.

In a Nash equilibrium of this model, the entry decisions of a firm, say $i$, should satisfy the following $L$ optimality conditions:

$$a_{i\ell} = 1\left\{ \mathbf{x}_{\ell}\, \beta_i + \gamma_{ij}\, a_{j\ell} - \theta_{i\ell}^{EC} + \frac{\theta^{ED}}{2} \sum_{\ell' \neq \ell} \frac{a_{i\ell'}}{d_{\ell\ell'}} \geq 0 \right\}$$

These conditions can be interpreted as the best response of firm $i$ in location $\ell$ given the other firm's entry decisions, and given also firm $i$'s entry decisions at locations other than $\ell$. We can write this system of conditions in a vector form as $\boldsymbol{a}_i = \mathrm{br}_i(\boldsymbol{a}_i, \boldsymbol{a}_j)$. Given $\boldsymbol{a}_j$, a fixed point of the mapping $\text{br}_i(., \boldsymbol{a}_j)$ is a (full) best response of firm $i$ to the choice $\boldsymbol{a}_j$ by firm $j$. With $\theta^{ED} > 0$ (i.e., economies of density), it is clear from equation (31) that the mapping $\text{br}_i$ is increasing in $\boldsymbol{a}_i$. By Topkis's Theorem, this increasing property implies that: (i) the mapping has at least one fixed point solution; (ii) if it has multiple fixed points they are ordered from the lowest to the largest; and (iii) the smallest (largest) fixed point can be obtained by successive iterations in the mapping $\text{br}_i$ using as starting value $\boldsymbol{a}_i = \mathbf{0}$ ($\boldsymbol{a}_i = \mathbf{1}$). Given these properties, Jia shows that the following algorithm provides the Nash Equilibrium that is most profitable for firm $i$:

-Step $[i]$: Given the lowest possible value for $\boldsymbol{a}_j$, i.e., $\boldsymbol{a}_j = (0,0, \dots 0)$, we apply successive iterations with respect to $\boldsymbol{a}_i$ in the fixed point mapping $\text{br}_i(., \boldsymbol{a}_j = \mathbf{0})$ starting at $\boldsymbol{a}_i = (1,1, \dots 1)$. These iterations converge to the largest best response of firm $i$, that we denote by $\boldsymbol{a}_i^{(1)} = BR_i^{(High)}(\mathbf{0})$.

-Step [$j$]: Given $\boldsymbol{a}_i^{(1)}$, we apply successive iterations with respect to $\boldsymbol{a}_j$ in the fixed point mapping $\mathrm{br}_j(.,\boldsymbol{a}_i^{(1)})$ starting at $\boldsymbol{a}_j = \mathbf{0}$. These iterations converge to the lowest best response of firm $j$, that we denote by $\boldsymbol{a}_j^{(1)} = BR_j^{(Low)}(\boldsymbol{a}_i^{(1)})$.

-We keep iterating in (Step [$i$]) and (Step [$j$]) until convergence.

At any iteration, say $k$, given $\boldsymbol{a}_j^{(k-1)}$ we first apply (Step [$i$]) to obtain $\boldsymbol{a}_i^{(k)} = BR_i^{(HIgh)}\left(\boldsymbol{a}_j^{(k-1)}\right)$, and then we apply (Step [ $j$ ) to obtain $\boldsymbol{a}_j^{(k)} = BR_j^{(Low)}\left(\boldsymbol{a}_i^{(k)}\right)$. The supermodularity of the game assures the convergence of this process and the resulting fixed point is the Nash equilibrium that most favors firm $i$. Jia combines this solution algorithm with a simulation of unobservables to estimate the parameters of the model using the method of simulated moments (MSM).

In his empirical study of convenience stores in Okinawa Island of Japan, Nishida (forthcoming) extends Jia's model in two directions. First, a firm is allowed to open multiple stores (up to four) in the same location. Second, the model explicitly incorporates some form of spatial competition: a store's revenue is affected not only by other stores in the same location but also by those in adjacent locations.

Although the approach used in these two studies is elegant and useful, its use in other applications is somewhat limited. First, supermodularity requires that the own network effect on profits is monotonic, i.e., the effect of $\sum_{\ell' \neq \ell} \frac{a_{i\ell'}}{d_{\ell\ell'}}$ is either always positive ($\theta^{ED} > 0$) or always negative ($\theta^{ED} < 0$). This condition rules out situations where the net effect of cannibalization and economies of density varies across markets. Second, the number of (strategic) players must be equal to two. For a game to be supermodular, players' strategies must be strategic complements. In a model of market entry, players' strategies are strategic substitutes. However, when the number of players is equal to two, any game of strategic substitutes can be transformed into one of strategic complements by changing the order of strategies of one player (e.g., use zero for entry and one for no entry). This trick no longer works when we have more than two players.

Ellickson et al. (2013, EHT hereafter) propose an alternative estimation strategy and apply it to data of U.S. discount store chains. Their estimation method is based on a set of inequalities that arise from the best response condition of a Nash equilibrium. Taking its opponents' decisions as given, a chain's profit associated with its observed entry decision must be larger than the profit of any alternative entry decision.

EHT consider particular deviations that relocate one of the observed stores to another location. Let $\boldsymbol{a}_i^*$ be the observed vector of entry decisions of firm $i$, and suppose that in this observed vector the firm has a store in location $\ell$ but not in location $\ell'$. Consider the alternative (hypothetical) choice $\boldsymbol{a}_i^{(\ell\to\ell')}$ that is equal to $\boldsymbol{a}_i^*$ except that the store in location $\ell$ is closed and relocated to location $\ell'$. Revealed preference implies that $\pi_i(\boldsymbol{a}_i^*) \geq \pi_i(\boldsymbol{a}_i^{(\ell\to\ell')})$. EHT further simplify this inequality by assuming that there are no economies of scope or density (e.g., $\theta^{ED} = 0$ ), and that there are no firm-location-specific factors unobservable to the researcher, i.e., $\varepsilon_{i\ell} = 0$ . Under these two assumptions, the inequality above can be written as the profit difference between two locations

$$[\mathbf{x}_\ell - \mathbf{x}_{\ell'}]\beta_i + \sum_{j \neq i} \gamma_{ij} \left[a^*_{j\ell} - a^*_{j\ell'}\right] + [\xi_\ell - \xi_{\ell'}] \geq 0$$

Now, consider another chain, say $k$, that has an observed choice $\boldsymbol{a}^*_k$ with a store in location $\ell'$ but not in location $\ell$. For this chain, we consider the opposite (hypothetical) relocation decision that for firm $i$ above: the store in location $\ell'$ is closed and a new store is open in location $\ell$. For this chain, revealed preference implies that $[\mathbf{x}_{\ell'} - \mathbf{x}_\ell]\beta_k + \sum_{j \neq k} \gamma_{kj} \left[a^*_{j\ell'} - a^*_{j\ell}\right] + [\xi_{\ell'} - \xi_\ell] \geq 0$. Summing up the inequalities for firms $i$ and $k$, we generate an inequality that is free from location fixed effects $\xi_\ell$.

$$[\mathbf{x}_\ell - \mathbf{x}_{\ell'}](\beta_i - \beta_k) + \sum_{j \neq i} \gamma_{ij} \left[a^*_{j\ell} - a^*_{j\ell'}\right] + \sum_{j \neq k} \gamma_{kj} \left[a^*_{j\ell'} - a^*_{j\ell}\right] \geq 0$$

EHT construct a number of inequalities of this type and obtain estimates of the parameters of the model by using a smooth maximum score estimator (Manski 1975, Horowitz, 1992, Fox, 2010).

Unlike the lattice theory approach of Jia and Nishida, the approach applied by EHT can accommodate more than two players, allows the researcher to be agnostic about equilibrium selections, and is robust to the presence of unobserved market heterogeneity. Their model, however, rules out any explicit interdependence between stores in different locations, including spatial competition, cannibalization and economies of density. Although incorporating such inter-locational interdependencies does not seem to cause any fundamental estimation issue, doing so can be difficult in practice as it considerably increases the amount of computation. Another possible downside of this approach is the restriction it imposes on unobservables. The only type of structural errors that this model includes are the variables $\xi_\ell$ that are common for all firms. Therefore, to accommodate observations that are incompatible with inequalities in (33) above, the model requires non-structural errors, which may be interpreted as firms' optimization errors.

## 3(f). Dynamics with Single-Store Firms

When the entry cost is partially sunk, firms' entry decisions depend on their incumbency status, and dynamic models become more relevant. The role of sunk entry costs in shaping market structure in an oligopoly industry was first empirically studied by Bresnahan and Reiss (1993). They estimate a two-period model using panel data of the number of dentists. Following recent developments in the econometrics of dynamic games of oligopoly competition, several studies have estimated dynamic games of market entry-exit in different retail industries.

Aguirregabiria and Mira (2007) estimate dynamic games of market entry and exit for five different retail industries: restaurants, bookstores, gas stations, shoe shops, and fish shops. They use annual data from a census of Chilean firms created for tax purposes by the Chilean Internal Revenue Service during the period 1994-1999. The estimated models show significant differences in fixed costs, entry costs, and competition effects across the five industries, and these three parameters provide a precise description of the observed differences in market structure and entry-exit rates between the five industries. Fixed operating costs are a very important component of total profits of a store in the five industries, and they range between 59% (in restaurants) to 85% (in bookstores) of the variable profit of a monopolist in a median market. Sunk entry costs are also significant in the five industries, and they range between 31% (in shoe shops) and 58% (in gas stations) of a monopolist variable profit in a median market. The estimates of the parameter that measures competition effect show that restaurants are the retailers with the smallest competition effects, that might explained by a higher degree of horizontal product differentiation in this industry.

Suzuki (2012) examines the consequence of tight land use regulation on market structure of hotels through its impacts on entry costs and fixed costs. He estimates a dynamic game of entry-exit of mid-scale hotels in Texas that incorporates detailed measures of land use regulation into cost functions of hotels. The estimated model shows that imposing stringent regulation increases costs considerably and has substantial effects on market structure and hotel profits. Consumers also incur a substantial part of the costs of regulation in the form of higher prices.

Dunne et al. (2013) estimate a dynamic game of entry and exit in the retail industries of dentists and chiropractors in US, and use the estimated model to evaluate the effects on market structure of subsidies for entry in small geographic markets, i.e., markets that were designated by the government as Health Professional Shortage Areas (HPSA). The authors compare the effects of this subsidy with those of a counterfactual subsidy on fixed costs, and they find that subsidies on entry costs are cheaper, or more effective for the same present value of the subsidy.

Yang (2014) extends the standard dynamic game of market entry-exit in a retail market by incorporating information spillovers from incumbent firms to potential entrants. In his model, a potential entrant does not know a market-specific component in the level of profitability of a market (e.g., a component of demand or operating costs). Firms learn about this profitability only when they actually enter that market. In this context, observing incumbents stay in this market is a positive signal for potential entrants about the quality of this market. Potential entrants use these signals to update their beliefs about the profitability of the market (i.e., Bayesian updating). These information spillovers from incumbents may contribute to explain why we observe retail clusters in some geographic markets. Yang estimates his model using data from the fast food restaurant industry in Canada, which goes back to the initial conditions of this industry in Canada. He finds significant evidence supporting the hypothesis that learning from incumbents induces retailers to herd into markets where others have previously done well in, and to avoid markets where others have previously failed in.

## 3(g). Dynamics and Spatial Competition with Retail Chains

A structural empirical analysis of economies of density, cannibalization, or spatial entry deterrence in retail chains requires the specification and estimation of models that incorporate dynamics, multi-store firms, and spatial competition. Some recent papers present contributions on this research topic.

Holmes (2011) studies the temporal and spatial pattern of store expansion by Wal-Mart during the period 1971-2005. He proposes and estimates a dynamic model of entry and store location by a multi-store firm similar to the one that we have described in Section 2.1(c) above. The model incorporates economies of density and cannibalization between Wal-Mart stores, though it does not model explicitly competition from other retailers or chains (e.g., Kmart or Target), and therefore it abstracts from dynamic strategic considerations such as spatial entry deterrence. The model also abstracts from price variation and assumes that Wal-Mart sets constant prices across all stores and over time. However, Holmes takes into account three different types of stores and plants in Wal-Mart retail network: regular stores that sell only general merchandise; supercenters, that sell both general merchandise and food; and distribution centers, which are the warehouses in the network, and that have also two different types, i.e., general and food distribution centers. The distinction between these types of stores and warehouses is particularly important to explain the evolution of Wal-Mart retail network over time and space. In the model, every year Wal-Mart decides the number and the geographic location of new regular stores, supercenters, and general and food distribution centers. Economies of density are channeled through the benefits of stores being close to distribution centers. The structural parameters of the model are estimated using the Moment Inequalities estimation method in Pakes et al. (2014). More specifically, moment inequalities are constructed by comparing the present value of profits from Wal-Mart's actual expansion decision with the present value from counterfactual expansion decisions which are slight deviations from the observed ones. Holmes finds that Wal-Mart obtains large savings in distribution costs by having a dense store network.

Igami and Yang (2014) study the trade-off between cannibalization and spatial pre-emption in the fast-food restaurant industry, e.g., McDonalds, Burger King, etc. Consider a chain store that has already opened its first store in a local market. Opening an additional store increases this chain's current and future variable profits by, first, attracting more consumers and, second, preventing its rivals' future entries (preemption). However, the magnitude of this increase could be marginal when the new store steals customers from its existing store (cannibalization). Whether opening a new store economically makes sense or not depends on the size of the entry cost. Igami and Yang estimate a dynamic structural model and find the quantitative importance of preemptive motives. However, they do not model explicitly spatial competition, or allow for multiple geographic locations within their broad definition of geographic market.

Schiraldi, Smith, and Takahashi (2013) study store location and spatial competition between UK supermarket chains. They propose and estimate a dynamic game similar to the one in Aguirregabiria and Vicentini (2012) that we have described in Section 2.1(c). A novel and interesting aspect of this application is that the authors incorporate the regulator's decision to approve or reject supermarkets' applications for opening a new store in a specific location. The estimation of the model exploits a very rich dataset from the U.K. supermarket industry on exact locations and dates of store openings/closings, applications for store opening, approval/rejection decisions by the regulator, as well as rich data of consumer choices and consumer locations.

# 4. Econometric Issues

## Multiple equilibria

Entry models with heterogeneous firms often generate more than one equilibria for a given set of parameters. Multiple equilibria pose challenges to the researcher for two main reasons. First, standard maximum likelihood estimation no longer works because the likelihood of certain outcomes is not well-defined without knowing the equilibrium selection mechanism. Second, without further assumptions, some predictions or counterfactual experiments using the estimated model are subject to an identification problem. These predictions depend on the type of equilibrium that is selected in an hypothetical scenario not included in the data.

Several approaches have been proposed to estimate an entry game with multiple equilibria. Which method works the best depends on assumptions imposed in the model, especially its information structure. In a game of complete information, there are at least four approaches. The simplest approach is to impose some particular equilibrium selection rule beforehand and estimate the model parameters under this rule. For instance, Jia (2008) estimates the model of competition between big-box chains using the equilibrium that is most preferable to K-mart. She also estimates the same model under alternative equilibrium selection rules to check for the robustness of some of her results. The second approach is to construct a likelihood function for some endogenous outcomes of the game that are common across all the equilibria. Bresnahan and Reiss (1991) estimate their model by exploiting the fact that, in their model, the total number of entrants is unique in all the equilibria.

A third approach is to make use of inequalities that are robust to multiple equilibria. One example is the profit inequality approach of Ellickson et al. (2013), which we described in Section 2.2(e) above. Another example is the method of moment inequality estimators proposed by Ciliberto and Tamer (2009). They characterize the lower and upper bounds of the probability of a certain outcome that are robust to any equilibrium selection rule. Estimation of structural parameters relies on the set of probability inequalities constructed from these bounds. In the first step, the researcher nonparametrically estimates the probabilities of equilibrium outcomes conditional on observables. The second step is to find a set of structural parameters such that the resulting probability inequalities are most consistent with the data. The application of Ciliberto and Tamer's approach to a spatial entry model may not be straightforward. In models of this class, the number of possible outcomes (i.e., market structures) is often very large. For example, consider a local market consisting of ten sub-blocks. When two chains decide whether they enter into each of these sub-blocks, the total number of possible market structures is 1,024 (=$2^{10}$). Such a large number of possible outcomes makes it difficult to implement this approach for two reasons. The first stage estimate is likely to be very imprecise even when a sample size is reasonably large. The second stage estimation can be computationally intensive because one needs to check, for a given set of parameters, whether each possible outcome meets the equilibrium conditions or not.

A fourth approach proposed by Bajari, Hong and Ryan (2010) consists in the specification of a flexible equilibrium selection mechanism and in the joint estimation of the parameters in this mechanism and the structural parameters in firms' profit functions. Together with standard exclusion restrictions for the identification of games, the key specification and identification assumption in this paper is that the equilibrium selection function depends only on firms' profits.

In empirical games of incomplete information, the standard way to deal with multiple equilibria is to use a two-step estimation method (Aguirregabiria and Mira, 2007, and Bajari, Hong, Krainer and Nekipelov, 2010). In the first step, the researcher estimates the probabilities of firms' entry conditional on market observables (called policy functions) in a nonparametric way, e.g., a sieves estimator. The second step is to find a set of structural parameters that are most consistent with the observed data and these estimated policy functions. A key assumption for the consistency of this approach is that, in the data, two markets with the same observable characteristics do not select different types of equilibria, i.e., same equilibria conditional on observables. Without this assumption, the recovered policy function in the first stage would be a weighted sum of firms' policies under different equilibria, making the second-stage estimates inconsistent. Several authors have recently proposed extensions of this method to allow for multiplicity of equilibria in the data for markets with the same observable characteristics.

## Unobserved Market Heterogeneity

Some market characteristics affecting firms' profits may not be observable to the researcher. For example, consider local attractions that spur the demand for hotels in a particular geographic location. Observing and controlling for all the relevant attractions are often impossible to the researcher. This demand effect implies that markets with such attractions should have more hotels than those without such attractions but with equivalent observable characteristics. Therefore, without accounting for this type of unobservables, researchers may wrongly conclude that competition boosts profits, or under-estimate the negative effect of competition on profits.

Unobserved market heterogeneity usually appears as an additive term ($\omega_\ell$) in the firm's profit function ($\pi_{i\ell}$) where $\omega_\ell$ is a random effect from a distribution known up to some parameters. The most common assumption (e.g., Seim, 2006, Zhu and Singh, 2009, Datta and Sudhir, 2013) is that these unobservables are common across locations in the same local market (i.e., $\omega_\ell = \omega$ for all $\ell$). Under this assumption the magnitude of unobserved market heterogeneity matters whether the firm enters some location in this market but not which location. Orhun (2013) relaxes this assumption by allowing unobserved heterogeneity to vary across locations in the same market.

In a game of complete information, accommodating unobserved market heterogeneity does not require a fundamental change in the estimation process. In a game of incomplete information, however, unobserved market heterogeneity introduces an additional challenge. Consistency of the two-step method requires that the initial nonparametric estimator of firms' entry probabilities in the first step should account for the presence of unobserved market heterogeneity. A possible solution is to use a finite mixture model. In this model, every market's $\omega_\ell$ is drawn from a distribution with finite support. Aguirregabiria and Mira (2007) show how to accommodate such market-specific unobservables into their nested pseudo likelihood (NPL) algorithm. Arcidiacono and Miller (2011) propose an expectation-maximization (EM) algorithm in a more general environment. An alternative way to deal with this problem is to use panel data with a reasonably long time horizon. In that way, we can incorporate market fixed effects as parameters to be estimated. This approach is popular when estimating a dynamic game (e.g., Ryan, 2012, and Suzuki, 2013). A necessary condition to implement this approach is that every market at least observes some entries during the sample period. Dropping markets with no entries from the sample may generate a selection bias.

## Computation

The number of geographic locations, $L$, introduces two dimensionality problems in the computation of firms' best responses in games of entry with spatial competition. First, in a static game, a multi-store firm's set of possible actions includes all the possible spatial configurations of its store network. The number of alternatives in this set is equal to $2^L$, and this number is extremely large even with modest values of $L$, such as a few hundred geographic locations. Without further assumptions, the computation of best responses becomes impractical. This is an important computational issue that has deterred some authors to account for multi-store retailers in their spatial competition models, e.g., Seim (2006), or Zhu and Singh (2009), among many others. As we have described in Section 2.2(e), two approaches that have been applied to deal with this issue are: (a) impose restrictions that guarantee supermodularity of the game (i.e., only two players, no cannibalization effects); (b) avoid the exact computation of best responses and use instead inequality restrictions implied by these best responses.

Looking at the firms' decision problem as a sequential or dynamic problem helps also to deal with the dimensionality in the space of possible choices. In a given period of time (e.g., year, quarter, month), we typically observe that a retail chain makes small changes in its network of stores, i.e., it opens a few new stores, or closes a few existing stores. Imposing these small changes as a restriction on the model implies a very dramatic reduction in the dimension of the action space such that the computation of best responses becomes practical, at least in a "myopic" version of the sequential decision problem.

However, to fully take into account the sequential or dynamic nature of a firm's decision problem, we also need to acknowledge that firms are forward looking. In the firm's dynamic programming problem, the set of possible states is equal to all the possible spatial configurations of a store network, and it has $2^L$ elements. Therefore, by going from a static model to a dynamic-forward-looking model, we have just "moved" the dimensionality problem from the action space into the state space. Recent papers propose different approaches to deal with this dimensionality problem in the state space. Arcidiacono et al. (2013) present a continuous-time dynamic game of spatial competition in a retail industry and propose an estimation method of this model. The continuous-time assumption eliminates the curse of dimensionality associated to integration over the state space. Aguirregabiria and Vicentini (2012) propose a method of spatial interpolation that exploits the information provided by the (indirect) variable profit function.

## Some ideas for further research

*Spillovers between different retail sectors*. Existing applications of games of entry and spatial competition in retail markets concentrate on a single retail industry. However, there are also interesting spillover effects between different retail industries. Some of these spillovers are positive, e.g., good restaurants can make a certain neighborhood more attractive for shopping. There are also negative spillovers effects through land prices, i.e., retail sectors with high value per unit of space (e.g., jewellery stores) are willing to pay higher land prices that supermarkets that have low markups and are intensive in the use of land. The consideration and measurement of these spillover effects is interesting in itself, and it can help to explain the turnover and reallocation of industries in different parts of a city. Relatedly, endogenizing land prices would also open the possibility of using these models for the evaluation of specific public policies at the city level.

*Richer datasets with store level information on prices, quantities, inventories*. The identification and estimation of competition effects based mainly on data of store locations has been the rule more than the exception in this literature. This approach typically requires strong restrictions in the specification of demand and variable costs. The increasing availability of datasets with rich information on prices and quantities at product and store level should create a new generation of empirical games of entry and spatial competition that relax these restrictions. Also, data on store characteristics such as product assortments or inventories will allow to introduce these important decisions as endogenous variables in empirical models of competition between retail stores.

*Measuring spatial pre-emption*. So far, all the empirical approaches to measure the effects of spatial pre-emption are based on the comparison of firms' actual entry with firms' behavior in a counterfactual scenario characterized by a change in either (i) a structural parameter (e.g., a store exit value), or (ii) firms' beliefs (e.g., a firm believes that other firms' entry decisions do not respond to this firm's entry behavior). These approaches suffer the serious limitation that they do not capture only the effect of pre-emption and are contaminated by other effects. The development of new approaches to measure the pure effect of pre-emption would be a methodological contribution with relevant implications in this literature

*Geography*. Every local market is different in its shape and its road network. These differences may have important impacts on the resulting market structure. For example, the center of a local market may be a quite attractive location for retailers when all highways go through there. However, it may not be the case anymore when highways encircle the city center (e.g., Beltway in Washington D.C.). These differences may affect retailers' location choices and the degree of competition in an equilibrium. The development of empirical models of competition in retail markets that incorporate, in a systematic way, these idiosyncratic geographic features will be an important contribution in this literature.