HAUTE ÉCOLE
D'INGÉNIERIE ET DE GESTION
DU CANTON DE VAUD
www.heig-vd.ch

# Business location choices in the Paris region: modeling and estimating a static discrete game

Matthieu de Lapparent & Sabina Buczkowska

23 Juin 2017

Hes·so
Haute Ecole Spécialisée
de Suisse occidentale
University of Applied Sciences and Arts
Western Switzerland

# Motivation

- There is no economic agent acting in a vacuum. Everyone consider strategic reactions of other players as well
- Agents are processing observed choices or anticipating choices of others when making decisions
  - Consumers :
    - Location choices and spatial seggregation patterns
    - Social interactions with family/peers/others ?
  - Firms :
    - Market Entry and Spatial Competition (lot of applications in retail Industries), spatial competition with endogenous location choices
    - Airline, car industries
    - ...

## Context

- Interrelated nature of many decisions suggests modeling them as strategic games
- Discrete games : game theory $\cap$ discrete choice econometrics
  - No general solution : precise structure of the game clearly depends on the particular application.
    - # of players
    - **Static** or dynamic ?
    - **Discrete** or continuous or mixed decisions ?
    - Complete or incomplete or **mixed** information settings ?
    - Timing of moves : Games with **simultaneous** vs sequential moves

Bresnahan and Reiss (1991), Seim (2006), Draganska et al. (2008), Zhu and Singh (2009), Ellickson and Misra (2011, 2012), Bajari et al. (2010, 2013), Aguirregabiria and Mira (2007, 2010), Aguirregabiria et al. (2016)

# A simple question

Consider RUM discrete choice modeling framework.
Given a population of players $i = 1, \cdots, n$ faced with alternatives $m = 1, \cdots, M$, what happens when moving from

$$U_{i,m} = V\left(\mathbf{x}_{i,m}, \mathbf{z}_i\right) + \varepsilon_{i,m}$$

to

$$U_{i,m} = V\left(\mathbf{x}_{i,m}, \mathbf{z}_i, \mathbf{y}_{-i}\right) + \varepsilon_{i,m},$$

where $\mathbf{y}$ are observed choices in the population of players ?

## Starting points

- Formulation of payoff functions
- Informational context : complete vs incomplete information (Bayesian Nash games)
- Equilibrium selection
- Econometric methods

## Formulation of payoff functions : a standard approach

In case of perfect information, the indicators of other players choices are observed :

$$U_{i,m} = V\left(\mathbf{x}_{i,m}, \mathbf{z}_i, \boldsymbol{\beta}\right) + \sum_k \sum_{j \neq i} \alpha_{i,j,m,k} y_{j,k} + \varepsilon_{i,m}$$

In case of imperfect information, one has to model beliefs of players : the indicators of other players choices are then replaced by their expectations, i.e. the probabilities of such choices

$$U_{i,m} = V\left(\mathbf{x}_{i,m}, \mathbf{z}_i, \boldsymbol{\beta}\right) + \sum_k \sum_{j \neq i} \alpha_{i,j,m,k} \operatorname{Pr}_{i,k} + \varepsilon_{i,m}$$

where $\operatorname{Pr}_{i,k} \equiv \operatorname{Pr}_{i,k}\left(\mathbf{x}, \mathbf{z}, \mathbf{Pr}, \boldsymbol{\theta}\right)$, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$. Note roles of $\boldsymbol{\alpha}$.

## Logit model and Bayes Nash Equilibrium

- Map expected utilities (conditional on beliefs characterized by Pr) into (ex ante) choice probabilities
- A Nash equilibrium is a strategy profile in which each player's strategy is a best reply to the others' strategies

Assuming that $\forall i, m, \varepsilon_{i,m}$ are iid EV1(0,1) + RUM :

$$\mathrm{Pr}_{i,m}(\mathbf{x}, \mathbf{z}, \mathbf{Pr}, \boldsymbol{\theta}) = \frac{\exp\Big(V\big(\mathbf{x}_{i,m}, \mathbf{z}_i, \boldsymbol{\beta}\big) + \sum_k \sum_{j \neq i} \alpha_{i,j,m,k} \mathrm{Pr}_{i,k}\Big)}{\sum_{l=1}^{M} \exp\Big(V\big(\mathbf{x}_{i,l}, \mathbf{z}_i, \boldsymbol{\beta}\big) + \sum_k \sum_{j \neq i} \alpha_{i,j,l,k} \mathrm{Pr}_{i,k}\Big)}$$

## Maximization program

Given. sample of observations, the objective is to :

$$\begin{cases} \max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \sum_{m=1}^{M} y_{i,n} \ln\left(\mathrm{Pr}_{i,m}\left(\mathbf{x}, \mathbf{z}, \mathbf{Pr}, \boldsymbol{\theta}\right)\right) \\ \text{s.t.} \forall i, m, \mathrm{Pr}_{i,m}\left(\mathbf{x}, \mathbf{z}, \mathbf{Pr}, \boldsymbol{\theta}\right) = \frac{\exp\left(V\left(\mathbf{x}_{i,m}, \mathbf{z}_i, \boldsymbol{\beta}\right) + \sum_k \sum_{j \neq i} \alpha_{i,j,m,k} \mathrm{Pr}_{i,k}\right)}{\sum_{l=1}^{M} \exp\left(V\left(\mathbf{x}_{i,l}, \mathbf{z}_i, \boldsymbol{\beta}\right) + \sum_k \sum_{j \neq i} \alpha_{i,j,l,k} \mathrm{Pr}_{i,k}\right)} \end{cases}$$

# Estimation procedures

- Nested Fixed Point (NFXP) FIML (Rust, 1985) :
  - Start from candidate values for the parameters
  - Inner loop : solve the fixed point problem (it is contracting for MEV & mixtures of MEV discrete choice models : can be done by successive iterations) for these values
  - Outer loop : update values of the parameters
  - Goto Inner loop step

# Estimation procedures

- Nested Pseudo Likelihood estimation (Aguirregabiria & Mira, 2010)
  - Start from candidate values for the parameters and the choice probabilities
  - Update values of the parameters by maximizing the log-likelihood function
  - Update the choice probabilities using new values of the parameters
  - Goto Update values step

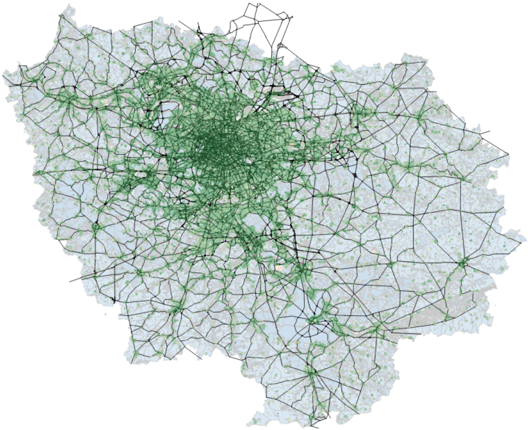
## Estimation procedures

- 2-step approach : Conditional Choice Probabilities + maximum likelihood (Aguirregabiria & Mira, 2007).
  - Eliminate the need to solve for a fixed point by recognizing that, at the "true" solution, the probabilities are simply (unknown) functions of the covariates.
  - Non- or semi- parametric estimation of "reduced form" choice probabilities ;
  - Plug them in the maximum likelihood estimation problem.
- Mathematical Programming with Equilibrium Constraints (Su & Judd, 2012)

## Equilibrium selection

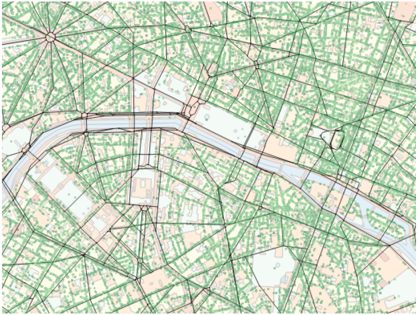
- Multiple equilibria are almost always present in incomplete information games.
- 4 main approaches to solve this this problem (Ellickson & Misra, 2011) :
  - aggregate to a different set of predictions which are robust to multiplicity (e.g. the number of players)
  - place restrictions on the model which guarantee a unique prediction (e.g. sequential moves),
  - specify a collective equilibrium selection rule (e.g. the equilibrium maximizes joint profits),
  - embrace the ambiguity and adopt a bounds approach

# Application : location choices of new establishments in the Paris region



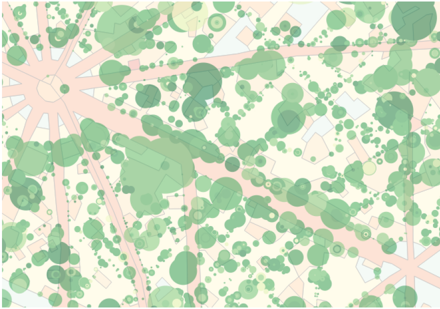
# Application

- Interactions in location choices of new establishments in the Paris region
- Focus on newly created establishments in 2006
- Location choices conditional to already existing establishments



# Data

- 2006 Census of establishments
- 1999 & 2006 Census of population
- 1980-2008 Land use survey
- Regional road and PT traffic model
- Land prices, real estate prices and rents
- 2001 & 2010 travel surveys



# Prototype model results

Table – Selected 7 types of newly-created establishments

| Establishment type | #New |
|---|---|
| Type 1 : Manufacturing | 3 296 |
| Type 2 : Retail | 10 899 |
| Type 3 : Wholesale | 8 572 |
| Type 4 : Transport, storage | 3 072 |
| Type 5 : Financial activities | 4 446 |
| Type 6 : Hotels, restaurants | 3 524 |
| Type 7 : Professional, scientific and technical activities | 15 282 |
| Other newly-created establishments | 38 883 |
| Total | 87 974 |

763 131 pre-existing establishments distributed across these types

# Formulation of the estimation problem : payoff functions

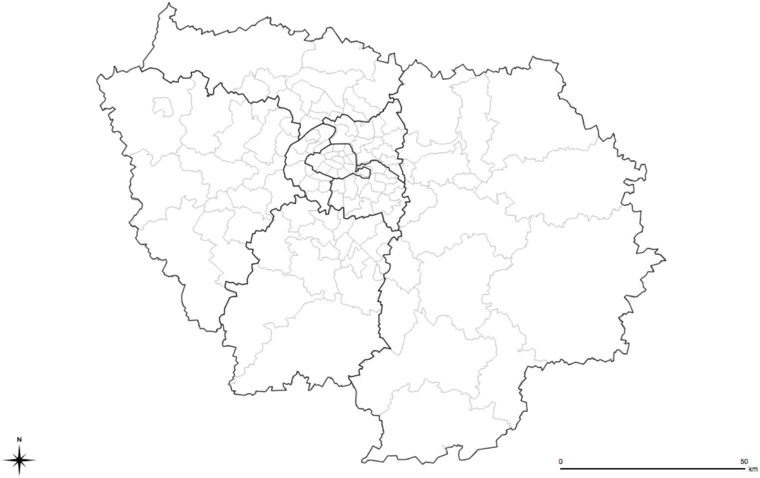
An establishment $i$ from sector $s$ locating at $l$ is endowed with the following expected profit function :

$$\pi_{s,i_s,l} = \mathbf{x}'_{s,l}\boldsymbol{\theta} + \sum_m \sum_k \sum_{j_k \neq i_s} \alpha_{s,k,l,m} \mathbb{E}\left(\mathbb{I}\left(y_{j_k,m} = 1\right)\right) + \xi_{s,l} + \varepsilon_{s,i_s,l}.$$

## Further assumptions

- Profit-maximizing establishments + "private shocks" $\varepsilon$ iid EV1 distributed
- Market unobservables are not correlated across sectors and locations
- Lack of variability in data :
    - Potential locations are tracts with available floorspace : finite discrete choice sets with $\leq 109$ alternatives
    - Within-group homogeneity : symmetric / exchangeable players when same industrial sector
    - Simultaneous moves of players : multiple equilibria even with imperfect information
    - Interaction terms $\alpha_{s,k,l,m} \equiv \alpha_{s,k}$

# Map of zones



## Bayes Nash Equilibrium : mixed Logit best response probability functions

$$\forall s, l, \Pr_{s,l}\left(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma} \middle| \mathbf{n}, \mathbf{x}, \mathbf{Pr}\right) =$$
$$\int_{\mathbb{D}(\boldsymbol{\xi}_s)} \frac{\exp\left(\mathbf{x}'_{s,l}\boldsymbol{\theta} + (n_s - 1)\alpha_{s,s}\Pr_{s,l} + \sum_{k \neq s} n_k \alpha_{s,k} \Pr_{k,l} + \xi_{s,l}\right)}{\sum_{m=1}^{L} \exp\left(\mathbf{x}'_{s,m}\boldsymbol{\theta} + (n_s - 1)\alpha_{s,s}\Pr_{s,m} + \sum_{k \neq s} n_k \alpha_{s,k} \Pr_{k,m} + \xi_{s,m}\right)} f\left(\boldsymbol{\xi}_s \middle| \boldsymbol{\sigma}\right) \mathrm{d}\boldsymbol{\xi}_s,$$

where $\Pr_{s,l} \equiv \Pr_{s,l}\left(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma} \middle| \mathbf{n}, \mathbf{x}, \mathbf{Pr}\right)$.

## Estimation of parameters

Aggregating observed locations of establishments by type, $d_{s,1}, \cdots, d_{s,L}, \sum_l d_{s,l} = n_s, \forall s = 1, \cdots, S$, the log-likelihood function is maximized wrt parameters $\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}$ subject to the fixed point problem :

$$\max_{\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\sigma}} \sum_s \sum_l d_{s,l} \ln\left(\mathrm{Pr}_{s,l}\left(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma} \middle| \mathbf{n}, \mathbf{x}, \mathbf{Pr}\right)\right)$$
$$\text{s.t. } \forall s, l, \mathrm{Pr}_{s,l}\left(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\sigma} \middle| \mathbf{n}, \mathbf{x}, \mathbf{Pr}\right) =$$
$$\int_{\mathbb{D}(\boldsymbol{\xi}_s)} \frac{\exp\left(\mathbf{x}'_{s,l}\boldsymbol{\theta} + (n_s - 1)\alpha_{s,s}\,\mathrm{Pr}_{s,l} + \sum_{k\neq s} n_k \alpha_{s,k}\,\mathrm{Pr}_{k,l} + \xi_{s,l}\right)}{\sum_{m=1}^{L} \exp\left(\mathbf{x}'_{s,m}\boldsymbol{\theta} + (n_s - 1)\alpha_{s,s}\,\mathrm{Pr}_{s,m} + \sum_{k\neq s} n_k \alpha_{s,k}\,\mathrm{Pr}_{k,m} + \xi_{s,m}\right)} f\left(\boldsymbol{\xi}_s \middle| \boldsymbol{\sigma}\right) \mathrm{d}\boldsymbol{\xi}_s.$$

## Estimation issues

- Choice probabilities approximated by MC integration
- Endogeneity : real estate rents in $\mathbf{x}$ correlated with market unobservables $\xi$ : IV approach
- Multiple BNE : since we have available exhaustive census of newly created establishments, we observe the target spatial equilibrium by industrial sectors, which we use in an initial NPL step before running NFXP estimation

# Model estimates, I

| | Retail | | HoRes | | Finan | | ProSci | | Whole | | TranSt | | Manuf | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Est | t-st | Est | t-st | Est | t-st | Est | t-st | Est | t-st | Est | t-st | Est | t-st |
| Paris | 0.0012 | 0.36 | -0.0003 | -0.04 | 0.3662 | 90.19 | 0.1673 | 36.06 | 0.1369 | 293.58 | 0.3098 | 50.50 | -0.0680 | -13.39 |
| La Défense | 0.0904 | 11.68 | 0.1160 | 86.65 | 0.3814 | 432.02 | 0.0284 | 8.14 | -0.0135 | -3.77 | 0.0908 | 34.37 | 0.2694 | 288.35 |
| New Cities (excl. La Défense) | 0.0592 | 11.02 | -0.0658 | -28.05 | 0.0433 | 18.57 | -0.1664 | -40.99 | 0.0747 | 34.56 | 0.2056 | 82.67 | 0.1246 | 160.80 |
| Outer suburbs (excl. New Cities) | 0.0370 | 5.55 | -0.0393 | -30.88 | -0.0456 | -13.52 | -0.0353 | -9.24 | -0.0120 | -3.33 | 0.0396 | 8.49 | 0.1572 | 48.41 |
| Zone's surface (log) | 0.1659 | 22.37 | 0.4644 | 895.97 | 0.1592 | 18.29 | 0.1554 | 81.79 | 0.3782 | 81.02 | 0.2797 | 58.36 | 0.4943 | 140.15 |
| Stock : The same estab. type | -0.0004 | -8.66 | -0.0001 | -1.28 | -0.0032 | -24.13 | -0.0005 | -41.31 | -0.0006 | -11.25 | ¡.0001 | 0.13 | -0.0005 | -2.87 |
| Stock : Commerce (G) | | | -0.0001 | -1.23 | | | | | | | -0.0003 | -4.02 | -0.0003 | -4.17 |
| Stock : Hotels, restaurants (I) | 0.0002 | 1.71 | | | | | | | 0.0008 | 8.15 | ¡-.0001 | -0.08 | 0.0006 | 9.62 |
| Stock : Finance, insurance (K) | | | | | | | -0.0001 | -0.84 | | | | | | |
| Stock : Pro., scien., tech. (M) | | | | | -0.0007 | -3.66 | | | | | | | | |
| Stock : Real estate (L) | | | | | 0.0018 | 4.75 | | | | | | | | |
| Stock : Manufacturing (C) | 0.0005 | 6.54 | | | | | | | 0.0008 | 8.05 | 0.0007 | 4.69 | | |
| Active density (labor force)/1000 | | | | | | | | | | | 0.2929 | 41.46 | 0.4722 | 262.70 |
| White-collar/manager (%tot emp) | | | | | 0.5029 | 90.90 | 0.5967 | 284.72 | | | | | | |
| Trips : Prof. meeting | | | | | 0.1549 | 33.71 | | | | | | | | |
| Trips : Prof. meeting/university | | | | | | | 0.0825 | 17.17 | | | | | | |
| Total pop. density/1000 | 0.1241 | 17.10 | 0.3855 | 246.60 | | | | | 0.3191 | 298.98 | | | | |
| Trips : Shopping purpose | 0.2557 | 187.95 | | | | | | | | | | | | |
| Trips : Restaurant visit | | | 0.0712 | 14.26 | | | | | | | | | | |
| Land : Shops (%zone) | 0.0426 | 16.82 | | | | | | | -0.0004 | -0.18 | | | | |
| Land : Industrial. economic act. | | | | | | | | | | | -0.0002 | -0.03 | -0.0184 | -2.78 |
| Land : Extraction of materials | | | | | | | | | 0.0099 | 5.85 | 0.0177 | 4.37 | 0.0098 | 23.50 |
| Road. rail terminal (%zone) | | | | | | | | | | | 0.0109 | 5.28 | | |
| Airport (%zone) | | | | | | | | | | | 0.0073 | 9.94 | | |
| Access to public transport (log) | 0.0371 | 16.73 | 0.0859 | 15.95 | 0.1410 | 21.16 | 0.1655 | 28.32 | 0.0136 | 3.63 | | | 0.0370 | 10.37 |
| Predicted rent (offices or shops) | -0.0387 | -18.10 | -0.1113 | -63.76 | -0.2468 | -117.43 | -0.2753 | -454.27 | -0.1560 | -64.55 | -0.0733 | -19.15 | -0.0322 | -6.07 |
| Instruments to predict rents [a] | Pop In,Co | Rev Le,Sp | Emp In,Co | Rev Le,Sp | Emp In,Co | Rev Le,Sp | Emp | Rev | Pop In,Co | Rev | Emp In,Co | Rev Le,Sp | Emp In,Co | Rev |

a. Proposed instruments : Pop/Emp : Population/Employment level in 1999 (log) ; Rev : Average net revenue per household in 1990 (log) ; In,Co/Le,Sp : Fraction of a zone's surface/100 dedicated to industry, commerce/leisure, sport facilities in 1990 (%).

# Model estimates, II

TABLE – Strategic interactions matrix

| | Retail | HoteRes | Finan | ProSci | Whole | TranSto | Manuf |
|---|---|---|---|---|---|---|---|
| Retail | 0.0067 [1] (15.22) | 0.0022 [2] (3.54 ) | -0.0004 (-1.92) | 0.0006 (5.42 ) | 0.0017 (5.38 ) | 0.0015 (3.15 ) | -0.0036 (-5.62) |
| HoteRes | -0.0006 (-0.61) | 0.0182 (36.62) | -0.0025 (-4.49) | 0.0006 (3.59 ) | 0.0005 (0.51 ) | 0.0003 (0.11 ) | 0.0024 (2.21 ) |
| Finan | 0.0005 (0.63 ) | -0.0114 (-5.96) | 0.0225 (29.03) | 0.0051 (4.69 ) | 0.0010 (2.04 ) | -0.0038 (-3.28) | 0.0031 (3.27 ) |
| ProSci | 0.0019 (6.61 ) | -0.0042 (-6.63) | -0.0057 (-10.58) | 0.0085 (107.56) | 0.0005 (2.84 ) | -0.0056 (-10.00) | -0.0019 (-3.66) |
| Whole | -0.0004 (-0.73) | 0.0013 (1.99 ) | -0.0009 (-2.10) | 0.0005 (3.48 ) | 0.0102 (26.49) | 0.00003 (0.06 ) | -0.0029 (-3.11) |
| TranSto | 0.0035 (5.93 ) | -0.0059 (-13.47) | 0.0023 (6.82 ) | 0.00002 (0.21 ) | -0.0021 (-6.59) | 0.0253 (18.54) | 0.0002 (0.15 ) |
| Manuf | -0.0021 (-2.79) | 0.0048 (5.04 ) | -0.0058 (-7.92) | 0.0015 (5.38 ) | 0.0029 (4.66 ) | 0.0020 (1.73 ) | 0.0206 (13.17) |

# Concluding remarks

- In NEIO, game theory is by far the most common tool used to model industries...
- ... but it can be applied to a very broad set of problems : labor, public finance, marketing, housing choices, driving behavior
- Calibration is very challenging and computationally intensive...
- ... but not accounting for strategic interaction or preselecting ad hoc values might strongly bias outcomes of the model if then used for simulation
- Often public data lacks good information to identify complex strategic behavior