Master Thesis

École Polytechnique Fédérale de Lausanne
School of Architecture, Civil and Environmental Engineering
Transport and Mobility Laboratory

Exploiting pedestrian WiFi traces for destination choice modeling

by

Loïc Tinguely

Under the supervision of Michel Bierlaire,
the assistance of Antonin Danalet & Matthieu de Lapparent
and the expertise of Andres Sevtsuk

June 2015

It is a genuine pleasure to express my sincere appreciation to Antonin Danalet and Matthieu de Lapparent for their availability, knowledge and advises, to Michel Bierlaire for his helpful feedback and, to my parents, Paul and Pierrette for their unconditional support during my studies.

Figure 1: Pedestrians in Cornavin's Railway Station (Geneva, Switzerland)

Abstract

My master thesis proposes a general methodology to model pedestrian destination choice from WiFi localization in multi-modal transport facilities (e.g., airports, railway stations). It is based on the output of Danalet *et al.* (2014) method to generate candidates of activity-episode sequences from WiFi measurements, locations of activities on a map and prior information. Destination choice is nested to the activity choice. An individual first chooses an activity (Danalet and Bierlaire, 2015b), and then selects the destination where to perform it. We propose an approach to model destination choice accounting for panel nature of data. We compare static, dynamic strict exogenous and dynamic model with two different agent effect corrections inspired by Wooldridge (2002) method.

In a case study using WiFi traces on EPFL campus, we focus on one activity type: catering. The choice set contains 21 alternatives on campus (restaurants, self-services, cafeterias,...). Our models reveal that the choice of a catering facility depends mostly on habits (e.g., where an individual ate the previous time), distance to walk from the previous activity-episode (calculated with a weighted shortest path algorithm) but less on destination specific determinants (e.g., price, capacity). The models are successfully validated using the same WiFi dataset and we forecast possible changes concerning catering destinations on the campus.

Contents

List of Figures

List of Tables

1 Introduction

My master thesis proposes a framework to model pedestrian demand in multi-modal transport hubs such as airports or train stations from WiFi localization. The use of those facilities increases, both for trains (425 million passengers in 2008 in Switzerland, 477 million in 2013, +12% (OFS, 2014)) and for planes (2 billion passengers in 2005 in the world, 3 billion in 2013, +50% (Worldbank, 2014)). Historically, paper-and-pencil or telephone surveys were conducted to collect data (information on people behavior and habits). They were expensive and could not be performed often. Nowadays, modern hubs all propose “free Wi-Fi”. Localization data from cell phones, tablets and computers can thus be collected from access points all around these hubs. These data are cheap and can cover the whole facility. Their quality is however poor.

Recently, Danalet *et al.* (2014) developed a methodology to deal with the quality of these data. Each measurement is associated to a point of interest (e.g., coffee shop, restaurant, ticket machine,. . .) in time using a measurement equation (Frejinger, 2008). Knowing people location in time permits to generate probabilistic candidates of activity episode sequences. They can be utilized to develop both an activity choice model (Danalet and Bierlaire, 2015b) and a destination choice model. These two models are sequential. Liu *et al.* (2014) suggest that they are explored together. Once an individual has chosen an activity, he selects the destination where to perform it (Bierlaire and Robin, 2009). This report discusses the second step of the sequence. It especially focuses on the development of a general methodology to describe and understand destination choice for pedestrians in multi-modal transport hubs. We propose to model and forecast people behavior based on WiFi data. It aims at optimizing these facilities (e.g., finding the optimal location for ticket machines or predicting the potential market shares of a new coffee shop).

To be more specific, the project is part of a collaboration between “Ecole Polytechnique Fédérale de Lausanne” (EPFL) and the Swiss Railway company (SBB) in the context of the project “Léman 2030” (CFF, 2014). It includes an increase in trains offer (100’000 travelers expected in 2030, 50’000 in 2013, +100%) and huge changes in Geneva and Lausanne train stations. The SBB is one of the most important property owners in Switzerland. Their lands have become a major source of income. RailCity is the name given to the largest railway stations because of their similarities to cities: more and more, train stations offer the opportunity not only to travel but also to eat, drink, shop, or entertain oneself. In 2009, the revenue of these infrastructures reached about 1.09 billion of Swiss francs (CFF, 2011).

In order to increase its revenues, the company wants to know how people behave when they visit the facility. A random pedestrian for example enters the station between 7:45 AM (it is the beginning of his activity-episode sequence) and 7:46 AM, buys a ticket between 7:49 AM and 7:51 AM, gets a sandwich between 7:54 AM and 7:57 AM and then reachs the platforms at 8:01 AM to take a train scheduled at 8:04 AM (it is the end of his activity-episode sequence). Once the activity is chosen (i.e., buying a ticket or getting a sandwich), a specific destination needs to be chosen (i.e., a specific ticket machine or a specific luncheonette). He performs a destination choice. As soon as the destination is known, a path needs to be defined. These nested episodes represent pedestrians tactical and strategical behavior (Hoogendoorn *et al.*, 2002). Similar studies have already been made for a destination choice model in railway stations (e.g., Ton (2014); Liu (2013)) or airports (e.g., Kalakou and Moura (2014b)). However, only Ton (2014) is developed from WiFi traces. The others are based on stated/revealed preference surveys. Pettersson (2011) also studies the behavior of pedestrians in train stations but he focuses on the factors (e.g., information, geometry, habits) influencing the waiting location of people on departure platforms.

The main goal of my thesis is to develop a general framework to model destination choice and to apply it to the EPFL campus. Indeed, WiFi traces from April 2012 to June 2012 are available (Danalet, 2015) and allow to track random employees and students on the campus and to define their probabilistic activity-episode sequences. Individuals may have multiple sequences over the observation period. A way to account for this panel nature of data to model individuals' habits is to be developed (i.e., according to our researches, agent effect correction for pedestrian's panel data has never been explored before). Also, the paper aims at proposing several variants of destination choice models on eating establishments (e.g., restaurants, self-services, cafeterias. . .) in order to explain the factors that influence pedestrians' decisions. The validation of these models in order to perform some forecasting on the campus (e.g., increase of prices and opening of a new establishment) is the final goal.

We review the existing literature concerning WiFi localization, pedestrians, activity and destination choices in Section 2. One explains the general equation, the data requirement, the data merging and methodologies in Section 3. After that methodological part comes a case study on the EPFL campus. Indeed, one applies the developed methodology and discuss several variants of destination choice models for pedestrians in Section 4. It includes a validation and forecasting. Recommendations for a future application in multi-modal transport hubs are available in Section 5. The conclusions of the thesis and future work are presented in Section 6.

2 Literature review

The literature review is separated into three parts. Section 2.1 explores methods to capture indoor pedestrians activities. In Section 2.2, existing destination choice models in multi-modal transport hubs are analyzed. Finally, Section 2.3 discusses how we are able to utilize this information.

2.1 Activity-episodes detection

In Ton (2014), WiFi and Bluetooth traces are involved. The methodology to transform signatures into activity sequences is not revealed. It is shortly discussed in Section 2.2. In Danalet *et al.* (2014), data requirement consists of timestamps and localization data coming from WiFi network traces and a semantically-enriched routing graph (SERG). A measurement is defined as:

$$\hat{m} = (\hat{x}, \hat{t}) \tag{1}$$

where $\hat{x} \in \mathbb{R}^3$ is the position of the measurement and $\hat{t}$ is the timestamps. The accuracy ξ defines the distribution of the Euclidean distance between location estimate $\hat{x}$ and actual location å:

$$\hat{x} = \text{å} + \xi \tag{2}$$

In order to associate activity-episodes (including stop detection and semantics of the stop) to these measurements, a semantically enriched routing graph (Goetz and Zipf, 2011) is defined as a set of nodes corresponding to the type of potential destinations i.e., room, restaurant, shop..., i.e., all Points Of Interest (*POI*).

The methodology to detect candidates of activity-episode sequences performed by pedestrians from the previously calculated digital traces follows a Bayesian approach explained in Danalet *et al.* (2014). An activity-episode is defined as

$$a = (x, t^-, t^+) \tag{3}$$

where x is the episode localization (or destination) and $t^+ - t^- \geq T_{min}$ the time spent at that location. A minimum threshold T_{min} of five minutes is set like in Bekhor *et al.* (2013). It permits to only keep activity-episodes longer than five minutes (and thus representing a destination and not only a crossing point). The output of this probabilistic method is defined as a set of L candidates of activity-episode sequences $a_{1:K_i}$ which are specific to an individual i. Basically an activity-episode sequence is a list of K_i activity-episodes performed (i.e., visited *POI*) by one tracked individual i during one day. Each candidate activity-episode sequence is associated with the probability of being the actual one. Danalet *et al.* (2014) define this probability as a Bayes formula (the subscript i is omitted to lighten the expressions):

$$P(a_{1:K}|\hat{m}_{1:J}) \propto P(\hat{m}_{1:J}|a_{1:K}) \cdot P(a_{1:K}) \tag{4}$$

It means that the activity probability $P(a_{1:K}|\hat{m}_{1:J})$ that $a_{1:K}$ is the actual activity-episodes sequence given the measurement $\hat{m}_{1:J}$ is proportional to the product of the measurement likelihood $P(\hat{m}_{1:J}|a_{1:K})$ with a prior knowledge $P(a_{1:K})$. As the goal is to compute the probability that the performed episodes generated the observed measurement sequence, the equation is decomposed:

$$P(\hat{m}_{1:J}|a_{1:K}) = \prod_{k=1}^{K} \prod_{j=1}^{J} P(\hat{x}_j^k|x_k) \tag{5}$$

It is assumed that the only measurement error is a localization error. Similar to the land use planning concept (Miller, 2010), the prior $P(a_{1:K})$ is defined based on a potential attractivity measure $S_{x,i}(t^-, t^+)$:

$$S_{x,i}(t^-, t^+) = \int_{t=t^-}^{t^+} \delta_{x,i}(t) \cdot A_i(x, t)dt \tag{6}$$

The idea is that the potential attractivity measure $S_{x,i}(t^-, t^+)$ between a start time t^- and an end time t^+ for $x \in POI$ and individual i is time dependent. It depends on the instantaneous potential activity and a dummy variable δ for time-constraints (e.g., opening hours, schedules...). The attractivity $A_i(x, t)$ defines the potential of a place (e.g., number of seating places for restaurants or number of workers per room for an office) as explained in Danalet *et al.* (2014). Then the prior can be calculated as

$$P(a_{1,K}) = \prod_{k=1}^{K} \frac{S_{x_k,i}(t_k^-, t_k^+)}{\sum_{x \in POI} S_{x,i}(t_k^-, t_k^+)} \tag{7}$$

It assumes that consecutive activity-episodes are independent.

Danalet *et al.* (2014) propose an algorithm to merge data from localization and pedestrian SERG to get candidates of activity-episode sequences. The generation of activity-episode sequences is divided in four steps. The first one introduces the concept of domain of data relevance (DDR) introduced in Bierlaire and Frejinger (2008). The DDR defines a physical area where a probabilistic measurement location linked to a POI is relevant. For each measurement $\hat{m}_j$, all possible activity-episodes sequences are generated for each individual. It leads to a recursively built network.

The second step consists in generating activity-episodes start and end times as soon as a sequence of potential episode locations is defined. The idea is to compare two consecutive measurements $\hat{m}_j$ and $\hat{m}_{j+1}$. Their timestamps and positions define a trip between them and thus a travel time. In that way, considering a maximum walking speed and a shortest path algorithm between both positions, bounds can be determined for the earliest and the latest start time and the earliest and the latest end time. Start and end times are considered to be uniformly distributed between these two bounds.

Third, once the distribution is known for the start and end times of each activity-episode, the duration is estimated. activity-episodes with a lower bound smaller as T_{min} are rejected. The last part of the procedure is the sequence elimination procedure. As the number of paths in a network growth exponentially with the number of measurements, there is a need for selection. Candidates with small probabilities of occurrence are rejected. The complete algorithm is available in Danalet *et al.* (2014). The methodology has been tested and validated on EPFL campus.

In Dalumpines (2014), the data requirement consists in GPS data. A GIS-based episode reconstruction toolkit (GERT) automatically extracts activity-episodes from GPS data and derives information related to these episodes. This kit generates an input for route choice modeling. The methodologies of Danalet *et al.* (2014) and Dalumpines (2014) are similar but the latter classifies activity-episodes into different types using multinomial logit models. Also, the first one deals with small scale problems (e.g., a multimodal facility, a campus...) whereas the second one fits better on a much larger framework (e.g., a transportation network).

2.2 Destination choice models for pedestrians

2.2.1 Influence of Space Syntax

Purpose and distance are intuitive factors used to explain a destination choice. When it comes to a pedestrian destination choice model, more determinants have to be accounted for. Kalakou and Moura (2014a,b) study the influence of space syntax (SS). SS is a theory and a set of methods about space reflecting both the objectivity of space and the intuitive engagement with it (Hillier, 2005). Important characteristics about space are connectivity, integration and visibility. Connectivity is a factor that expresses the number of “neighbors” of each space. Integration is the relation of one space with all others. According to Zhang *et al.* (2012) visibility is one of the most influential factors in people’s behavior when moving in commercial areas. Ueno *et al.*

(2009) find out that the visibility, the number of turns and the distance affect pedestrians' path choice in railway stations.

2.2.2 An analogy with route choice modelling

Hoogendoorn and Bovy (2004) distinguish three levels of choices: the strategic level (Activity pattern choice and departure time choice), the tactical level (Activity scheduling, destination choice and route choice) and the operation level (walking behavior). An activity may be performed at multiple destinations. Also, Daamen (2004) considers the tactical level and in particular the prevailing conditions of the network (travel time to reach each destination, queues…) to model destination choice.

According to Hoogendoorn and Bovy (2004), the choice of an activity area is based on factors such as the directness (number of sharp turns and rapid directional changes (Helbing, 1997)), the distance and the level-of-service of the route, the necessity of performing that activity (e.g., is it mandatory?) and personal preferences. Furthermore, the choice of a route and the choice of a destination are done simultaneously thus factors influencing both choices are considered. This point may be subject to discussion. Unlike Hoogendoorn and Bovy (2004), Daamen (2004) considers that the choice of an activity precedes the choice of a destination. We may contemplate both approaches.

2.2.3 Destination choice models in airports

Kalakou and Moura (2014b) made a survey in Lisbon Portela's airport and collected information about space syntax and travelers' habits. A discrete choice model was built to capture the significant parameters that influence the choice of a destination. Four coffee shops were selected as potential destinations. Space syntax parameters were introduced in the model. Visibility from a mandatory place to visit (check-in, entrance) has a significant impact on the choice of a destination. The integration level of the activity location adds value to a place for passengers who only choose one coffee shop. Similarly, places having a good connectivity are more likely chosen after the check-in.

Liu (2013) also studies pedestrian behavior in an airport on the basis of both revealed and stated preference survey data. She develops an activity-destination choice model. Traveled distance, congestion or the type of service have a significant impact on people's decisions. Models validated by Liu (2013) are used for forecasting: in more than 50% of the cases, the prediction fits the observation.

2.2.4 Destination choice models in railway stations

Ton (2014) studies the route and activity location choice behavior of departing pedestrians in the Utrecht railway station in Netherlands. Using WiFi and Bluetooth traces, she builds both destination and path choice models. Her work is based on a framework proposed by Hoogendoorn *et al.* (2002). It focuses on the strategical and tactical levels when faced with discrete choices in a train station.

Ton (2014) defines an activity as a punch. The movement of a pedestrian contains several punches (e.g., enter the station, visit a Burger King, leave via platform…). Therefore the possible activities are caught in a punch card. However, this list only tells if a pedestrian was seen at one place or not (binary observation). It means that the sequence cannot be directly derived from the punch card. Thus, the activity sequence must be determined. Ton (2014) does not explain how she defines the chronological order of the punches. One limitation of the data is that the list of activities performed by an individual is only available for one day because everyone receives a new identification number everyday to respect privacy.

Using these activity sequences, Ton (2014) applies a binary discrete model to a choice of a coffeeshop. Two Starbucks are selling coffees in the railway station and the aim is to capture the factors that influence pedestrian destination choice. Traveled time from entrance to coffeeshop, total distance covered and having to take a detour are robust parameters. It is interesting to note that the orientation is also significant. The fact that a coffeeshop is located on the right hand side of the railway station (from the main entrance) increases its utility because pedestrians are used to walk on the right.

2.3 Critics and comments

We discuss how we are able to account for the ideas developed in reviewed literature:

- Liu (2013) and Kalakou and Moura (2014b) are based on both SP and RP surveys.
  - Socio-economic parameters can easily be taken into account with surveys, not with WiFi traces since the data are partially anonymized.
- Kalakou and Moura (2014b), Liu (2013) and Ton (2014) destination choice models were developed for destinations in only one building.
  - Impact of SS in larger facilities (e.g., a campus) is unknown.
- Liu (2013) and Daamen (2004) emphasize the impact of congestion in destination choice but using different approaches (CCTV and tactical-network models respectively).

  - We have to find other indicators to consider congestion.
- The methodology developed by Ton (2014) is limited because the route choice is dependent of the punch card's simplicity and does not measure pedestrian's habits.
  - WiFi traces from Danalet *et al.* (2014) seem more relevant, and are thus used.
- Factors such as directness (Hoogendoorn and Bovy, 2004) or the works realized by Helbing (1997) and Ueno *et al.* (2009) are mainly discussing route choice.
- Alternative specific parameters (e.g., the price, the quality, the well-being, the comfort or aesthetic indicators) are barely described in pedestrian activity/destination choice's context but have an influence on people's decision making (Deutsch *et al.*, 2014).

The methodology developed by Danalet *et al.* (2014) is well fitted to create a destination choice model, but it has some limitations. The algorithm defined by Danalet *et al.* (2014) associates the WiFi measurements with *POI*[1] inside a zone. Points of interest are represented as points while they are areas in reality. It creates a problem when the accuracy of the measurement is good and the "zone of interest" is large. In this case, the point of interest might not be inside the domain of data relevance (*DDR*[2]). Thus, the actual point of interest, representing the possible activity performed by the receiver, might not be considered.

In the case of data collected with the Cisco Context Aware Mobility API with the Cisco Mobility Services Engine (MSE) (Cisco, 2011), the domain of data relevance is defined as a square around the measurement with sides of size $2 * cF$, where cF is called the confidence factor. The WiFi device is estimated to be in this square with 95% probability. The minimum observed cF is 16 meters (see Figure 2(c)) on EPFL campus. Some *POI* on campus clearly have a surface bigger than a $16 * 16$ square. In this case, the intersection between the *DDR* (i.e., the square with side $2 * cF$) and the point representing the *POI* might be empty, and so the actual activity-episode is not detected.

This limitation is observed in the case study (see Section 4 for a detailed description). The data collected in the library of the Rolex Learning Center (RLC) are good due to the lack of walls or obstacles and due to the large number of WiFi antennas[3](Sen *et al.*, 2013; Nandakumar *et al.*, 2012). Figure 2(b) shows that the level of accuracy in the library is higher than on the rest of the campus. Figure 2(c) show that some points of the library effectively lead to an empty intersection between *DDR* and the *POI*. It is a limitation of the methodology but it can be corrected by using an area instead of a point for representing *POI*.

[1]POI: Point Of Interest, see Section 2.1

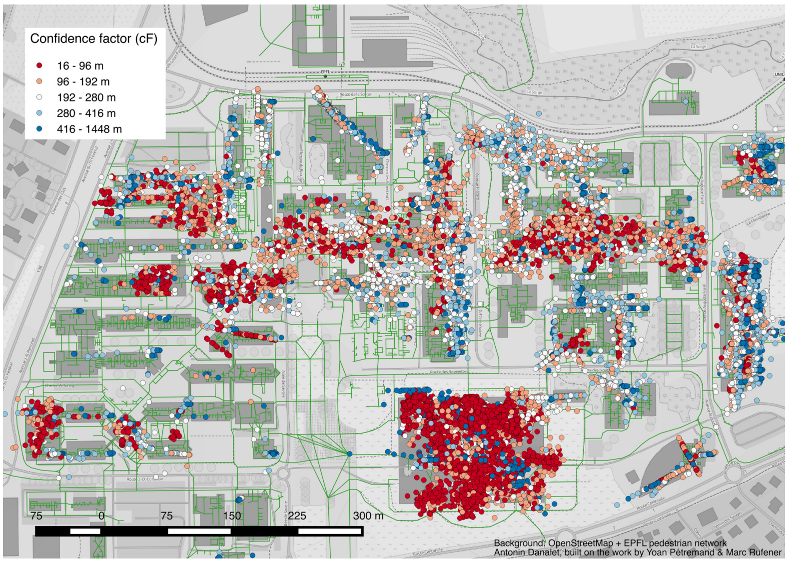
[2]Domain of Data Relevance, see Section 2.1

[3]On Figure 2(a), the density does not look higher than in other buildings on the campus. However, the RLC has only one floor, while all other buildings have more. This is just a visual effect due to projection on the map. In reality, density of WiFi access points is actually higher

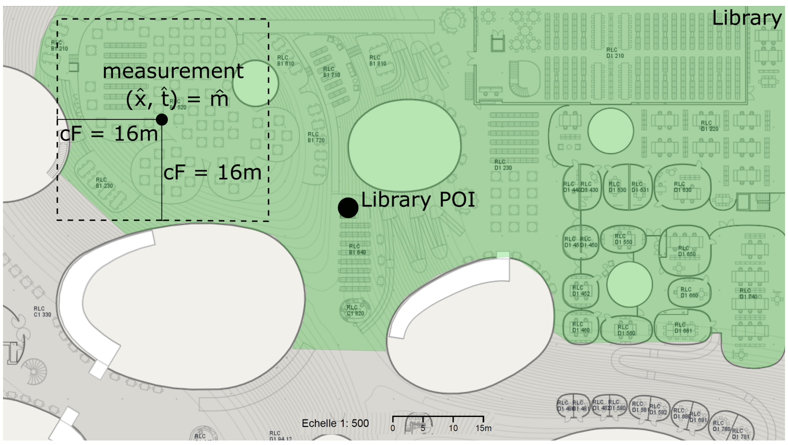
Figure 2: WiFi antennas and confidence factor (cF) on the EPFL campus

(a) WiFi antennas on the EPFL campus (`http://map.epfl.ch`)

(b) Confidence factor on the EPFL campus (Danalet, 2015)



(c) No intersection between DDR and POI (`http://map.epfl.ch`



3 Methodology

In Section 3.1, we focus on the general model which considers nested choices of activity and destination. Then, Section 3.2 describes the output of the algorithm developed by Danalet *et al.* (2014) and characterizes activity-episode sequences. Finally, Section 3.3 presents suggestions of destination choice model specification and an approach to account for the panel nature of data.

3.1 General model

The complete methodology developed by Danalet *et al.* (2014) aims at modelling activity-episode sequences. The choice of an activity-episode sequence is decomposed in two nested choices: first, the choice of an activity pattern, and then, conditional on the activity pattern, the choice of a destination. These decisions are not independent and we consider them together. The probability of matching a set of J measurements $\hat{m}_{1:J}$ (e.g., WiFi traces) for one individual i is (Danalet and Bierlaire, 2015a):

$$P_i(\hat{m}_{1:J}) = \sum_{a_{1:\Psi} \in C} P(\hat{m}_{1:J} | a_{1:\Psi}) \cdot P(a_{1:\Psi}), \tag{8}$$

where

- $P(\hat{m}_{1:J} | a_{1:\Psi})$ is the conditional probability that a set of measurements $\hat{m}_{1:J}$ is generated from an activity-episode sequence $a_{1:\Psi} = (a_1, \ldots, a_{\psi_i}, \ldots, a_{\Psi_i})$. A specific activity-episode, for an individual i, is $a_{\psi_i} = (x, t^-, t^+)$ (e.g., Café A, 12:05 PM, 13 PM). Frejinger (2008) and Chen (2013) call this term a measurement equation. It permits to take into account measurement errors due to the poor quality of WiFi localization, and
- $P(a_{1:\Psi})$ is the probability of performing an activity-episode sequence $a_{1:\Psi}$ (see Section 2.1).

We develop Equation (8) to decompose the choice of an activity-episode sequence $a_{1:\Psi}$ between the choice of an activity pattern $A_{1:\Psi}$ and the choice of a destination x:

$$P_i(\hat{m}_{1:J}) = \sum_{a_{1:\Psi} \in C} \prod_{j=1}^{J} P(\hat{x}_j | x_{\psi}^{j}) \cdot P(A_{1:\Psi}) \cdot \prod_{\psi=1}^{\Psi} P(x | A_{1:\Psi}), \tag{9}$$

where

- $\prod_{j=1}^{J} P(\hat{x}_j | x_{\psi}^{j})$, is the product of conditional probabilities that a measurement location is $\hat{x}_j$ knowing the episode location $x^j_\psi \in$ Point Of Interest (*POI*) (Danalet *et al.*, 2014) corresponding to measurement $\hat{m}_j$,

- $P(A_{1:\Psi})$ is the probability of performing an activity pattern $A_{1:\Psi} = (A_1, \ldots, A_\psi, \ldots, A_{\Psi_i})$ of Ψ activities. Also, an activity $A_\psi = (\mathcal{A}_k, t^-, t^+)$ (e.g., Café, 12:05 PM, 1 PM) is characterized by its start and end times and indexed by its type k, and
- $P(x|A_{1:\Psi})$ is the conditional probability of choosing a specific destination x for a specific activity-episode a_ψ knowing the complete activity pattern $A_{1:\Psi}$.

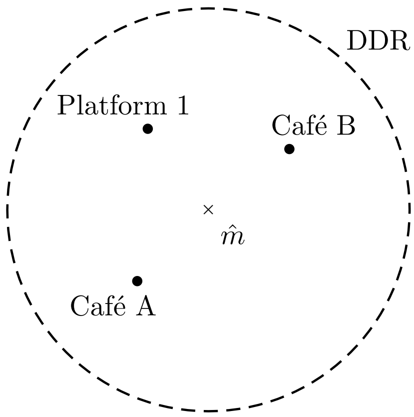
Basically, the probability of reproducing the observations depends on the destination choice which is nested to the activity choice and a measurement equation. The goal is to model the activity-episode sequence $a_{1:\Psi}$ when observing $\hat{m}_{1:J}$ from antennas. Here, we develop a framework to model destination choice considering that the activity is known.

Figure 4 shows a simple example. A measurement in a railway station has three equidistant possible *POI* (at distance d) inside the Domain of Data Relevance (DDR, see Section 2.1) defined by measurement $\hat{m}$. Café A and Café B are considered as cafés whereas the other one is a platform. According to Equation (9), the probability of observing the measurement is:

$$\begin{aligned} P(\hat{m}) \quad = \quad & \frac{1}{2\pi\sigma^2} e^{-\frac{d^2}{2\sigma^2}} \cdot \Big(P(\text{Café}) \cdot \Big(P(\text{Café A}|\text{Café}) + P(\text{Café B}|\text{Café}) \Big) \quad + \\ & P(\text{Platform}) \cdot P(\text{Platform 1}|\text{Platform}) \Big), \end{aligned} \tag{10}$$

so that the probability depends on two destination choice models (i.e., one for each activity type) which are nested with an activity choice model and a measurement equation. We assume that error in latitude and longitude are independently and normally distributed (Danalet *et al.*, 2014).

Figure 4: A measurement and its decision's alternatives in a railway station framework



3.2 Activity-episode sequences

Section 3.2.1 describes activity-episode sequences $a_{1:\Psi}$. Then, Section 3.2.2 shows how we can use them and Section 3.2.3 especially focuses on the distances' calculation.

3.2.1 Description of activity-episode sequences

Danalet *et al.* (2014) develop a framework for detecting pedestrian mobility pattern from WiFi traces (see Section 2.1). The methodology explained in the paper is used to create candidate's lists of activity-episode sequences from WiFi traces. They are then used to develop a destination choice model for pedestrians.

An activity-episode sequence $a_{1:\Psi}$ has several characteristics (sequence specific attributes). An example is described in Figure 5 and Table 1. Each sequence is associated to an individual (with a unique ID) tracked during one day and a probability of occurrence defined with its log-likelihood (i.e., from the measurement equation). Activity-episode sequences also contain several socio-economic (e.g., age, gender, or typology of visitor) and time specific attributes (e.g., the day of the week and year of the sequence). As sequences may be calculated during a period of several months, each individual has potentially more than one observation.

Within the sequence, there are one or more activity-episodes a_ψ. Each activity-episode is related to a point of interest (*POI*, see Section 2.1). It is described by its start and end times bounds (following a uniform distribution). Each *POI* associated to an activity-episode defines an activity A_ψ and a destination x. The activity is grouping destinations in categories. Typical categories, or activity types, are *working*, *eating*, *shopping*, etc. Destinations are more detailed. They have a name, coordinates and floor. Each type of destination is subject to an independent choice model.

Figure 5: An activity-episode sequence pattern in a campus framework

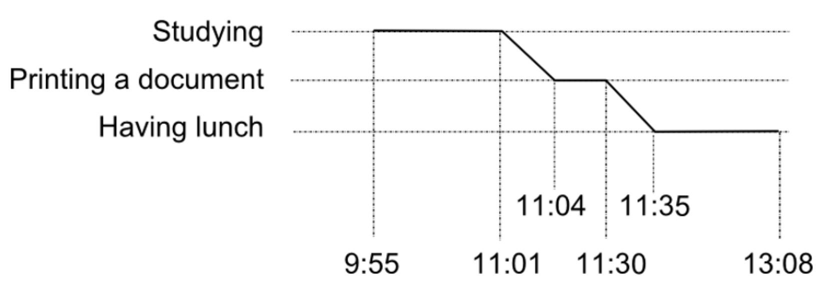


Table 1: This sequence measured from a second year bachelor student (ID=10001) in civil engineering contains 3 activity-episodes performed the June 29, 2012. This student has been seen 112 times by the Cisco WiFi device (only the destinations are kept). Each activity-episode is related to a point of interest. In that case the student first visited the library, then printed something (still in the library) and finally went to eat at the library's self-service. These sequences are the input of our methodology. Each activity-episode has an upper and lower bound for both start and end times (replaced by their mean on this figure). The probability of occurrence of the sequence is defined by its log-likelihood.

Nb of observations: 112, Date: 2012-06-29, loglikelihood = -125.5				
Start_time	End_time	Floor	Name	Type
09:55:01	11:01:30	1	Library_name	Library
11:04:39	11:30:03	1	Printer_Lib	Printer
11:35:23	13:08:04	1	Self-service_Lib	Restaurant

3.2.2 Characterization of activity-episode sequences

Each activity type corresponds to several possible destinations. For each destination, three types of attributes exist: sequence attributes (it corresponds to attributes specific to the whole one-day sequence), activity-episode attributes (it stands for attributes relative to one activity-episode only) and alternative attributes (they are the destination specific attributes, they need to be collected). They are defined and imaged with examples in Table 2.

Table 2: Table of attributes

Sequence attributes ($a_{1:\Psi}$)	Activity-episode attributes (a_ψ)	Destination attributes (x)
Day of the year	Activity-type	Capacity
Log-likelihood	Start/end times	Price/Quality
Socio-economic attributes	Coordinates	Integration
Distance	Floor	Opening hours

3.2.3 Calculating distances

By comparing two consecutive activity-episodes of a same activity-episode sequence, one can calculate the distances between the two destinations (and similarly for all elements in the choice set in a discrete choice context). There are two possibilities to calculate the distances. First, one can simply compute Euclidean distance between the consecutive activity-episodes of an activity-episode sequence $a_{\psi,t-1}$ and $a_{\psi,t}$ using the (x, y) coordinates of the points.

$$d(a_{\psi,t-1}, a_{\psi,t}) = \sqrt{(a_{\psi,t-1,x} - a_{\psi,t,x})^2 + (a_{\psi,t-1,y} - a_{\psi,t,y})^2}. \tag{11}$$

It means that we relax the assumption of anisotropy (Kim and Hespanha, 2003) and thus pedestrians can reach each point with a straight line path. A better way to calculate the distances is by using a shortest path algorithm. It may already have been constructed if the methodology explained by Danalet *et al.* (2014) has been strictly followed (indeed the path generation is based on it). It takes into account the network anisotropy and thus we obtain realistic distances.

3.3 Modelling

Section 3.3.1 proposes some recommendations for the modelling whereas Section 3.3.2 describes in detail agent effect correction.

3.3.1 A destination choice model

We develop a multinomial logit model with a linear-in-parameters formulation. The probability for an individual i of choosing a destination x compared to the other destinations $j \in \{1, \ldots, J\}$, $j \in POI$ knowing the activity type $\mathcal{A}_k$, and the activity pattern $A_{1:\Psi}$ is defined as:

$$P(x|A_{1:\Psi}) = \frac{e^{\mu V_{ix}}}{\sum_{j=1}^{J} e^{\mu V_{ij}}}. \tag{12}$$

We decompose the utility in three categories: sequence attributes, activity-episode attributes and destination attributes (see Table 2). Activity-episode sequences specific attributes are mainly represented by the distances between the consecutive performed activities. The distance parameter is generic when the destinations all offer the same type of offer (e.g., a same type of ticket machine). If the destinations studied are more heterogeneous, one uses alternative specific parameters (e.g., for eating establishments).

Furthermore we suggest to split the parameters depending on the period of the day if time of the day may change the purpose (e.g., one individual may visit a pub at 12 AM probably to have lunch but at 11 PM to drink beers) or the reason of the visit (e.g., one individual visits the same cafeteria every morning because it is the nearest destination but is disposed to change his lunch place everyday). We propose to call that daily seasonality.

Still from the sequence, socio-economic parameters are difficult to take into account because the data are usually partially anonymous. We suggest that the gender, the age and the type of visitor are collected and introduced in the model as dummy variables to alleviate the alternative specific constants.

The timestamps we propose to introduce in the distance calculation (see Section 3.2.3) are activity-episode specific parameters. There are only few factors from this category that can be added in the utility function. The activity type permits to select a specific type of activity and the destination x to perform the selected activity which represents the choice that the individual made. The floor of the destination is introduced in the case of a place without elevators.

Alternative specific parameters can be variables representing the congestion (capacity, queues), the quality/price ratio, the space syntax (visibility, integration, directness, detour), the type of services offered, aesthetics aspects, safety or the advertising (communication, information, directional sign). The case study presented in Section 4 gives an example in the context of catering destinations.

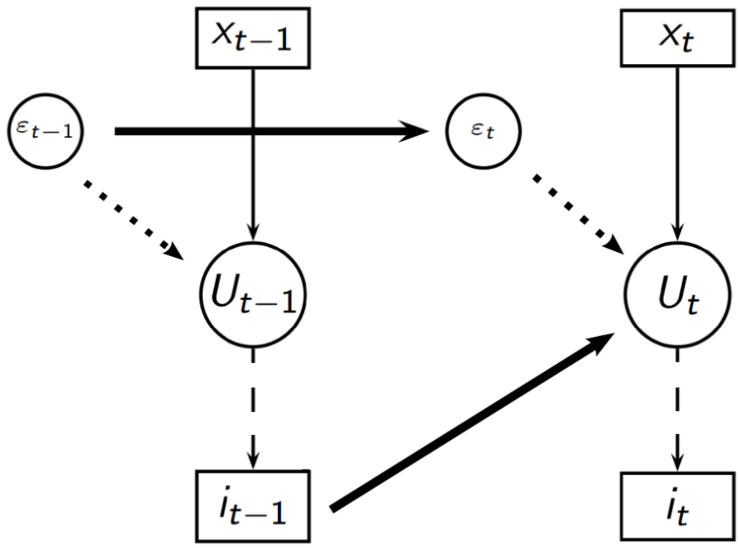
3.3.2 Accounting for panel nature of data

If the network traces are collected without anonymizing the identity of the individual too often, activity-episode sequences are available for a long enough period to observe repeated destination choices for the same activity type and a same individual. Thus it is possible to take into account the habits of each individual $i \in I$ (where I is the total sampled population of individuals). Wooldridge (2002) develops a general methodology to deal with unobserved individual heterogeneity in dynamic panel data with discrete dependent variables. We apply it to a pedestrian destination choice context.

The habits of an individual i are considered as the previous choice for the same type of activity performed at a similar time of the day. It is represented as a dummy variable that takes the value 1 for the previously chosen alternative, 0 otherwise and -1 if no previous choice is available. There is no strict and regular periodicity between 2 consecutive choices: it can be one day, two weeks or several months, and it may change from individual to individual and from observation to observation. The difficulty of considering activity-episode sequences over time is that the problem becomes dynamic (Bierlaire, 2014).

The utility function at time t takes into account the choice performed at time $t - 1$. It means that the observations and the error terms are not independent anymore. Figure 6 shows the interaction

Figure 6: Dynamic Markov model with correlation:x are explanatory variables, U utility functions and i the revealed choices. ε represent error terms and t is time. The solid arrows are causal relationships, the dashed arrows are measurement relationships, dotted arrows are disturbance relationships and black arrows are correlation relationships. Round boxes are latent variables and square boxes are observed variables (Walker, 2001; Aurélie Glerum, 2014).



Source: (Bierlaire, 2014)

between error terms, utility functions and choices performed. It leads to serial correlation and agent effect issues (also known as one-way effect, i.e. time-invariant unobserved terms). We here consider that the error terms are defined as the sum of two unobserved components. The first is a time-invariant unobserved effect (i.e., σ_i in Equation (15)) and the second is an error term that is independent and identically distributed over time and individuals (i.e., $u_{i,t}$ in Equation (15)).

If we assume that current choices are influenced by past choices, the individual error terms are correlated over time. We thus need to correct for this correlation issue. According to Wooldridge (2002), it is possible to manage this issue by defining a function c_i that is (1) conditional to the initial choice and (2) time-invariant observed characteristics of the individual. We consider the following distribution:

$$c_i|y_{i,0}, z_i \sim Normal(\alpha_1 + \alpha_0 y_{i,0} + \alpha_2 z_i, \sigma_i^2). \tag{13}$$

We rewrite the function c_i as:

$$c_i = \alpha_0 y_{i,0} + \alpha_2 z_i + \sigma_i. \tag{14}$$

σ_i is a parameter to be determined, normally distributed and independent of y_{i0} and z_i. y_{i0} is the first choice ever made by an individual i (i.e., the anchor point of the autoregressive dynamic). z_i reveals the individual behavior among the past period (e.g., average distance covered, most frequently chosen destination…). Thus the choice of the alternative x at time t performed by i is rewritten as:

$$y_{x,i,t} = \beta z_{x,i,t} + \rho y_{i,t-1} + \alpha_0 y_{i,0} + \alpha_2 z_i + \sigma_i + u_{i,t}. \tag{15}$$

Basically, the choice that the individual i does depends on some parameters observed at time t, his choice made at time $t-1$ and is corrected with his first choice ever performed, some observed habits among the past observations, a normally distributed zero centered error distribution and a common error term. The model is thus mixed in errors. It takes into account a panel effect specific to each individual. The parameters $\beta, \rho, \alpha_0, \alpha_2$ and σ_i are estimated.

As suggested by Pirotte (1996), one has to consider short-term (within an individual variability) and long-term effects (between individuals variability). It means that some parameters that used to be significant in the short-term (without panel effect) should be left in the model even if they are not anymore.

Table 3: Definition of static and dynamic models. AE stands for Agent Effect. In the case study (see Section 4), one decides to split the dynamic model with agent effect correction into two submodels to fully understand the influence of each term of the Wooldridge (2002) correction (e.g., α_2 is equal to zero in one model).

Static model	Dynamic strict exogenous model	Dynamic with AE correction model
$\rho = 0$	$\rho \neq 0$	$\rho \neq 0$
$\alpha_0 = 0$	$\alpha_0 = 0$	$\alpha_0 \neq 0$
$\alpha_2 = 0$	$\alpha_2 = 0$	$\alpha_2 \neq 0$ or $\alpha_2 = 0$
$\sigma_i = 0$	$\sigma_i = 0$	$\sigma_i \neq 0$

We consider and compare three situations: a static model (no previous choice considered at all), a dynamic strict exogenous with the period model (previous choice considered but with the assumption that individuals have no memory on short observation periods. It means that the choice is exogenous within a short period, but endogenous over time) and a dynamic situation with panel data and agent effect model (previous choice considered and agent-effect issue corrected). These cases are explained in Table 3.

The strict utility function may have the following shape:

$$\begin{aligned}V_{i,x,t} = ASC_x + &\beta_{socio-eco} * SOCIOECO_i + \beta_{altspecific} * ALTSPECIFIC_x + \\ &\beta_{distance} * DISTANCE_x + \rho * CHOICE_{i,t-1} + \\ &\alpha_0 * CHOICE_{i,t_0} + \alpha_2 * SOMEHABITS_{i,\bar{t}} + \sigma_i,\end{aligned} \quad (16)$$

where i is an individual, x a destination and t is the time. From an activity type to another (e.g., buying a ticket, visiting a shop, drinking a coffee. . .), a specific model must be developed with a specific panel of attributes. In this paper, we make the strong assumption that choices of destinations for different types of activities are independent: sequences of activities are series of independent choices.

In the event of a dynamic based on different seasonnalities (see Section 3.3.1) inside a same model (e.g., morning habits, lunch habits. . .), one suggests that each time period has its own dynamic parameters (i.e., ρ, α_0, α_2) and that values of σ_i of each seasonnality are compared to decide if individuals' unobserved heterogeneity changes during the day (different values depending on the day period), or not (similar values all day).

4 A case study on EPFL campus

We perform a case study on the Ecole Polytechnique Fédérale de Lausanne (EPFL) campus (see Section 4.1). The methodology developed by Danalet *et al.* (2014) converts WiFi localizations collected from students and employees into activity-episode sequences. These data are dated spring 2012. Due to privacy issues, they are partially confidential (see Section 4.2). In Section 4.3, some descriptive statistics on the data are reported. The models are introduced and discussed in Section 4.4. A validation and imaginary previsions are proposed in Sections 4.5 and 4.6. Issues encountered are presented in Section 4.7 and recommendations in Section 4.9.

4.1 The EPFL campus

We work on the catering facilities destination choice and with the most likely candidate of activity-episode sequences only (see Sections 3.1 and 3.2). Also, one considers that the destination and activity choice models are independent. Thus, $P(\hat{x}_j|x^j_\psi)$ and $P(A_{1:\Psi})$ are equal to 1 in Equation (9) if k = *Restaurant*, 0 otherwise. These assumptions are possible in this particular case because:

1. The attractivity (see Section 2.1) of destination type *Restaurant* is higher than the attractivity of the other activity types. It means that if one considers more than one candidate of activity-episode sequences, the *Restaurants* are likely to remain in the sequence. Changes will probably occur for offices;
2. It is not likely that there are trades between two possible destinations (see Figure 7);

The catering destinations represent 21 possible alternatives (destinations) on the campus. Their locations and types are represented on Figure 7. They are separated in 5 categories (restaurants, self-services, cafeterias, caravans and others) depending on the sort of service they propose (see Table 4). We follow the methodology introduced in the previous chapter (see Section 3).

More information is required in order to explain people destination choice. These factors are related to the destination (destination specific attribute) and not to the individual (socio-economic attribute). Services' availability is described in Table 5. Factors such as prices, outside/inside capacities, opening hours or quality surveys have been collected from the EPFL restauration service. Collected data are explained below.

The activity-episode sequences contain socio-economic information such as the individual anonymized and unique ID and the occupation (student or employee, see Section 4.2). They also

Table 4: Table of types of destinations

Destination	Type
Cafe Le Klee	Cafeteria
BC	Self-service
BM	Other
ELA	Cafeteria
INM	Cafeteria
MX	Cafeteria
PH	Other
L'Arcadie	Cafeteria
L'Atlantide	Self-service
Le Copernic	Restaurant
Le Corbusier	Self-service
Le Giacometti	Cafeteria
Le Parmentier	Self-service
Le Vinci	Self-service
L'Esplanade	Self-service
L'Ornithorynque and Cybercafé	Self-service
Pizza	Caravan
Kebab	Caravan
Satellite	Cafeteria
Le Hodler	Self-service
Table de Vallotton	Restaurant

collect the day of the year and the start and end times of the full sequence. Activity-episodes contain start and end times and the location of the activity (destination). We compare two consecutive activity-episodes of a same day to calculate the distance between all the possible destinations (see Section 3.2.3). As people are tracked during a period of three months each individual has several observations (activity-episode sequences). We use them to measure their habits (previous, first and most frequent choices as explained in Section 3.3.2).

Cafeterias mostly offer coffee and sandwiches and can usually be used as workspaces outside lunch hours. Self-services have at least one hot lunch menu and may also propose pizzas, meat or pastas. Restaurants have several menus, propose a table service and are more expensive than the other catering destinations.

Caravans sell kebabs, pizzas and French-fries. They can be considered as fast-foods. The *other* catering areas are tables with an automatic coffee machine and a microwave. They are

Figure 7: Localization of destinations on the EPFL campus (`http://map.epfl.ch`). Destinations circled in red opened more recently (see Section 4.6.2)

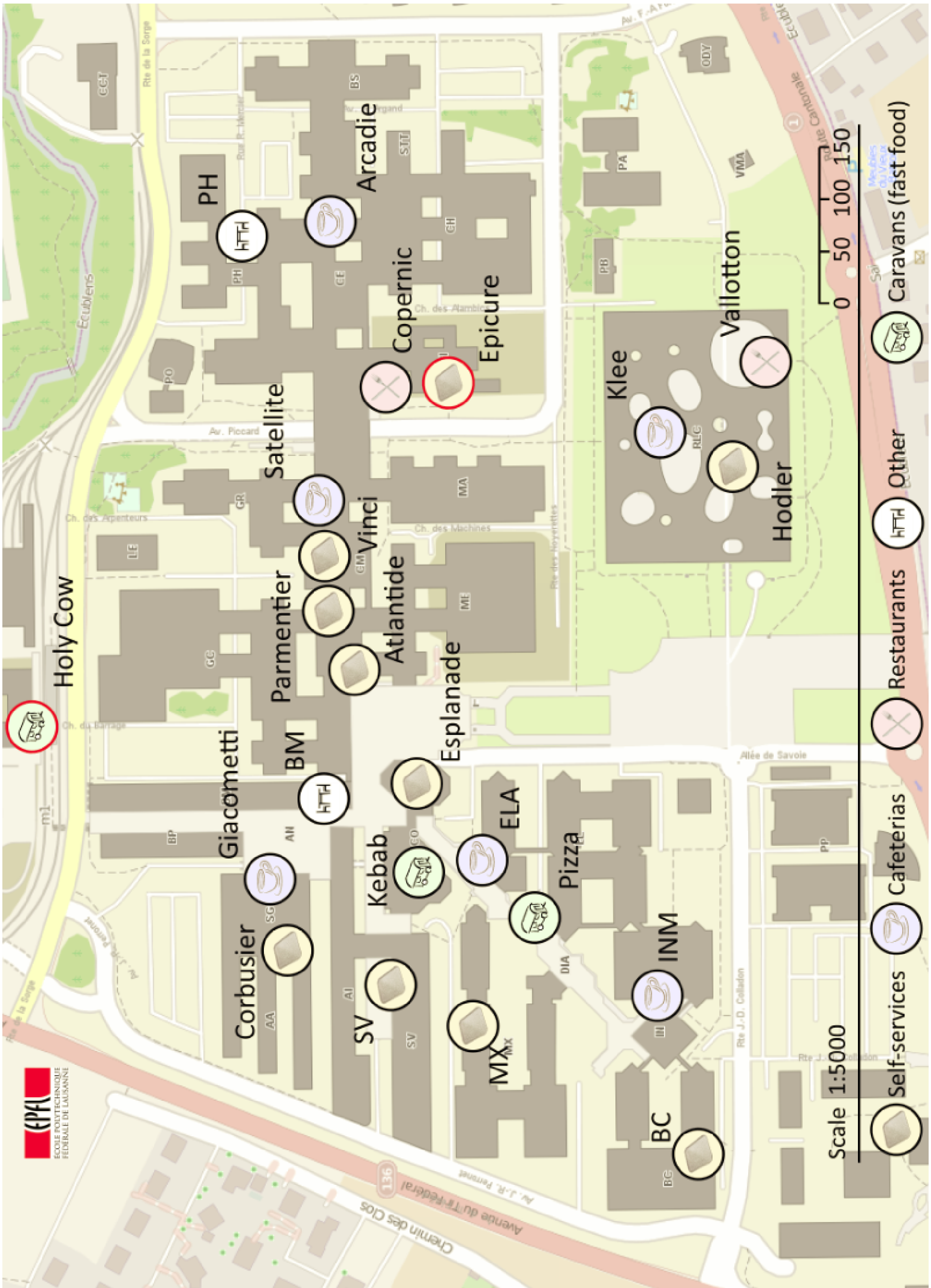


Table 5: Table of services availability

Destination	Coffee	Hot meal	Tables service	Visibility	Terrace	Workspace	Green fork	Dinner	Sandwiches	Selecta	Food	Tap beer	Fidelity card	% Av.
Cafeteria Cafe Le Klee	X	0	0	X	0	X	0	0	X	X	X	X	0	53,8%
Self-service BC	X	X	0	0	X	X	X	0	0	X	X	0	0	53,8%
Other BM	X	0	0	0	X	X	0	0	0	X	0	0	0	30,8%
Cafeteria ELA	X	0	0	0	X	X	0	0	X	0	X	0	0	38,5%
Cafeteria INM	X	0	0	X	X	X	0	0	X	X	X	0	0	53,8%
Cafeteria MX	X	X	0	0	X	X	X	0	X	X	X	0	0	61,5%
Other PH	X	0	0	0	0	X	0	0	0	X	0	0	0	23,1%
Cafeteria L'Arcadie	X	X	0	X	X	X	0	0	X	0	X	X	0	61,5%
Self-service L'Atlantide	X	X	0	X	X	X	0	0	X	0	X	0	0	53,8%
Restaurant Le Copernic	0	X	X	X	X	0	0	0	0	0	X	0	0	38,5%
Self-service Le Corbusier	0	X	0	0	X	0	X	0	0	0	X	0	0	30,8%
Cafeteria Le Giacometti	X	0	0	X	X	X	0	0	X	X	X	0	0	53,8%
Self-service Le Parmentier	0	X	0	X	X	0	X	X	0	0	X	0	0	46,2%
Self-service Le Vinci	0	X	0	X	X	0	X	0	0	0	X	0	0	38,5%
Self-service L'Esplanade	X	X	0	X	X	X	X	X	X	X	X	0	0	76,9%
Self-service L'Ornithorynque	0	X	0	0	X	0	0	0	0	0	X	0	0	23,1%
Caravan Pizza	0	X	0	X	X	0	0	X	0	0	X	0	X	46,2%
Caravan Kebab	0	X	0	X	X	0	0	X	0	0	X	0	X	46,2%
Cafeteria Satellite	X	0	0	X	X	X	0	0	X	0	X	X	X	61,5%
Self-service Le Hodler	0	X	0	0	0	X	X	0	0	X	X	0	0	38,5%
Restaurant Table de Vallotton	0	X	X	0	0	0	0	0	0	0	X	0	0	23,1%
% Availability	57,1%	66,7%	9,5%	57,1%	81%	61,9%	33,3%	19%	42,9%	42,9%	90,5%	14,3%	14,3%	

used for coffee breaks. Thus, catering destinations are not necessary visited with intent to have lunch. As Table 6 suggests, some of them are open all day as others are only open for a couple of hours during lunch time. The lunch period is the only moment of the day where all the eating establishments on the campus are open.

Table 7 shows the maximum and minimum prices for a hot meal at each destination. One can see that self-services all have a 7 CHF menu for students (Self-services get subsidies from EPFL) except for self-service L'Ornithorynque and self-service L'Atlantide who have a menu for about 10 CHF. Restaurants are more expensive. Their cheapest meal is 18.5 CHF for Restaurant Le Copernic and 25 CHF for Restaurant La Table de Vallotton. Caravans sell Kebabs for 7 CHF and pizza (without fillings) for 8 CHF.

The only gap in these prices is between Restaurants and the rest of the destinations. Restaurants are mainly frequented by visitors, professors and employees. The maximum prices of self-services and caravans still are below the restaurants' cheapest menu's price. Employees must pay an additional amount of 1 CHF for self-services 7 CHF meals. There are no prices differences on the other menus. Students pay at least 7 CHF for a hot meal and personnel at least 8 CHF (except if they order a kebab).

The capacity (see Table 8) varies a lot between all the destinations. It is necessary to separate the inside capacity from the outside capacity as they are not available in winter or when it rains. Only caravans do not have an inside seating area. The inside capacity fluctuates between 25 and 320 seats. Furthermore some self-services offer up to 180 seats on their terrace. They are the destinations with the highest capacities.

Since the campus is outside the city center, they need to accommodate all students and employees[4] for lunch with affordable menus and a large capacity. In 2012, the food service (restauration.epfl.ch) from EPFL made a survey (on both pen-and-paper and Internet supports) concerning the quality of the food on the campus. People were asked to grade the quality of food and to answer some questions about their habits and destination choice's factors. The results show that people choose their lunch destination because of determinants such as the proximity, the price, the meal itself (which we could unfortunately not collect) or the time they are willing to spend. These factors and the grades given to each destination are used in the model.

All the destinations got a grade superior to the mean (4). Furthermore, destinations with higher prices have a better evaluation which means that the price reflects the quality of the food

[4]about 7100 students and 5400 employees in 2012 (vpri.epfl.ch)

Table 6: Opening hours and availability of destinations

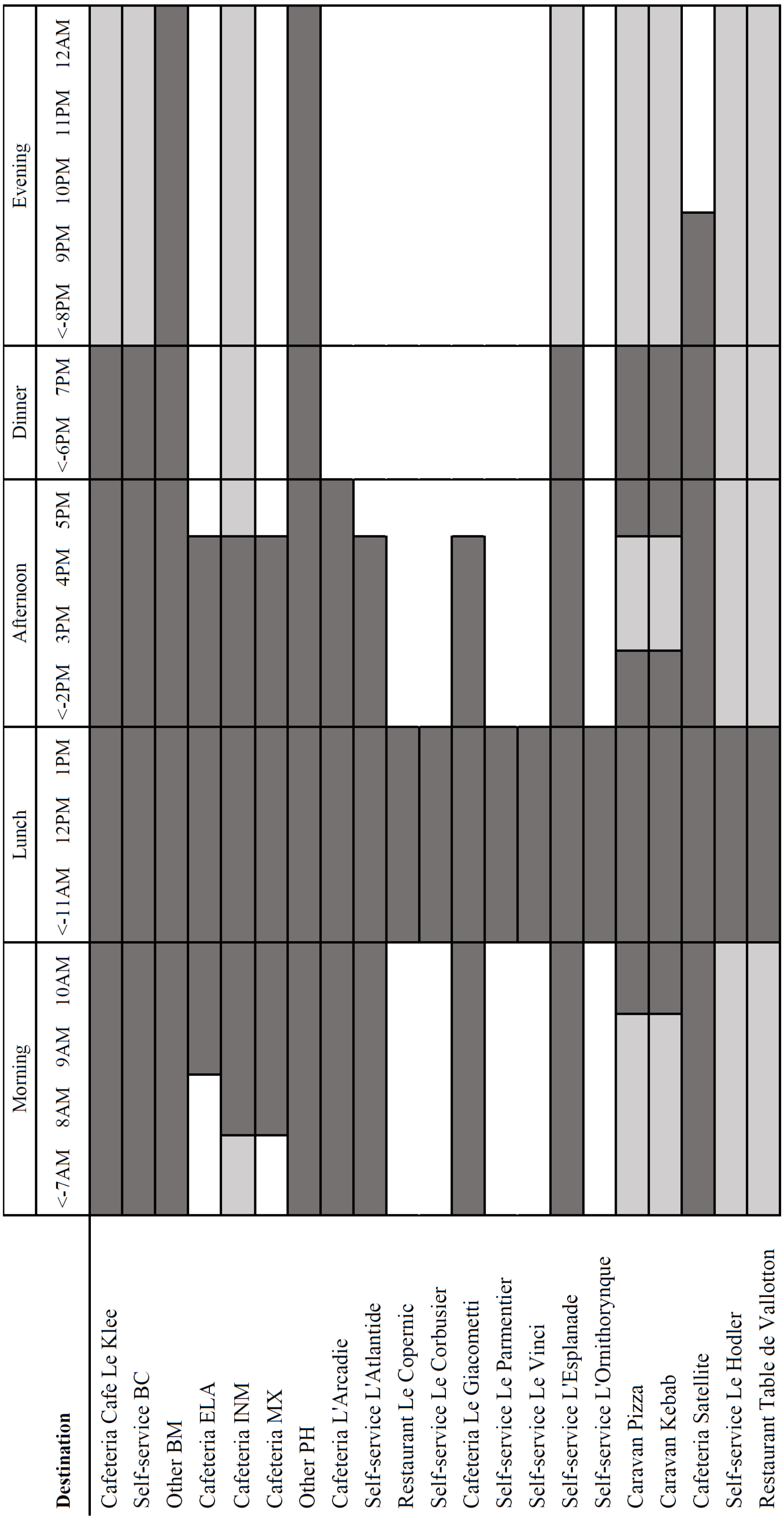


Table 7: Table of student prices for a hot meal

Destination	Cheapest	Most expensive
Cafeteria Cafe Le Klee	-	-
Self-service BC	7	12
Other BM	-	-
Cafeteria ELA	-	-
Cafeteria INM	-	-
Cafeteria MX	7	7
Other PH	-	-
Cafeteria L'Arcadie	9.9	9.9
Self-service L'Atlantide	9.8	9.8
Restaurant Le Copernic	18.5	27
Self-service Le Corbusier	7	11
Cafeteria Le Giacometti	-	-
Self-service Le Parmentier	7	12
Self-service Le Vinci	7	12
Self-service L'Esplanade	7	9
Self-service L'Ornithorynque	7.65	11.05
Caravan Pizza	8	12
Caravan Kebab	7	10
Cafeteria Satellite	-	-
Self-service Le Hodler	7	14
Restaurant Table de Vallotton	25	31

and of the service. Small cafeterias also have good grades although they do not sell hot meals. According to the survey, these destinations have a good relation with customers.

Each destination has several additional services. They are summarized in Table 5. Most of the eating establishments have a terrace and sell food (of any kind) but only 67% offer a hot meal. The majority of them is selling coffee and proposes a workspace. Also, 57% of places are visible from the common sidewalk. Just 43% of the places sell sandwiches or have a Selecta (automatic vending machine). One third of the destinations are part of the "Green Fork" (a quality label) deal and only 14% of them sell tap beers or have a fidelity card.

Self-service L'Esplanade is the most complete catering destination. Nearly all services are available and it is located in the middle of the campus. Only table service, tap beer and fidelity card are missing. On the other hand, restaurants and self-service L'Ornithorynque only have half of all the presented services. There is not much heterogeneity between destinations of a same type.

Table 8: Table of capacities

Destination	Inside	Outside
Cafeteria Cafe Le Klee	70	0
Self-service BC	82	119
Other BM	60	10
Cafeteria ELA	98	68
Cafeteria INM	20	14
Cafeteria MX	50	25
Other PH	15	0
Cafeteria L'Arcadie	60	100
Self-service L'Atlantide	125	50
Restaurant Le Copernic	105	50
Self-service Le Corbusier	228	100
Cafeteria Le Giacometti	90	30
Self-service Le Parmentier	320	52
Self-service Le Vinci	240	52
Self-service L'Esplanade	225	180
Self-service L'Ornithorynque	250	120
Caravan Pizza	0	15
Caravan Kebab	0	0
Cafeteria Satellite	200	30
Self-service Le Hodler	128	0
Restaurant Table de Vallotton	80	0

4.2 WiFi traces on the campus

In their case study, Danalet *et al.* (2014) explain the nature of EPFL WiFi data (the data are available in Danalet (2015)). People working or studying on the campus can connect to the WiFi network (see Figure 2(a)) for free using their username. The authentication is made through WiFi Protected Access using a radius server. It processes accounting by allowing to associate a MAC address with the username.

In order to anonymize the data, the username and the MAC address are replaced by a individual and unique ID and a socio-economic attribute: the category of users. They are shown in Table 9.

Also, the number of observations per occupation is specified. Employees represent the majority of the total sample. The number of visits in eating establishments varies between 54 for students in master of computer science and 152 for life sciences bachelor students. Note that these

Table 9: Category of traced individuals

	Students		Employees
Section	Semester	Number of observations	Number of observations
Civil engineering	4	141	1323
Computer science	4	89	
Computer science	8	54	
Mathematics	2	109	
Life science engineering	2	152	
Physics	2	140	
Total: 685 observations performed by 59 students and 1323 observations by 130 employees			

activity-episodes are performed by 189 different individuals.

4.3 Descriptive statistics on activity-episodes

We compute some descriptive statistics about destination choice. The aim is to capture factors that reveal people's decision logic. Table 10 show that self-service L'Esplanade is the most visited eating establishment on the campus. It makes sense since this destination is strategically placed (in the middle of the school and surrounded with auditoriums). Then come the other self-services and cafeterias. They are followed by the caravans and the restaurants.

Catering facilities located in the Rolex Learning Center (RLC) and Self-service Le Vinci do not have many visits. Danalet *et al.* (2014) explain that it is, in particular, due to the higher attractivity of the surrounding places (e.g., the library in the RLC and self-service Le Parmentier next to the Vinci. See Section 2.1 and Section 2.3). Indeed eating and working areas are (nearly) melted in the RLC and the seated capacity of the working area is about ten times bigger. Thus, activity-episode sequences measured in the library are slightly biased due to the low precision of the attractivity measure in the library (the number of seats is used as an aggregate measure of occupation).

We present the catering destinations per daily seasonalities (i.e., morning, lunch, ..., see Section 3.3.1)[5] in Table 11. Lunch time is the most attractive period in average. More than one third of the visits are made between 11 AM and 2 PM. Note that some destinations are less visited during this period. It is the case for self-service L'Atlantide, cafeteria Satellite, cafeteria

[5]Note that in Section 4.4, we change the definition of the seasonalities (periods of the day). Afternoon, dinner and evening become *afternoon*

Table 10: Observed choices per destination

Destination	Nb picks
Cafeteria Cafe Le Klee	4
Self-service BC	172
Other BM	47
Cafeteria ELA	145
Cafeteria INM	13
Cafeteria MX	86
Other PH	85
Cafeteria L'Arcadie	38
Self-service L'Atlantide	146
Restaurant Le Copernic	6
Self-service Le Corbusier	73
Cafeteria Le Giacometti	182
Self-service Le Parmentier	139
Self-service Le Vinci	2
Self-service L'Esplanade	448
Self-service L'Ornithorynque	102
Caravan Pizza	65
Caravan Kebab	68
Cafeteria Satellite	142
Self-service Le Hodler	36
Restaurant Table de Vallotton	9

MX or PH (others) which are destinations where it is common to take coffee breaks. Similar observations can be done in the afternoon. Destinations that are visited out of the lunch time all have a working space and/or additional services (e.g., coffee or tap beers). We consider now more specifically the lunch period. As students courses usually finish at 11 AM, 12 PM and 1 PM, one can expect several peaks in the demand. Destinations are aggregated by types (see Figure 8).

The lunch demand is separated into 3 peaks. There is one small peak between 11 AM and 12 PM because most of the self-services and restaurants only open at 11:30 AM. People reach a catering facility during this period to avoid queues and get a table more easily. The biggest peak is between 12 PM and 1 PM as the majority of students and people of the personnel lunch during this period. Then the third peak between 1 PM and 2 PM concerns students that finish their courses late and some employees. Cafeterias reach their maximum attendance during that period. It is possibly due to the fact that some people drink a coffee after their lunch. Also, individuals going to a restaurant do not move before 12 PM because their table is probably reserved.

Table 11: Choices performed depending on the time of the day

	7AM-11:30AM	11:30AM-2PM	2PM-6PM	6PM-8PM	8PM-11PM	Total
Cafeteria Cafe Le Klee	1	1	2			4
Self-service BC	50	69	42	11		172
Other BM	11	14	16	5	1	47
Cafeteria ELA	37	55	53			145
Cafeteria INM	2	7	4			13
Cafeteria MX	38	22	26			86
Other PH	35	16	26	6	1	84
Cafeteria L'Arcadie	11	19	8			38
Self-service L'Atlantide	72	18	56			146
Restaurant Le Copernic		6				6
Self-service Le Corbusier		73				73
Cafeteria Le Giacometti	45	56	81			182
Self-service Le Parmentier		82		55	2	139
Self-service Le Vinci		2				2
Self-service L'Esplanade	95	148	162	44		449
Self-service L'Ornithorynque		102				102
Caravan Pizza	12	35	5	13		65
Caravan Kebab	11	19	24	14		68
Cafeteria Satellite	37	14	74	11	6	142
Self-service Le Hodler		36				36
Restaurant Table de Vallotton		8			1	9
Total	457	802	579	159	11	2008

Figure 8: Demand peaks during lunch hours (one hour periods)

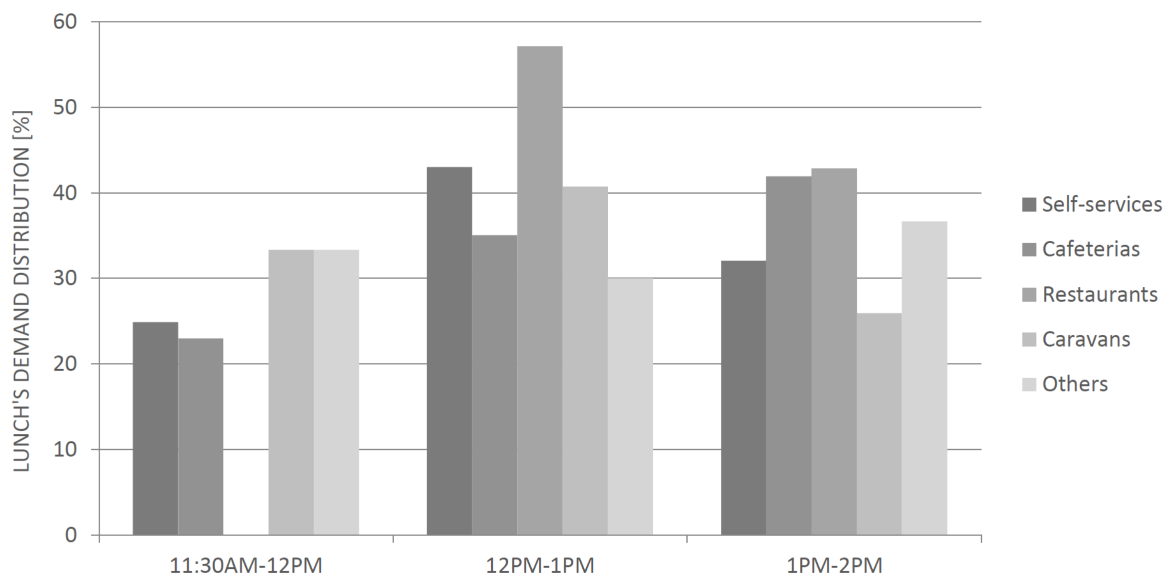
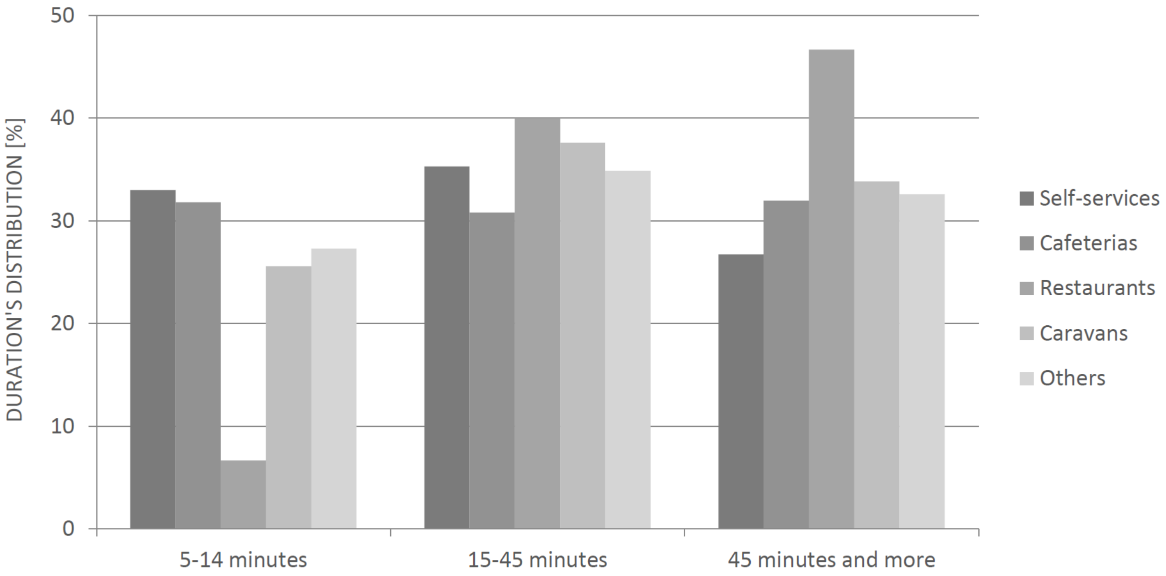


Figure 9: Durations (in minutes) of observations depending on the type of destination

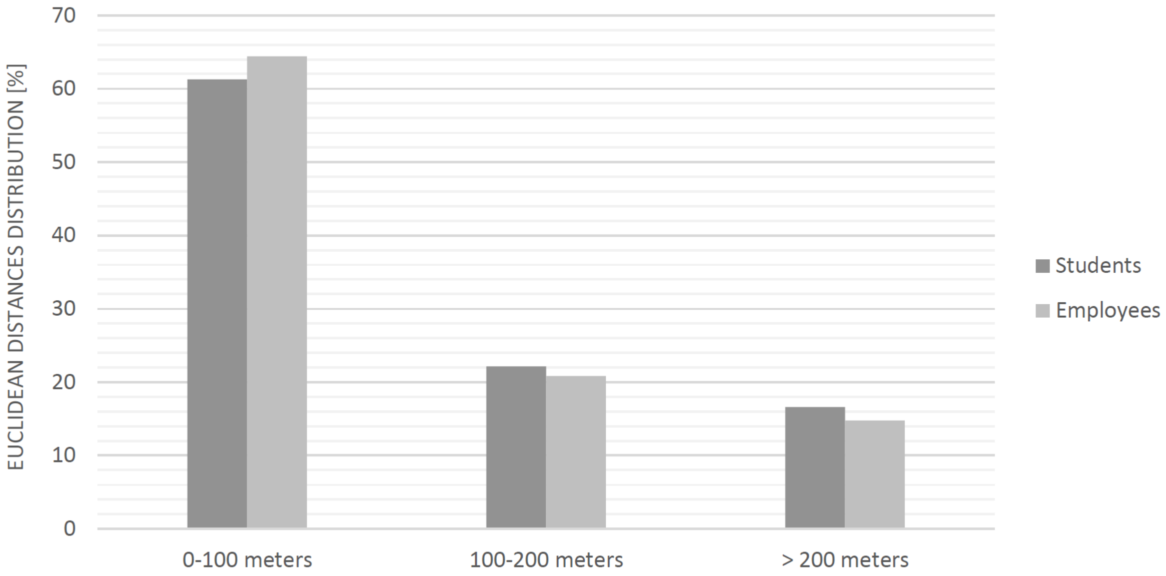


The durations of activity-episodes depending on the destination type is shown in Figure 9. They have been separated into three nearly equal categories. The first one reflects short visits (between 5 and 14 minutes). They can be interpreted as short breaks or as visits to buy a snack or a drink. The second one represents long breaks (between 15 and 45 minutes) to perform activities such as having lunch or spend an hour to rest. The last one stands for long activities. One can see on the figure that visiting a restaurant may take more than 45 minutes. Also, studying for a course or spending free time in a cafeteria can take more than an hour.

We take a look at the choices performed by some individuals (see Table 12). Civil engineering students have some habits. Indeed nearly all individuals have a preference for one or several destinations. This is also true for students from other sections and for employees. The repetition of the same catering destination choice over time for a same individual motivates to consider habits. The choice of a catering destination at time t was the same as the choice of a catering destination at time $t-1$ in 79% of the cases during the morning, in 40% of the cases during lunch hours and in 33% of the cases after lunch (i.e., only when a previous choice was performed).

According to the literature review, the distance to walk has a significant impact in both route and destination choices. On the campus, if a student finishes his course at the extreme east (BS) and decides to lunch at the extreme southwest (BC) he has to walk about 1200 meters if he takes the shortest path (only 700 in Euclidean distance). By looking at Euclidean distances, students and

Figure 10: Distribution of Euclidean distances



employees have a preference for short distances but may change their habits sometimes. Indeed the average Euclidean distance covered is 110 meters (109 for students and 100 for employees).

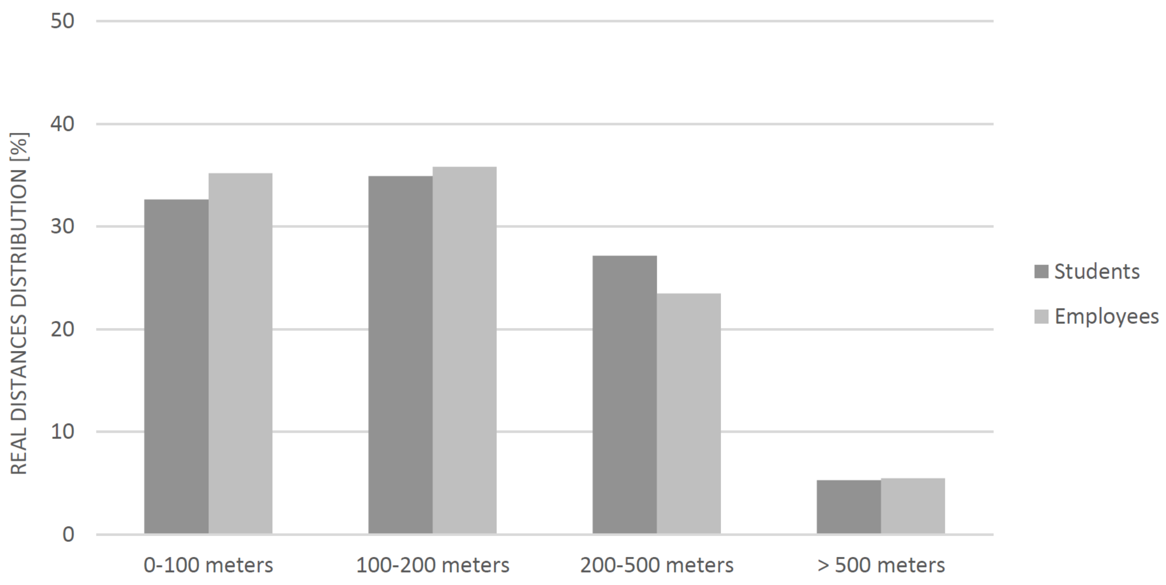
Since a pedestrian network is available, realistic shortest path can be calculated between two destinations. They are more realistic than Euclidean distances. We compare the Euclidean and real distances covered to reach the chosen destination.The Euclidean distances reduces all the non-null paths (i.e. paths shorter than 20 meters are omitted) by 90% in average compared to paths calculated with a weighted shortest path algorithm (the complete algorithm is available in Danalet *et al.* (2014)). However, the standard deviation is high (around 100%). Using such algorithms takes into account the anisotropy of the place (Kim and Hespanha, 2003). Figures 10 and 11 represent the distribution of Euclidean and real distances walked by the individuals.

The trends are similar as before except that the distances to reach a catering destination are longer. In average, both students and employees walk 175 meters to visit an eating establishment. 5% of individuals cover a distance longer than 500 meters to reach their catering destination. This weighted shortest path algorithm does not provide distances between all points of the pedestrian network. This is due to the coding of the network (i.e., some doors need an access card and are assumed to be closed). This study uses a sample of 4,5 millions paths. About 10% of the possible distances are not calculated. However the distance to reach the chosen catering destination is always available.

Table 12: Choices performed by civil engineering students: the bold numbers represent the most frequently chosen destination of one individual and the italic numbers, its first chosen destination.

	Choices made by individuals over a 3 months period												
ID	BM	ELA	MX	PH	L'Atlantide	Le Giacometti	Le Parmentier	L'Esplanade	L'Ornithorynque	Pizza	Kebab	Satellite	Le Hodler
10000		1	1				1	**9**	*3*	1		5	
10001			1				*5*	4				**6**	4
10003					1			***3***		1		2	1
10004							7	**10**			1	1	1
10006								***4***					
10007						***7***							1
10008							***1***						
10009		1	1				1	***5***		2			1
10010	1						*8*	4			1	**12**	
10013	1							***2***					
10017		1		1									***2***

Figure 11: Distribution of real distances



We consider the weather since we have daily data from Meteosuisse. During the case study, the average maximum temperature was $15°C$ and two third of the days were dry. It was a typical Swiss spring.

Individual's choices are related to two important factors: the distance and the habits. Indeed people seem to prefer a catering destination close to their previous location and a destination they know well (they have already visited). Also, students and employees do not necessary visit a destination for the purpose of eating. More characteristics such as offering work places, coffee or tap beers may influence people's choice of catering destination.

4.4 Modelling of destination choice

4.4.1 Description of the models

Considering the points highlighted in Section 4.3, we develop linear in parameter Multinomial and Mixed Logit Models. Before we describe these models in detail, one needs to define the dynamic variables. One proposes that the previous choice is the previous catering destination visited by a same individual during a similar period of the day (see Table 14). It means that the time interval between the activity-episode sequences varies (e.g., it can be one day or weeks depending on the availability of information and the frequency of observations) and that we have daily seasonalities (see Section 3.3.1). We decide to merge afternoon, dinner and night time periods to have enough observations; we call that period *afternoon*. The daily seasonalities are thus: morning, lunch and afternoon. Models focused on the specific lunch period are available in Tinguely *et al.* (2015).

Similarly, the first choice is the first catering destination ever visited by this same individual during the selected day period. Finally, we propose to use the most frequent choice to describe one individual average behavior among the past period (see Section 3.3.2). The most frequent choice stands for the most visited catering destination, during one period of the day, before the actual choice. In the event of a tie, the most visited destination is randomly selected among the destinations with the same number of visits. We consider three main variants as defined in Table 13 and we also add two submodels to the dynamic with agent effect correction's variant.

1. A static model with no previous choice considered where each observation is independent;
2. A dynamic strict exogenous model where the previous choice is considered but with the assumption that individuals have no memory on short observation periods (thus, the choice is based on exogenous factors within a short period, but on endogenous determinants over time);
3. Two dynamic models with panel data and agent effect correction. The previous choice is considered and two approaches are used to correct for agent effect issue using the principles described in Section 3.3.2:
    a) Only the first choice is considered to correct agent effect issue;
    b) The first and most frequent choices are considered to correct agent effect issue;

Table 13: Definition of static and dynamic models for the case study. Both first models are Multinomial Logit whereas the ones with agent effect correction are Mixed Logit

Static model	Dynamic strict exogenous model	Dynamic models with agent effect correction	
		First choice	First and most frequent choices
$\rho = 0$	$\rho \neq 0$	$\rho \neq 0$	$\rho \neq 0$
$\alpha_0 = 0$	$\alpha_0 = 0$	$\alpha_0 \neq 0$	$\alpha_0 \neq 0$
$\alpha_2 = 0$	$\alpha_2 = 0$	$\alpha_2 = 0$	$\alpha_2 \neq 0$
$\sigma = 0$	$\sigma = 0$	$\sigma \neq 0$	$\sigma \neq 0$

There are 21 catering destinations on the EPFL campus, thus 21 utility functions. Table 14 clarifies the variables introduced in the models. We estimate the parameters for all four models using Python Biogeme software (Bierlaire, 2003; Bierlaire and Fetiarison, 2009) in Section 4.4.2.

Table 14: Specification table: each variable has possibly 21 different values. The daily seasonalities are: morning is from 7 AM to 11:29 AM, lunch is from 11:30 AM to 2 PM, afternoon is from 2 PM to 6 PM, dinner is from 6 PM to 8 PM, night is from 8 PM to 11 PM. If the day period constraint is false, the variable is 0. Distances are measured in meters. A good weather is a dry day and at least a max daily temperature of $20°C$. $\beta_{DISTANCE\_LUNCH_{TYPE}}$ and $\beta_{EVALUATION_{TYPE}}$ are type specific (see Table 4). α_{FIRST_CHOICE}, α_{MOST_CHOSEN} are day period specific and are only considered in the dynamic models with agent effect correction. $\rho_{PREVIOUS\_CHOICE}$ are day period specific and null in the static model (see Table 13). Missing values are equal to -1 in the dataset. ASC_x and σ_x are alternative specific

Parameter	Variable	Variable description	Season
ASC_x	1	1 if not Esplanade, 0 otherwise	
$\beta_{DIST\_LUNCH_{TYPE}}$	*lunch_distance*	distance from the previous activity-episode 0 otherwise	lunch
$\beta_{DIST\_MORNING}$	*morning_distance*	distance from the previous activity-episode 0 otherwise	morning
$\beta_{DIST\_AFTERNOON}$	*afternoon_distance*	distance from the previous activity-episode 0 otherwise	afternoon
$\beta_{NO\_DISTANCE\_AV}$	*distance_not_av*	1 if no distance is available 0 otherwise	
$\beta_{EVALUATION_{TYPE}}$	*evaluation_survey*	average quality grade on a [1;6] scale 0 otherwise	lunch
$\beta_{PRICE\_STUDENT}$	*price_min_student*	price for the cheapest hot meal if student 0 otherwise	lunch
$\beta_{PRICE\_EMPLOYEE}$	*price_min_employee*	price for the cheapest hot meal if employee 0 otherwise	lunch
β_{TAP_BEER}	*beer_av*	1 if tap beer is available 0 otherwise	afternoon
β_{DINNER}	*dinner_av*	1 if dinner is available 0 otherwise	dinner
$\beta_{CAPACITY\_TERRACE}$	*capacity_terrace*	outside number of seats if the weather is good 0 otherwise	lunch
$\beta_{CAPACITY\_INSIDE}$	*capacity_inside*	inside number of seats 0 otherwise	lunch
$\rho_{PREV\_MORNING\_CHOICE}$	*previous_choice*	1 if the destination was the previous destination 0 otherwise	morning
$\alpha_{MOST\_CHOSEN\_MORNING}$	*most_freq_choice*	1 if the destination was the most frequented 0 otherwise	morning
$\alpha_{FIRST\_MORNING\_CHOICE}$	*first_choice*	1 if the destination was the first destination 0 otherwise	morning
$\rho_{PREV\_LUNCH\_CHOICE}$	*previous_choice*	1 if the destination was the previous destination 0 otherwise	lunch
$\alpha_{MOST\_CHOSEN\_LUNCH}$	*most_freq_choice*	1 if the destination was the most frequented 0 otherwise	lunch
$\alpha_{FIRST\_LUNCH\_CHOICE}$	*first_choice*	1 if the destination was the first destination 0 otherwise	lunch
$\sigma_{morning_x}$	1	1 if not Esplanade, 0 otherwise	morning
σ_{lunch_x}	1	1 if not Esplanade, 0 otherwise	lunch

4.4.2 Estimation of the models

One shows a summary of the results in Table 15. The complete results (also containing Alternative Specific Constants (ASC) and σ values) and one typical utility function are available in Appendices A and B. The afternoon habits are not significant and thus removed from the models. Also, the values of $\sigma_{morning_x}$ and σ_{lunch_x} for a specific destination are not similar at all so we keep both terms (see Section 3.3.2). It means that individuals have a different behavior depending on the period of the day when they chose a catering destination. It makes sense since the criterion and sensibility to chose an establishment to drink a coffee are likely not to be the same as the ones to chose for a place to lunch. Also, the destination specific parameters shown in Table 5 and not represented in the models are not significant.

Table 15: Table of estimates. Number of observations = 1868. Parameters without superscript are significant with a 95% confidence. * stands for a confidence level between 70% and 95% and ** for a confidence level lower than 70%. Double star parameters are thus not significant

	Static		Strict exogenous		First choice		First and most freq	
Parameters	Value	*t*-test	Value	*t*-test	Value	*t*-test	Value	*t*-test
$\beta_{DIST\_LUNCH\_CAFET}$	-0.00703	-16.69	-0.00629	-14.49	-0.00412	-8.23	-0.00348	-6.79
$\beta_{DIST\_LUNCH\_REST}$	-0.00276	-2.18	-0.00271	-2.11	-0.00173	-0.82**	-0.00292	-1.35*
$\beta_{DIST\_LUNCH\_SELF}$	-0.00646	-19.99	-0.00555	-16.66	-0.00393	-10.51	-0.00339	-9.06
$\beta_{DIST\_MORNING}$	-0.00379	-5.97	-0.0028	-4.32	-0.00241	-2.84	-0.00237	-2.68
$\beta_{DIST\_AFTERNOON}$	-0.00061	-1.31*	-0.0011	-2.35	-0.00076	-1.3*	-0.0012	-1.94*
$\beta_{NO\_DISTANCE\_AV}$	-4.89	-13.84	-4.31	-12.32	-3.45	-8.59	-3.13	-7.75
$\beta_{EVALUATION\_CAFET}$	1.79	9.98	1.74	9.58	2.28	8.44	2.54	7.71
$\beta_{EVALUATION\_SELF}$	1.88	9.66	1.8	9.04	2.37	8.05	2.78	7.74
$\beta_{PRICE\_STUDENT}$	-0.0681	-2.07	-0.0512	-1.51*	-0.0776	-1.45*	-0.0776	-1.16*
$\beta_{PRICE\_EMPLOYEE}$	-0.0054	-0.18**	-0.0049	-0.16**	-0.024	-0.48**	-0.010	-0.17**
β_{TAP_BEER}	0.669	3.62	0.49	2.55	0.759	3.02	0.742	2.89
β_{DINNER}	0.943	3.35	0.865	3.09	0.634	1.73*	0.786	2.16
$\beta_{CAPACITY\_TERRACE}$	0.00162	1.84*	0.00215	2.29	0.00258	2.39	0.00149	1.33*
$\beta_{CAPACITY\_INSIDE}$	0.00277	1.29*	0.00386	1.69*	0.00365	1.28*	0.00439	1.36*
$\rho_{PREVIOUS\_LUNCH\_CHOICE}$	0	0	1.78	16.44	0.507	3.6	0.0143	0.09**
$\alpha_{MOST\_FREQ\_LUNCH\_CHOICE}$	0	0	0	0	0	0	1.77	14.26
$\alpha_{FIRST\_LUNCH\_CHOICE}$	0	0	0	0	1.03	7.91	0.827	6.33
$\rho_{PREVIOUS\_MORNING\_CHOICE}$	0	0	3.21	19.4	0.721	3.03	0.143	0.52**
$\alpha_{MOST\_FREQ\_MORNING\_CHOICE}$	0	0	0	0	0	0	2.43	10.27
$\alpha_{FIRST\_MORNING\_CHOICE}$	0	0	0	0	1.67	8.37	0.417	1.63*
$\mathcal{L}(0)$	-5037.914		-5037.914		-5037.914		-5037.914	
$\mathcal{L}(\hat{\beta})$	-3238.926		-2870.976		-2352.137		-2182.172	
ρ^2	0.357		0.43		0.533		0.567	

The values and signs of exogenous parameters are close between all models. The static model (Table 20) is the *restricted* version of the dynamic strict exogenous model (Table 21) which also is the *restricted* version of both dynamic with agent effect correction models (Tables 22 and 24). The addition of the previous lunch's choice (at time $t-1$) decreases the t-test of the parameters related to the choice of the catering destination at time t. A similar effect is observed with the addition of both agent effect issue's corrections. As suggested by Pirotte (1996) (see Section 3.3.2), long-term parameters measure the variability within individuals and thus alleviate the weight of short-term significants (i.e., reduce their t-tests).

In each model, the opening hours are considered as the availability of the destination (closed catering destinations cannot be visited even if pedestrians could technically reach them). Now, we examine parameters' sign and t-test to describe and analyze the results of the models. Capacities (number of seats) of terraces and inside spaces have a positive parameter's sign. It means that people have a preference for catering destinations with a bigger capacity. It makes sense since having an important number of places increases the chance to find a seat. Also, the destinations with terraces are more likely to be visited when the weather is sunny.

The distance from the previous activity-episode is significant in the choice of an establishment. The sign is negative independently of the period of the day which represents the fact that people prefer a close destination. In the morning, the main activity that can be performed in a catering destination is having a coffee. In the afternoon, it can be several things like having a coffee, working or drinking a beer. The comparison between the parameters of both these time periods shows that individuals prefer to walk less in the morning than in the afternoon. A possible explanation is that coffee is available nearly everywhere but descent workspaces or tap beers are much rarer so people accept to travel longer. Another possible reason is that people tend to have a coffee next to their following activity-episode (instead of next to the previous activity-episode). This has not been explored. Other possible explanations include looking for a sunny terrace or for a place selling ice creams, since the data collection took place in the beginning of the summer.

At lunch time, the distance covered from the last activity-episode depends on the type of destination chosen. For example, individuals are more likely to walk when going to a restaurant as the choice set is small for this destination type (only two restaurants on campus). On the other hand, students and employees prefer a near self-service or cafeteria compared to a far one. The fact that this kind of destinations is distributed everywhere on the campus can be an explanation.

Note that individuals visiting a caravan or another catering destinations (PH and BM) are not sensitive to distance (i.e., the parameters are not significant). It is not a surprise since those places have their own distinctive offers. People accept to cover more distance if they want a specific type of meal. The parameter accounting for the non-availability of distances is negative as well. It means that catering destinations that are the least connected to the network are less likely to be visited.

The minimum price for a hot meal is not significant in dynamic models for both students and employees but we decide to keep it anyway because we expected it to be. As explained in Section 3.3.2, it may be the fact that the price is considered as a short-term determinant in our models. Moreover, prices have low variability on the campus; this also explains why cost is not significant in our models. We give an explanation to these parameters anyway.

Price has a negative sign for students. It makes sense as they are not willing to spend 25 CHF to go to the restaurant and prefer catering destinations with 7 CHF meals or caravans. Employees may look for eating establishments with higher prices because the price is connected with the food quality. Therefore, the price is not significant at all for them. Also, working people earn a salary and bills can be attributed to the company expenses.

Evaluations have a positive sign for both cafeterias and self-services. It means that individuals choose a cafeteria or a self-service as a destination depending on the average quality of the offer. Evaluations are not significant for caravans and restaurants. Eating establishment that offer dinner are more likely to be visited between 6 PM and 8 PM.

The availability of tap beer after midday increases the utility of a catering destination. Indeed, some individuals may want to relax more than work in the afternoon and the evening. Only three destinations offer tap beers on the campus; the well-known Satellite bar, the cafeteria of the Rolex Learning Center (Klee) and cafeteria L'Arcadie.

Habits are significant in all dynamic models. The previous choice made by people at lunch time has a parameter with a positive sign. It means that students and employees have some habits when choosing for an eating destination. As an example, if the previous time they ate on the campus for lunch, they chose to eat at self-service Le Corbusier, they are more likely to pick this alternative again.

Also, the correcting terms have a positive sign and a strong *t*-test. However, the previous choice becomes non-significant with the double agent effect correction which may mean that average behavior among the observation period is stronger as the previous choice (it better explains the long-term behavior). We think that the choice of a destination for lunch may appear as an erratic phenomenon around one anchor destination (e.g., the first choice).

Similar observations are done concerning the morning behavior. The *t*-tests are even higher than the ones for the lunch period. It means that individuals have strong habits in the morning. They tend to always chose the same destination to drink a coffee.

The most robust explanatory variables of the models are the distance and the morning and lunch habits. Prices or services availability are less robust determinants. Probably because prices are relatively cheap and uniform (except for restaurants) and because a same type of catering destinations usually proposes the same services in every destinations. Also, both corrections of agent effect seem to improve the models since their ρ^2 parameter are much higher. We verify this impression in the next chapter (see Section 4.4.3).

4.4.3 Comparison of the models

All four models estimated in Section 4.4.2 have close values of parameters. We compare these models to find which one fits the data the best. A log-likelihood ratio test is performed. We can use this test because the models are nested. The static model is the *restricted* model of the dynamic strict exogenous model which is the *restricted* model of both dynamic with panel data and agent effect issue correction models. Also, the model considering the first choice is the *restricted* version of the one accounting for both the first and the most frequent choices. The statistic

$$-2(\mathcal{L}(\hat{\beta_R}) - \mathcal{L}(\hat{\beta_U})) \tag{17}$$

is χ^2 distributed, with degrees of freedom equal to

$$K_U - K_R \tag{18}$$

with K, the number of parameters of each model (*Unrestricted* and *Restricted*). If the result of Equation (17) is bigger than the percentile of the chi square distribution, then we can reject the null hypothesis (at a chosen level of confidence) and the *unrestricted* model is preferred to the *restricted* one. We perform the log-likelihood ratio test on each model according to the specification made in Table 13. Table 16 presents the results.

Both models, accounting for panel nature of data and correcting agent effect, are statistically better (with more than 95% confidence) than the second one which is statistically better than the static one as well.

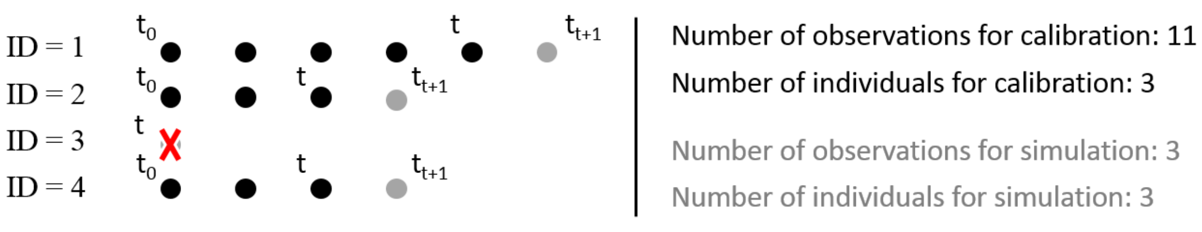
Table 16: Table of likelihood ratio test. DSE stands for Dynamic Strict Exogenous, DAEC stands for Dynamic with Agent Effect Correction.

	Static	DSE	DAEC first choice	DAEC first and most frequent choices
$\mathcal{L}(\hat{\beta})$	-3238.926	-2870.976	-2352.137	-2182.172
Nb of parameters	34	36	78	80
Loglikelihood ratio test				
Static vs DSE:$-2(-3238.9 + 2870.9) = 736 > 5.99$				
DSE vs AEC (first choice):$-2(-2870.9 + 2352.137) = 1036 > 58.12$				
DAEC (first choice) vs DAEC (first and most frequent choices):$-2(-2352.137 + 2182.172) = 340 > 5.99$				

4.5 Validation

We perform an aggregated validation on our models. The dataset is separated into two subsamples: one to calibrate the models, the second one to simulate the future destination choices and compare the output of the models with the actual choices. The first sample represents the past choices of individuals and the second sample contains their most recent observation. Basically we use people's past choices to estimate the models (first sample) and we forecast their most recent observation of a destination to have lunch or coffee (second sample). People with only one observation are removed because they do not fulfill the dynamic conditions (thus the dataset is not exactly the same that the one used for estimation in Table 15. We keep 1512 observations to calibrate the models and 144 to simulate future choices). An example of sample separation is given on Figure 12.

Figure 12: Separation of the total sample for calibration and simulation: the black dots represent the activity-episodes used for calibration whereas gray dots represent activity-episodes used for simulation.



Dynamic models with agent effect correction are simulated as Mixed Logit Models because they have two error terms and one of them is normally distributed (see Section 3.3.2) whereas both static and dynamic strict exogenous models only have a single error term and are thus simulated as Multinomial Logit Models. Table 17 summarizes the results. A similar validation

Table 17: Validation of the models (morning and lunch hours). Observed and estimated choices performed by 144 individuals on their most recent activity episode

Observed			Predicted							
			Static		Strict exo		First choice		First and most freq	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
Cafeteria Cafe Le Klee	0	0%	0	0.2%	0	0.1%	0	0.2%	0	0.1%
Self-service BC	15	10.4%	10	7%	10	6.9%	10	6.7%	11	7.5%
Other BM	1	0.7%	3	2.2%	2	1.6%	2	1.5%	3	2%
Cafeteria ELA	14	9.7%	8	5.3%	7	4.8%	8	5.7%	7	4.9%
Cafeteria INM	1	0.7%	1	0.8%	1	0.7%	2	1.2%	2	1.7%
Cafeteria MX	6	4.2%	6	4.3%	6	4.4%	4	2.9%	6	3.9%
Other PH	6	4.2%	5	3.8%	5	3.7%	4	2.9%	4	2.4%
Cafeteria L'Arcadie	6	4.2%	2	1.1%	3	1.8%	2	1.3%	2	1.5%
Self-service L'Atlantide	7	4.9%	10	7.1%	10	6.9%	7	4.7%	6	4.5%
Restaurant Le Copernic	1	0.7%	1	0.7%	2	1.1%	2	1.3%	2	1.4%
Self-service Le Corbusier	4	2.8%	12	8.5%	10	7.2%	10	6.9%	10	7%
Cafeteria Le Giacometti	13	9%	12	8.1%	12	8.1%	12	8.3%	14	9.4%
Self-service Le Parmentier	8	5.6%	13	8.8%	12	8.6%	14	9.6%	11	7.5%
Self-service Le Vinci	1	0.7%	0	0.1%	0	0.1%	0	0.1%	0	0.1%
Self-service L'Esplanade	23	16%	26	18%	25	17.7%	27	18.7%	27	18.7%
Self-service L'Ornithorynque	15	10.4%	15	10.5%	17	11.7%	17	12%	17	11.7%
Caravan Pizza	6	4.2%	3	2.3%	4	2.5%	4	2.8%	4	2.9%
Caravan Kebab	5	3.5%	3	2.1%	3	2%	3	2.4%	3	1.9%
Cafeteria Satellite	5	3.5%	7	4.8%	7	5.1%	8	5.5%	7	4.8%
Self-service Le Hodler	6	4.2%	5	3.6%	6	4.3%	6	4.2%	7	4.9%
Restaurant Table de Vallotton	1	0.7%	1	0.7%	1	0.6%	1	1%	1	1%

is performed on lunch hours in Tinguely *et al.* (2015) and two validations on the morning period and the complete day are available in Appendix B, Tables 26 and 27 respectively.

The trends are similar between observations and estimated choices. These results are positive since they show that even a basic static model simulates reasonable forecasting on a small validation sample. The errors mainly come from the estimation of self-services. The number of destination type's choices (e.g., Self-service, cafeteria…) is accurate for each model. It means that our models are good at forecasting the destination type choice but then are less accurate to select a specific destination. The reason could be that the variability of services' availability for destinations of a same type is narrow. Also, the fact that catering destinations are relatively evenly distributed on the campus (see Figure 7) does that individuals usually have equidistant possible destinations of a same type. This latter point is especially true for self-services.

We expected the accuracy to be better for both dynamic models with panel data and agent effect correction as they are statistically better than both other models (see Section 4.4.3) but according to Table 17, it seems that it is not the case. We propose to use a least squares' method to measure objectively the accuracy of each model:

$$S_m = \sum_{x=1}^{21} (O_x - E_{x,m})^2 \tag{19}$$

where O_d is the percentage of Observations for destination x and $E_{x,m}$ is the expected number of visitors based on the choice probabilities for destination x and model m. The best model is the one that minimizes the least squares' method (S_m). The results are shown in Table 18.

Table 18: Least squares' method. DAEC stands for Dynamic with Agent Effect Correction

Static	Strict exogenous	DAEC first choice	DAEC first and most frequent choices
$S_{static} = 104$	$S_{strict\_exogenous} = 89$	$S_{first\_choice} = 94$	$S_{first\_and\_most\_frequent\_choices} = 83$

The model that minimizes the difference between observations and estimated choices is the dynamic with both agent effect corrections (first and most frequent choices) which is also the one that fits the data the best (Section 4.4.3). The static and strict exogenous models show accurate forecasting as well. The dynamic model with only one agent effect correction performs less for validation than other models. The first choice may not be very representative of individuals' habits on short periods. We see that the first and most frequent choices is a better measurement of habits.

With these results, one considers that our models validate successfully the methodology and recommend the use of the first and most frequent choices as representative of people's habits in destination choices. The first choice only, as described by Wooldridge (2002) does not perform well.

4.6 Forecasting

Between 2012 and 2015, some changes occurred on the EPFL campus. We forecast the variations in market shares related to the self-services prices' increase in Section 4.6.1 and to the programmed opening of new catering destinations in Section 4.6.2.

4.6.1 Sensibility to the price

The price policy on the campus obliges some self-services to serve a subsidized menu. Between 2008 and summer 2014, the price was set to 7 CHF for the students and 8 CHF for the employees. Since the beginning of the 2014/2015 academic year, both these prices increased by 1 CHF. We apply this change to the most complete model (model with both first choice and most frequently chosen destinations corrections) to see the impact of this measure on the market shares. We expect a small variation due to the low significance of the prices for both employees and students.

Figure 13: Market shares after the increase of prices in self-services

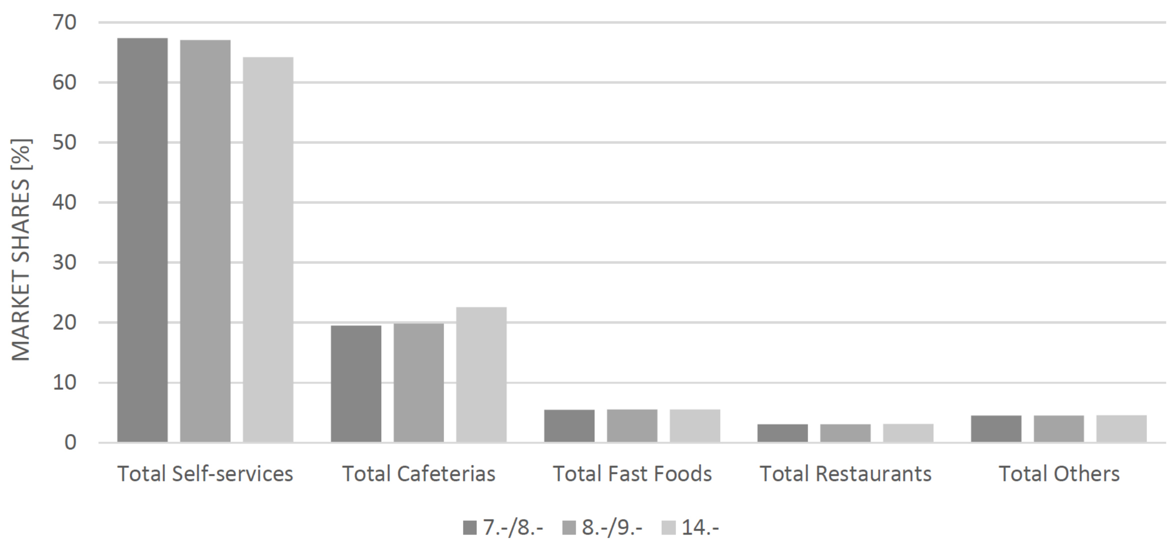


Figure 13 shows the market shares, during lunch hours, for each type of destination. The variation is negligible. Few people are more likely to select a cafeteria instead of a self-service. Within self-services, the price of L'Ornithorynque remains unchanged (i.e., 7.65 CHF) and thus, the destination attracts slightly more individuals (+2%).

Even if the increase of prices on the campus did probably not change the market shares importantly, we think that the price variable in the model is not robust enough to allow accurate price policy forecasts. Significant changes in market shares only occur with radical changes (e.g., doubling the prices starts to motivate individuals to chose another destination type, often a cafeteria).

4.6.2 Opening of a fast food

The EPFL is constantly evolving. In summer 2013, new facilities opened north of the campus (the Arcades) and on the campus itself. It includes a conference center, a hotel, shops and new catering facilities. One of these facilities that used to sell cold food and tee (Tekoe) closed and lets its place to a Swiss fast food company named Holy Cow. It serves “good and healthy” burgers and is popular among students. The details of the new destination are available in Table 29 in Appendix B.

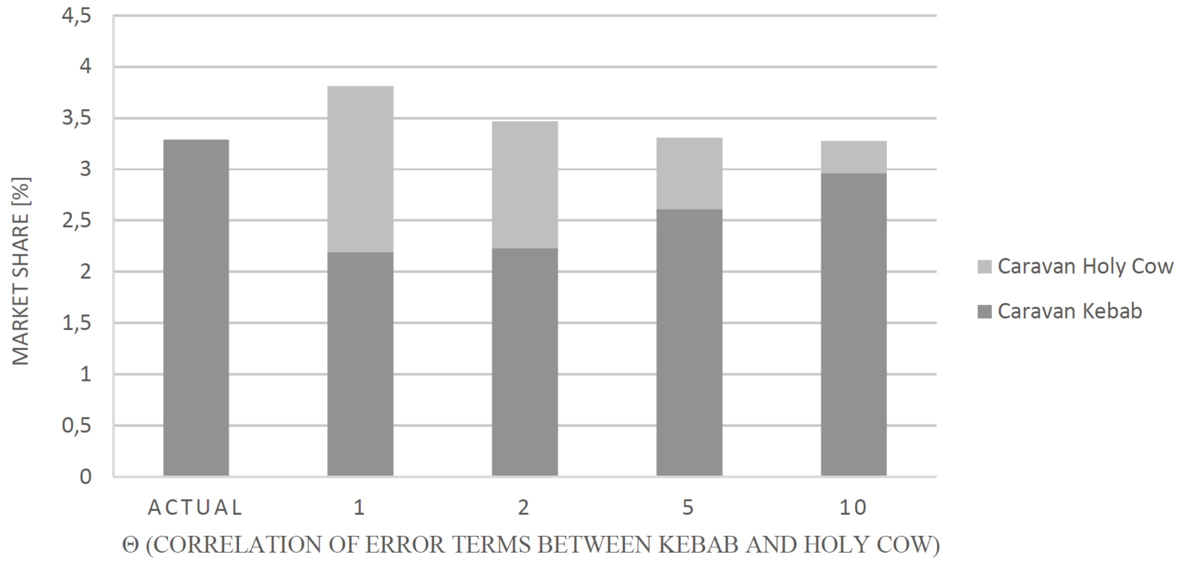
We propose to apply our best model (model with both first choice and most frequently chosen destinations corrections), adding this new alternative to simulate its estimated market share among the population, according to our model specification. Usually, forecasting is performed via revealed (RP) and stated (SP) preference surveys. Having these two sets of data permit to adjust the alternative specific constant (ASC) of the existing and predicted alternatives (Cherchi and de Dios Ortúzar, 2006). Since the reproduction of market shares can be sensitive to the ASC, it is necessary to address this issue.

We haven’t performed any SP data collection and we perform forecast based uniquely on RP data from WiFi traces. It shows the opportunities of such data in forecasting market shares. We first assume that the type of the opening destination is “caravan” (i.e., fast food). Also, we suppose that its alternative specific constant and sigmas are the same as the Kebab caravan (which also used to sell burgers) because it is the most look-alike existing catering destination. However, this approach leads to an issue: the errors terms of the new alternative and the borrowed existing alternative may be correlated.

We suggest that these two destinations are included in a nest (thus a Nested Logit (NL) specification). Since there is no way of estimating the nest parameter θ, we try an interval of values between 1 (MNL with independent error terms) and $+\infty$ (NL with perfectly correlated error terms). With $\theta = 1$, the model is a logit, and the new alternative gathers shares from all other alternatives. When $\theta \to \infty$, the new burger restaurant will mostly gather shares from the other burger alternative. We cannot know the exact value of θ and must evaluate the impact of all its different values on market shares predictions.

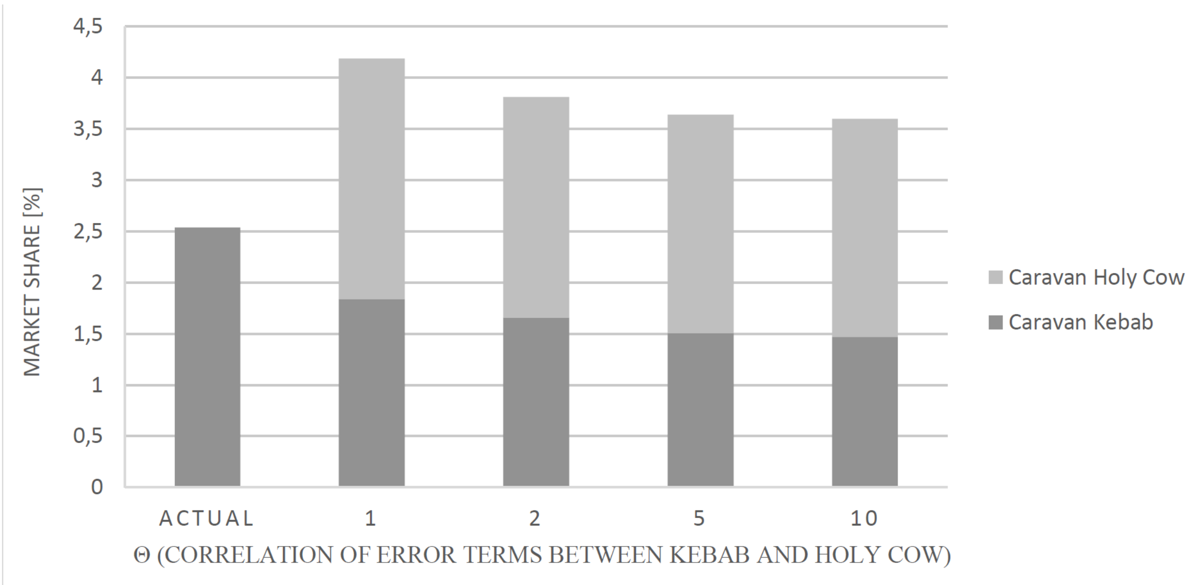
We compare the estimated market shares for different values of θ in Figure 14 during lunch hours. One can see that the more correlated the error terms are (i.e., the largest θ is), the least individuals visit Holy Cow (because if individuals have to choose between two identical destinations, they will select the cheapest). As we do not know the real value of the parameter, one estimates that the new Burger can attract between 0.3% ($\theta = 10$) and 1.6% ($\theta = 1$) people among the sample.

Figure 14: Market shares of destinations Kebab and Holy Cow (lunch hours) depending on the value of θ



It corresponds to a market share similar to the ones of restaurants and caravans. As the fast food is open from 11AM to 10PM, one also estimates the market share during the complete day.

Figure 15: Market shares of destinations Kebab and Holy Cow (all day) depending on the value of θ

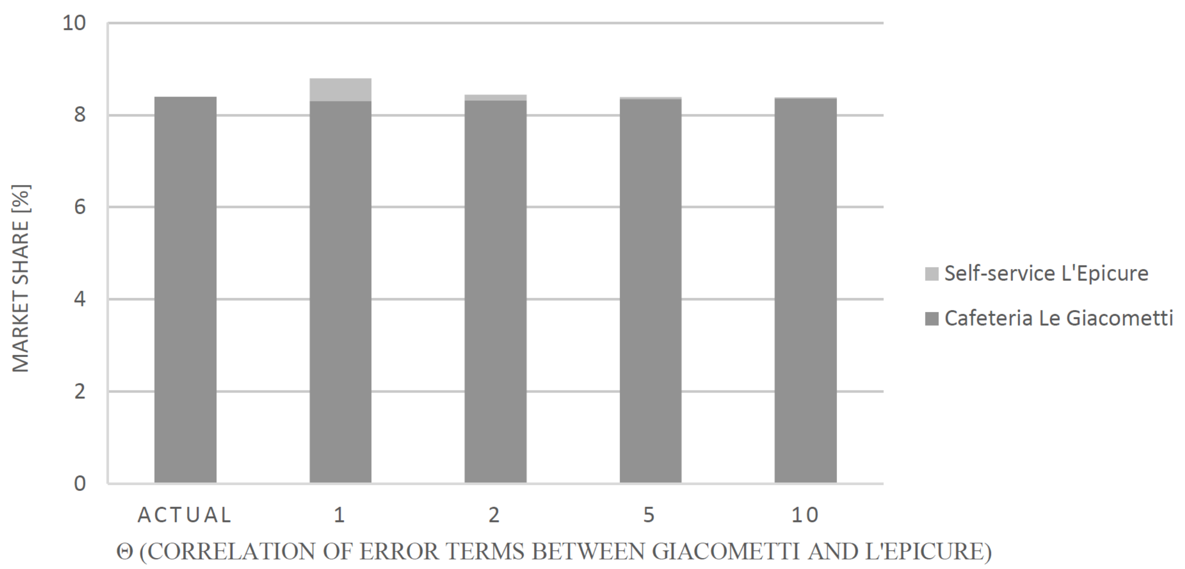


The results are different. Holy Cow is capable to maintain a market share of 2.2% for the complete day, independently of the correlation with the caravan Kebab whereas they were highly correlated during lunch hours. The large market share for the full day compared to lunch break is due to the availability of dinner and proximity from the previous activity episode, both significant in the model. We think that it may be realistic to consider that Holy Cow attracts a lot of people for dinner.

4.6.3 Opening of a self-service

We also forecast the impact on the market shares of the opening of a new self-service named L'Epicure. The model has more determinants concerning cafeterias and self-services. L'Epicure serves salads (menu), snacks, coffee and deserts thus, it looks more like a cafeteria. The details of the new self-service are available in Table 29 in Appendix B. We apply the same methodology as before. We propose to borrow the alternative specific parameters from cafeteria Le Giacometti because they are both destinations selling the same kind of food and similar opening hours.

Figure 16: Market shares of destinations Le Giacometti and L'Epicure (lunch) depending on the value of θ



One sees that the market share of self-service L'Epicure varies between 0.15% and 0.5% during lunch hours. It is small but understandable because the number of similar establishments is high. Also, the price is relatively expensive for a salad (8.5 CHF) which makes individuals chose another destination. We suppose that self-service l'Epicure and cafeteria Le Giacometti are not too correlated, thus the expected market share is close from 0.5%. Figure 17 considers the market shares during the complete day.

Figure 17: Market shares of destinations Le Giacometti and L'Epicure (all day) depending on the value of θ

If one considers the complete day, the market share may reach between 0.6% and 1.6%, with a likely value close to 1%. It is small as well. The destination is far from the offices which can explain why it does not attract many visitors.

For both Holy Cow and L'Epicure, the expected market shares are sensitive to the correlation of their error terms with the borrowed ones. Also, the fact that these destinations are not part of people's habits makes that they are less likely to visit them.

4.6.4 Absolute market shares

We estimate that Holy Cow may reach a daily market share of 2.2% whereas L'Epicure may only reach 1%. Now, we make the assumption that the forecast samples are representative of the population on the campus and that everybody visits a catering destination to eat. Market shares are multiplied by the complete population (i.e., 12'500 people) to represent the results as an absolute number of visitors.

According to Table 19, Holy Cow would serve between 266 and 293 hamburgers per day. It

Table 19: Absolute number of visitors: all 12'500 people on the campus eat at a catering destination. The daily demand repartition and the market shares of the forecast samples are representative of the total population.

Catering destination	Period of the day	θ=1	θ=2	θ=5	θ=10
Caravan Holy Cow	Lunch only	202	155	87	39
	Complete day	293	270	266	266
Self-service L'Epicure	Lunch only	65	16	5	3
	Complete day	198	111	85	81

is relatively small compared to the 750 daily menus they produce in their other restaurants[6]. We do not have similar numbers for catering destinations on the campus. We assume that they also have more daily visitors than the 85 to198 visitors predicted at Self-service L'Epicure. The results show that L'Epicure will have more difficulties to find its costumers than Holy Cow. Forecasts during lunch period are sensitive to the correlation between error terms and thus hard to interpret.

4.7 Limitations of the case study

The validation revealed encouraging results. However, we emphasize some limitations that we encountered during the case study. They are related to the number of observations and individuals, the nature of data and the models specification.

The number of observations in catering destinations is relatively small (i.e., about 2000 observations performed by 200 individuals). We built weekly models (i.e., one for each day of the week) and daily seasonal models (i.e., one for each period of the day, see Section 3.3.1) but the number of observations was too small for the models to converge properly. This is due to two reasons: (1) some destinations are absent from some submodels and (2) most robust parameters in the complete model are not/less significant in those submodels because of the lack of observations. We managed this situation by working on full models only. The problem of these models is that their runtime is long (i.e., between ten hours and two days for the models with agent effect correction, depending on the computational power and the complexity of the correction) and that they do not perform so well when the density of observations is small (i.e., during the afternoon). In our case, the seasonality in term of period of the day is strong (i.e., the behavior of individuals changes depending of the time of the day), so that building models for each period of the day would have been much more convenient.

[6] 2250 hamburgers served per day in 2012, in 3 restaurants (Source: `http://go.epfl.ch/holycow`)

The fact that the dataset is small also influences some of our specification. It is especially true with the Wooldridge correction terms (see Section 4.4.1). The first choice made by individuals is significant in our model with single agent effect correction. It means that the first observation usually fits the actual choice. It may be due to luck because we think that this parameter do not make that much sense in our case study (i.e., it would have been more convenient to observe the first destination choice of the semester). What's more we made a short sensitivity analysis on the second agent effect correction (i.e., the most frequented destination). When two destinations are previously selected the same number of times, we randomly select the most frequently chosen. It occurs quite often in our dataset, so that we decided to generate five samples. After we estimate our model with every samples, it appeared that the final log-likelihood varied significantly.

The nature of data does that we do not know exactly the purpose of each visit. It is especially true after lunch when the only significant variables introduced in the model are the distance and the availability of tap beer. Also, we could not obtain the list of menus that were served during the observation period (because of privacy and missing information) so we admitted that individuals would order the cheapest hot meal.

There are still some points that may be improved concerning the definition of the previous choice. As explained in Section 4.4.1, the definition of time is not strict between activity-episode sequences. The time interval between actual, previous and first choices varies. Fixing time intervals between consecutive activity-episodes (e.g., one day or one week) was not significant (because of the lack of observations?). One suggests that a better solution is found to clearly model the influence of habits over time.

Furthermore, some destinations still have a strong alternative specific constant t-test (see Appendix B). The number of determinants for these alternatives is too small to describe properly the possible reasons of the choice. Also, activity-episodes happening after lunch lack of explanatory variables. As explained in Section 4.4.2, the problem is that destination specific parameters defining ancillary services and habits are not significant during this period of the day.

To perform the validations we split the total sample into two small samples (see Table 17). Thus, we are only able to say that our models are accurate during the three months of the observation period. It would be interesting to collect similar data from a latest period to measure the ability at forecasting of the models with a new dataset.

4.8 Future utilization of the data

Although the goal of the presented case study is to test and validate the methodology developed in Section 3, we can think of some direct applications on the EPFL campus. In Section 4.6.1 and Section 4.6.2, we show how we are able to perform some basic forecasting. Here, one proposes a quick overview of possible future applications:

1. A decision help tool for EPFL restauration service (new policy): we think that including point-of-sale data ("camipro")[7] could highly improve the models because one could know what tracked people actually bought in a catering destination. The restauration service could use it in order to predict some price policies or change on the Campus. It would however require a lot of authorizations to obtain this information.
2. A model to estimate the influence of queues among students and employees: Pocketcampus[8] is developing a "No queue" smart phone application. It aims at tracking how big the queues (via Bluetooth) at catering destinations are, so that individuals can make a decision. Using these data could permit to include congestion in future development.

The data (a dataset and a dynamic model) are available online in Tinguely (2015).

4.9 Recommendations

We base on the experience and limitations obtained from this case study (see Section 4.7) to formulate some recommendations. They are related to the improvement of the existing destination choice models.

- Add more socio-economic attributes to the WiFi traces.
  - The age, the occupation of employees or the gender would permit to improve the models and to measure the possible heterogeneity within the population.
- Collect data for a one year period.
  - It is likely that the habits change every semester for students because of the schedules. Also, destinations with high inside capacities are probably more visited during winter.
- Try to enrich the data with additional tools.
  - As specified in Section 4.8, adding Camipro data or information on queues would permit to improve significantly the models.

[7] Students and employees can pay the majority of their transactions with a personal electronic card.

[8] Pocketcampus is a smart phone application developed by a team of graduate students. It proposes information about the campus (map, events), academic life (schedule, moodle) or catering destinations (evaluation, menus), `http://www.pocketcampus.org/`

5 Recommendations

We formulate some recommendations in order to simplify the future application (e.g., in multimodal transport hubs) of the methodology developed in Section 3. Section 5.1 advises on data collection, Section 5.2 on data processing, Section 5.3 on destination choice modelling and Section 5.4 on forecasting.

5.1 Collection of data

The first step is the data collection. It is necessary to directly specify the future need in data in order to get what is needed. We propose to:

- Add socio-economic attributes to the measurements.
  - Although measurements from WiFi or Bluetooth are usually anonymized, one suggests that as many socio-economic attributes as possible are integrated (e.g., age, gender, occupation, ...). It permits to alleviate the alternative specific constants and to consider segmental market shares (e.g., commuters vs leisures vs employees, young people vs old people,...) in the modelling part. This information can simply be asked during the (mandatory?) registration to the wireless network.
- Collect data for at least a complete year.
  - We consider that the influence of weather and holidays must be accounted for. It permits to observe seasonal variations and change of habits over time. The longer, the better.
- Track at least 1000 individuals.
  - One recommends to track as many individuals as possible to perform accurate estimating and forecasting. Also, it is important to have a representative sample of the population (i.e., commuters, travelers, employees,...).

5.2 Data processing

When the data are collected, one suggests to apply the methodology developed by Danalet *et al.* (2014) to generate candidates of activity-episode sequences from the measurements. Then, one advises to:

- Consider areas of interest (*AOI*) instead of points of interest (*POI*)

  - As discussed in Section 2.3, the intersection between the domain of data relevance (*DDR*) and *POI* can be empty. It means that some measurements are not connected to any *POI*. It can happen when the confidence factor (cF) is high. One recommends to use *AOI* instead of *POI* to address this issue.
- Merge and improve the data.
  - Follow the methodology developed in Section 3 to translate candidates of activity-episode sequences into destination type specific datasets. It requires to collect destination specific attributes for each destination from each destination type. It may take time to gather these attributes, especially if there are many destinations.
- Perform descriptive statistics.
  - It is absolutely necessary to visualize the data before modelling destination choice. Performing simple data processing (e.g., pivot table, graphics,...) can directly show which variables have an influence on the destination choice. It also allows to verify if the sample is representative of the total population.

5.3 Destination choice modelling

Once the datasets are ready, we propose to start the destination choice modelling. In particular, we suggest to:

- Model the measurement equation
  - In the context of railway stations or airports, one suggests that the measurement equation is used to account for the probabilistic nature of the Bayesian approach and the low quality of measurements. Indeed, different activities and different destinations can be really close of each other (e.g., see Figure 4) which may lead to conflicting candidates of activity-episode sequence. This was unlikely in the case study presented in Section 4. The use of a measurement equation addresses this issue.
- Account for the panel nature of data.
  - We emphasize in this paper the significance of habits. Intuitively, it may also be the case in a multi-modal transport framework. The approach developed in Section 3.3.2 may be followed. Also, one suggests to explore different types of routine (i.e., routine concerning the period of the day and the day of the week). Strong seasonalities are likely to be observed, particularly among employees and commuters. We propose to impose a periodicity between successive activity-episode sequences to have a strict definition of time.
- Develop the Wooldridge correction.

  - As shown in Section 4.5, the addition of the first and most frequent choices not only corrects for agent effect issue but also improves significantly the models. It is recommended to add more of these average variables as discussed in Section 3.3.2 (e.g., average distance covered by each individual).

5.4 Forecasting

After understanding pedestrians behavior, forecasting is the second purpose of destination choice modelling. We propose to:

- Conduct a stated preference (SP) survey to improve forecasting.
  - We show in Section 4.6 that forecasting is subject to some strong hypothesis. We can relax them by asking people their opinions (e.g., on the opening of new infrastructures). SP surveys also allow to better understand the reasons of people's decisions (e.g., in the case study (see Section 4), we assume that individuals order the cheapest menus which have a low variability. This assumption makes that the price only has little impact in the models).
- Explore segmental forecasting.
  - The acceptation of a new price policy or the move of a ticket machine can lead to different reactions among the population. Thus, it is necessary to perform the most disaggregated forecasts as possible.

6 Conclusion

My master thesis proposes a framework to model pedestrians destination choice in multi-model facilities from WiFi localization. It is part of Danalet *et al.* (2014) research to model activity-episode sequences when observing measurements from antennas. The probability of reproducing the observations depends on the destination choice which is nested to the activity choice and a measurement equation. Here, one develops a methodology (see Section 3) to model destination choice considering that the activity is known.

The input of my work consists of activity-episode sequences. One activity type (e.g., eating) represents one destination choice model. Once it is selected, all the possible destinations to perform this activity are considered. The attributes to explain destination choice are collected. These attributes are either sequence specific (e.g., ID, distance, socio-economic determinants), activity-episode specific (e.g., location, start and end times) or destination specific (e.g., opening hours, capacity, price).

Individuals may have multiple sequences over the observed time period. Panel nature of data and how to correct agent effect issue are accounted for using Wooldridge (2002) approach. Three types of models are developed: a *static model*, a *dynamic strict exogenous model* and two variants of *dynamic model with panel data and agent effect correction* (thus, a total of four models). They reveal the importance of the previous choice, and of habits among the observation period. We emphasize that taking into account the previous choice and correcting for agent effect issue contribute to improve significantly the fitting of destination choice models for pedestrians on the data.

We present a case study on the EPFL campus (see Section 4). Eating is considered as the activity type. 21 eating establishments represent the destination choices for this activity. These destinations are decomposed into types (i.e., cafeteria, self-service,...) depending on the services they propose. We develop two Multinomial and two Mixed Logit models following the steps described before.

The models reveal four major points (see Section 4.4.2). First, individuals prefer destinations close to their previous activity-episode. It means that they reduce the distance to walk for reaching an eating establishment. This is especially observed when people need to chose for a destination to have coffee in the morning and to have lunch in a cafeteria or a self-service. Individuals are more willing to walk if they want to visit a destination serving a specific type of food. Second, the choice of a catering destination to lunch or take a coffee in the morning at time t is connected to the previous catering destination choice performed at time $t-1$. Indeed, if one eating establishment has been visited the previous time, it is more likely to be chosen again.

The results show that correcting agent effect leads to restore the orthogonality between error terms but also and above all to improve significantly the fit of the models with the data. In particular, we suggest that the agent effect correction is as complete as possible. The most frequented destination before the actual choice is a particularly significant correction. Third, we observe strong seasonalities during the day. Individuals have different behaviors in the morning, lunch and afternoon because the purpose of the visit is not the same. Fourth and final point, ancillary services (e.g., selling sandwiches, having a fidelity card. . .) and prices do not seem to influence people's choice because destinations of a same type all propose the same range of services (and prices).

We split our dataset into two samples to validate the models (see Section 4.5). The first sample contains the past observations of the individuals and is used to estimate the parameters. The second one represents the choice that is about to be done by the individuals and permits to simulate the decisions, according to the estimated models. We consider that the models are successfully validated since the predicted market shares of the catering destinations are similar to the observed ones. The model that performs the best (i.e., the model that minimizes the difference between observations and predictions) is the one with the most complete agent effect correction (i.e., first and most frequent choices).

We keep this model to perform some forecasting on the EPFL campus (see Section 4.6). The impact of a recent change in price policy and the opening of two new catering destinations (a fast food and a self-service) is estimated. The non-significance of prices makes that the increase of prices does not affect market shares much. This is an issue of our models. It could be solved by complementing WiFi data with a stated preference (SP) survey. Concerning the opening of the new facilities, we are able to predict an interval of market shares depending of the correlation with the most lookalike destination. It is however hard to determine if those results are realistic. A stated preference survey would address this problem.

Some limitations have been revealed during the case study (see Section 4.7). The number of observations and individuals is relatively small. It prevents from doing submodels. This is an issue in case of distinctive behavior depending on the day of the week or the period of the day. Furthermore, the nature of data makes that one does not know exactly the purpose of each visit. Thus, some variables are destination specific determinants (e.g, the price, evaluations) whereas they should ideally be individual specific (i.e., individuals do not all visit a destination to order the cheapest meal and do not have the same feeling on destinations). Some periods of the day still lack of specifications.

Of course, it is still possible to improve the models (see Section 4.8). Destination choice models usually consider Space Syntax parameters. We did not implement such determinants in the models but we think that they may be significant. One could merge WiFi localization with point-of-sale data (called Camipro on EPFL campus) in order to model specifically the goal of the visit of a catering destination (e.g., buy a sandwich, a coffee or a menu). This would however invade individuals' privacy. Also, considering queues would permit to account for congestion.

In future works, the methodology itself should be enhanced.As a first approach, we assumed in the case study that time interval between consecutive choices was undefined (mainly because of the lack of data). Time between activity-episode sequences should be clearly defined to measure the impact of time (see Section 4.4.1). We present a full equation in Section 3.1. It considers the nested activity and destination choices and the measurement equation together. One should model everything together in a simple case (i.e., with only two activity types and only few destinations).

My thesis reveals that WiFi localization is suitable to model destination choice for pedestrians. We should consider applying the methodology to a multi-modal facility context. Railway stations, airports, stores or public buildings are as many new opportunities to understand and model pedestrian destination choice using WiFi localization.

We think that the methodology is well fitted to be adapted in those facilities (see Section 5). First, it would allow to understand pedestrians' behavior (e.g., sensitivity to the distance, daily routine,...). Second, it is useful to optimize the organization of existing infrastructure (e.g., what happens if we move the location of a ticket machine?) and to design effective new facilities (e.g., from our model we find out that people are really sensitive to travel time, so we suggest to add treadmills). The utilization of wireless measurements is likely to be a widespread practice in future years.

7 References

Aurélie Glerum (2014) Static and dynamic mathematical models of behavior, Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne.

Bekhor, S., Y. Cohen and C. Solomon (2013) Evaluating long-distance travel patterns in Israel by tracking cellular phone positions, *Journal of Advanced Transportation*, **47** (4) 435–446, June 2013, ISSN 01976729.

Bierlaire, M. (2003) BIOGEME: a free package for the estimation of discrete choice models, paper presented at the *Proceedings of the 3rd Swiss Transportation Research Conference*, Monte Verità, Ascona, Switzerland.

Bierlaire, M. (2014) Discrete panel data, *Mathematical modeling of behavior - EPFL course*.

Bierlaire, M. and M. Fetiarison (2009) Estimation of discrete choice models: extending biogeme, *Swiss Transport Research Conference (STRC)*, (September).

Bierlaire, M. and E. Frejinger (2008) Route choice modeling with network-free data, *Transportation Research Part C*, **16** (2) 187–198, April 2008, ISSN 0968090X.

Bierlaire, M. and T. Robin (2009) Pedestrians Choices, paper presented at the *Pedestrian Behavior. Models, Data Collection and Applications*, 1–26, ISBN 978-1-84855-750-5.

CFF (2011) La gare, une VIP: au Coeur des CFF, `http://www.cff.ch/content/dam/sbb/fr/pdf/fr_sbb-konzern/fr_ueber-die-sbb/fr_organisation/Nos+gares.pdf`.

CFF (2014) Léman 2030, `http://www.cff.ch/groupe/entreprise/projets/extension-du-reseau-ferroviaire/leman-2030/apercu/leman.html`.

Chen, J. (2013) Modeling route choice behavior using smartphone data, Ph.D. Thesis, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Cherchi, E. and J. de Dios Ortúzar (2006) On fitting mode specific constants in the presence of new options in RP/SP models, *Transportation Research Part A: Policy and Practice*, **40**, 1–18, ISSN 09658564.

Cisco (2011) Cisco MSE API Specification Guide - Location Service, Release 7.1, *Technical Report*, Cisco, April 2011.

Daamen, W. (2004) Modelling Passenger Flows in Public Transport Facilities, Ph.D. Thesis, Delft University of Technology, The Netherlands.

Dalumpines, R. (2014) GIS-based episode reconstruction using gps data for activity analysis and route choice modeling, Thesis, McMaster.

Danalet, A. (2015) A Bayesian Approach to Detect Pedestrian Destination-Sequences from WiFi Signatures: Data, `http://zenodo.org/record/15798`.

Danalet, A. and M. Bierlaire (2015a) Activity path size for correlation between activity paths.

Danalet, A. and M. Bierlaire (2015b) Importance sampling for activity path choice, paper presented at the *15th Swiss Transport Research Conference (STRC)*, Monte Verità, Ascona, Switzerland.

Danalet, A., B. Farooq and M. Bierlaire (2014) A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures, *Transportation Research Part C: Emerging Technologies*, **44**, 146–170.

Deutsch, K., S. Ravulaparthy and K. Goulias (2014) Place happiness: it's constituents and the influence of emotions and subjective importance on activity type and destination choice, **12**, 1323–1340, ISSN 15729435.

Frejinger, E. (2008) Route choice analysis: Data, models, algorithms and applications, Ph.D. Thesis, EPF Lausanne, Lausanne, April 2008.

Goetz, M. and A. Zipf (2011) Formal definition of a user-adaptive and length-optimal routing graph for complex indoor environments, *Geo-spatial Information Science*, **14** (2) 119–128, January 2011, ISSN 1009-5020.

Helbing, D. (1997) Traffic Dynamics: New Physical Modeling Concepts (In German), *Springer-Verlag*.

Hillier, B. (2005) The art of place and the science of space, 24–34, ISSN 0956-9758.

Hoogendoorn, S. P. and P. H. L. Bovy (2004) Pedestrian route-choice and activity scheduling theory and models, *Transportation Research Part B: Methodological*, **38**, 169–190, ISSN 01912615.

Hoogendoorn, S. P., P. H. L. Bovy and W. Daamen (2002) Microscopic pedestrian wayfinding and dynamics modelling, in M. Schreckenberg and S. D. Sharma (eds.) *Pedestrian and Evacuation Dynamics*, 123–155, Springer.

Kalakou, S. and F. Moura (2014a) Bridging the Gap in Planning Indoor Pedestrian Facilities, *Transport Reviews*, **34** (May 2015) 474–500, ISSN 0144-1647.

Kalakou, S. and F. Moura (2014b) Effects of terminal planning on passenger choices, paper presented at the *14th Swiss Transport Research Conference (STRC)*, Monte Verità, Ascona, Switzerland.

Kim, J. and J. Hespanha (2003) Discrete approximations to continuous shortest-path: application to minimum-risk path planning for groups of UAVs, *42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475)*, **2** (0), ISSN 0191-2216.

Liu, X. (2013) Activity-based pedestrian behavior simulation inside intermodal facilities, Ph.D. Thesis, Mississippi State University.

Liu, X., J. M. Usher and L. Strawderman (2014) An analysis of activity scheduling behavior of airport travelers, *Computers and Industrial Engineering*, **74**, 208–218, ISSN 03608352.

Miller, H. J. (2010) Measuring Space-Time Accessibility Benefits within Transportation Networks: Basic Theory and Computational Procedures, *Geographical Analysis*, **31** (1) 1–26, September 2010, ISSN 00167363.

Nandakumar, R., K. K. Chintalapudi and V. N. Padmanabhan (2012) Centaur: locating devices in an office environment, *Proc.18th Ann. Int. Conf. Mob. Comput. Netw. (MobiCom'12)*, 1–12.

OFS (2014) Prestations de transport de personnes, `http://www.bfs.admin.ch/bfs/portal/fr/index/themen/11/04/blank/01.html`.

Pettersson, P. (2011) Passenger waiting strategies on railway platforms - Effects of information and platform facilities, Master thesis, KTH.

Pirotte, A. (1996) Estimation de relations de long terme sur données de panel : nouveaux résultats, *Économie & Prévision*, **126**, 143–161, ISSN 0249-4744.

Sen, S., J. Lee, K.-h. Kim and P. Congdon (2013) Avoiding Multipath to Revive Inbuilding WiFi Localization, *Proc. of MobiSys*, 249.

Tinguely, L. (2015) Destination Choice Model including panel datausing WiFi localization in a pedestrian facility: data, `https://zenodo.org/record/18528`.

Tinguely, L., M. de Lapparent, A. Danalet and M. Bierlaire (2015) Destination Choice Model including panel data using WiFi localization in a pedestrian facility, *Swiss Transport Research Conference (STRC)*, (April).

Ton, D. (2014) NAVISTATION: a study into the route and activity location choice behaviour of departing pedestrians in train stations, Master thesis, Delft University of Technology.

Ueno, J., A. Nakazawa and T. Kishimoto (2009) An Analysis of Pedestrian Movement in Multilevel Complex by Space Syntax Theory - In the Case of Shibuya Station, *Proceedings of the 7th International Space Syntax Symposium*, 1–12.

Walker, J. L. (2001) Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables, Ph.D. Thesis, Massachusetts Institute of Technology.

Wooldridge, J. M. (2002) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, *Journal of applied econometrics*, 44, ISSN 1753-9196.

Worldbank (2014) Air transport, passengers carried, `http://data.worldbank.org/indicator/IS.AIR.PSGR/countries?display=graph`.

Zhang, L., Y. Zhuang and X. Dai (2012) A configurational study of pedestrian flows in multilevel commercial space. Case study Shanghai, *Eighth International Space Syntax Symposium*, 8044:1–16.

A Detailed utility functions

We show here a generic utility function (Eq. (20)). The types are either self-service, restaurant, cafeteria, caravan or others. The 21 alternative description and variables are available in Section 4.1. Two forecasting scenarios (see Section 4.6.2) include a 22nd alternative (Holy Cow and L'Epicure). The daily seasonalities are morning (7 AM to 11:30 AM), lunch (11:30 AM to 2 PM) and afternoon (2 PM to 10 PM).

$$
\begin{aligned}
V_d = ASC_d &+ \beta_{DISTANCE\_LUNCH_{TYPE}} * lunch\_distance\_d \\
&+ \beta_{DISTANCE\_MORNING} * morning\_distance\_d \\
&+ \beta_{DISTANCE\_AFTERNOON} * afternoon\_distance\_d \\
&+ \beta_{NO\_DISTANCE\_AV} * distance\_not\_av\_d \\
&+ \beta_{EVALUATION_{TYPE}} * evaluation\_survey\_2013\_d \\
&+ \beta_{PRICE\_STUDENT} * lunch\_price\_min\_student\_d \\
&+ \beta_{PRICE\_EMPLOYEE} * lunch\_price\_min\_employee\_d \\
&+ \beta_{TAP\_BEER\_AFTER\_LUNCH} * beer\_after\_lunch\_filter\_d \\
&+ \beta_{DINNER} * dinner\_filter\_d \\
&+ \beta_{METEO\_TERRACE} * meteo\_terrace\_filter\_d \\
&+ \beta_{CAPACITY\_INSIDE} * cap\_inside\_filter\_d \\
&+ \rho_{PREVIOUS\_CHOICE\_MORNING} * previous\_choice\_filter\_morning\_d \\
&+ \alpha_{MOST\_FREQUENT\_CHOICE\_MORNING} * most\_frequent\_choice\_filter\_morning\_d \\
&+ \alpha_{FIRST\_CHOICE\_MORNING} * first\_choice\_filter\_morning\_d + \mathcal{N}(0, \sigma^2_{morning_d}) \\
&+ \rho_{PREVIOUS\_CHOICE\_LUNCH} * previous\_choice\_filter\_lunch\_d \\
&+ \alpha_{MOST\_FREQUENT\_CHOICE\_LUNCH} * most\_frequent\_choice\_filter\_lunch\_d \\
&+ \alpha_{FIRST\_CHOICE\_LUNCH} * first\_choice\_filter\_lunch\_d + \mathcal{N}(0, \sigma^2_{lunch_d}) \\
&+ \rho_{PREVIOUS\_CHOICE\_AFTERNOON} * previous\_choice\_filter\_afternoon\_d \\
&+ \alpha_{MOST\_FREQUENT\_CHOICE\_AFTERNOON} * most\_frequent\_choice\_filter\_afternoon\_d \\
&+ \alpha_{FIRST\_CHOICE\_AFTERNOON} * first\_choice\_filter\_afternoon\_d + \mathcal{N}(0, \sigma^2_{afternoon_d})
\end{aligned} \tag{20}
$$

Other models were developed. In particular, daily models (i.e., one for Mondays, Tuesdays…) and daily seasonal models (i.e., one for the morning, lunch and afternoon). The small sample size makes that they do not work well (problem of convergence or non-significance of important parameters). If there is enough data, one recommends to develop smaller (disaggregated) models.

B Detailed results

B.1 Detailed estimations

Table 20: Static model

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC_ARC	-1.47	0.318	-4.60	0.00
2	ASC_ATL	-0.966	0.325	-2.97	0.00
3	ASC_BC	-0.369	0.397	-0.93	0.35
4	ASC_BM	0.666	0.324	2.06	0.04
5	ASC_COP	1.03	0.590	1.74	0.08
6	ASC_COR	-0.235	0.141	-1.67	0.10
7	ASC_ELA	-1.33	0.435	-3.06	0.00
8	ASC_GIA	0.204	0.392	0.52	0.60
9	ASC_HOD	-0.130	0.393	-0.33	0.74
10	ASC_INM	-2.92	0.608	-4.81	0.00
11	ASC_KEB	0.770	0.247	3.11	0.00
12	ASC_KLE	-3.34	0.647	-5.17	0.00
13	ASC_MX	-1.34	0.351	-3.81	0.00
14	ASC_ORN	-0.797	0.134	-5.93	0.00
15	ASC_PAR	-0.381	0.268	-1.42	0.15
16	ASC_PH	1.36	0.323	4.23	0.00
17	ASC_PIZ	0.980	0.237	4.14	0.00
18	ASC_SAT	-1.32	0.473	-2.79	0.01
19	ASC_VAL	1.49	0.734	2.02	0.04
20	ASC_VIN	-4.02	0.715	-5.62	0.00
21	BETA_CAPACITY_INSIDE	0.00277	0.00257	1.08	0.28
22	BETA_DINNER	0.943	0.289	3.26	0.00
23	BETA_DISTANCE_AFTERNOON	-0.000606	0.000545	-1.11	0.27
24	BETA_DISTANCE_LUNCH_CAF	-0.00703	0.000506	-13.88	0.00
25	BETA_DISTANCE_LUNCH_REST	-0.00276	0.00128	-2.16	0.03
26	BETA_DISTANCE_LUNCH_SELF	-0.00646	0.000418	-15.45	0.00
27	BETA_DISTANCE_MORNING	-0.00379	0.000826	-4.59	0.00
28	BETA_EVALUATION_CAFET	1.79	0.0929	19.26	0.00
29	BETA_EVALUATION_SELF	1.88	0.125	15.04	0.00
30	BETA_METEO_TERRACE	0.00162	0.000878	1.85	0.07
31	BETA_NO_DISTANCE_AV	-4.89	0.420	-11.66	0.00
32	BETA_PRICE_EMPLOYEE	-0.00537	0.0333	-0.16	0.87
33	BETA_PRICE_STUDENT	-0.0681	0.0369	-1.85	0.06
34	BETA_TAP_BEER_AFTER_LUNCH	0.669	0.180	3.71	0.00

Summary statistics

Number of observations = 1868

Number of estimated parameters = 34

$$\mathcal{L}(\beta_0) = -5035.914$$
$$\mathcal{L}(\hat{\beta}) = -3238.926$$
$$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 3593.005$$
$$\rho^2 = 0.357$$
$$\bar{\rho}^2 = 0.350$$

Table 21: Dynamic strict exogenous model

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC_ARC	-1.23	0.347	-3.53	0.00
2	ASC_ATL	-1.01	0.363	-2.78	0.01
3	ASC_BC	-0.227	0.458	-0.50	0.62
4	ASC_BM	0.823	0.340	2.42	0.02
5	ASC_COP	1.25	0.580	2.16	0.03
6	ASC_COR	-0.200	0.156	-1.28	0.20
7	ASC_ELA	-1.00	0.457	-2.20	0.03
8	ASC_GIA	0.394	0.414	0.95	0.34
9	ASC_HOD	0.110	0.397	0.28	0.78
10	ASC_INM	-2.42	0.647	-3.74	0.00
11	ASC_KEB	1.07	0.289	3.70	0.00
12	ASC_KLE	-2.83	0.661	-4.28	0.00
13	ASC_MX	-1.27	0.400	-3.19	0.00
14	ASC_ORN	-0.823	0.147	-5.61	0.00
15	ASC_PAR	-0.429	0.303	-1.42	0.16
16	ASC_PH	1.42	0.344	4.12	0.00
17	ASC_PIZ	1.15	0.259	4.43	0.00
18	ASC_SAT	-1.12	0.487	-2.29	0.02
19	ASC_VAL	1.88	0.733	2.57	0.01
20	ASC_VIN	-3.71	0.716	-5.19	0.00
21	BETA_CAPACITY_INSIDE	0.00386	0.00295	1.31	0.19
22	BETA_DINNER	0.865	0.281	3.07	0.00
23	BETA_DISTANCE_AFTERNOON	-0.00111	0.000558	-1.98	0.05
24	BETA_DISTANCE_LUNCH_CAF	-0.00629	0.000523	-12.02	0.00
25	BETA_DISTANCE_LUNCH_REST	-0.00271	0.00126	-2.16	0.03
26	BETA_DISTANCE_LUNCH_SELF	-0.00555	0.000429	-12.92	0.00
27	BETA_DISTANCE_MORNING	-0.00280	0.000775	-3.61	0.00
28	BETA_EVALUATION_CAFET	1.74	0.0982	17.68	0.00
29	BETA_EVALUATION_SELF	1.80	0.138	12.98	0.00
30	BETA_METEO_TERRACE	0.00215	0.000947	2.27	0.02
31	BETA_NO_DISTANCE_AV	-4.31	0.398	-10.83	0.00
32	BETA_PRICE_EMPLOYEE	-0.00490	0.0325	-0.15	0.88
33	BETA_PRICE_STUDENT	-0.0512	0.0354	-1.45	0.15
34	BETA_TAP_BEER_AFTER_LUNCH	0.490	0.183	2.67	0.01
35	RHO_PREVIOUS_LUNCH_CHOICE	1.78	0.114	15.61	0.00
36	RHO_PREVIOUS_MORNING_CHOICE	3.21	0.176	18.19	0.00

Summary statistics

Number of observations = 1868

Number of estimated parameters = 36

$$\mathcal{L}(\beta_0) = -5037.914$$
$$\mathcal{L}(\hat{\beta}) = -2870.976$$
$$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 4333.877$$
$$\rho^2 = 0.430$$
$$\bar{\rho}^2 = 0.423$$

Table 22: Dynamic model with agent effect correction (first choice only) (1): here the results with 250 draws (results are similar with more draws).

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ALPHA_FIRST_LUNCH_CHOICE	1.03	0.180	5.73	0.00
2	ALPHA_FIRST_MORNING_CHOICE	1.67	0.262	6.38	0.00
3	ASC_ARC	-3.47	0.830	-4.18	0.00
4	ASC_ATL	-2.84	1.02	-2.79	0.01
5	ASC_BC	-1.74	0.818	-2.13	0.03
6	ASC_BM	-1.33	0.899	-1.48	0.14
7	ASC_COP	-0.873	1.15	-0.76	0.45
8	ASC_COR	-0.707	0.260	-2.71	0.01
9	ASC_ELA	-1.97	0.729	-2.70	0.01
10	ASC_GIA	-0.198	0.599	-0.33	0.74
11	ASC_HOD	-0.181	0.634	-0.28	0.78
12	ASC_INM	-5.42	1.42	-3.82	0.00
13	ASC_KEB	1.12	0.456	2.45	0.01
14	ASC_KLE	-4.50	1.12	-4.03	0.00
15	ASC_MX	-5.06	0.811	-6.24	0.00
16	ASC_ORN	-1.43	0.312	-4.57	0.00
17	ASC_PAR	-1.21	0.450	-2.69	0.01
18	ASC_PH	0.698	0.488	1.43	0.15
19	ASC_PIZ	1.08	0.436	2.49	0.01
20	ASC_SAT	-2.68	0.727	-3.69	0.00
21	ASC_VAL	2.06	1.81	1.14	0.26
22	ASC_VIN	-16.4	22.3	-0.74	0.46
23	BETA_CAPACITY_INSIDE	0.00365	0.00394	0.93	0.35
24	BETA_DINNER	0.634	0.360	1.76	0.08
25	BETA_DISTANCE_AFTERNOON	-0.000765	0.000621	-1.23	0.22
26	BETA_DISTANCE_LUNCH_CAF	-0.00412	0.000669	-6.17	0.00
27	BETA_DISTANCE_LUNCH_REST	-0.00173	0.00222	-0.78	0.44
28	BETA_DISTANCE_LUNCH_SELF	-0.00393	0.000485	-8.09	0.00
29	BETA_DISTANCE_MORNING	-0.00241	0.00101	-2.39	0.02
30	BETA_EVALUATION_CAFET	2.28	0.206	11.09	0.00
31	BETA_EVALUATION_SELF	2.37	0.244	9.73	0.00
32	BETA_METEO_TERRACE	0.00258	0.00110	2.35	0.02
33	BETA_NO_DISTANCE_AV	-3.45	0.550	-6.28	0.00
34	BETA_PRICE_EMPLOYEE	-0.0238	0.0568	-0.42	0.68
35	BETA_PRICE_STUDENT	-0.0776	0.0572	-1.36	0.17
36	BETA_TAP_BEER_AFTER_LUNCH	0.759	0.249	3.05	0.00
37	RHO_PREVIOUS_LUNCH_CHOICE	0.507	0.162	3.12	0.00
38	RHO_PREVIOUS_MORNING_CHOICE	0.721	0.263	2.74	0.01

Table 23: Dynamic model with agent effect correction (first choice only) (2)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
39	SIGMA_LUNCH_ARC	1.53	0.664	2.31	0.02
40	SIGMA_LUNCH_ATL	2.21	0.521	4.25	0.00
41	SIGMA_LUNCH_BC	0.953	0.206	4.64	0.00
42	SIGMA_LUNCH_BM	-0.123	0.579	-0.21	0.83
43	SIGMA_LUNCH_COP	-3.47	0.893	-3.89	0.00
44	SIGMA_LUNCH_COR	0.847	0.454	1.87	0.06
45	SIGMA_LUNCH_ELA	-0.865	0.273	-3.17	0.00
46	SIGMA_LUNCH_GIA	0.584	0.251	2.33	0.02
47	SIGMA_LUNCH_HOD	-0.941	0.345	-2.73	0.01
48	SIGMA_LUNCH_INM	-2.40	0.636	-3.77	0.00
49	SIGMA_LUNCH_KEB	-2.84	0.470	-6.03	0.00
50	SIGMA_LUNCH_KLE	-1.88	0.296	-6.35	0.00
51	SIGMA_LUNCH_MX	0.0871	0.300	0.29	0.77
52	SIGMA_LUNCH_ORN	0.544	0.262	2.08	0.04
53	SIGMA_LUNCH_PAR	-1.02	0.245	-4.16	0.00
54	SIGMA_LUNCH_PH	1.86	0.560	3.32	0.00
55	SIGMA_LUNCH_PIZ	0.367	0.259	1.42	0.16
56	SIGMA_LUNCH_SAT	0.183	0.661	0.28	0.78
57	SIGMA_LUNCH_VAL	2.04	1.98	1.03	0.30
58	SIGMA_LUNCH_VIN	-5.68	9.44	-0.60	0.55
59	SIGMA_MORNING_ARC	-1.83	0.352	-5.20	0.00
60	SIGMA_MORNING_ATL	-0.124	0.779	-0.16	0.87
61	SIGMA_MORNING_BC	-2.07	0.295	-7.02	0.00
62	SIGMA_MORNING_BM	3.66	0.756	4.84	0.00
63	SIGMA_MORNING_COP	-0.735	0.697	-1.06	0.29
64	SIGMA_MORNING_COR	1.14	0.263	4.35	0.00
65	SIGMA_MORNING_ELA	0.571	0.337	1.70	0.09
66	SIGMA_MORNING_GIA	-1.15	0.272	-4.23	0.00
67	SIGMA_MORNING_HOD	-0.340	0.618	-0.55	0.58
68	SIGMA_MORNING_INM	-0.304	0.370	-0.82	0.41
69	SIGMA_MORNING_KEB	-1.79	0.793	-2.26	0.02
70	SIGMA_MORNING_KLE	0.321	0.547	0.59	0.56
71	SIGMA_MORNING_MX	3.28	0.463	7.08	0.00
72	SIGMA_MORNING_ORN	-0.538	0.273	-1.97	0.05
73	SIGMA_MORNING_PAR	-1.24	0.387	-3.20	0.00
74	SIGMA_MORNING_PH	0.915	0.928	0.99	0.32
75	SIGMA_MORNING_PIZ	-1.44	0.465	-3.10	0.00
76	SIGMA_MORNING_SAT	2.03	0.307	6.61	0.00
77	SIGMA_MORNING_VAL	-1.50	2.77	-0.54	0.59
78	SIGMA_MORNING_VIN	-7.22	11.4	-0.63	0.53

Summary statistics

Number of observations = 1868

Number of estimated parameters = 78

$$\mathcal{L}(\beta_0) = -5037.914$$
$$\mathcal{L}(\hat{\beta}) = -2352.137$$
$$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 5371.553$$
$$\rho^2 = 0.533$$
$$\bar{\rho}^2 = 0.518$$

Table 24: Dynamic model with agent effect correction (first and most frequent choices) (1): here the results with 250 draws (results are similar with more draws).

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ALPHA_FIRST_LUNCH_CHOICE	0.827	0.203	4.07	0.00
2	ALPHA_FIRST_MORNING_CHOICE	0.417	0.289	1.44	0.15
3	ALPHA_MOST_CHOSEN_LUNCH	1.77	0.126	14.03	0.00
4	ALPHA_MOST_CHOSEN_MORNING	2.43	0.278	8.74	0.00
5	ASC_ARC	-6.71	1.44	-4.66	0.00
6	ASC_ATL	-4.52	1.05	-4.30	0.00
7	ASC_BC	-1.30	0.609	-2.14	0.03
8	ASC_BM	-0.793	1.00	-0.79	0.43
9	ASC_COP	2.13	1.01	2.11	0.04
10	ASC_COR	-0.756	0.263	-2.87	0.00
11	ASC_ELA	-1.19	0.688	-1.73	0.08
12	ASC_GIA	0.501	0.688	0.73	0.47
13	ASC_HOD	0.0118	0.539	0.02	0.98
14	ASC_INM	-5.54	1.57	-3.54	0.00
15	ASC_KEB	1.37	0.445	3.08	0.00
16	ASC_KLE	-6.64	1.63	-4.06	0.00
17	ASC_MX	-3.03	0.632	-4.79	0.00
18	ASC_ORN	-1.44	0.247	-5.82	0.00
19	ASC_PAR	-0.864	0.397	-2.18	0.03
20	ASC_PH	-0.0246	0.642	-0.04	0.97
21	ASC_PIZ	1.14	0.505	2.26	0.02
22	ASC_SAT	-1.48	0.578	-2.55	0.01
23	ASC_VAL	0.700	2.34	0.30	0.76
24	ASC_VIN	-7.52	2.34	-3.21	0.00
25	BETA_CAPACITY_INSIDE	0.00439	0.00338	1.30	0.19
26	BETA_DINNER	0.786	0.362	2.17	0.03
27	BETA_DISTANCE_AFTERNOON	-0.00116	0.000638	-1.81	0.07
28	BETA_DISTANCE_LUNCH_CAF	-0.00348	0.000683	-5.09	0.00
29	BETA_DISTANCE_LUNCH_REST	-0.00292	0.00147	-1.98	0.05
30	BETA_DISTANCE_LUNCH_SELF	-0.00339	0.000453	-7.49	0.00
31	BETA_DISTANCE_MORNING	-0.00237	0.00112	-2.12	0.03
32	BETA_EVALUATION_CAFET	2.54	0.244	10.41	0.00
33	BETA_EVALUATION_SELF	2.78	0.276	10.06	0.00
34	BETA_METEO_TERRACE	0.00149	0.00119	1.24	0.21
35	BETA_NO_DISTANCE_AV	-3.13	0.532	-5.88	0.00
36	BETA_PRICE_EMPLOYEE	-0.0102	0.0779	-0.13	0.90
37	BETA_PRICE_STUDENT	-0.0776	0.0835	-0.93	0.35
38	BETA_TAP_BEER_AFTER_LUNCH	0.742	0.251	2.95	0.00
39	RHO_PREVIOUS_LUNCH_CHOICE	0.0143	0.137	0.10	0.92
40	RHO_PREVIOUS_MORNING_CHOICE	0.143	0.339	0.42	0.67

Table 25: Dynamic model with agent effect correction (first and most frequent choices) (2)

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
41	SIGMA_LUNCH_ARC	-4.21	0.676	-6.23	0.00
42	SIGMA_LUNCH_ATL	-0.686	0.187	-3.67	0.00
43	SIGMA_LUNCH_BC	-0.199	0.206	-0.96	0.33
44	SIGMA_LUNCH_BM	3.27	0.887	3.68	0.00
45	SIGMA_LUNCH_COP	0.881	0.453	1.95	0.05
46	SIGMA_LUNCH_COR	-0.695	0.299	-2.33	0.02
47	SIGMA_LUNCH_ELA	-0.668	0.428	-1.56	0.12
48	SIGMA_LUNCH_GIA	-1.10	0.372	-2.96	0.00
49	SIGMA_LUNCH_HOD	0.801	0.283	2.83	0.00
50	SIGMA_LUNCH_INM	-1.79	0.658	-2.73	0.01
51	SIGMA_LUNCH_KEB	0.501	0.276	1.81	0.07
52	SIGMA_LUNCH_KLE	-1.29	0.573	-2.25	0.02
53	SIGMA_LUNCH_MX	-1.69	0.248	-6.83	0.00
54	SIGMA_LUNCH_ORN	0.177	0.251	0.71	0.48
55	SIGMA_LUNCH_PAR	-0.444	0.620	-0.72	0.47
56	SIGMA_LUNCH_PH	3.61	0.660	5.47	0.00
57	SIGMA_LUNCH_PIZ	-3.21	0.726	-4.42	0.00
58	SIGMA_LUNCH_SAT	-0.123	0.195	-0.63	0.53
59	SIGMA_LUNCH_VAL	4.26	1.57	2.72	0.01
60	SIGMA_LUNCH_VIN	2.68	0.643	4.16	0.00
61	SIGMA_MORNING_ARC	-2.56	0.598	-4.29	0.00
62	SIGMA_MORNING_ATL	3.10	0.633	4.91	0.00
63	SIGMA_MORNING_BC	1.50	0.226	6.64	0.00
64	SIGMA_MORNING_BM	-0.414	1.21	-0.34	0.73
65	SIGMA_MORNING_COP	-0.401	0.324	-1.24	0.22
66	SIGMA_MORNING_COR	0.755	0.253	2.98	0.00
67	SIGMA_MORNING_ELA	-0.491	0.396	-1.24	0.21
68	SIGMA_MORNING_GIA	-1.28	0.256	-5.02	0.00
69	SIGMA_MORNING_HOD	0.154	0.150	1.03	0.30
70	SIGMA_MORNING_INM	2.93	0.780	3.76	0.00
71	SIGMA_MORNING_KEB	-3.75	0.562	-6.67	0.00
72	SIGMA_MORNING_KLE	-3.19	0.321	-9.93	0.00
73	SIGMA_MORNING_MX	0.130	0.363	0.36	0.72
74	SIGMA_MORNING_ORN	-0.327	0.278	-1.18	0.24
75	SIGMA_MORNING_PAR	0.575	0.307	1.87	0.06
76	SIGMA_MORNING_PH	-1.30	0.254	-5.13	0.00
77	SIGMA_MORNING_PIZ	-1.21	0.266	-4.56	0.00
78	SIGMA_MORNING_SAT	1.32	0.178	7.40	0.00
79	SIGMA_MORNING_VAL	-0.157	1.58	-0.10	0.92
80	SIGMA_MORNING_VIN	-1.76	0.510	-3.45	0.00

Summary statistics

Number of observations = 1868

Number of estimated parameters = 80

$$\mathcal{L}(\beta_0) = -5037.914$$
$$\mathcal{L}(\hat{\beta}) = -2182.172$$
$$-2[\mathcal{L}(\beta_0) - \mathcal{L}(\hat{\beta})] = 5711.485$$
$$\rho^2 = 0.567$$
$$\bar{\rho}^2 = 0.551$$

B.2 Complementary results and tables

Table 26: Validation of the models (morning). Choices performed by 88 individuals

Observed			Predicted							
			Static		Strict exo		First choice		First and most freq	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
Cafeteria Cafe Le Klee	0	0%	0	0.1%	0	0.1%	0	0.1%	1	1.6%
Self-service BC	7	8%	6	6.8%	7	7.7%	6	7%	8	9.3%
Other BM	2	2.3%	2	2.6%	2	2.1%	2	2.1%	3	3.1%
Cafeteria ELA	10	11.4%	6	6.6%	5	5.6%	6	6.4%	5	5.8%
Cafeteria INM	1	1.1%	1	1.1%	1	0.7%	2	2.8%	4	4.3%
Cafeteria MX	6	6.8%	7	7.8%	7	8.5%	3	3.5%	6	6.4%
Other PH	5	5.7%	5	5.8%	5	5.2%	4	4.4%	3	3.6%
Cafeteria L'Arcadie	3	3.4%	1	1.7%	2	2.5%	3	2.9%	2	2.4%
Self-service L'Atlantide	9	10.2%	11	12.6%	11	12.3%	8	9.3%	7	8.5%
Restaurant Le Copernic	0	0%	0	0.5%	0	0.2%	1	0.8%	1	0.7%
Self-service Le Corbusier	0	0%	2	1.8%	1	1.4%	2	1.8%	2	1.9%
Cafeteria Le Giacometti	11	12.5%	9	10.8%	12	13.2%	11	12.3%	11	12.9%
Self-service Le Parmentier	0	0%	2	2.3%	1	1.6%	2	2.4%	2	1.8%
Self-service Le Vinci	0	0%	0	0%	0	0%	0	0.1%	0	0.1%
Self-service L'Esplanade	18	20.5%	22	24.7%	20	22.4%	23	26.4%	19	21.8%
Self-service L'Ornithorynque	0	0%	2	1.8%	1	1.5%	2	1.8%	2	2.3%
Caravan Pizza	4	4.5%	3	3.9%	3	3.7%	4	4.2%	3	2.9%
Caravan Kebab	5	5.7%	2	2.8%	3	3.2%	2	1.8%	2	2.7%
Cafeteria Satellite	7	8%	5	5.9%	7	7.9%	8	9.1%	6	7%
Self-service Le Hodler	0	0%	0	0.2%	0	0.2%	0	0.5%	1	0.7%
Restaurant Table de Vallotton	0	0%	0	0.2%	0	0.1%	0	0.3%	0	0.1%

Table 27: Validation of the models (all day). Choices performed by 175 individuals

Observed			Predicted							
			Static		Strict exo		First choice		First and most freq	
	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
Cafeteria Cafe Le Klee	0	0%	0	0.2%	0	0.2%	1	0.5%	1	0.3%
Self-service BC	23	13.1%	14	8.2%	14	7.8%	12	6.9%	13	7.4%
Other BM	3	1.7%	5	2.8%	4	2.4%	3	1.9%	4	2.1%
Cafeteria ELA	15	8.6%	9	5.2%	9	5.2%	9	5.4%	8	4.6%
Cafeteria INM	1	0.6%	1	0.8%	1	0.8%	2	1.4%	3	1.4%
Cafeteria MX	5	2.9%	7	3.8%	7	4.2%	9	5.1%	6	3.7%
Other PH	8	4.6%	9	4.9%	9	5.1%	5	3%	6	3.6%
Cafeteria L'Arcadie	7	4%	3	1.5%	4	2.2%	4	2.5%	4	2.3%
Self-service L'Atlantide	5	2.9%	12	6.7%	9	5%	7	4.1%	8	4.6%
Restaurant Le Copernic	1	0.6%	1	0.5%	1	0.8%	2	1%	1	0.8%
Self-service Le Corbusier	4	2.3%	8	4.7%	7	3.8%	8	4.4%	8	4.6%
Cafeteria Le Giacometti	18	10.3%	16	9%	17	9.5%	17	9.6%	17	9.7%
Self-service Le Parmentier	12	6.9%	14	8%	13	7.3%	13	7.3%	10	5.9%
Self-service Le Vinci	1	0.6%	0	0.1%	0	0.1%	1	0.6%	0	0.1%
Self-service L'Esplanade	31	17.7%	38	21.5%	38	21.9%	39	22.6%	43	24.7%
Self-service L'Ornithorynque	11	6.3%	11	6.5%	13	7.5%	12	6.8%	13	7.3%
Caravan Pizza	4	2.3%	5	3.1%	5	3.1%	5	3%	6	3.2%
Caravan Kebab	6	3.4%	5	2.7%	5	2.6%	6	3.3%	5	2.7%
Cafeteria Satellite	13	7.4%	12	7.1%	13	7.2%	13	7.4%	12	7%
Self-service Le Hodler	5	2.9%	4	2.2%	5	2.6%	4	2.5%	5	2.7%
Restaurant Table de Vallotton	2	1.1%	1	0.6%	1	0.6%	1	0.7%	2	1.2%

Table 28: Destination specific attributes 1

Type	Attribute	1 Cafe Le Klee	2 Cafeteria BC	3 Cafeteria BM	4 Cafeteria ELA	5 Cafeteria INM	6 Cafeteria MX	7 Cafeteria PH	8 L'Arcadie	9 L'Atlantide	10 Le Copernic	11 Le Corbusier
Students	Cheapest Hot Meal	0	7	0	0	0	7	0	9.9	9.8	18.5	7
	Most expensive HM	0	12	0	0	0	7	0	9.9	9.8	27	11
Personnel	Cheapest HM	0	8	0	0	0	8	0	9.9	9.8	18.5	8
	Most expensive HM	0	12	0	0	0	8	0	9.9	9.8	27	11
Availability	opening	7	8		8	7	8		7	7	11	11
	closing	20	18		17	17	17		18	17	14	14
	opening2											
	closing2											
	Closed door	No	No	No	Yes	No	No	No	No	Yes	Yes	Yes
Coordinates	x coordinate	533197.1709	532723.3408	532981.8687	532935.0428	532831.2319	532851.1903	533264.172	533300.8256	533100.9355	533247.3611	532895.3665
	y coordinate	152236.3728	152269.3764	152505.5316	152413.6977	152293.9709	152361.7704	152536.4189	152493.9743	152498.4882	152476.1802	152535.3014
	Vertex (SQL)	31041	10125	11809	14365	25514	33631	33095	6079	7522	29345	16895
Capacity	Inside	70	82	60	98	20	50	15	60	125	105	228
	Outside	0	119	10	68	14	25	0	100	50	50	100
Evaluation	Service and meal	4.35	4.92	-1.00	5.34	5.25	5.54	-1.00	4.82	4.94	5.32	4.63
Availabilites services	Coffee	1	1	1	1	1	1	1	1	1	0	0
	Hot meal	0	1	0	0	0	1	0	1	1	1	1
	Table service	0	0	0	0	0	0	0	0	0	1	0
	Visibility	1	0	0	0	1	0	0	1	1	1	0
	Terrace	0	1	1	1	1	1	0	1	1	1	1
	Workspace	1	1	1	1	1	1	1	1	1	0	0
	Fourchette verte	0	1	0	0	0	1	0	0	0	0	1
	Dinner offer	0	0	0	0	0	0	0	0	0	0	0
	Sandwich	1	0	0	1	1	1	0	1	1	0	0
	Selecta	1	1	1	0	1	1	1	0	0	0	0
	Food	1	1	0	1	1	1	0	1	1	1	1
	Tap beer	1	0	0	0	0	0	0	1	0	0	0
	Fidelity card	0	0	0	0	0	0	0	0	0	0	0
	Micro-wave	1	1	1	1	1	1	1	0	0	0	0
	Price coffee	1.7	1.7	1.3	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7

Table 29: Destination specific attributes 2

12	13	14	15	16	17	18	19	20	21	22	23
Le Giacometti	Le Parmentier	Le Vinci	L'Esplanade	L'Ornithorynque	Roulotte Diagonale	Roulotte Esplanade	Satellite	Self-service Le Hodler	Table de Vallotton	Caravan Holy Cow	Self-service L'Epicure
0	7	7	7	7.65	8	7	0	14	31	13.9	8.5
0	12	12	9	11.05	10	9	0			20.9	8.5
0	8	8	7	8.35	10	7	0	8	25	15.9	8.5
0	12	12	9	12	10	9	0	14	31	20.9	8.5
7	11	11	7	11	10	10	7	11	11	11	7
18	14	14	20	14	15	15	21	14	14	22	18
	18				17	17					
	20				20	20					
Yes	Yes	Yes	No	Yes	No	No	Yes	No	No	Yes	Yes
532918.7947	533126.88	533152.5659	532990.7136	532853.4989	532868.4258	532961.8839	533176.9273	533197.3543	533255.8414	533022.6	533249.5
152530.1958	152500.4118	152502.5327	152462.8814	152458.8984	152347.8048	152465.7978	152497.7843	152223.1355	152210.4827	152713.3	152415
26732	7526	7383	24713	31262	30648	30647	7375	31050	30967		
90	320	240	225	250	0	0	200	128	80	50	60
30	52	52	180	120	15	0	30	0	0	30	10
4.60	4.68	4.74	4.59	5.05	4.96	4.97	4.99	4.54	5.37	-1	-1
1	0	0	1	0	0	0	1	0	0	0	1
0	1	1	1	1	1	1	0	1	1	1	1
0	0	0	0	0	0	0	0	0	1	0	0
1	1	1	1	0	1	1	1	0	0	1	1
1	1	1	1	1	1	1	1	0	0	1	1
1	0	0	1	0	0	0	1	1	0	1	1
0	1	1	1	0	0	0	0	1	0	0	0
0	1	0	1	0	1	1	0	0	0	1	0
1	0	0	1	0	0	0	1	0	0	1	1
1	0	0	1	0	0	0	0	1	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	1	1	1	0	0	1	0
1	1	1	1	1	0	0	0	1	0	0	1
1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7	0	1.7

C Future utilization of data

We give here the codes and variables' definition in case of a future utilization of the data. Also, a simplified version is available online (Tinguely, 2015).

C.1 Codes for Pythonbiogeme

Here the specifications for Pythonbiogeme. Again, for one alternative only. Complete codes and expressions' definition are available in the electronic version.

```
####Destination choice model on EPFL campus from Wifi traces####

# Author: Loic Tinguely, EPFL
# Date: Wed Jun 17 18:24:40 2015
from biogeme import *
from headers import *
from loglikelihood import *
from statistics import *

######## Parameters to be estimated ##########
# Arguments:
#   1  Name for report. Typically, the same as the variable
#   2  Starting value
#   3  Lower bound
#   4  Upper bound
#   5  0: estimate the parameter, 1: keep it fixed

# Alternative specific constant
ASC_1 = Beta('ASC_1', 0, -100, 100, 0 )

# Alternative specific parameters
BETA_PRICE_STUDENT = Beta('BETA_PRICE_STUDENT', 0, -100, 100, 0 )
BETA_PRICE_EMPLOYEE = Beta('BETA_PRICE_EMPLOYEE', 0, -100, 100, 0 )
BETA_DINNER = Beta('BETA_DINNER', 0, -100, 100, 0 )
BETA_TAP_BEER_AFTER_LUNCH = Beta('BETA_TAP_BEER_AFTER_LUNCH', 0, -100,
  100, 0 )
BETA_EVALUATION_SELF = Beta('BETA_EVALUATION_SELF', 0, -100, 100, 0 )
BETA_EVALUATION_CAFET = Beta('BETA_EVALUATION_CAFET', 0, -100, 100, 0 )
BETA_EVALUATION_REST = Beta('BETA_EVALUATION_REST', 0, -100, 100, 1 )
BETA_EVALUATION_CARAVAN = Beta('BETA_EVALUATION_CARAVAN', 0, -100,
  100, 1 )
BETA_EVALUATION_OTHER = Beta('BETA_EVALUATION_OTHER', 0, -100, 100, 1 )
BETA_METEO_TERRACE = Beta('BETA_METEO_TERRACE', 0, -100, 100, 0 )
```

```
BETA_CAPACITY_INSIDE = Beta('BETA_CAPACITY_INSIDE', 0, -100, 100, 0 )

# Activity-episode sequence specific parameters
BETA_DISTANCE_LUNCH_REST = Beta('BETA_DISTANCE_LUNCH_REST', 0, -100,
  100, 0 )
BETA_DISTANCE_LUNCH_SELF = Beta('BETA_DISTANCE_LUNCH_SELF', 0, -100,
  100, 0 )
BETA_DISTANCE_LUNCH_CAF = Beta('BETA_DISTANCE_LUNCH_CAF', 0, -100,
  100, 0 )
BETA_DISTANCE_LUNCH_OTHER = Beta('BETA_DISTANCE_LUNCH_OTHER', 0, -100,
  100, 1 )
BETA_DISTANCE_LUNCH_CARAVAN = Beta('BETA_DISTANCE_LUNCH_CARAVAN', 0,
  -100, 100, 1 )
BETA_DISTANCE_MORNING = Beta('BETA_DISTANCE_MORNING', 0, -100, 100, 0 )
BETA_DISTANCE_AFTERNOON = Beta('BETA_DISTANCE_AFTERNOON', 0, -100,
  100, 0 )
BETA_NO_DISTANCE_AV = Beta('BETA_NO_DISTANCE_AV', 0, -100, 100, 0 )

# Wooldridge correction parameters
ALPHA_MOST_CHOSEN_LUNCH = Beta('ALPHA_MOST_CHOSEN_LUNCH', 0, -100,
  100, 0 )
ALPHA_MOST_CHOSEN_MORNING = Beta('ALPHA_MOST_CHOSEN_MORNING', 0, -100,
  100, 0 )
RHO_PREVIOUS_LUNCH_CHOICE = Beta('RHO_PREVIOUS_LUNCH_CHOICE', 0, -100,
  100, 0 )
RHO_PREVIOUS_MORNING_CHOICE = Beta('RHO_PREVIOUS_MORNING_CHOICE', 0,
  -100, 100, 0 )
RHO_PREVIOUS_AFTERNOON_CHOICE = Beta('RHO_PREVIOUS_AFTERNOON_CHOICE',
  0, -100, 100, 1 )
ALPHA_FIRST_LUNCH_CHOICE = Beta('ALPHA_FIRST_LUNCH_CHOICE', 0, -100,
  100, 0 )
ALPHA_FIRST_MORNING_CHOICE = Beta('ALPHA_FIRST_MORNING_CHOICE', 0,
  -100, 100, 0 )
ALPHA_FIRST_AFTERNOON_CHOICE = Beta('ALPHA_FIRST_AFTERNOON_CHOICE', 0,
  -100, 100, 1 )
SIGMA_LUNCH_1 = Beta('SIGMA_LUNCH_1', 0, -100, 100, 0 )
SIGMA_MORNING_1 = Beta('SIGMA_MORNING_1', 0, -100, 100, 0 )
SIGMA_AFTERNOON_1 = Beta('SIGMA_AFTERNOON_1', 0, -100, 100, 1 )
EC_SIGMA_LUNCH_1 = SIGMA_LUNCH_1 * bioNormalDraws('EC_SIGMA_LUNCH_1',
  'ID')
EC_SIGMA_MORNING_1 = SIGMA_MORNING_1 *
  bioNormalDraws('EC_SIGMA_MORNING_1', 'ID')

########## Expressions ###########

# Period of the day
```

```
one  = 1
lunch11 = (H_START == 11 * M_START > 29.9) > 0
lunch12 = H_START == 12
lunch13 = H_START == 13
dinner18 = H_START == 18
dinner19 = H_START == 19
morning7 = H_START == 7
morning8 = H_START == 8
morning9 = H_START == 9
morning10 = H_START == 10
morning11 = (H_START == 11 * M_START < 29.8) > 0
afternoon14 = H_START == 14
afternoon15 = H_START == 15
afternoon16 = H_START == 16
afternoon17 = H_START == 17
night20 = H_START == 20
night21 = H_START == 21
night_end21 = H_END == 21
night_end22 = H_END == 22
night = (night20 + night21 + night_end21 + night_end22) > 0
afternoon = (afternoon15 + afternoon16 + afternoon17 + afternoon14) > 0
morning = (morning8 + morning9 + morning10 + morning7 + morning11) > 0
dinner = (dinner18 + dinner19) > 0
lunch = (lunch11 + lunch12 + lunch13) > 0
after_lunch = afternoon + night + dinner
rain_and_cold_max_20 = SUN_AND_HEAT_MIN_20 == 0

# Socio-economic
first_year = SEMESTER == 2
perso = SECTION_ID == 5
student_true = STUDENT == 1

# Alternative specific
capacity_at_lunch_inside_av_1 =  lunch   * CAPACITY_INSIDE_1
capacity_at_lunch_outside_av_1 = lunch  * CAPACITY_OUTSIDE_1
meteo_terrace_av_1 = ( SUN_AND_HEAT_MIN_20 * TERRACE_AV_1 ) *
  capacity_at_lunch_outside_av_1
cap_inside_av_1 = capacity_at_lunch_inside_av_1
morning_coffee_1 = morning * CAFE_AV_1
distance_filter_1 = DISTANCE_1 > -1
distance_no_av_1 = DISTANCE_1 == -1
lunch_distance_1 = lunch * (( distance_filter_1 * DISTANCE_1 ) + (
  distance_no_av_1 * 0 ) )
morning_distance_1 = mroning * (( distance_filter_1 * DISTANCE_1 ) + (
  distance_no_av_1 * 0 ) )
```

```
afternoon_distance_1 = after_lunch * (( distance_filter_1 * DISTANCE_1
  ) + ( distance_no_av_1 * 0 ) )
lunch_hot_meal_av_1 = lunch * HOT_MEAL_AV_1
lunch_price_min_1 = ( lunch * MIN_PRICE_1 ) * lunch_hot_meal_av_1
beer_in_the_afternoon_dinner_night_av_1 = after_lunch * TAP_BEER_AV_1
service_at_table_lunch_av_1 = ( lunch * SERVICE_TABLE_AV_1 ) * perso
sandwich_lunch_av_1 = lunch * SANDWICH_AV_1
evaluation_filter_1 = EVALUATION_2013_1 > -1
evaluation_2013_1 = ( evaluation_filter_1 * EVALUATION_2013_1 ) * lunch
dinner_av_1 = DINNER_HOT_MEAL_AV_1 * dinner
lunch_price_min_student_1 = ( lunch * MIN_PRICE_1 ) * student_true
lunch_price_min_employee_1 = ( lunch * MIN_PRICE_1 ) * perso

# Wooldridge correction
last_choice_filter_1 = LAST_CHOICE_LAST_TIME_TRUE_1 > -1
no_previous_choice_filter_1 = LAST_CHOICE_LAST_TIME_TRUE_1 == -1
last_choice_true_1 = ( last_choice_filter_1 *
  LAST_CHOICE_LAST_TIME_TRUE_1 ) + ( no_previous_choice_filter_1 * 0 )
previous_choice_morning_filter_1 = PREVIOUS_CHOICE_MORNING_TRUE_1 > -1
no_previous_choice_morning_filter_1 = PREVIOUS_CHOICE_MORNING_TRUE_1
  == -1
previous_choice_morning_1 = ( previous_choice_morning_filter_1 *
  PREVIOUS_CHOICE_MORNING_TRUE_1 ) + (
  no_previous_choice_morning_filter_1 * 0 )
previous_choice_afternoon_filter_1 = PREVIOUS_CHOICE_AFTERNOON_TRUE_1
  > -1
no_previous_choice_afternoon_filter_1 =
  PREVIOUS_CHOICE_AFTERNOON_TRUE_1 == -1
previous_choice_afternoon_1 = ( previous_choice_afternoon_filter_1 *
  PREVIOUS_CHOICE_AFTERNOON_TRUE_1 ) + (
  no_previous_choice_afternoon_filter_1 * 0 )
first_choice_true_filter_1 = FIRST_CHOICE_TRUE_1 > -1
first_choice_filter_1 = FIRST_CHOICE_TRUE_1 *
  first_choice_true_filter_1 * lunch
first_choice_true_morning_1 = FIRST_CHOICE_AFTERNOON_TRUE_1 > -1
first_choice_morning_1 = FIRST_CHOICE_MORNING_TRUE_1 *
  first_choice_true_morning_1 * morning
first_choice_true_afternoon_1 = FIRST_CHOICE_AFTERNOON_TRUE_1 > -1
first_choice_afternoon_1 = FIRST_CHOICE_AFTERNOON_TRUE_1 *
  first_choice_true_afternoon_1 * after_lunch
most_chosen_filter_1 = MOST_CHOSEN_1 > -1
most_chosen_filter_lunch_1 = MOST_CHOSEN_1 * most_chosen_filter_1 *
  lunch
most_chosen_morning_1 = MOST_CHOSEN_MORNING_1 > -1
most_chosen_filter_morning_1 = MOST_CHOSEN_MORNING_1 *
  most_chosen_morning_1 * morning
```

```
####### Utility function (fixed to 0 parameters are removed) ########
V1  = ASC_1 \
  + ALPHA_FIRST_LUNCH_CHOICE * first_choice_filter_1 \
  + ALPHA_FIRST_MORNING_CHOICE * first_choice_morning_1 \
  + ALPHA_MOST_CHOSEN_LUNCH * most_chosen_filter_lunch_1 \
  + ALPHA_MOST_CHOSEN_MORNING * most_chosen_filter_morning_1 \
  + RHO_PREVIOUS_MORNING_CHOICE * previous_choice_morning_1 \
  + RHO_PREVIOUS_LUNCH_CHOICE * last_choice_true_1 \
  + BETA_DISTANCE_LUNCH_CAF * lunch_distance_1 \
  + BETA_DISTANCE_MORNING * morning_distance_1 \
  + BETA_DISTANCE_AFTERNOON * afternoon_distance_1\
  + BETA_NO_DISTANCE_AV * distance_no_av_1 \
  + BETA_TAP_BEER_AFTER_LUNCH *
    beer_in_the_afternoon_dinner_night_av_1 \
  + BETA_EVALUATION_CAFET * evaluation_2013_1 \
  + BETA_DINNER * dinner_av_1 \
  + BETA_PRICE_STUDENT * lunch_price_min_student_1 \
  + BETA_PRICE_EMPLOYEE * lunch_price_min_employee_1 \
  + BETA_METEO_TERRACE * meteo_terrace_av_1 \
  + BETA_CAPACITY_INSIDE * cap_inside_av_1 \
  + EC_SIGMA_MORNING_1 * morning \
  + EC_SIGMA_LUNCH_1 * lunch

############ Specifications for estimation #############

# Associate utility functions with the numbering of alternatives
V = {1: V1}
# Associate the availability conditions with the alternatives
av = {1 : OPEN_AV_1}
# The choice model is a Logit, with availability conditions
prob = bioLogit(V,av,CHOICE)
# Iterator on individuals, that is on groups of rows.
metaIterator('personIter','__dataFile__','panelObsIter','ID')
# For each item of personIter, iterates on the rows of the group.
rowIterator('panelObsIter','personIter')
#Iterator on draws for Monte-Carlo simulation
drawIterator('drawIter')
#Conditional probability for the sequence of choices of an individual
condProbIndiv = Prod(prob,'panelObsIter')
# Integration by simulation
probIndiv = Sum(condProbIndiv,'drawIter')
# Likelihood function
loglikelihood = Sum(log(probIndiv),'personIter')
BIOGEME_OBJECT.ESTIMATE = loglikelihood
# Parameters
```

```
BIOGEME_OBJECT.PARAMETERS['NbrOfDraws'] = "250"
BIOGEME_OBJECT.PARAMETERS['RandomDistribution'] = "HALTON"
BIOGEME_OBJECT.PARAMETERS['optimizationAlgorithm'] = "CFSQP"
BIOGEME_OBJECT.PARAMETERS['checkDerivatives'] = "0"
BIOGEME_OBJECT.PARAMETERS['numberOfThreads'] = "6"

############### Specifications for the validation ##############

# The choice model is a mixture Logit, with availability conditions
prob1 =  bioLogit(V,av,1)
log1  = mixedloglikelihood(prob1)
# Defines an iterator on the data
rowIterator('obsIter')
simulate = {'Prob. KLE': prob1}
# Simulation
simulate_mixture = {'Prob. KLE': log1}
BIOGEME_OBJECT.SIMULATE = Enumerate(simulate_mixture,'obsIter')
# Statistics
nullLoglikelihood(av,'obsIter')
choiceSet = [1]
cteLoglikelihood(choiceSet,CHOICE,'obsIter')
availabilityStatistics(av,'obsIter')
# Parameters
BIOGEME_OBJECT.PARAMETERS['NbrOfDraws'] = "250"
BIOGEME_OBJECT.PARAMETERS['numberOfThreads'] = "1"
BIOGEME_OBJECT.PARAMETERS['optimizationAlgorithm'] = "CFSQP"
BIOGEME_OBJECT.PARAMETERS['dumpDrawsOnFile'] = "1"

###### Specifications for forecasting with new alternative ######

from biogeme import *
from headers import *
from nested import *
from loglikelihood import *
from statistics import *

#Nest parameter (1 = independent, infinite = correlated)
MU = Beta('MU', 5, -100, 100, 0 )

#Definition of nests:
# 1: nests parameter
# 2: list of alternatives. The alternative 3 is new and borrows the
  parameters of alternative 2.
nonnested =  1.0 , [1]
nested = MU , [2,3]
nests = nonnested,nested
```

```
# The choice model is a mixturelogit, with availability conditions
prob1  =  nested(V,av,nests,1)
prob2 =  nested(V,av,nests,2)
prob3 =  nested(V,av,nests,3)

log1  = mixedloglikelihood(prob1)
log2 = mixedloglikelihood(prob2)
log3 = mixedloglikelihood(prob3)

# Defines an itertor on the data
rowIterator('obsIter')

simulate = {'Prob. KLE': prob1,
      'Prob. BC': prob2,
      'Prob. BM': prob3,}

simulate_mixture = {'Prob. KLE': log1,
      'Prob. BC': log2,
      'Prob. BM': log3}


BIOGEME_OBJECT.SIMULATE = Enumerate(simulate_mixture,'obsIter')

# Statistics
nullLoglikelihood(av,'obsIter')
choiceSet = [1,2,3]
cteLoglikelihood(choiceSet,CHOICE,'obsIter')
availabilityStatistics(av,'obsIter')

BIOGEME_OBJECT.PARAMETERS['NbrOfDraws'] = "250"
BIOGEME_OBJECT.PARAMETERS['checkDerivatives'] = "1"
BIOGEME_OBJECT.PARAMETERS['numberOfThreads'] = "1"
BIOGEME_OBJECT.PARAMETERS['moreRobustToNumericalIssues'] = "0"
BIOGEME_OBJECT.PARAMETERS['optimizationAlgorithm'] = "CFSQP"
BIOGEME_OBJECT.PARAMETERS['dumpDrawsOnFile'] = "1"
```

C.2 Strict definition of variables

Figure 18: Definition of the variables in the dataset (1)

Variables description							
Sequence specific attributes							
Name in dataset	Description	Average	Median	Standard Error	Range	Min	Max
ID	Each student/employee is identified by a unique number (the first number of the ID represents the section whereas the last one represent the number of the individual, see below)	46933	50138	14091	60011	10000	70011
SECTION_ID	The section where the student is registered takes a number between 1 and 7 and employees take number 5. (GCBA4 = 1, INBA4 = 2, INMA2 = 3, MABA2 = 4, employee = 5, PHBA2 = 6, SVBA2 = 7)	4.7	5	1.41	6	1	7
STUDENT	Takes a value of 1 if the observation is from a student and 0 if it's an employee	0.34	0	0.48	1	0	1
DAY_WEEK	The day of the week is identified by a number between 1 = Monday and 7 = Sunday)	2.96	3	1.3	5	1	6
SEMESTER	Students are described by there section but also their semester. (BA4 = 4, BA2 = 2, MA2 = 8, Employees = 0)	1.41	0	2.39	8	0	8
Meteo attributes							
TEMPERATURE	The daily average temperature. Unity: °C	17.5	19.1	3.4	17.9	7.7	25.6
MAX_TEMP	The daily maximum temperature. Unity: °C	21.4	20.2	4.3	21.3	10.2	31.5
RAIN	The daily rain record. Unity: mm	1.9	0.0	5.0	23.9	0.0	23.9
SUNNY_DAY_AV	1 if no rain was recorded, 0 otherwise	0.7	1.0	0.5	1.0	0.0	1.0
SUN_AND_HEAT_MIN_20	1 if no rain was recorded and the maximum temperature exceeded 20°C, 0 otherwise	0.5	1.0	0.5	1.0	0.0	1.0
Activity-episode specific attributes							
H_START	H_START represents the average start hour of the observation. For example, if the observation takes place at 7:16AM, the value will be 7. If it happens at 3:57PM the value if 15	12.88	13	2.99	22	1	23
M_START	M_START represents the average start minute of the observation. For example if the observation happens at 7:16AM, the value will be 16	27.89	27	17.48	59	0	59
H_END	The idea is the same as H_START but with the average end hour of the	13.74	13	2.97	22	1	23
M_END	The idea is the same as M_START but with the average end hour of the observation	30.08	29	17.76	59	0	59
DURATION	The variable calculates the duration (in minutes) of the observation. If the activity begins at 7:15AM and ends at 8:19AM, the duration will be 64 minutes	53.83	20.67	85.43	618.39	5	619.65
Alternative specific attributes (x=1 to x=21)							
Price							
MIN_PRICE_x	This variable represents the minimum price for a hot meal. Unity: CHF	See prices table					
PRICE_COFFEE_x	This variable shows the price of a coffee. Unity:CHF						
Quality							
EVALUATION_x	The evaluation from the 2013 restauration survey (-1 if not available)	See evaluation table					
Distance from previous activity-episode							
DISTANCE_x	The real distance from the previous activity-episode. It's calculated in meters (-1 if not available)						
Capacity							
CAPACITY_INSIDE_x	This variable represents the inside capacity (total number of seating places) (exists for all alternatives)	See Capacity table					
CAPACITY_OUTSIDE_x	This variable represents the outside capacity of the terraces (total number of seating places)						
Services availability (1)							
CAFE_AV_x	This dummy takes the value 1 if the destination sells coffee and 0 otherwise (exists for all alternatives)	See availabilities' table					
HOT_MEAL_AV_x	This dummy takes the value 1 if the destination sells hot meals for lunch and 0 otherwise (exists for all alternatives)						
SERVICE_TABLE_AV_x	This dummy takes the value 1 if the destination has an "at table" service for lunch and 0 otherwise (exists for all alternatives)						

Figure 19: Definition of the variables in the dataset (2)

Services availability (2)		
VISIBILITY_x	This dummy takes the value 1 if the destination can be seen from the common sidewalk and 0 otherwise (exists for all alternatives)	See availabilities' table
TERRACE_AV_x	This dummy takes the value 1 if the destination has a terrace, 0 otherwise (exists for all alternatives)	
WORKSPACE_AV_x	This dummy takes the value 1 if the destination lets the possibility to work, 0 otherwise (exists for all alternatives)	
FOURCHETTE_VERTE_AV_x	This dummy takes the value 1 if the destination is part of the "Fourchette Verte" deal and 0 otherwise (exists for all alternatives)	
DINNER_HOT_MEAL_AV_x	This dummy takes the value 1 if the destination sells hot food for dinner (between 6PM and 7:59PM) and 0 otherwise (exists for all alternatives)	
SANDWICH_AV_x	This dummy takes the value 1 if the destination sells sandwiches (exists for all alternatives)	
SELECTA_AV_x	This dummy takes the value 1 if a Selecta is available (exists for all alternatives)	
FIDELITY_CARD_x	This dummy takes the value 1 if the destination offers a fidelity card (exists for all alternatives)	
FOOD_AV_x	This dummy takes the value 1 if the destination sells any kind of food (exists for all alternatives)	
OPEN_AV_x	This dummy variable take the value 1 if the destination is open.	
TAP_BEER_AV_x	This dummy takes the value 1 if the destination sells tap beer and 0 otherwise. (exists for all alternatives)	
Choice and panel data (these variables are ID specific)		
LAST_CHOICE_LAST_MEAL_x	This variable shows the previous choice that was made by one individual when choosing for a lunch (between 11:30AM and 1:59PM) destination. It takes the value 0 if the previous choice is different from the actual choice. A value of 1 if it's the same and a value -1 if the information is not available	See descriptive statistics
FIRST_CHOICE_TRUE_x	This variable shows the first ever choice that was made when choosing for a lunch (between 11:30AM and 1:59PM) destination. It takes the value 0 if the first choice is different from the actual choice. A value of 1 if it's the same and a value -1 if the information is not available	
MOST_CHOSEN_x	This variable shows the most frequently chosen destination for a lunch (between 11:30AM and 1:59PM). It takes the value 0 if the first choice is different from the actual choice. A value of 1 if it's the same and a value -1 if the information is not available	
PREVIOUS_CHOICE_MORNING_TRUE_x	This variable shows the previous choice that was made by one individual when choosing for a morning (between 7AM and 11:29AM) destination. It takes the value 0 if the previous choice is different from the actual choice. A value of 1 if it's the same and a value -1 if the information is not available	
MOST_CHOSEN_MORNING_x	This variable shows the most frequently chosen destination for a morning (between 7AM and 11:29AM). It takes the value 0 if the first choice is different from the actual choice. A value of 1 if it's the same and a value -1 if the information is not available	
FIRST_CHOICE_MORNING_TRUE_x	This variable shows the first ever choice that was made when choosing for a morning (between 7AM and 11:29AM) destination. It takes the value 0 if the first choice is different from the actual choice. A value of 1 if it's the same and a value -1 if the information is not available	
PREVIOUS_CHOICE_AFTERNOON_TRUE_x	This variable shows the previous choice that was made by one individual when choosing for an afternoon (between 2PM and 10PM) destination. It takes the value 0 if the previous choice is different from the actual choice. A value of 1 if it's the same and a value -1 if the information is not available	
FIRST_CHOICE_MORNING_TRUE_x	This variable shows the first ever choice that was made when choosing for an afternoon (between 2PM and 10PM) destination. It takes the value 0 if the first choice is different from the actual choice. A value of 1 if it's the same and a value -1 if the information is not available	
HOURS_FROM_PREVIOUS_CHOICE	The number of hours between the actual lunch choice and the previous lunch choice. -1 if no previous choice is recorded	
CHOICE	This variable shows the choice that was made in the observation. It takes a value between 0 and 21. (Cafe Le Klee = 1, Cafeteria BC = 2, Cafeteria BM = 3, Cafeteria ELA = 4, Cafeteria INM = 5, Cafeteria MX = 6, Cafeteria PH = 7, L'Arcadie = 8, L'Atlantide = 9, Le Copernic = 10, Le Corbusier = 11, Le Giacometti = 12, Le Parmentier = 13, Le Vinci = 14, L'Esplanade = 15, L'Ornithorynque = 16, Roulotte Diagonale = 17, Roulotte Esplanade = 18, Satellite = 19, Self-service Le Hodler = 20, Table de Vallotton = 21)	