

ECO 310: Empirical Industrial Organization

Lecture 12: Models of Market Entry: Spatial Location

Victor Aguirregabiria (University of Toronto)

December 1st, 2022

Models of Market Entry [Cont.]: Outline

1. **Bresnahan & Reiss [Empirical Results]**
2. **Models of Firms' Spatial Location (Seim, 2006)**

1. **Bresnahan & Reiss (JPE, 1991)**

Market entry with homogeneous firms

- We start with an empirical model of entry in an homogeneous product industry and where all the firms have the same costs.
- There are several reasons why we start with this case.
- 1. This is the simpler empirical model of entry, and where this literature started with the seminal work by Bresnahan & Reiss (JPE, 1990).
- 2. The model with heterogeneous firms typically has multiple equilibria, and this makes the estimation more complicated.
- 3. Sometimes we have very limited information about firms' heterogeneous characteristics.

Market entry with homogeneous firms: Data

- Suppose the researcher has data from M markets in the same industry.
- For instance, the supermarket industry. The M markets are M neighborhoods from different Canadian cities.
- Markets are indexed by m .
- The dataset consists of:

$$\text{Data} = \{ n_m, X_m : m = 1, 2, \dots, M \}$$

n_m = number of active firms;

X_m = other market characteristics affecting demand or costs.

Market entry with homogeneous firms: Model

- All the potential entrants in a market have the same profit function:
 - Same costs, and same demand (homogenous product).
- The profit function of a firm in market m is:

$$V_m(n) - F_m$$

where $V_m(n)$ is the variables profit, F_m is the fixed cost, and n is the number of active firms in the market.

- We describe below the specification of $V_m(n)$ and F_m in terms of observable variables and unobservables.
- A key feature is that $V_m(n)$ is a strictly decreasing function of n .

Market entry with homogeneous firms: Model [2]

- Under Nash-equilibrium, we have the following conditions:

$$V_m \left(1 + \sum_{j \neq i} a_{jm} \right) - F_m \geq 0 \quad \text{for firms with } a_{im} = 1$$

$$V_m \left(1 + \sum_{j \neq i} a_{jm} \right) - F_m < 0 \quad \text{for firms with } a_{im} = 0$$

- Then, n_m is an equilibrium iff:

$$V_m(n_m) - F_m \geq 0 \quad \text{Active firms are in their best response}$$

$$V_m(1 + n_m) - F_m < 0 \quad \text{Inactive firms are in their best response}$$

Market entry with homogeneous firms: Model [3]

- We can also write the equilibrium conditions as:

$$n_m = \begin{cases} 0 & \text{if } F_m > V_m(1) \\ 1 & \text{if } V_m(2) < F_m \leq V_m(1) \\ 2 & \text{if } V_m(3) < F_m \leq V_m(2) \\ \vdots & \vdots \\ N-1 & \text{if } V_m(N) < F_m \leq V_m(N-1) \\ N & \text{if } F_m \leq V_m(N) \end{cases}$$

Market entry with homogeneous firms: Model [4]

- Suppose that F_m has the following structure:

$$F_m = X_m \beta^f + \varepsilon_m$$

where ε_m is unobservable to the researcher and it is i.i.d. $N(0, \sigma^2)$.

- Then, the model implies that:

$$n_m = \begin{cases} 0 & \text{if } \varepsilon_m > V_m(1) - X_m \beta^f \\ 1 & \text{if } V_m(2) - X_m \beta^f < \varepsilon_m \leq V_m(1) - X_m \beta^f \\ \vdots & \vdots \\ N-1 & \text{if } V_m(N) - X_m \beta^f < \varepsilon_m \leq V_m(N-1) - X_m \beta^f \\ N & \text{if } \varepsilon_m \leq V_m(N) - X_m \beta^f \end{cases}$$

Market entry with homogeneous firms: Model [5]

- Taking into account that $\varepsilon_m \sim N(0, \sigma^2)$, we have that for every integer value $n \geq 1$:

$$\Pr(n_m \leq n | X_m) = \Phi \left(\frac{V_m(n) - X_m \beta^f}{\sigma} \right)$$

- This equation from the equilibrium of the model provides restrictions that allow us to identify (estimate) the parameters of the model.
- The most common approach to estimate these models is maximum likelihood (This is an ordered probit model and it can be easily estimated using maximum likelihood).
- For simplicity, we consider here a simple two-step method that can be implemented using simply OLS methods.

Two-step Least squares method

- In a first step, we estimate the probability distribution $\Pr(n_m \leq n | X_m)$. For instance, this can be implemented using a simple frequency estimator. For any integer $n \geq 1$:

$$\hat{\Pr}(n_m \leq n | X) = \frac{\# \text{ obs. with } n_m \leq n \text{ \& } X_m = X}{\# \text{ obs. with } X_m = X}$$

- Let $y_m(n)$ be $\Phi^{-1}(\hat{\Pr}(n_m \leq n | X_m))$, where $\Phi^{-1}(\cdot)$ is the inverse function of the CDF $\Phi(\cdot)$.
- Then, we have the regression model for any market m and any $n \geq 1$:

$$y_m(n) = \theta^v V_m(n) + X_m \theta^f + e_m(n)$$

with $\theta^v = 1/\sigma$, and $\theta^f = -\beta^f/\sigma$, and $e_m(n)$ is an unobservable that comes from the estimation error in step 1.

Estimation: Two main cases

$$y_m(n) = \theta^v V_m(n) + X_m \theta^f + e_m(n)$$

- We can distinguish two main cases for the estimation of this model.
- Case 1. $V_m(n)$ is known to the researcher.
The researcher has data on prices and quantities such that demand and variable costs have been estimated and the indirect variable profit $V_m(n)$ is known.
- Case 2. $V_m(n)$ is unknown to the researcher.
The researcher does not know $V_m(n)$. This function should be specified and estimated together with the parameters in the fixed entry costs.

Estimation: $V(n)$ is known to the researcher

$$y_m(n) = \theta^v V_m(n) + X_m \theta^f + e_m(n)$$

- In this case, we have a simple OLS estimation.
- We can estimate θ^v and θ^f by OLS.
- Since $\theta^v = 1/\sigma$ and $\theta^f = -\beta^f/\sigma$, we have that σ and β^f are identified:

$$\sigma = 1/\theta^v \quad \text{and} \quad \beta^f = -\theta^f/\theta^v$$

- Note that σ and β^f are measured in the same units as $V_m(n)$, in dollars.

Estimation: $V(n)$ is known to the researcher

- In this case, the researcher needs to make an assumption about the functional form of $V_m(n)$.
- Ideally, we want expressions for $V_m(n)$ that are linear in parameters, such that we still have a linear in parameters regression model.
- For instance, under Cournot competition with linear demand ($P = A - B Q$) and constant marginal costs (c), we have that:

$$V_m(n) = \frac{(A - c)^2}{B} \frac{1}{(n + 1)^2}$$

such that:

$$y_m(n) = \tilde{\theta}^v \frac{1}{(n + 1)^2} + X_m \theta^f + e_m(n)$$

with $\tilde{\theta}^v = \frac{(A - c)^2}{\sigma B}$ and $\theta^f = -\beta^f / \sigma$.

Estimation: $V(n)$ is known to the researcher [2]

$$y_m(n) = \tilde{\theta}^v \frac{1}{(n+1)^2} + X_m \theta^f + e_m(n)$$

- Using OLS, we can estimate $\tilde{\theta}^v$ and θ^f .
- From here we cannot identify σ and β^f separately. But the estimated parameters provide useful information.
- For instance, for every possible value of n

$$\frac{X_m \theta^f}{\tilde{\theta}^v \frac{1}{(n+1)^2}} = \frac{F_m}{V_m(n)}$$

% of the fixed entry cost relative to the variable profit of a monopolist, duopolist, etc.

Specification of the variable profit function

- We can consider much more flexible specifications for the variable profit $V_m(n)$.
 - Including exogenous market characteristics X_m ;
 - More flexible functional forms with respect to n (it can be even nonparametric).
- Bresnahan and Reiss (JPE, 1991) do not model explicitly the form of price/quantity competition and consider a flexible model for the variable profit.

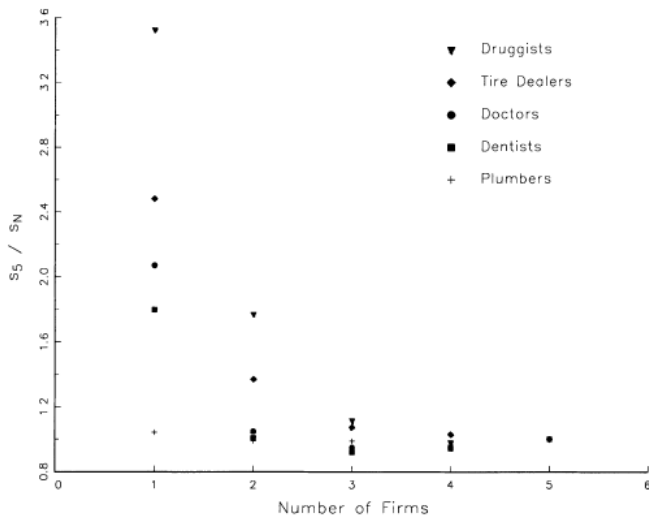
$$V_m(n) = S_m [X_m^\nu \beta^\nu - \alpha(n)]$$

- S_m represents market size.
- X_m^ν is a vector of observable market characteristics affecting variable profits, e.g., income, prices of variable inputs, and β^ν is a vector of parameters.
- The parameters $\alpha(1), \alpha(2), \dots$ capture the competitive effect. We expect:

Bresnahan & Reiss (JPE, 1991): Empirical results

- $M = 202$ local markets (small towns)
- Five industries: dentists, doctors, drug stores, plumbers and tire dealers.
- Main Findings:
 - Entry thresholds converge quite fast after the second entrant.
 - After three or four firms, an additional entrant doesn't affect much competition.

Bresnahan Reiss (JPE 1990)

FIG. 4.—Industry ratios of s_5 to s_N by N

Bresnahan & Reiss (JPE, 1990)

- **[Question]** What can we learn about the "nature of competition" in an industry from the empirical relationship between market size (S_m) and market structure/concentration (n_m)?
- $\Pi(n) = S [p(n) - AVC(q(n))] q(n) - F(n)$
- $\Pi(n) \geq 0$ implies $S \geq S^*(n)$ where the threshold market size $S^*(n)$ is:

$$S^*(n) \equiv \frac{F(n)}{[p(n) - AVC(q(n))] q(n)}$$

- How does $S^*(n)$ depends on n ? e.g., under the hypothesis of **contestable markets** $\frac{S^*(n)}{n}$ becomes constant for $n \geq n^*$ with a small value for n^* .

Bresnahan & Reiss (JPE, 1990)

TABLE 5
A. ENTRY THRESHOLD ESTIMATES

PROFESSION	ENTRY THRESHOLDS (000's)					PER FIRM ENTRY THRESHOLD RATIOS			
	S_1	S_2	S_3	S_4	S_5	s_2/s_1	s_3/s_2	s_4/s_3	s_5/s_4
Doctors	.88	3.49	5.78	7.72	9.14	1.98	1.10	1.00	.95
Dentists	.71	2.54	4.18	5.43	6.41	1.78	.79	.97	.94
Druggists	.53	2.12	5.04	7.67	9.39	1.99	1.58	1.14	.98
Plumbers	1.43	3.02	4.53	6.20	7.47	1.06	1.00	1.02	.96
Tire dealers	.49	1.78	3.41	4.74	6.10	1.81	1.28	1.04	1.03

B. LIKELIHOOD RATIO TESTS FOR THRESHOLD PROPORTIONALITY

Profession	Test for $s_4 = s_5$	Test for $s_3 = s_4 = s_5$	Test for $s_2 = s_3 = s_4 = s_5$	Test for $s_1 = s_2 = s_3 = s_4 = s_5$
Doctors	1.12 (1)	6.20 (3)	8.33 (4)	45.06* (6)
Dentists	1.59 (1)	12.30* (2)	19.13* (4)	36.67* (5)
Druggists	.43 (2)	7.13 (4)	65.28* (6)	113.92* (8)
Plumbers	1.99 (2)	4.01 (4)	12.07 (6)	15.62* (7)
Tire dealers	3.59 (2)	4.24 (3)	14.52* (5)	20.89* (7)

2. Models of Firms' Spatial Location

Models of Firms' Spatial Location

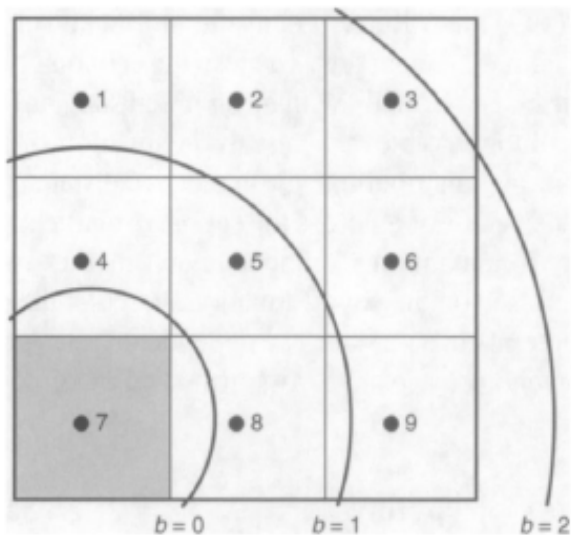
- Consider the decision of retail firm of where to open a new store within a city, e.g., a coffee shop, restaurant, supermarket, department store, etc.
- Different factors can play an important role:
 - Demand: what is the consumer traffic at different locations;
 - Rental prices
 - Location of competitors
- Geographic distance can be an important source of product differentiation. *Ceteris paribus*, a firm's profit increases with its distance to competitors.
- How profits decline when stores get closer?

Model: The city (1)

- From a geographical point of view, a market (city) is a set, for instance **a rectangle**, in the space \mathbb{R}^2 .
- Suppose that we divide this city/rectangle into L small squares, each one with its center.
- We can call each of these squares a submarket, or neighborhood, or location.
- A market/city can have hundreds of these submarkets/locations, e.g., $L = 200$.
- We index these locations by $\ell \in \{1, 2, \dots, L\}$

FIGURE 1

IMPACT ON PROFITS OF COMPETITORS' LOCATIONS: ILLUSTRATION

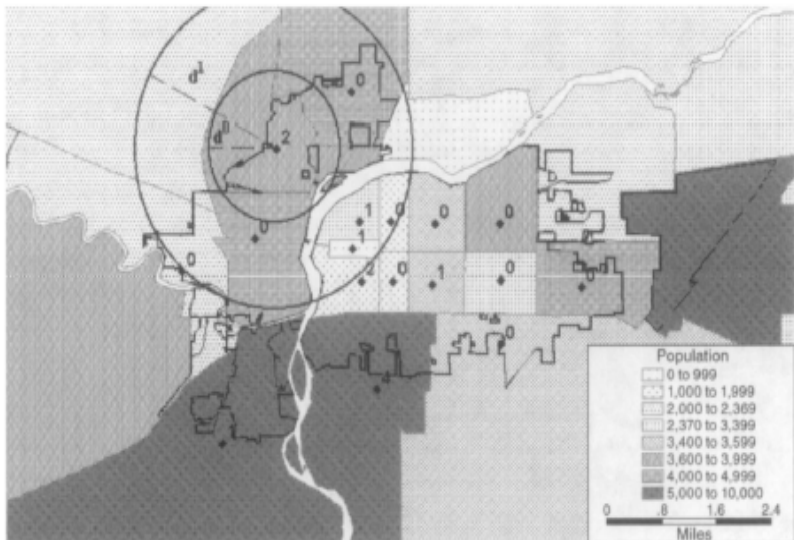


Model: The city (2)

- Each location has some characteristics that can affect demand and costs of a firm in that location:
 - Population; demographic characteristics of the population; rental prices.
- We represent the exogenous characteristics of location ℓ using the vector \mathbf{x}_ℓ .
- Therefore, we can see a city as a landscape of the characteristics \mathbf{x}_ℓ over the L locations.

FIGURE 2

SAMPLE MARKET: GREAT FALLS, MONTANA



Model: Firms

- There are N potential entrants in this industry (e.g., supermarkets) and city (Toronto).
- In the simpler version of the model, each potential entrant has only one possible store: no multi-store firms (chains).
- We consider this simpler version.
- Let a_i represent the entry / location decision of firm i .

$$a_i \in \{0, 1, \dots, L_m\}$$

- $a_{im} = 0$ represents "no entry";
- $a_{im} = \ell > 0$ represents entry in location ℓ .

Model: Profit function

- What is the profit of firm i if it opens a store in location ℓ ?
- In principle, we could consider a model of consumer choice of where to purchase (e.g., logit), a model of price competition between active firms; obtain the Bertrand equilibrium of that game, and the corresponding equilibrium profits.
- This approach has several important complications, and it requires having data on prices and quantities at every location.
- Instead, Seim (2006) considers a convenient shortcut.
- Her model does not explicitly specifies consumer choices and price competition, but it incorporates the idea that consumers face transportation costs and this implies that geographic distance with competitors (spatial differentiation) can increase a firm's profit.

Model: Profit function [2]

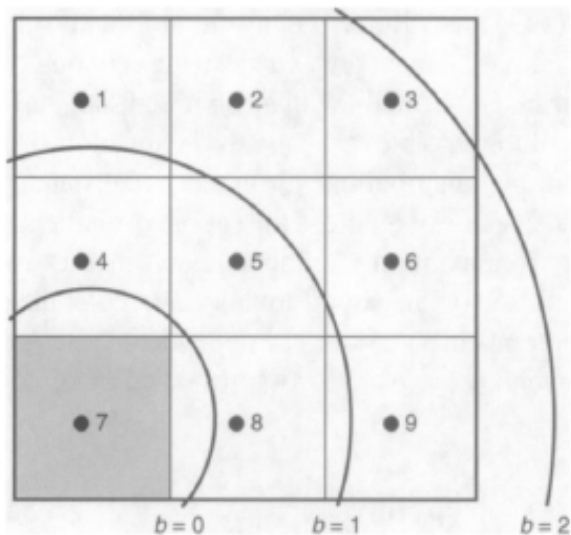
- Suppose that we draw a circle of radius d around the center point of location ℓ , e.g., a radius of 1km.
- From the point of view of a store located at ℓ , we can divide its competitors in two groups:
 - Close competitors: within the circle of radius d .
 - Far away competitors: outside the circle of radius d .
- Let $N_{\ell}(close)$ and $N_{\ell}(far)$ be the number of close and far away competitors relative to location ℓ .
- We can consider a profit function that depends on $\gamma_{close} N_{\ell}(close) + \gamma_{far} N_{\ell}(far)$, where γ_{close} and γ_{far} are parameters to estimate.
- We expect $\gamma_{close} < \gamma_{far} < 0$. The difference between γ_{close} and γ_{far} tell us how important is geographic distance as a form of differentiation to increase profits.

Model: Profit function [3]

- We can generalize this idea to allow for multiple circles, with different radius, around a location the center point of a location ℓ .
- Let $d_1 < d_2 < \dots < d_B$ be B different radius of increasing magnitude, e.g., $d_1 = 0.2 \text{ km}$, $d_2 = 0.5 \text{ km}$, $d_3 = 1 \text{ km}$, $d_4 = 2 \text{ km}$, ..., $d_{10} = 20 \text{ km}$.
- Given these radii, we can construct the number of firms with each of the bands defined by these radii:
 - $N_\ell(1)$ = Number of firms within the circle of radius d_1 ;
 - $N_\ell(2)$ = Number of firms within the band defined by the circles with radii d_1 and d_2 ;
 - $N_\ell(3)$ = Number of firms within the band defined by the circles with radii d_2 and d_3 ;
 - ...
 - $N_\ell(B+1)$ = Number of firms outside the circle with radius d_B .

FIGURE 1

IMPACT ON PROFITS OF COMPETITORS' LOCATIONS: ILLUSTRATION



Model: Profit function [4]

- Profit of an active firm at location ℓ is:

$$\Pi_i(\ell) = x_\ell \beta + \xi_\ell + \sum_{b=1}^B \gamma_b N_\ell(b) + \varepsilon_{i\ell}$$

- We expect:

$$\gamma_1 < \gamma_2 < \dots < \gamma_B < 0$$

- ξ_ℓ represents attributes of location ℓ which are known to firms but unobserved to the researcher.
- $\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iL}$ are assumed iid over firms and locations with extreme value distribution.

Model: Equilibrium (1)

- Suppose that firm i knows the actions of the other firms such that she knows, the landscape of firms: $\varepsilon_{i\ell}$, $N_\ell(1)$, ..., $N_\ell(B)$ for every location $\ell = 1, 2, \dots, L$.
- Best response of firm i is to choose location ℓ that maximizes:

$$\Pi_i(\ell) = x_\ell \beta + \xi_\ell + \sum_{b=1}^B \gamma_b N_\ell(b) + \varepsilon_{i\ell}$$

including the possibility of no entry, $a_i = 0$ with $\Pi_i(0) = 0$.

- Given the logit assumption on $\varepsilon_i(\ell)$'s the proportion or share of firms with a **best response** of locating in ℓ is:

$$s_\ell = \frac{n_\ell}{N} = \frac{\exp \left\{ x_\ell \beta + \xi_\ell + \sum_{b=1}^B \gamma_b N_\ell(b) \right\}}{1 + \sum_{j=1}^L \exp \left\{ x_j \beta + \xi_j + \sum_{b=1}^B \gamma_b N_j(b) \right\}}$$

Model: Equilibrium (2)

- The equilibrium of the model is described by L simultaneous equations, one for the share of each location ℓ .
- The L are simultaneously determined: Note that $N_\ell(b)$ is just equal to N times the sum of shares s_j in locations j within the band b around location ℓ .
- In equilibrium, a change in x_ℓ in a single location affects the shares s_j at every location in the city.
- Example: Policy that encourages entry in location 1.
 - Direct effect of substitution from other locations.
 - Indirect equilibrium effect: that has the form of the waves generated by a water drop.

Data and Estimation

- Suppose that we have data from an industry (e.g., supermarkets) in a city (Toronto). We observe:

$$\text{Data} = \{x_\ell, n_\ell : \ell = 1, 2, \dots, L\}$$

We also know the potential number of entrants, N .

- Given these data, we can construct the shares: $s_\ell : \ell = 1, 2, \dots, L$, with:

$$s_\ell = \frac{n_\ell}{N} \quad \text{and} \quad s_0 = \frac{N - n_1 - \dots - n_L}{N}$$

- The logit model implies that, for locations with $n_\ell > 0$:

$$\ln \left(\frac{s_\ell}{s_0} \right) = x_\ell \beta + \sum_{b=1}^B \gamma_b N_\ell(b) + \xi_\ell$$

- This is a linear regression model with regressors $x_\ell, N_\ell(1), \dots, N_\ell(B)$, and error term ξ_ℓ .

Inconsistency of OLS

$$\ln \left(\frac{s_\ell}{s_0} \right) = x_\ell \beta + \sum_{b=1}^B \gamma_b N_\ell(b) + \xi_\ell$$

- Regressors $N_b(\ell)$ are endogenous: they are determined in the equilibrium of the model.
- $N_\ell(b)$ is correlated with ξ_ℓ . OLS estimator is biased.
- We expect: $\text{cov}(N_1(\ell), \xi_\ell) > 0$ and
 $\text{cov}(N_1(\ell), \xi_\ell) > \text{cov}(N_2(\ell), \xi_\ell) > \dots > \text{cov}(N_B(\ell), \xi_\ell)$
- This implies that OLS estimator of γ_1 is upward biased, and
 $\text{bias}(\gamma_1) > \text{bias}(\gamma_2) > \dots > \text{bias}(\gamma_B)$
- We might wrongly conclude that distance does not affect competition.
 Example: True γ 's: $\gamma_1 = -2$, $\gamma_2 = -1$, $\gamma_3 = -0.5$, and OLS estimates: $\gamma_1^{OLS} = -0.5$, $\gamma_2 = -0.5$, $\gamma_3 = -0.5$.

Instrumental variables estimation

$$\ln \left(\frac{s_\ell}{s_0} \right) = x_\ell \beta + \sum_{b=1}^B \gamma_b N_\ell(b) + \xi_\ell$$

- The model implies instruments for the endogenous regressors $N_\ell(b)$.
- Market characteristics x_j in locations other than ℓ do not enter in the equation for location ℓ but affect the equilibrium values $N_\ell(b)$.
- Let $\bar{x}_\ell(b)$ be the mean value of x_j in the those locations that belong to the band b around location ℓ :

$$\bar{x}_\ell(b) = \frac{\sum_{j=1}^L 1\{\text{location } j \text{ belongs to band } b \text{ around } \ell\} x_j}{\sum_{j=1}^L 1\{\text{location } j \text{ belongs to band } b \text{ around } \ell\}}$$

- We can use $\bar{x}_\ell(b)$ as an instrument for $N_\ell(b)$.

Entry and store location: Results

- Seim (2006) finds very significant results of spatial differentiation (γ parameters decline very significantly with distance)
- Market structure and spatial structure of stores under two different scenarios of city growth.
 - Growth in population but keeping city boundaries.
 - Growth in population and in city boundaries
- The model can be used to study how changes in the exogenous characteristics x_ℓ of a single location (e.g., new amenities, schools, new local regulations, transportation, developments) can affect the landscape of firms in the whole city.