Location Choice Equilibrium - pedestrian demand analysis at EPFL

Tianyang Dong

03.06.2019

Outline

Outline

Project Overview

This project aims at:

- derive a pedestrian location choice model on EPFL campus about catering places during lunch
- compute the aggregated demand at each location
- extend the model by taking congestion effect into consideration
- estimate the extended model using 2-step Pseudo Maximum Likelihood Estimation and Nested Pseudo Likelihood Estimation

Outline

Data Description

The original data are WiFi traces collected in 2012. It is then cleaned by Danalet et al.(2014) using a Bayesian procedure to detect activity-episode sequence.

The data used for modelling contains 715 observations.

Each observation corresponds to a catering location choice, including:

- choice results: chosen location, duration, ...
- attributes of alternatives: evaluation, capacity, ...
- socio-economic characteristics: employee or student, ...

Basic Model

$$
\begin{aligned}
V_{ni} = ASC_i &+ \beta_{PRICE\_STUDENT} \cdot lunch\_min\_price\_student_{ni} \\
&+ \beta_{DISTANCE} \cdot distance_{ni} + \beta_{EVA\_X} \cdot eva_i \\
&+ \beta_{DISTANCE\_NO\_AV} \cdot distance\_no\_av_{ni}
\end{aligned}
$$

$i \in C = \{1, 2, 4, 5, 6, 8, 9, ..., 21, OPTOUT\}$

Since alternative 3 and 7 do not serve meals during lunch, these two alternatives are taken as an OPTOUT choice.

Variable	Definition
$lunch\_min\_price\_student_{ni}$	MIN_PRICE_i if student
$distance_{ni}$	$DISTANCE_{ni}/100$ if $DISTANCE_{ni} > 0$
$distance\_no\_av_{ni}$	1 if $DISTANCE_{ni} = -1$
eva_i	$EVALUATION\_2013_i$ if > 0

Main Results

Number of estimated parameters	26
Sample size	715
Excluded observations	0
Init log likelihood	-2141.949
Final log likelihood	-1600.269

Description	Coeff. estimate	Std err	*t*-stat
BETA_DISTANCE	-0.617	0.0334	-18.5
BETA_EVA_CAFET	0.306	0.0536	5.7
BETA_EVA_CARA	0.367	0.0628	5.85
BETA_EVA_REST	0.245	0.0789	3.1
BETA_EVA_SELF	0.762	0.0601	12.7
BETA_NO_DISTANCE_AV	-3.49	0.278	-12.5
BETA_PRICE_STUDENT	-0.0494	0.0253	-1.95

Outline

Building-based Aggregation

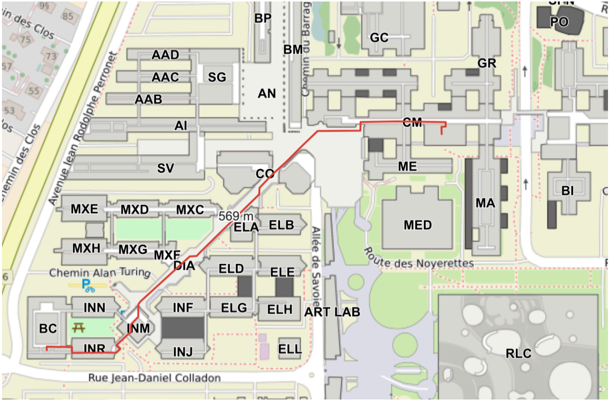
To aggregate the market share:

- the population is divided into two groups: student and employee
- suppose that the whole population will have lunch on campus
- suppose that in each building, everyone leaves from the same location
- the number of students leaving from one building is proportional to number of students subscribed to the classes
- the number of employees leaving from one building is proportional to the capacity of offices.
- ASCs are removed from utility function for prediction.

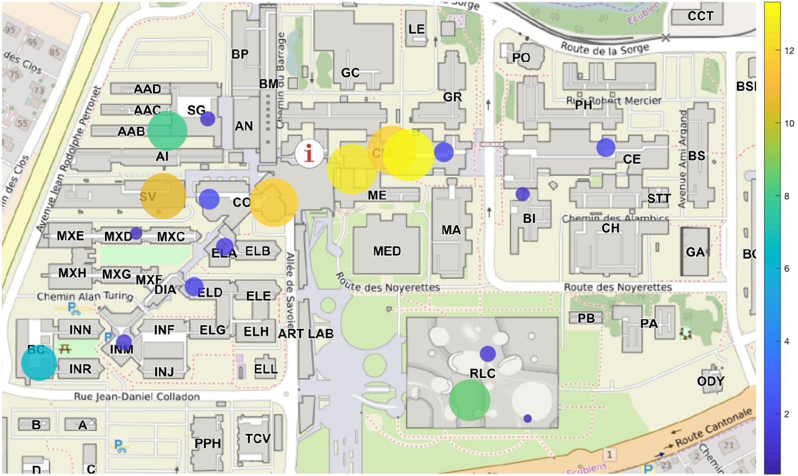
According to the figures of 2012, there are 7117 students and 5493 employees.

Computing distance to the restaurants

The following image gives an example of how to compute the distance between CM and Cafeteria BC using EPFL Géoportail.



Market Share of EPFL Catering Locations during Lunch



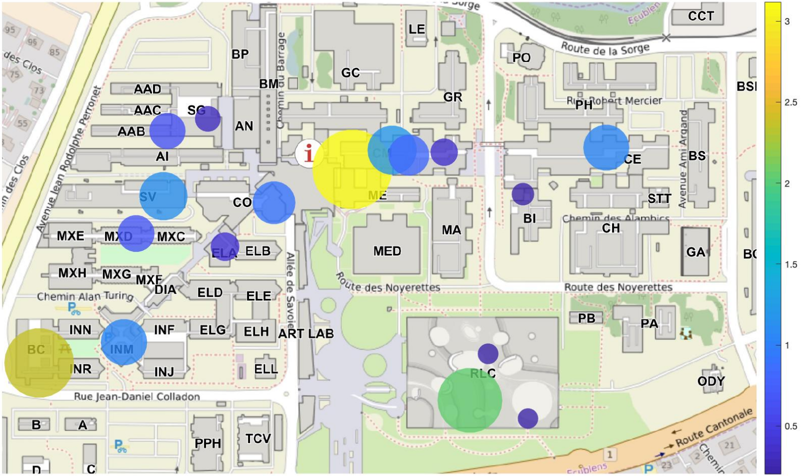
Definition of Congestion

When computing the congestion, the capacity of each catering location is taken into consideration.
T: total time during lunch (150min).
$\bar{t}_i$: average time individuals spent at each location.
c_i: indoor capacity at location i.
$d_i(\hat{W}_i)$: demand at location i, it can be computed using the predicted market share $\hat{W}_i$ and the size of the population on campus.

$$congestion_i = \frac{t_i d_i}{T c_i}$$

For those locations with $congestion_i > 1$, demand is larger than the capacity, leading to over congestion.

Congestion Condition of EPFL Catering Locations during Lunch



Outline

Model with Congestion Effect

$$U_{n.i} = V_{n,i} + \alpha f(\bar{p}_{n,i}) + \epsilon_{n,i}$$

N individuals, I alternatives:1, ..., I

$V_{n,i}$: deterministic part $\quad\quad$ $\epsilon_{n,i}$: error term.

$f(\bar{p}_{n,i})$: social utility.

$\bar{p}_{n,i}$: individual n's expected percentage of people choosing location i.

α: parameter to be estimated.

Under Logit model:

$$\mathbb{P}_n(i|\, V_{n,i}, \bar{p}_{n,i}, \forall i) = \frac{\exp(\, V_{n,i} + \alpha f(\bar{p}_{n,i}))}{\sum_{i=1}^{I} \exp(\, V_{n,i} + \alpha f(\bar{p}_{n,i}))}$$

Suppose that all individuals make rational expectations:

$\bar{p}_{n,i} = \bar{p}_i = \hat{W}(i)$

In our case: the social utility can be written as $f(\bar{p}_{n,i}) = congestion_i$

Estimation Methods

Different approaches have been used to estimate the model:

- Nested Fixed Point(Rust, 1985)
- 2-step Pseudo Maximum Likelihood Estimation(PML)(Aguirregabiria and Mira, 2007)
- Nested Pseudo Likelihood Estimation(NPL)(Aguirregabiria and Mira, 2007)
- Mathematical program with equilibrium constraints(Su and Judd, 2007)

Outline

PML

- Derive a base model without congestion effect.
- Predict the marketshare using building-based aggregation.
- Plug them in the model with congestion effect and estimate the parameters.

Description	Coeff. estimate	Std err	t-stat
ALPHA	-0.541	0.0933	-5.79
BETA_DISTANCE	-0.617	0.0334	-18.5
BETA_EVA_CAFET	0.301	0.0537	5.6
BETA_EVA_CARA	0.404	0.0619	6.54
BETA_EVA_REST	0.203	0.0806	2.52
BETA_EVA_SELF	0.793	0.06	13.2
BETA_NO_DISTANCE_AV	-3.49	0.278	-12.5
BETA_PRICE_STUDENT	-0.0494	0.0253	-1.95

PML drawbacks

- **asymptotically insufficient** because its asymptotic variance depends on the variance of the initial non-parametric estimator
- initial estimator may be **imprecise** in the small samples, this can generate serious finite sample biases.
- under certain circumstances, a **consistent estimator is not feasible**.

NPL addresses these drawbacks.

NPL

1. Start from candidate values for the parameters and market share
2. Update values of the parameters by maximizing the log-likelihood function
3. Update the choice probabilities using new values of the parameters
4. If not converge, go to to step 2.

NPL with sample-based aggregation

Number of estimated parameters	27
Sample size	715
Excluded observations	0
Init log likelihood	-2141.949
Final log likelihood	-1600.269

Description	Coeff. estimate	Std err	*t*-stat
ALPHA	1.1	0.127	8.64
BETA_DISTANCE	-0.617	0.0334	-18.5
BETA_EVA_CAFET	0.0629	0.0635	0.991
BETA_EVA_CARA	0.138	0.0629	2.19
BETA_EVA_REST	0.176	0.0782	2.25
BETA_EVA_SELF	0.45	0.0618	7.27
BETA_NO_DISTANCE_AV	-3.48	0.278	-12.5
BETA_PRICE_STUDENT	-0.0494	0.0253	-1.95

Remark

- The sign of α is opposed to our expectation.

Potential reason

- The dataset used is biased.
- The congestion defined above might actually indicate the popularity of the alternatives.

Remedy

- Collect additional information for market share prediction. Use building-based aggregation to compute the market share.

NPL with building-based aggregation

Number of estimated parameters	27
Sample size	715
Excluded observations	0
Init log likelihood	-2141.949
Final log likelihood	-1600.269

Description	Coeff. estimate	Std err	t-stat
ALPHA	1.0	0.137	7.28
BETA_DISTANCE	-0.617	0.0334	-18.5
BETA_EVA_CAFET	0.282	0.0542	5.2
BETA_EVA_CARA	0.37	0.0628	5.89
BETA_EVA_REST	0.377	0.084	4.49
BETA_EVA_SELF	0.727	0.059	12.3
BETA_NO_DISTANCE_AV	-3.49	0.278	-12.5
BETA_PRICE_STUDENT	-0.0494	0.0253	-1.95

Remark

- The sign of α stays unchanged.
- The predicted market share is still somehow biased. (Market share of Le Vinci is 0.4%, opposed to our knowledge)

Potential Reason

- Some ASC has dominant influence in the utility function, compared with parameters estimated above.

Description	Value	t-stat
ASC_VIN	-2.74	-4.31
ASC_ARC	-1.41	7.28
ASC_ATL	-1.48	-5.1
ASC_GIA	0.866	4.82

Remedy

- Exclude the influence of ASC from the utility function used for prediction.

NPL with building-based aggregation without ASC

The convergence can not be reached when ASCs are excluded.

Iteration	1	2	3	4	5	6
α	-0.541	-0.0865	-0.554	-0.0591	-0.55	-0.066
Iteration	7	8	9	10	11	...
α	-0.532	-0.0136	-0.538	-0.0269	-0.539	...

Potential reason

Convergence condition of NPL (Aguirregabiria and Mira, 2002):
Under the assumption that

- the utility function $U_{n,i}$
- the distribution of error term p_ϵ
- the choice probability $\mathbb{P}_n(i|\, V_{n,i}, \bar{p}_{n,i}, \forall i)$

is continuous and twice differentiable, the probability that NPL converges locally goes to 1 as the sample size increases.

Remedy

- Use a larger data set for estimation.
- Try a different way to reduce the influence of ASC: fix the ASC so that other parameters can capture more features.

NPL with fixed ASC

Number of estimated parameters	8
Sample size	715
Excluded observations	0
Init log likelihood	-2212.942
Final log likelihood	-1600.27

Description	Coeff. estimate	Std err	t-stat
ALPHA	-0.000463	0.785	-0.00589
BETA_DISTANCE	-0.617	0.0334	-18.5
BETA_EVA_CAFET	0.306	0.0572	5.34
BETA_EVA_CARA	0.368	0.0647	5.68
BETA_EVA_REST	0.245	0.0813	3.01
BETA_EVA_SELF	0.762	0.0598	12.7
BETA_NO_DISTANCE_AV	-3.49	0.275	-12.7
BETA_PRICE_STUDENT	-0.0494	0.0231	-2.14

Remark:α's value is small and insignificant.

Work could be done in the future

- Estimate with a larger dataset.
- Derive a model that captures the features better, for example, nested logit model.

Conclusions

- 2-step PML has a small computational cost, yet it has a high demand of the consistency of the choice probability estimator.
- NPL addresses the drawback above with a rather small additional computational cost, yet the convergence is not guaranteed when the sample size is small.