Notice: This is a preprint of a work that was accepted for publication in the Journal of Choice Modelling. Peer review is not included in this document. Changes have been made during the publishing process. A definitive version was published in the Journal of Choice modelling:

Danalet, A., Tinguely, L., de Lapparent, M. and Bierlaire, M. (2016). Location choice with longitudinal WiFi data, *Journal of Choice Modelling* **18**: 1–17.
URL: *http://dx.doi.org/10.1016/j.jocm.2016.04.003*

Location choice with longitudinal WiFi data

Antonin Danalet * Loïc Tinguely * Matthieu de Lapparent *
Michel Bierlaire *

19th November 2015

Report TRANSP-OR 151110
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
`transp-or.epfl.ch`

Abstract

While moving from diary survey to location-aware technologies, recent data collection techniques provide new insights about location choices. Only few dynamic models of location choice exist in the literature, and none of them to our knowledge correct for serial correlation. In this paper, we apply a method proposed by Wooldridge (2005) to deal with the initial values problem on the choice of catering locations on a campus using WiFi traces. Cross-validation, price elasticity and simulation of a scenario predicting the opening of a new catering location are presented. Predicted market shares of the new catering location correspond to point-of-sale data of the first week of opening.

Key words

location choice; panel data; pedestrians; dynamic model; initial conditions problem

The data, model specification files and results of our case study are freely available in Danalet et al. (2015).

*ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory (TRANSP-OR), antonin.danalet@epfl.ch, loic.tinguely@epfl.ch, matthieu.delapparent@epfl.ch, michel.bierlaire@epfl.ch

Contents

1 Introduction

Properly modeling destination choices is important to understand travel behavior and travel demand, both at the urban scale and in pedestrian facilities. In transportation, destination choice models are used by local and national authorities for planning future infrastructures and policies (e.g., Fox et al.; 2014) and for the planning and design of multimodal transport hubs (Hoogendoorn and Bovy; 2004). In tourism, the choice of destinations is important for demand for holidays locations (e.g., Yang et al.; 2013) and for the management of pedestrian flows in museums (Yoshimura et al.; 2014) and in parks (O'Connor et al.; 2005). In all these contexts, destination choice models infer on the relevant factors that influence the decisions and allow to test policies when building new infrastructures or optimizing current ones. Demand management strategies can be evaluated.

Destination choice models rely mostly on static frameworks with cross-sectional data, collected at one point in time (e.g., Ben-Akiva and Lerman; 1985; Zhu and Timmermans; 2011; Scott and He; 2012; Kalakou et al.; 2014). Panel data are difficult and expensive to collect using standard survey techniques (Yang and Timmermans; 2015), and sometimes inexistant, e.g., for the analysis of induced trafic at an aggregate level (Weis and Axhausen; 2009). In absence of actual panel data, pseudo panel data are constructed by grouping individuals from cross sectional sectional data into cohorts and by considering behavior of cohorts as individuals (Deaton; 1985; Weis and Axhausen; 2009; McDonald; 2015). However, actual panel data from new technologies are more and more used (Carrion et al.; 2014; Kazagli et al.; 2014). Network traces (e.g., WiFi traces or cell tower data) are increasingly available for location choices (see Section 2.2). Compared to traditional surveys, network traces follow individuals over longer periods (see Section 2.1). Thus, it becomes possible to collect sequences of activity locations covering several days, weeks or months. Location choice models must be adapted to use these data.

In this article, we apply a method proposed by Wooldridge (2005) to deal with the initial values problem in modeling dynamics of location choices conditional to an activity type. Accounting for panel data nature in location choices has never been treated in the literature before. It allows to correct for serial correlation, while understanding people's habits in their decision process. The methodology is applied to sequences of catering locations on a campus collected using WiFi access points (Danalet et al.; 2014).

We present a literature review in Section 2, the methodology in Section 3 and apply it to a pedestrian case study in Section 4, including cross-validation and forecasting. We conclude in Section 5.

2 Literature review

2.1 From diary surveys to location-aware technologies

One recent trend in travel demand modeling is the usage of data from location-aware technologies (Chen and Yang; 2014; Danalet et al.; 2014; Miller; 2014; Carrel et al.; 2015). Traditionally, collected data are revealed preferences about activity and travel patterns from diary surveys, where people describe 1-2 past days (Ettema; 1996; Carrel et al.; 2015). The largest panel surveys include a six-week period for 317 participants (Axhausen et al.; 2002), a six-week period for 261 participants (Axhausen et al.; 2007) and a twelve-week period for 71 participants (Schlich; 2004). Most long-term surveys cover a maximum of 7 days and are not panel data (Ortúzar et al.; 2011; Carrel et al.; 2015). GPS-based prompted recall activity-travel survey allows for longitudinal surveys, using GPS devices carried by respondents (Frignani et al.; 2010; Yang and Timmermans; 2015). Recall methods can be implemented on mobile devices (Rindfüser et al.; 2003; Cottrill et al.; 2013).

Location-aware technologies help improving the quality of explicit surveying. They can also be used alone, from the communication infrastructure side, such as cell tower traces or WiFi access points traces (Bekhor et al.; 2013; Calabrese et al.; 2013; Danalet et al.; 2014), or from the individuals' devices (Etter et al.; 2012; Buisson; 2014; Chen and Yang; 2014; Carrel et al.; 2015). Etter et al. (2012) show that it is possible to predict up to 60% of next visited places from passive smartphone data.

2.2 Location choice

Location choice models are common in studies of urban transportation policies and planning. Ben-Akiva and Lerman (1985) mention three of them, for the Paris region and Maceio, Brazil. Very often, such models are applied to the choice of location for grocery shopping (Timmermans; 1996; Dellaert et al.; 1998; Fox et al.; 2004; Scott and He; 2012). Location choice models are applied to several other different contexts, such as the choice of a departure airport (Furuichi and Koppelman; 1994), the choice of a hospital for patients by general practitioner (primary care physicians) (Whynes et al.; 1996), the choice of touristic destinations (Woodside and Lysonski; 1989; Um and Crompton; 1990; Eymann and Ronning; 1997; Oppermann; 2000; Seddighi and Theocharous; 2002; Bigano et al.; 2006; Chi and Qu; 2008; Gössling et al.; 2012; Yang et al.; 2013) and in particular recreational outdoor facilities (Fesenmaier; 1988; Scarpa and Thiene; 2005; Thiene and Scarpa; 2009), the choice of migrants (Fotheringham; 1986) and the optimal allocation of charging stations for electric vehicles (He et al.; 2013).

Regarding pedestrians, Zhu and Timmermans (2011) propose heuristics rules including principles of bounding rationality and compare them to discrete choice models. The models are validated on the same sample used for estimation and no cross validation is performed. Ton (2014) studies route and location choice in train stations based on tracking and counting data. Counting data come from infrared scanners and tracking data come from WiFi and Bluetooth scanners. Counting data allow to apply the model to pedestrians without smartphones. The choice is between locations for a given activity type. Kalakou et al. (2014) apply a similar approach for location choice for a given activity type (which coffee shop knowing that the individual is visiting one) in an airport.

2.2.1 Attributes of the choice for a location

The main attributes in location choices in urban context are travel time, travel cost and distance (Cambridge Systematics Europe; 1984; Ben-Akiva and Lerman; 1985; Whynes et al.; 1996). Other variables are used: park-seek time, parking cost, type of neighborhood, number of different services (banks, post offices, medical facilities, offices, shops, etc.) in the zone (Cambridge Systematics Europe; 1984; Ben-Akiva and Lerman; 1985). Another typical attribute is the *size* in the context of aggregation of alternatives (see Section 2.2.2). It represents the number of elemental alternatives in the considered aggregate alternatives (subsets of the choice set). The interpretation of this attribute is complicated, since it absorbs both the preference for a large set of destinations compared to a small one and the correlation between destinations in the set. The expected sign is opposite in the two situations (Frejinger and Bierlaire; 2007). In shop patronage, the main attributes are the retail floor space, the accessibility and the price (Arnold et al.; 1983; Scott and He; 2012). Other attributes include parking facilities, number of speciality stores, number of retail employees, having different services (foodcourts, cinemas) (Zhu et al.; 2006; Shobeirinejad et al.; 2013) or symbolic acts (support of community charities, front-door greeters, patriotic displays) (Arnold et al.; 1996).

In the pedestrian context, the main attributes of location choice are the attractiveness of the location and travel time. More specifically, models include floor space (Borgers and Timmermans; 1986), pedestrian environment in neighborhood, employment (Eash; 1999) as measures of attractiveness and distance as an approximation of travel time (Borgers and Timmermans; 1986; Ton; 2014). Kalakou et al. (2014) include space syntax in the specification of the utility through "integration", i.e., a measure of accessibility.

2.2.2 Location choice models

In an urban context, models are often based on tours, characterized by a travel mode and a destination. Models of joint choice of travel mode and destination aggregate destinations into zones (Cambridge Systematics Europe; 1984; Ben-Akiva and Lerman; 1985). Stratified importance sampling is used, dividing the destination choice set into non-overlapping strata based on the origin zone. In the Paris Region example, this procedure decreases the choice set from 595 destinations $\times$ 4 travel modes to 7 sampled alternatives for each trip (Cambridge Systematics Europe; 1984; Ben-Akiva and Lerman; 1985). In a pedestrian context, the choice set is often smaller, due to the smaller study area (e.g., Ton; 2014; Kalakou et al.; 2014, with 2 to 4 alternatives). Most destination choice models are logit models (Arnold et al.; 1983; Zhu et al.; 2006; Scott and He; 2012; Kalakou et al.; 2014; Ton; 2014). Probit models have been used (e.g., Whynes et al.; 1996).

Panel data are common in transportation research (Golob et al.; 1997) and habits are often observed in travel behavior (Gärling and Axhausen; 2003), in particular in route choice (Aarts and Dijksterhuis; 2000; Bamberg et al.; 2003; Thøgersen; 2006; Eriksson et al.; 2008; Verplanken et al.; 2008; Gardner; 2009; Schwanen et al.; 2012) and in car ownership (Jong et al.; 2004). In Markov models of destination choices, transition matrix represents the probability of choosing a destination given the choice of destination at the previous stop. Markov models are criticized for being descriptive, replicating the data, and not being sensitive to behavioral changes (Kitamura; 1990; Timmermans et al.; 1992). Dynamic models using panel data increase statistical efficiency, improve predictions and allow to study behavioral dynamics (Kitamura; 1990). In 1990, Kitamura (1990) considered the inclusion of lag terms in discrete choice models not well advanced. Unresolved issues in the estimation of dynamic models using panel data included the representation of the initial conditions and the correlated error term in dynamic models. In 2001, McFadden (2001) highlighted the importance of panel data in discrete choice models (Carrel et al.; 2015). Yang et al. (2013) model the choice of a second touristic destination after visiting a first one using a nested logit. The panel nature of the data is not explicitly taken into account in their model, similarly to Wu (2012, ch. 5.2), since the previous destination characteristics are not included in their model. For pedestrian destination models, Timmermans et al. (1992) mention in their review the "issue of whether a pedestrian tends to always buy certain items in the same store", i.e., the question of loyalty, as future research.

Few authors explicitly include lagged variables in location models. For non-work activity location choice, Sivakumar and Bhat (2007) include the previous location choices of the individual and the frequency of past visits in the same location in the utility. They do not deal with endogeneity issues due to the presence of lagged utility functions and assume the first location choice to be exogenous. In tourism literature, Grigolon et al. (2014) include the previous vacation length choice in the choice of the current vacation length. They compare a logit, a mixed logit and a dynamic mixed logit and show that the dynamic mixed logit is the best in estimation and forecasting. In their dynamic mixed logit, by assuming that the error term is independent of the variables (i.e., exogenous), and in particular independent of the lagged variable, they assume that unobserved attributes do not persist over time for a given individual. This can lead to bias in the estimation of the model, in particular when the choice of the vacation length of a stay is influenced by variables not included in their model. In the choice of a shop in a pedestrian street, Zhu et al. (2006) also face serial correlation and mention independence issues as a technical challenge for future research.

In light of the above literature review, we emphasize that location-aware technologies allow to collect panel data in the long term. These data must be used in location choice models and lagged variables must be included in the utility function of locations. This leads to bias in the estimation of the model and serial correlation. Wooldridge (2005) proposes a solution to the problem of serial correlation (see Section 3 for details). It is mostly applied to binary probit (Arulampalam and Stewart; 2009). Our contribution in this paper develops a location choice model using panel data from localization-aware technologies. We include a lagged variable and Wooldrige's correction for endogeneity. To our knowledge, this correction has never been applied to dynamic location choice. We apply it to a logit model with 21 locations in the choice set.

3 Methodology

We consider that an individual n repeatedly visits locations. For each individual n, we assume a sequence of events $\{1, ..., t_n, ..., T_n\}$. This sequence is exogenous and individual specific. At each event, a location choice is made. The indicator y_{int_n} is 1 if individual n selects location i for event t_n. The time interval between two events vary, as well as the number T_n of events per individual. To make the notation light, we use t instead of t_n in the following developments.

A sequence of events with varying time intervals between the decisions is typical for the choice of buying or selling for investors in the stock market (e.g., Robin and Bierlaire; 2012). It is also common when considering the activity location choice conditional on an activity type (e.g., Kalakou et al.; 2014; Ton; 2014). The modeling and forecasting of choices of activity type and time intervals between events is covered in Danalet and Bierlaire (2015).

We use a logit model for the choice of a location i. We present three models: a *static* model, a *dynamic model without agent effect* and a *dynamic model with agent effect*.

We associate a utility U_{int} with a location i:

$$U_{int} = V_{int} + \varepsilon_{int} \tag{1}$$

where $i \in \mathcal{C}_{nt}$ and $\mathcal{C}_{nt}$ is the choice set of all available locations at time t for individual n. This model is simple to estimate when we assume that $\varepsilon_{int} \overset{iid}{\sim} EV(0,1)$ across i, n and t, i.e., a static logit model. It ignores two aspects: dynamics and serial correlation.

First, the choice at a certain event t may depend on previous choices. Individuals tend to have state dependence towards already visited locations. For simplification purpose, we make three additional assumptions. First, we assume a dynamic process of order one: the current level of utility of location i partly depends on the previously chosen location for the same type of activity. Second, the state dependence is location specific: utility for a location depends only on previous choice of this location. Third, we assume that the weight ρ of this state dependence is the same for every individuals n and every locations i (the assumption is restrictive and could be relaxed by considering variations across locations and individuals):

$$U_{int} = V_{int} + \rho y_{in(t-1)} + \varepsilon_{int} \tag{2}$$

where $y_{in(t-1)}$ is a dummy variable with value one if location i was chosen by individual n as the previous location choice, and 0 otherwise. The coefficient ρ measures the effect of previous experience of the location on its current utility. ρ can be specific to the time of day, e.g., the choice of a catering location for a coffee break *in the afternoon* depends on the previous catering location choice *in the afternoon*, ignoring the other catering activity locations in-between.

We assume that the time interval between two events does not change the impact of the previous experience, i.e., the memory of the previous activity location choice. The choice probability of an activity location is only influenced by a previous visit at the same activity location. Duration between two events does not affect choice probability.

We initially assume that the previous choice $y_{in(t-1)}$ is independent of the error term ε_{int} (strict exogeneity assumption) and that ε_{int} are independent and identically distributed across i, n and t. We term such a model a *dynamic model without agent effect*.

The error terms ε_{int} model the unobserved factors. In the *static* and the *dynamic model without agent effect*, we assume that they are independently distributed over time, individuals and locations. In practice, it is very likely that they share time-invariant components associated with the decision-maker, thereby generating *serial correlation*. This raises the second issue of the static model. For example, in the successive choice of a restaurant, taste for healthy food is usually unobserved (Burton et al.; 2014; Chen and Yang; 2014). In our context, it can be considered as an unobserved time-invariant factor[1].

As a consequence, the lagged variable $y_{in(t-1)}$ and the unobserved factors ε_{int} are correlated since they both depend on the time-invariant factor, also known as *agent effects*. This is called *endogeneity*. It has to be taken into account to avoid bias in the estimation of the parameters of the model.

We relax the independence assumption of error terms $\varepsilon_{in(t-1)}$ and ε_{int} by replacing the original single error term ε_{int} by the sum of two error terms: $\alpha_{in} + \varepsilon'_{int}$. α_{in} is the agent effect. It is time-invariant and represents the long-term preferences of individual n over time for location i. The agent effect α_{in} does not vary over time but varies across individuals (*inter-individual variability*). ε'_{int} is the unobserved heterogeneity and represents the short-term variation of preferences of individual n (*intra-individual variability*). ε'_{int} are independent across time and individuals. The utility function becomes:

$$U_{int} = V_{int} + \rho y_{in(t-1)} + \alpha_{in} + \varepsilon'_{int}. \tag{3}$$

In classical dynamic panel data models with agent effects and lagged dependent variables, solving endogeneity bias in estimation by some maximum likelihood techniques requires the computation of the marginal/steady state choice probability for the first observed outcome of the dependent variable (Wooldridge; 2005, p.40: "use the joint distribution of all outcomes on the response—including that in the initial time period—conditional on unobserved heterogeneity and observed strictly exogenous explanatory variables"). This is often referred as the *initial conditions problem* in econometrics (Heckman; 1981; Hsaio; 2003; Train; 2003; Wooldridge; 2005). Computation of such marginal probability is intractable except for some simple binary models (see Bhargava and Sargan; 1983, Hsaio;

[1] We agree that taste may change in the lifecycle of an individual, but not during the time horizon of the data we use for the application.

2003 (Section 4.3) and Wooldridge; 2005). Several authors have proposed circumventing strategies to solve this problem (see Hsaio; 2003 and Wooldridge; 2005 for reviews). We here build up on the Wooldridge (2005) correction method.

3.1 Correcting endogeneity for dynamic discrete choice models

In general, endogeneity must be corrected to get consistent estimates (Train; 2003, Ch. 13). *Control functions* capture the relationship between the unobserved factors and the observed variables and "absorb" endogeneity (Heckman; 1978).

Wooldridge (2005) proposes to model the distribution of the agent effect α_{in} conditional on the initial value and any exogenous explanatory variables:

$$\alpha_{in} = a + by_{in0} + c'\bar{x}_n + \xi_{in} \tag{4}$$

where ξ_{in} is normally distributed, $\xi_{in} \sim N(0, \Sigma_\alpha)$, with Σ_α is a matrix of parameters to be estimated[2], and $\bar{x}_n$ is a vector of time-invariant explanatory variables (i.e., long-term preferences, socioeconomic characteristics). The utility of the *dynamic model with agent effect* is:

$$U_{int} = V_{int} + \rho y_{in(t-1)} + a + by_{in0} + c'\bar{x}_n + \xi_{in} + \varepsilon'_{int}. \tag{5}$$

The endogeneity issue is addressed with this utility function, given assumption in Eq. 4 is valid (see Wooldridge (2005) for a detailed discussion). The contribution of a series of observations y_{int} at times $t = 1, ..., T$ for individual n to the likelihood function, conditional on the initial value y_{in0} and the agent effects $\alpha_n = \{\alpha_{in}, \forall i\}$, is:

$$P(y_{in1}, y_{in2}, ..., y_{int} | y_{in0}, \alpha_n) = \prod_{t=1}^{T} P(y_{int} | y_{in0}, y_{in(t-1)}, \alpha_n). \tag{6}$$

Note that we do not model the first choice y_{in0}. Given our assumptions, it turns out that our estimator is a conditional maximum likelihood estimator. It is asymptotically equivalent to the full information maximum likelihood estimator. Only efficiency is affected.

When integrating out the agent effects $\alpha_n \in \mathbb{R}^{dim(i)}$, as for any mixture model, Eq. 6 becomes:

$$P(y_{in1}, y_{in2}, ..., y_{int} | y_{in0}) = \int_{\alpha_n} \prod_{t=1}^{T} P(y_{int} | y_{in0}, y_{in(t-1)}, \alpha_n) f(\alpha_n | y_{in0}, \bar{x}_n) d\alpha_n. \tag{7}$$

[2]Note that Wooldridge (2005) is more general in his approach and other distributions might be used. Here we assume $\Sigma_{\alpha_i} = \sigma^2_{\alpha_i} I$. In the current developments, the parameters of the normal distribution σ_{α_i} are location specific, but the i subscript is omitted to make the notation light.

Here, $P(y_{int}|y_{in0}, y_{in(t-1)}, \alpha_n)$ is a logit model. $f(\alpha_n|y_{in0}, \bar{x}_n)$ is normally distributed, following Eq. 4. Endogeneity is corrected.

Table 1 summarizes the three different models presented in Section 3.

Static model	Dynamic model without agent effect	Dynamic model with agent effect
$\rho = 0$ $a, b, c, \sigma_\alpha^2 = 0$	$\rho \neq 0$ $a, b, c, \sigma_\alpha^2 = 0$	$\rho \neq 0$ $a, b, c, \sigma_\alpha^2 \neq 0$

Table 1: Description of static model, dynamic model without agent effect and dynamic model with panel effect as a function of Eq 5.

4 Pedestrian case study for EPFL catering locations

We present results for the three models presented in Section 3, summarized in Table 1, in the context of location choice on the EPFL campus. We focus on the choice of catering facilities during their opening hours. The choice set $\mathcal{C}$ contains 21 alternatives corresponding to the services available in 2012 (Figure 1). We use WiFi traces to detect sequences of activity episodes. WiFi traces are merged with map information (localization of points of interest), attractivity (aggregate measures of occupancy, e.g., from point-of-sale data) and time constraints (e.g., shop opening or class schedules), as described in Danalet et al. (2014), with $L = 1$. This Bayesian approach merges data, detects stops and give semantics to the WiFi traces. The raw data are available in Danalet (2015). The processed data, model specification files for Pythonbiogeme (Bierlaire; 2003; Bierlaire and Fetiarison; 2009) and all results presented in this section are available in Danalet et al. (2015). Note that we have access to some socio-economic attributes in this dataset. We associated MAC addresses to usernames using the radius server, and then usernames to employee or class attributes using LDAP. Finally, usernames and MAC addresses have been deleted (Danalet et al.; 2014).

4.1 Model specification and estimation

The explanatory variables used for the location choice are (1) attributes varying with the alternatives: distance from the previous activity location in the sequence, duration, cost, time of the day, opening hours, quality evaluation of the catering location, its capacity, its type of offer, and (2) characteristics describing the choice context, constant across

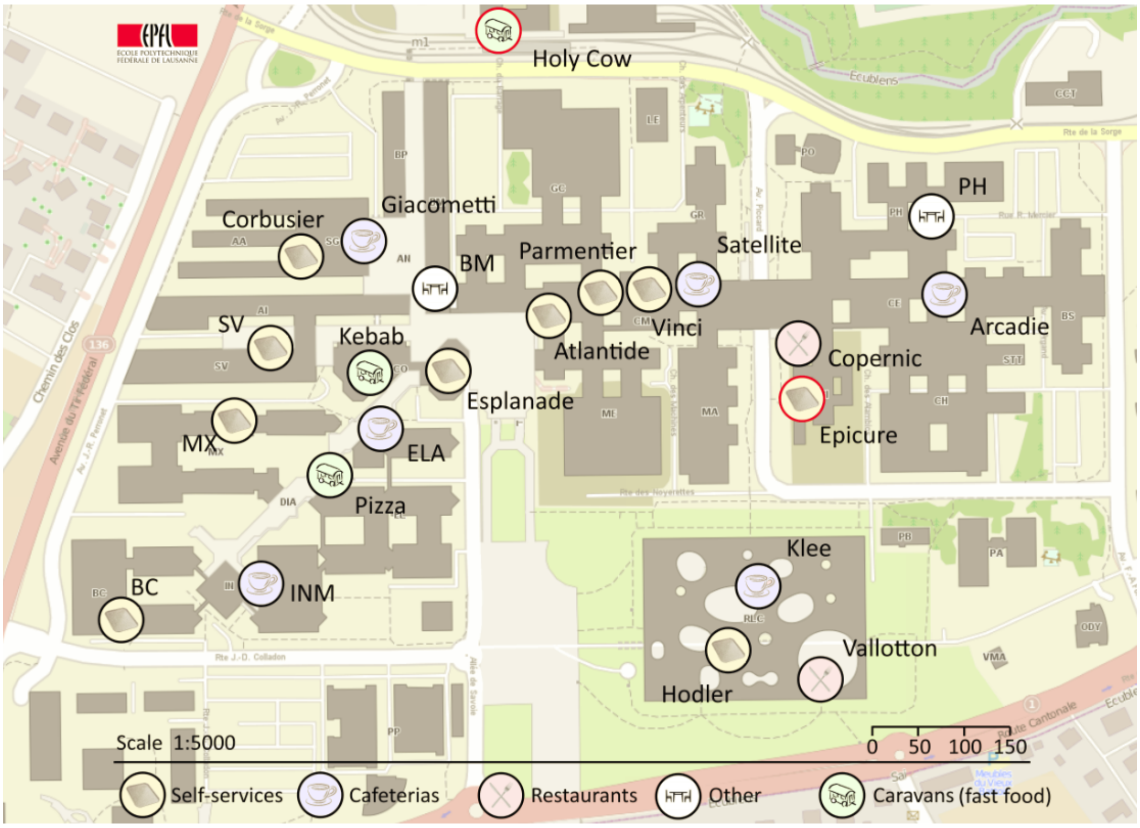

Figure 1: Catering facilities on EPFL campus with different categories: self-services, cafeterias, restaurants, caravans (fast food) and others. The alternatives in red circles did not exist in 2012, when WiFi traces were collected. Image: Tinguely (2015).

alternatives: weather conditions, day of the year, socio-economic attributes. Descriptive statistics on the collected data are available in Appendix A.1.

In the dynamic models, habits are assumed only for the morning and lunch break. Two lagged variables $y_{in(t-1)}$ are defined in the dynamic models, one for the morning and one for the lunch break. Thus, the dynamic Markov process is over individuals and periods of the day. Equations 2 and 3 become:

$$U_{int} = V_{int} + \rho_{\text{morning}} y^{\text{morning}}_{in(t-1)} + \rho_{\text{lunch}} y^{\text{lunch}}_{in(t-1)} + \varepsilon_{int} \tag{8}$$

$$U_{int} = V_{int} + \rho_{\text{morning}} y^{\text{morning}}_{in(t-1)} + \rho_{\text{lunch}} y^{\text{lunch}}_{in(t-1)} + \alpha^{\text{morning}}_{in} + \alpha^{\text{lunch}}_{in} + \varepsilon'_{int} \tag{9}$$

The specification of the agent effect distribution must be correct to get consistent estimates (Wooldridge; 2005). Therefore, we propose two different specifications for α_{in}. The first specification corresponds to $c = 0$ in Eq. 4. We assume the agent effect to depend only on the *first choice*:

$$\alpha_{in} = a + b y_{in0} + \xi_n. \tag{10}$$

The second specification for the agent effect includes the count y_{int}^{count} of previous choices of alternative i by individual n up to the event t of the current choice: $y_{int}^{\text{count}} = \sum_{t'=1}^{t-1} I(y_{int'})$. Note that in the definition of the count of previous choices, the first observation y_{in0} is not included and the summation start at $t' = 1$. It allows to avoid biases (Rabe-Hesketh and Skrondal; 2013).

Eq. 4 becomes:

$$\alpha_{in} = a + by_{in0} + cy_{int}^{\text{count}} + \xi_n. \tag{11}$$

Since the lagged variable $y_{in(t-1)}$ is interacted with the period of the day, the count of previous choices is also specified for each period of the day.

Consequently, we estimate 4 models: the static model (utility defined in Eq. 1), a dynamic model without agent effect (utility defined in Eq. 2) and two dynamic models with agent effect: one with a first choice agent effect specification (Eq. 10) and one with a first choice and frequency specification (Eq. 11).

Static model	**Dynamic model without agent effect**	**Dynamic model with agent effect**	
		First choice	First choice and frequency
$\rho = 0$	$\rho \neq 0$	$\rho \neq 0$	$\rho \neq 0$
$a = 0$	$a = 0$	$a \neq 0$	$a \neq 0$
$b = 0$	$b = 0$	$b \neq 0$	$b \neq 0$
$c = 0$	$c = 0$	$c = 0$	$c \neq 0$
$\sigma_\alpha^2 = 0$	$\sigma_\alpha^2 = 0$	$\sigma_\alpha^2 \neq 0$	$\sigma_\alpha^2 \neq 0$

Table 2: Description of static model, dynamic model without agent effect and two dynamic models with panel effect used in the case study as a function of Eq 5.

We use a linear specification for the different models, whose variables are described in Table 3. Table 4 summarizes the estimation results for the 4 models of Table 2.

Parameters	Variables	Description of the variable
ASC_i	1_i	Alternative specific constant for catering location i
$\beta_{\text{dist, cat},ToD}$	$\text{dist}_{\text{cat},ToD}$	Distance from the previous activity episode (previous stop, not necessarily a catering location)
$\beta_{\text{no dist}}$	$1_{\text{dist }NA}$	Variable for missing data about distance
β_{eval}	eval_i	Evaluation of catering location from survey data (grade between 1 and 6)
$\beta_{\text{cost, student}}$	$\text{cost}_{\text{students}}$	Cost of the cheapest meal for students
$\beta_{\text{cost, employees}}$	$\text{cost}_{\text{employees}}$	Cost of the cheapest meal for employees
β_{beer}	1_{beer}	Availability of beer after 14:00
β_{dinner}	1_{dinner}	Availability of dinner
β_{capacity}	$\text{capacity}_{\text{outdoor}}$	Number of seats in the catering location
ρ_{ToD}	$y_{it(t-1),ToD}$	Indicator variable with value 1 if the previous catering location in the same time of day (ToD) is the same as the current catering location
c_{ToD}	y_{int}^{count}	Variable counting the frequency of visit to catering location i in the same time of day (ToD)
$\sigma_{i,ToD}$	1_{ToD}	Variance of ξ for each time of day (ToD) and each catering location i

Table 3: Description of the variables in the catering location choice model. Some variables are interacted with the category of catering location (cat), i.e. the categories of restaurant presented in Figure 1, or are interacted with time of day (ToD), divided in morning (until 11:29), lunch break (11:30-13:59), afternoon (14:00-17:59), dinner (18:00-19:59) and night (from 20:00).

Parameters	Static model Value	Static model t-test	Dynamic model without agent effect Value	Dynamic model without agent effect t-test	Dynamic model with agent effect: First choice Value	First choice t-test	First choice and frequency Value	First choice and frequency t-test
$ASC_{\text{Le Klee}}$	−3.26	−5.52	−2.91	−4.82	−4.90	−3.75	−5.24	−3.83
$ASC_{\text{Cafétéria BC}}$	0.387	0.97*	0.481	1.12*	−1.09	−1.62*	−0.682	−0.88*
ASC_{BM}	0.450	1.29*	0.453	1.33*	−0.147	−0.24*	−0.320	−0.49*
$ASC_{\text{Cafétéria ELA}}$	−0.823	−2.42	−0.579	−1.59*	−1.08	−1.68*	−0.919	−1.67*
$ASC_{\text{Cafétéria INM}}$	−2.19	−3.97	−1.82	−3.13	−1.64	−1.52*	−1.81	−1.75*
$ASC_{\text{Cafétéria MX}}$	−0.461	−1.22*	−0.514	−1.23*	−1.89	−2.05	−1.78	−2.57
ASC_{PH}	1.28	3.48	1.11	2.99	0.298	0.62*	0.704	1.39*
$ASC_{\text{L'Arcadie}}$	−0.738	−2.08	−0.684	−1.85*	−1.98	−1.85*	−1.70	−1.81*
$ASC_{\text{L'Atlantide}}$	−0.143	−0.47*	−0.285	−0.88*	−1.23	−2.21	−0.731	−1.23*
$ASC_{\text{Le Copernic}}$	2.83	2.04	2.67	2.29	2.59	0.75*	1.88	1.25*
$ASC_{\text{Le Corbusier}}$	−0.278	−2.05	−0.259	−1.74*	−1.05	−2.52	−0.585	−2.28
$ASC_{\text{Le Giacometti}}$	0.323	1.12*	0.398	1.26*	0.760	1.47*	0.685	1.34*
$ASC_{\text{Le Parmentier}}$	−0.846	−3.22	−0.883	−3.14	−1.44	−3.63	−1.60	−3.61
$ASC_{\text{Le Vinci}}$	−4.11	−5.77	−3.81	−5.35	−8.24	−2.58	−4.97	−3.42
$ASC_{\text{L'Esplanade}}$	0.0	-	0.0	-	0.0	-	0.0	-
$ASC_{\text{L'Ornithorynque}}$	−0.631	−4.81	−0.641	−4.55	−1.26	−6.48	−1.24	−5.74
$ASC_{\text{Caravan Pizza}}$	−1.97	−3.40	−1.84	−3.23	−2.47	−2.89	−1.91	−2.79
$ASC_{\text{Caravan Kebab}}$	−2.73	−4.42	−2.51	−4.16	−3.12	−3.39	−2.64	−3.21
$ASC_{\text{Bar Satellite}}$	−1.60	−4.34	−1.42	−3.72	−2.27	−3.51	−2.65	−4.52
$ASC_{\text{Le Hodler}}$	0.995	2.07	0.954	2.10	2.40	3.33	2.76	3.86
$ASC_{\text{Table de Vallotton}}$	4.25	2.10	4.02	2.56	0.987	0.67*	1.34	0.80*
$\beta_{\text{dist, lunch, cafet}}$	−0.006 89	−13.47	−0.006 12	−11.64	−0.004 06	−6.37	−0.003 97	−6.71
$\beta_{\text{dist, lunch, rest}}$	−0.001 38	−0.63*	−0.001 27	−0.62*	−0.000 498	−0.29*	0.001 66	0.75*
$\beta_{\text{dist, lunch, self}}$	−0.006 38	−15.45	−0.005 43	−12.88	−0.003 94	−8.91	−0.004 00	−9.32
$\beta_{\text{dist, lunch, fast food}}$	−0.009 53	−9.55	−0.008 81	−9.06	−0.006 72	−5.50	−0.006 76	−5.31
$\beta_{\text{dist, lunch, other}}$	−0.001 87	−2.20	−0.001 00	−1.40*	0.000 738	0.79*	0.000 190	0.17*
$\beta_{\text{dist, morning}}$	−0.005 57	−5.74	−0.004 48	−4.59	−0.004 05	−3.88	−0.003 90	−3.60
$\beta_{\text{dist, after lunch}}$	−0.000 453	−0.76*	−0.001 07	−1.84*	−0.001 01	−1.67*	−0.001 07	−1.72*
$\beta_{\text{no dist}}$	−5.07	−12.70	−4.48	−11.79	−3.82	−6.98	−3.66	−7.66
$\beta_{\text{eval, cafet}}$	1.18	12.27	1.10	11.91	1.92	10.59	1.90	7.72
$\beta_{\text{eval, self}}$	1.21	9.25	1.09	8.45	2.12	8.28	2.02	6.55
$\beta_{\text{eval, fast food}}$	1.69	11.81	1.60	11.85	2.71	10.78	2.58	8.65
$\beta_{\text{cost, student}}$	−0.245	−3.50	−0.189	−3.01	−0.471	−4.00	−0.538	−4.47
$\beta_{\text{cost, employees}}$	−0.128	−2.26	−0.102	−1.97	−0.352	−3.20	−0.368	−3.64
β_{beer}	0.722	4.07	0.539	3.02	1.05	3.85	1.14	3.93
β_{dinner}	1.04	3.34	1.03	3.39	0.997	2.60	0.795	2.01
β_{capacity}	0.006 80	2.62	0.007 49	2.69	0.0104	2.71	0.0119	2.82
ρ_{morning}	0.0	-	3.06	17.48	0.591	1.09*	0.476	1.69*
b_{morning}	0.0	-	0.0	-	1.80	3.10	1.46	4.82
c_{morning}	0.0	-	0.0	-	0.0	-	0.450	2.76
ρ_{lunch}	0.0	-	1.78	15.45	0.644	4.36	0.355	1.95*
b_{lunch}	0.0	-	0.0	-	1.19	5.50	1.07	5.22
c_{lunch}	0.0	-	0.0	-	0.0	-	0.618	3.36
$\sigma_{\text{Klee, morning}}$	0.0	-	0.0	-	−2.55	−3.51	2.17	3.28
$\sigma_{\text{BC, morning}}$	0.0	-	0.0	-	1.76	5.91	1.61	3.90
$\sigma_{\text{BM, morning}}$	0.0	-	0.0	-	−0.578	−1.04*	0.195	0.51*
$\sigma_{\text{ELA, morning}}$	0.0	-	0.0	-	1.72	2.69	1.14	2.80
$\sigma_{\text{INM, morning}}$	0.0	-	0.0	-	−1.01	−2.31	0.725	0.87*
$\sigma_{\text{MX, morning}}$	0.0	-	0.0	-	−0.0850	−0.18*	1.17	1.91*
$\sigma_{\text{PH, morning}}$	0.0	-	0.0	-	0.246	0.79*	−0.352	−1.44*
$\sigma_{\text{Arcadie, morning}}$	0.0	-	0.0	-	1.41	1.77*	−0.726	−1.19*
$\sigma_{\text{Atlantide, morning}}$	0.0	-	0.0	-	−1.85	−6.59	1.21	4.39
$\sigma_{\text{Copernic, morning}}$	0.0	-	0.0	-	−0.922	−0.40*	0.378	0.86*
$\sigma_{\text{Corbusier, morning}}$	0.0	-	0.0	-	1.85	2.74	−1.74	−2.59
$\sigma_{\text{Giacometti, morning}}$	0.0	-	0.0	-	−0.007 21	−0.02*	−0.147	−0.49*
$\sigma_{\text{Parmentier, morning}}$	0.0	-	0.0	-	0.967	1.69*	−1.43	−1.98
$\sigma_{\text{Vinci, morning}}$	0.0	-	0.0	-	−0.0815	−0.23*	0.396	0.74*
$\sigma_{\text{Esplanade, morning}}$	0.0	-	0.0	-	0.0	-	0.0	-
$\sigma_{\text{Ornithorynque, morning}}$	0.0	-	0.0	-	0.0261	0.05*	0.237	0.59*
$\sigma_{\text{Pizza, morning}}$	0.0	-	0.0	-	1.60	2.74	−0.932	−6.03
$\sigma_{\text{Kebab, morning}}$	0.0	-	0.0	-	1.82	5.98	−1.79	−5.59
$\sigma_{\text{Satellite, morning}}$	0.0	-	0.0	-	2.02	5.35	−2.32	−6.41
$\sigma_{\text{Hodler, morning}}$	0.0	-	0.0	-	1.71	2.42	0.290	0.41*

continued . . .

	Static model		Dynamic model without agent effect		Dynamic model with agent effect			
					First choice		First choice and frequency	
Parameters	Value	t-test	Value	t-test	Value	t-test	Value	t-test
$\sigma_{\text{Vallotton, morning}}$	0.0	-	0.0	-	0.578	0.53*	0.292	0.75*
$\sigma_{\text{Klee, lunch}}$	0.0	-	0.0	-	−2.59	−5.44	2.71	7.08
$\sigma_{\text{BC, lunch}}$	0.0	-	0.0	-	2.06	6.11	−2.20	−7.48
$\sigma_{\text{BM, lunch}}$	0.0	-	0.0	-	2.33	3.52	−2.50	−4.18
$\sigma_{\text{ELA, lunch}}$	0.0	-	0.0	-	−1.05	−3.92	−0.789	−2.60
$\sigma_{\text{INM, lunch}}$	0.0	-	0.0	-	0.883	1.47*	−1.33	−2.83
$\sigma_{\text{MX, lunch}}$	0.0	-	0.0	-	−2.06	−6.27	1.66	9.55
$\sigma_{\text{PH, lunch}}$	0.0	-	0.0	-	2.63	7.13	−2.30	−3.39
$\sigma_{\text{Arcadie, lunch}}$	0.0	-	0.0	-	2.97	5.74	2.46	5.25
$\sigma_{\text{Atlantide, lunch}}$	0.0	-	0.0	-	1.85	5.44	−1.54	−6.82
$\sigma_{\text{Copernic, lunch}}$	0.0	-	0.0	-	5.78	2.92	6.06	4.32
$\sigma_{\text{Corbusier, lunch}}$	0.0	-	0.0	-	−1.27	−3.89	−0.855	−3.16
$\sigma_{\text{Giacometti, lunch}}$	0.0	-	0.0	-	−1.31	−6.13	1.24	6.35
$\sigma_{\text{Parmentier, lunch}}$	0.0	-	0.0	-	0.961	3.75	−1.19	−2.57
$\sigma_{\text{Vinci, lunch}}$	0.0	-	0.0	-	3.56	1.91*	−1.37	−1.36
$\sigma_{\text{Esplanade, lunch}}$	0.0	-	0.0	-	0.0	-	0.0	-
$\sigma_{\text{Ornithorynque, lunch}}$	0.0	-	0.0	-	0.128	0.49*	−0.258	−1.23*
$\sigma_{\text{Pizza, lunch}}$	0.0	-	0.0	-	−1.24	−5.42	1.29	5.15
$\sigma_{\text{Kebab, lunch}}$	0.0	-	0.0	-	0.677	3.00	−1.11	−4.48
$\sigma_{\text{Satellite, lunch}}$	0.0	-	0.0	-	0.776	5.26	−1.20	−4.13
$\sigma_{\text{Hodler, lunch}}$	0.0	-	0.0	-	1.05	3.51	−0.910	−1.91*
$\sigma_{\text{Vallotton, lunch}}$	0.0	-	0.0	-	10.7	5.52	−10.8	−7.20
Nb of observations				1868				
$\mathcal{L}(0)$				−5037.914				
Nb estim. param.	36		38		80		82	
$\mathcal{L}(\hat{\beta})$	−3446.109		−3092.106		−2631.929		−2623.843	
Adjused rho square $\bar{\rho}^2$	0.309		0.379		0.462		0.480	
Likelihood ratio test		354.003 (> 5.99)		920.354 (> 58.12)		16.172 (> 5.99)		

Table 4: Summary of estimation results for the 4 models of Table 2. 1868 observations are used for estimation. Parameters without stars are significantly different from zero with a 95 % confidence level. A likelihood ratio test is performed between the static model and the dynamic model without agent effect, between the dynamic model without agent effect and the dynamic model with agent effect (first choice specification), and between the dynamic model with agent effect (first choice specification) and the dynamic model with agent effect (first choice and frequency). The numbers in parenthesis for the likelihood ratio tests are the percentiles of the χ^2 distribution.

Lagged variables ρ_{lunch} and ρ_{morning} have positive signs, showing habits and repeated choices. Their value decreases when the dynamic model includes the agent effect (as compared to the model when it is considered exogenous). It has been reported in Monte Carlo simulations that ρ is overestimated in dynamic models without agent effect as compared to dynamic models with agent effect (Akay; 2012). This is due to the double nature of the lagged variable ρ: the previous choice impacts the current choice because the past experience modifies the current preferences and because the past and current choices both depend on the same time-persistent unobserved parameters. These two factors are called *true state dependence* and *spurious state dependence*, respectively, by Heckman (1978, 1981) (see also Hsaio; 2003, Section 7.5.4). The agent effect, and in particular the *first* *choice and frequency* version of it, is absorbing the time-persistent unobserved preferences.

The parameters have expected signs. Indoor capacity (number of seats) has a positive impact on the choice of visiting a catering location. Distance from the previous activity episode has a negative impact on the propensity to visit a catering location. This effect is strong in the morning and during lunch time for cafeterias, while it is not significant in the afternoon and during lunch time for restaurants (there are not many restaurants on campus, and consequently longer distances to walk). The cost parameters have a negative sign and their magnitude is larger for student than for employees. This is explained by the fact that employees have salaries and thus a higher purchasing power and a lower sensitivity to price. Annual evaluations by students (as a proxy for average quality), offering meals for dinner and beers after 14:00 all have a positive impact on the choice of catering locations.

The dynamic model without agent effect, the dynamic model with agent effect (*first choice* correction) and the dynamic model with agent effect (*first choice and frequency* correction) are unrestricted versions of the previous, simpler model in Table 2 (i.e., static model, dynamic model without agent effect, dynamic mode with agent effect (*first choice* correction), resp.). Table 4 (last line) shows the results of three likelihood ratio tests. In all cases, we can reject the null hypothesis at a 95 % confidence level and the unrestricted model is preferred to the restricted one.

4.2 Validation

Cross-validation has been performed, partitioning the data in an *estimation dataset* containing past observations $i_1, i_2, ..., i_{T_n-1}$ of individuals n and a *validation dataset* with their most recent choice i_{T_n}. Models presented in Section 4.1 are applied to observations in the morning and during lunch break, in order to test the dynamics. The estimation dataset contains 1512 observations. The model is then applied to the validation dataset (containing 144 observations), using the parameter estimates from the previous step. Aggregate average number of visits across individuals' most recent choices from observations and from the model output are compared in Table 5.

In order to compare the performance of the different models over all catering locations in Table 5, we compute the sum of the squares of the errors: $S_m = \sum_i (O_i - E_{i,m})^2$, where O_i is the observed average number of visits for location i and $E_{i,m}$ is the expected average number of visits based on the choice probabilities for location i assuming model m.

Observed and predicted average number of visits show similar tendencies, even for the static model, meaning that the specification of the model is generally good. The model

	Observed		Predicted							
			Static model		Dynamic model without agent effect		Dynamic model with agent effect			
							First choice		First choice and frequency	
Catering locations	Nb	%	Nb	%	Nb	%	Nb	%	Nb	%
Cafet. Le Klee	0	0.0	0.4	0.3	0.3	0.2	0.4	0.3	0.3	0.2
Cafet. ELA	14	9.7	7.6	5.3	6.9	4.8	8.0	5.5	8.0	5.6
Cafet. INM	1	0.7	1.2	0.9	1.1	0.8	2.2	1.5	2.1	1.4
Cafet. MX	6	4.2	6.3	4.4	6.4	4.4	5.3	3.7	5.8	4.0
Cafet. L'Arcadie	6	4.2	1.4	1.0	2.4	1.7	1.5	1.1	1.7	1.2
Cafet. Le Giacometti	13	9.0	12.0	8.3	11.8	8.2	12.8	8.9	12.2	8.5
Cafet. Satellite	5	3.5	7.2	5.0	7.6	5.3	7.8	5.4	7.5	5.2
Self BC	15	10.4	9.7	6.7	9.5	6.6	10.8	7.5	10.8	7.5
Self L'Atlantide	7	4.9	10.8	7.5	10.6	7.4	8.2	5.7	8.1	5.6
Self Le Corbusier	4	2.8	12.6	8.7	10.6	7.4	9.4	6.5	10.8	7.5
Self Le Parmentier	8	5.6	13.1	9.1	12.9	9.0	13.1	9.1	13.2	9.1
Self Le Vinci	1	0.7	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1
Self L'Esplanade	23	16.0	26.1	18.2	25.9	18.0	24.2	16.8	24.4	17.0
Self L'Ornithorynque	15	10.4	15.0	10.4	16.4	11.4	15.6	10.8	15.7	10.9
Self Le Hodler	6	4.2	5.2	3.6	6.1	4.3	5.7	4.0	6.2	4.3
Rest. Le Copernic	1	0.7	1.0	0.7	1.4	1.0	3.4	2.4	3.3	2.3
Rest. Table de Vallotton	1	0.7	1.3	0.9	1.1	0.8	0.6	0.4	0.5	0.3
Caravan Pizza	6	4.2	4.2	2.9	4.5	3.1	4.5	3.1	4.8	3.4
Caravan Kebab	5	3.5	3.6	2.5	3.7	2.6	3.5	2.4	3.8	2.6
Other BM	1	0.7	1.8	1.2	1.2	0.8	1.6	1.1	1.3	0.9
Other PH	6	4.2	3.2	2.2	3.6	2.5	3.3	2.3	3.3	2.3
S_m			232.95		204.01		184.16		173.85	

Table 5: Aggregate average number of visits of the observations and of the different models, from the 144 most recent observations for each individual in the morning and during lunch break. For the observations and for each model, the number of visitors ("Nb") and the proportion of visitors ("%") are presented for each catering location. "Rest." stands for restaurant, "Self" for self-service, "Cafet." for cafeteria.

minimizing the sum of the squares of the errors is the dynamic model with agent effect using the first choice and the frequency. It is also the model that fits the data the best (Table 4). It is an evidence that Wooldridge's approach is valid, and it performs better when the specification of the agent effect distribution includes the frequency of visits.

4.3 Elasticity to price

Aggregate direct elasticity of cost denotes the percent change in the number of visits for each catering location with respect to a change of 1 % in the cost of a meal. Aggregate direct elasticities of cost are presented for each restaurant, for students and employees, in Table 8 in Appendix A.2. Figure 2 summarizes the distribution of aggregate direct elasticities of cost as box-plots for each model, across students and employees.

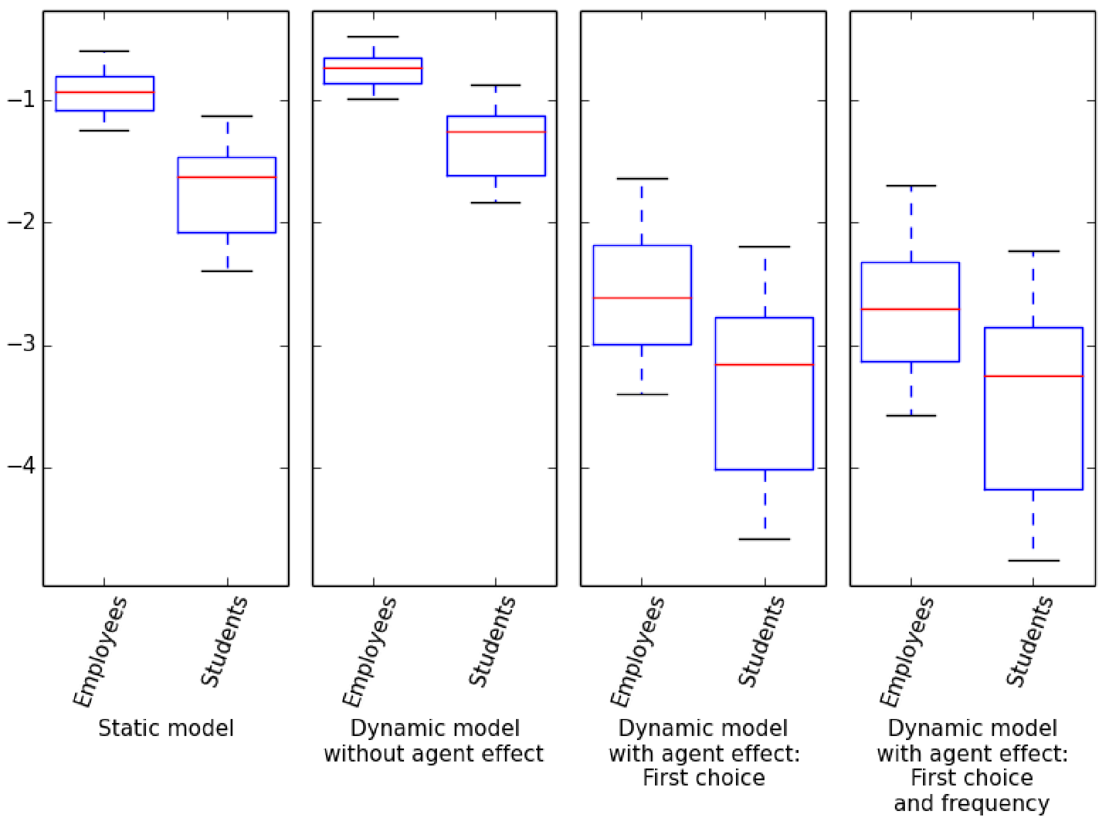

Figure 2: Distribution of aggregate direct elasticities of cost for different models, for students and employees.

Demand for catering locations for students is more elastic to a change in the cost of a meal as compared to employees. This is explained by the higher purchasing power of employees. With the static model and the dynamic model without agent effect, employees mostly show a inelastic demand (< 1 in absolute value) and students show an elastic demand (> 1 in absolute value). With the dynamic models with agent effect, using the first choice and the frequency of choices, the absolute values of elasticities increase and employees have an elastic demand with respect to the cost of a meal. Generally, models ignoring the dynamics are less sensitive to cost. A possible analogy is the presence of unobserved variables, such as quality of the service or of the meal (Train; 2003, ch. 13). Decision makers prefer cheap meals, but also like quality meals. When endogeneity is not corrected for, β_{cost} absorbs both effects and its absolute value is attenuated. When endogeneity is accounted for, β_{cost} is more negative, including only the taste for cheap meals. Here, in the static model and the model without agent effect, β_{cost} absorbs a taste for cheap meals and other unobserved factors positively correlated with cost, such as a warm atmosphere or any attribute of quality for a meal. In the models with agent effect, unobserved factors are absorbed by the agent effect.

4.4 Forecasting visits when opening a new catering location

Data used for estimation have been collected in 2012. We forecast the average number of visits for 2013 after the opening of a new self-service.

In this scenario, habits regarding the new catering location are not considered. The new alternative is not part of people's habits in the model: the previous catering location and the frequency of visits are null ($y_{in(t-1)} = 0$ and $y_{int}^{\text{count}} = 0$ when i is the new catering location).

A new self-service, *L'Epicure*, opened in October 2013. The four models of Table 2 are applied to this new choice set. The parameters for the new self-service are the same as *Le Giacometti*, since it is the most similar existing catering location on campus and no stated preference is available.

The error term of the new alternative and the error term of the most similar existing alternative might be correlated. Indeed, if the new catering location does not share any unobserved attribute with the most similar catering location, a logit specification is valid. On the contrary, if unobserved attributes are shared, the two locations should be included in a nest and a nested logit specification is used for forecasting. Since we don't know the value of the nest parameter θ, an interval of values is used from 1 (i.e., logit model and independent error terms) to $+\infty$ (i.e., perfectly correlated error terms) when applying the model to forecast average number of visits. Results are presented in Fig. 3.

When using a static model, the predicted average frequency of visits varies between 0.7 % and 2.0 % for the full day. When correcting for endogeneity and using frequency of visits in the specification of the agent effect, the predicted average frequency of visits varies between 0.4 % and 1.1 %. It shows that correcting for endogeneity when using panel data has a significant impact when predicting the destination choices of people. The effect of the unknown level of correlation between the new catering location and its most similar alternative also seems lower when using the dynamic models with agent effect.

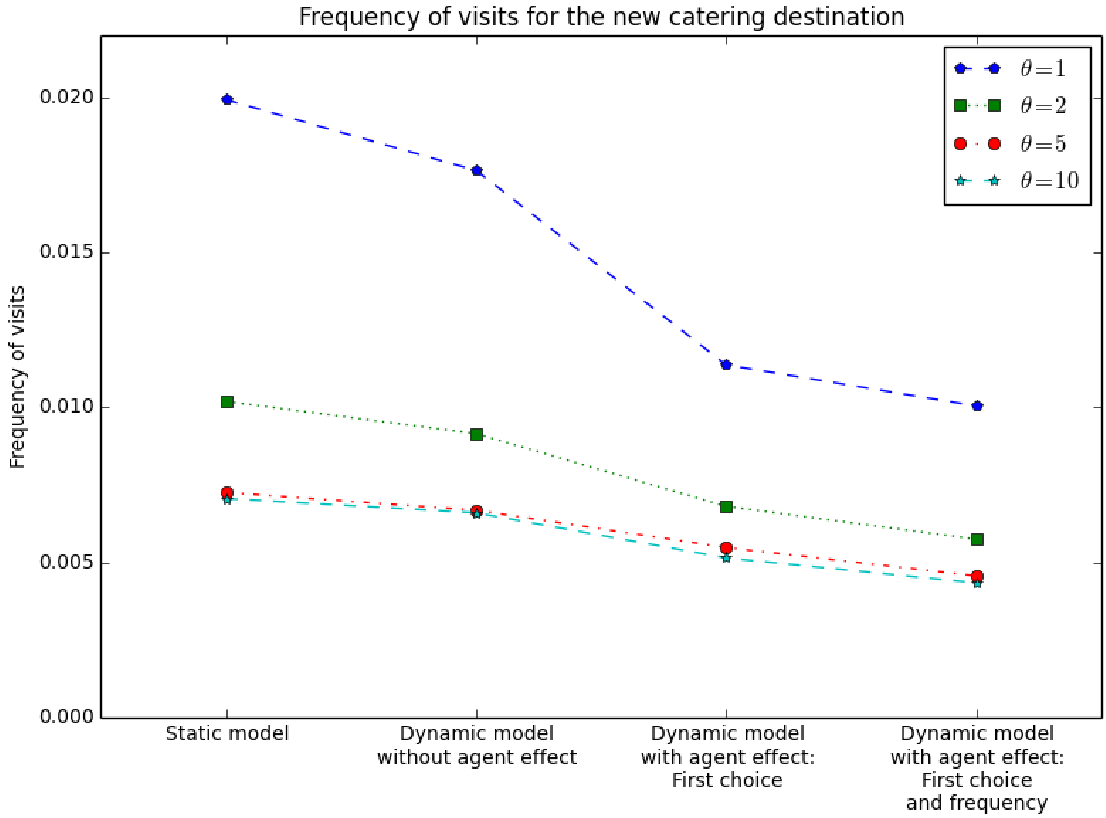

Figure 3: Average frequency of visits for the new self-service for the different models, as a function of θ.

According to point-of-sale data collected from October 21 to 23, 2013, the frequency of financial transactions in the new self-service is 1.5 %, that has a level of magnitude consistent with the values predicted by the model.

Results presented here are only valid in the short term, since the model has been applied only once. For forecasting in the long term, accounting for the habits and routines, the model should be applied several times so that habits for the new location are establishing as an output of the model.

5 Conclusion

In this article, we model location choice conditional on the choice of activity type in activity episodes from WiFi traces. WiFi traces provide panel data. We estimate dynamic models, including lagged variables. They express the habits that could appear in repeated choices.

Including lagged variables in a discrete choice model generates endogeneity. The error term and the explanatory variable representing the previous choice are serially correlated. The so-called initial conditions problem is solved using a control function proposed by Wooldridge (2005). The error term is decomposed in an agent effect and an independent error term. The conditional distribution of the agent effect, knowing the first choice, is approximated.

The approach of Section 3 has been applied to a case study on a campus (Section 4), based on actual WiFi traces, preprocessed as in Danalet et al. (2014). Campus members tend to visit catering locations that are closer, with large capacities, that offer beers and serve meals for dinner. Students are more sensitive to cost than employees. The previous choice significantly impacts the choice of the current catering location in the morning and in the lunch break in a dynamic model without agent effect, without the correction for endogeneity. When controlling for true state dependence and spurious state dependence, time-persistent unobserved effects are detected and the previous choice becomes not significant anymore. A likelihood ratio test has been performed between the different models, and the null hypothesis of a restricted model can always be rejected when comparing two consecutive models in terms of complexity: a dynamic model without agent effect (without correction for endogeneity) is preferred to a static model, a dynamic model with agent effect is preferred to the dynamic model without agent effect, and an agent effect including the first choice and the frequency in its specification is preferred to an agent effect specified with the first choice only.

Models are validated in Section 4.2 and the models seem correctly specified, reproducing the observations of the validation dataset. In terms of predictive power, dynamic models outperform static models, and the agent effect including the first choice and the frequency of visits performs the best.

Elasticity to the cost of a menu and forecasting in the case of the opening of new catering locations are presented in Section 4.3 and 4.4, respectively. Elasticity to the cost of a menu increases with dynamic models with agent effect. In the scenario of the opening of a new catering location, predicted average number of visits correspond to point-of-sale data.

This model can be applied in pedestrian facilities to estimate demand for specific locations. Wooldridge's approach is easy to implement for discrete choice models with many alternatives and improves the estimation and predictive power of the model. Our model specification could be extended towards more complex discrete choice models (e.g., a nested logit where categories of catering locations would be the nests in our case study). Collection of more socioeconomic data would also improve the specification and prove useful for marketing purposes. On campuses, in transportation hubs or music festivals, information on congestion at location (i.e., queues for a service) is likely to be significant in explaining people's behavior. Endogeneity in the model due to congestion could also be corrected, using the occupation rates for each location as measures of queues and congestion at these locations. Some endogeneity could also be related to group effects, when a group chooses a location together instead of each individual independently (Louviere et al.; 2005, Section 2). This could be corrected using proximity as a measure of social networks. Finally, space syntax has been used in recent research and could help in formalizing intuitions such as "visibility" in public spaces.

Acknowledgments

This research is supported by the Swiss National Science Foundation Grant 200021-141099 "Pedestrian dynamics: flows and behavior". The authors would like to thank Etienne Marclay and Nils Rinaldi for providing the point-of-sale data.

References

Aarts, H. and Dijksterhuis, A. (2000). The automatic activation of goal-directed behavioour: The case of travel habits, *Journal of Environmental Psychology* **20**(1): 75–82.
URL: *http://dx.doi.org/10.1006/jevp.1999.0156*

Akay, A. (2012). Finite-sample comparison of alternative methods for estimating dynamic panel data models, *Journal of Applied Econometrics* **27**(7): 1189–1204.
URL: *http://dx.doi.org/10.1002/jae.1254*

Arnold, S. J., Handelman, J. and Tigert, D. J. (1996). Organizational legitimacy and retail store patronage, *Journal of Business Research* **35**(3): 229–239.
URL: *http://dx.doi.org/10.1016/0148-2963(95)00128-X*

Arnold, S. J., Oum, T. H. and Tigert, D. J. (1983). Determinant Attributes in Retail Patronage: Seasonal, Temporal, Regional, and International Comparisons, *Journal of Marketing Research* **20**(2): 149.
URL: *http://dx.doi.org/10.2307/3151681*

Arulampalam, W. and Stewart, M. B. (2009). Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators, *Oxford Bulletin of Economics and Statistics* **71**(5): 659–681.
URL: *http://dx.doi.org/10.1111/j.1468-0084.2009.00554.x*

Axhausen, K. W., Löchl, M., Schlich, R., Buhl, T. and Widmer, P. (2007). Fatigue in long-duration travel diaries, *Transportation* **34**(2): 143–160.
URL: *http://dx.doi.org/10.1007/s11116-006-9106-4*

Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfüser, G. and Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary, *Transportation* **29**(2): 95–124.
URL: *http://dx.doi.org/10.1023/A:1014247822322*

Bamberg, S., Ajzen, I. and Schmidt, P. (2003). Choice of Travel Mode in the Theory of Planned Behavior: The Roles of Past Behavior, Habit, and Reasoned Action, *Basic and Applied Social Psychology* **25**(3): 175–187.
URL: *http://dx.doi.org/10.1207/S15324834BASP2503_01*

Bekhor, S., Cohen, Y. and Solomon, C. (2013). Evaluating long-distance travel patterns in Israel by tracking cellular phone positions, *Journal of Advanced Transportation* **47**(4): 435–446.
URL: *http://dx.doi.org/10.1002/atr.170*

Ben-Akiva, M. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA.

Bhargava, A. and Sargan, J. D. (1983). Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods, *Econometrica* **51**(6): 1635.
URL: *http://dx.doi.org/10.2307/1912110*

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

Switzerland.
URL: *http://infoscience.epfl.ch/record/117133/files/bierlaire.pdf*

Bierlaire, M. and Fetiarison, M. (2009). Estimation of discrete choice models: extending BIOGEME., *Swiss Transport Research Conference (STRC)*, Monte Verità, Ascona, Switzerland.
URL: *http://www.strc.ch/conferences/2009/Bierlaire_3.pdf*

Bigano, A., Hamilton, J. M. and Tol, R. S. J. (2006). The Impact of Climate on Holiday Destination Choice, *Climatic Change* **76**(3-4): 389–406.
URL: *http://dx.doi.org/10.1007/s10584-005-9015-0*

Borgers, A. and Timmermans, H. (1986). A Model of Pedestrian Route Choice and Demand for Retail Facilities within Inner-City Shopping Areas, *Geographical Analysis* **18**(2): 115–128.
URL: *http://dx.doi.org/10.1111/j.1538-4632.1986.tb00086.x*

Buisson, A. (2014). *Individual activity-travel analysis based on smartphone WiFi data*, Master thesis, EPFL.
URL: *http://infoscience.epfl.ch/record/209106*

Burton, S., Tangari, A. H., Howlett, E. and Turri, A. M. (2014). How the Perceived Healthfulness of Restaurant Menu Items Influences Sodium and Calorie Misperceptions: Implications for Nutrition Disclosures in Chain Restaurants, *Journal of Consumer Affairs* **48**(1): 62–95.
URL: *http://dx.doi.org/10.1111/joca.12015*

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J. and Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example, *Transportation Research Part C* **26**(0): 301–313.
URL: *http://dx.doi.org/10.1016/j.trc.2012.09.009*

Cambridge Systematics Europe (1984). Estimation and Application of Disaggregate Models of Mode and Destination Choice, *Technical report*, Régie Autonome des Transports parisien, Paris.

Carrel, A., Sengupta, R. and Walker, J. L. (2015). The San Francisco Travel Quality Study: Tracking Trials and Tribulations of a Transit Taker, *Technical report*, University

of California, Berkeley, Berkeley, Ca.
URL: *http://www.joanwalker.com/uploads/3/6/9/5/3695513/carrel_ et_ al_ 2015_ sftqs_ .pdf*

Carrion, C., Pereira, F., Ball, R., Zhao, F., Kim, Y., Nawarathne, K., Zheng, N., Zegras, C. and Ben-Akiva, M. (2014). Evaluating FMS: A Preliminary Comparison with a Traditional Travel Survey, *Transporation Research Board 93rd Annual Meeting*, Washington D.C.

Chen, X. and Yang, X. (2014). Does food environment influence food choices? A geographical analysis through "tweets", *Applied Geography* **51**: 82–89.
URL: *http://dx.doi.org/10.1016/j.apgeog.2014.04.003*

Chi, C. G.-Q. and Qu, H. (2008). Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach, *Tourism Management* **29**(4): 624–636.
URL: *http://dx.doi.org/10.1016/j.tourman.2007.06.007*

Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M. and Zegras, P. (2013). Future Mobility Survey, *Transportation Research Record: Journal of the Transportation Research Board* **2354**(07): 59–67.
URL: *http://dx.doi.org/10.3141/2354-07*

Danalet, A. (2015). A Bayesian Approach to Detect Pedestrian Destination-Sequences from WiFi Signatures: Data (Transp. Res. Part C, 2014).
URL: *http://dx.doi.org/10.5281/zenodo.15798*

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

Danalet, A., Farooq, B. and Bierlaire, M. (2014). A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures, *Transportation Research Part C* **44**: 146–170.
URL: *http://dx.doi.org/10.1016/j.trc.2014.03.015*

Danalet, A., Tinguely, L., de Lapparent, M. and Bierlaire, M. (2015). Location choice with longitudinal WiFi data: Research Data.
URL: *http://dx.doi.org/10.5281/zenodo.33973*

Danalet, A., Tinguely, L., de Lapparent, M. and Bierlaire, M. (2016). Location choice with longitudinal WiFi data, *Journal of Choice Modelling* **18**: 1–17.
URL: *http://dx.doi.org/10.1016/j.jocm.2016.04.003*

Deaton, A. (1985). Panel data from time series of cross-sections, *Journal of Econometrics* **30**(1-2): 109–126.
URL: *http://dx.doi.org/10.1016/0304-4076(85)90134-4*

Dellaert, B. G. C., Arentze, T. A., Bierlaire, M., Borgers, A. W. J. and Timmermans, H. J. P. (1998). Investigating Consumers' Tendency to Combine Multiple Shopping Purposes and Destinations, *Journal of Marketing Research* **35**(2): 177.
URL: *http://dx.doi.org/10.2307/3151846*

Eash, R. (1999). Destination and Mode Choice Models for Nonmotorized Travel, *Transportation Research Record: Journal of the Transportation Research Board* **1674**: 1–8.

Eriksson, L., Garvill, J. and Nordlund, A. M. (2008). Interrupting habitual car use: The importance of car habit strength and moral motivation for personal car use reduction, *Transportation Research Part F: Traffic Psychology and Behaviour* **11**(1): 10–23.
URL: *http://dx.doi.org/10.1016/j.trf.2007.05.004*

Ettema, D. (1996). *Activity based Travel Demand Modelling*, PhD thesis, Eindhoven Technical University, Holland.

Etter, V., Kafsi, M. and Kazemi, E. (2012). Been There , Done That: What Your Mobility Traces Reveal about Your Behavior, *Nokia Mobile Data Challenge 2012 Workshop*, June 18-19, Newcastle, UK, pp. 1–6.

Eymann, A. and Ronning, G. (1997). Microeconometric models of tourists' destination choice, *Regional Science and Urban Economics* **27**(6): 735–761.
URL: *http://dx.doi.org/10.1016/S0166-0462(97)00006-9*

Fesenmaier, D. (1988). Integrating Activity Patterns into Destination Choice Models, *Journal of Lesiure Research* **20**(3): 175–191.

Fotheringham, A. S. (1986). Modelling hierarchical destination choice, *Environment and Planning A* **18**(3): 401–418.
URL: *http://dx.doi.org/10.1068/a180401*

Fox, E. J., Montgomery, A. L. and Lodish, L. M. (2004). Consumer Shopping and Spending across Retail Formats, *The Journal of Business* **77**(S2): S25–S60.
URL: *http://dx.doi.org/10.1086/381518*

Fox, J., Daly, A., Hess, S. and Miller, E. (2014). Temporal transferability of models of mode-destination choice for the Greater Toronto and Hamilton Area, *Journal of Transport and Land Use* **7**(2): 41.
URL: *http://dx.doi.org/10.5198/jtlu.v7i2.701*

Frejinger, E. and Bierlaire, M. (2007). Capturing correlation with subnetworks in route choice models, *Transportation Research Part B* **41**(3): 363–378.
URL: *http://dx.doi.org/10.1016/j.trb.2006.06.003*

Frignani, M. Z., Auld, J., Mohammadian, A. K., Williams, C. and Nelson, P. (2010). Urban Travel Route and Activity Choice Survey (UTRACS): An Internet-Based Prompted Recall Activity Travel Survey using GPS Data, *Transportation Research Record: Journal of the Transportation Research Board* **2183**: 19–28.

Furuichi, M. and Koppelman, F. S. (1994). An analysis of air travelers' departure airport and destination choice behavior, *Transportation research Part A* **28**(3): 187–195.

Gardner, B. (2009). Modelling motivation and habit in stable travel mode contexts, *Transportation Research Part F: Traffic Psychology and Behaviour* **12**(1): 68–76.
URL: *http://dx.doi.org/10.1016/j.trf.2008.08.001*

Gärling, T. and Axhausen, K. W. (2003). Introduction: Habitual travel choice, *Transportation* **30**(1): 1–11.
URL: *http://dx.doi.org/10.1023/A:1021230223001*

Golob, T. F., Kitamura, R. and Long, L. (eds) (1997). *Panels for Transportation Planning*, Transportation Research, Economics and Policy, Springer US, Boston, MA.
URL: *http://dx.doi.org/10.1007/978-1-4757-2642-8*

Gössling, S., Scott, D., Hall, C. M., Ceron, J.-P. and Dubois, G. (2012). Consumer behaviour and demand response of tourists to climate change, *Annals of Tourism Research* **39**(1): 36–58.
URL: *http://dx.doi.org/10.1016/j.annals.2011.11.002*

Grigolon, A. B., Borgers, A. W., Kemperman, A. D. and Timmermans, H. J. (2014). Vacation length choice: A dynamic mixed multinomial logit model, *Tourism Management* **41**: 158–167.
URL: *http://dx.doi.org/10.1016/j.tourman.2013.09.002*

He, F., Wu, D., Yin, Y. and Guan, Y. (2013). Optimal deployment of public charging stations for plug-in hybrid electric vehicles, *Transportation Research Part B: Methodological* **47**: 87–101.
URL: *http://dx.doi.org/10.1016/j.trb.2012.09.007*

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system, *Econometrica* **46**(4): 931–959.
URL: *http://dx.doi.org/10.2307/1909757*

Heckman, J. (1981). The incidental parameters problem and the problem of initial condition in estimating a discrete time-discrete data stochastic process, *in* C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA, pp. 179–185.

Hoogendoorn, S. P. and Bovy, P. H. L. (2004). Pedestrian route-choice and activity scheduling theory and models, *Transportation Research Part B* **38**(2): 169–190.
URL: *http://dx.doi.org/10.1016/S0191-2615(03)00007-9*

Hsaio, C. (2003). *Analysis of Panel Data*, Cambridge University Press, Cambridge, UK.

Jong, G. D., Fox, J., Daly, A., Pieters, M. and Smit, R. (2004). Comparison of car ownership models, *Transport Reviews* **24**(4): 379–408.
URL: *http://dx.doi.org/10.1080/0144164032000138733*

Kalakou, S., Bierlaire, M. and Moura, F. (2014). Effects of terminal planning on passenger choices, *14th Swiss Transport Research Conference (STRC)*, Monte Verità, Ascona, Switzerland.
URL: *http://www.strc.ch/conferences/2014/Kalakou_Moura.pdf*

Kazagli, E., Chen, J. and Bierlaire, M. (2014). Individual Mobility Analysis Using Smartphone Data, *in* S. Rasouli and H. Timmermans (eds), *Mobile Technologies for Activity-Travel Data Collection and Analysis*, IGI Global, chapter 12, pp. 187–208.
URL: *http://dx.doi.org/10.4018/978-1-4666-6170-7.ch012*

Kitamura, R. (1990). Panel analysis in transportation planning: An overview, *Transportation Research Part A: General* **24**(6): 401–415.
URL: *http://dx.doi.org/10.1016/0191-2607(90)90032-2*

Louviere, J., Train, K., Ben-Akiva, M., Bhat, C., Brownstone, D., Cameron, T. A., Carson, R. T., Deshazo, J. R., Fiebig, D., Greene, W., Hensher, D. and Waldman, D. (2005). Recent Progress on Endogeneity in Choice Modeling, *Marketing Letters* **16**(3-4): 255–265.
URL: *http://dxd.oi.org/10.1007/s11002-005-5890-4*

McDonald, N. (2015). Assessing the Travel of the Millennial Generation Using Pseudo-Panels, *14th International Conference on Travel Behaviour Research (IATBR)*, Windsor.

McFadden, D. (2001). Economic Choices, *The American Economic Review* **91**(3): 351–378.
URL: *http://www.jstor.org/stable/2677869*

Miller, H. J. (2014). Activity-Based Analysis, *in* M. M. Fischer and P. Nijkamp (eds), *Handbook of Regional Science*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 705–724.
URL: *http://dx.doi.org/10.1007/978-3-642-23430-9_106*

O'Connor, A., Zerger, A. and Itami, B. (2005). Geo-temporal tracking and analysis of tourist movement, *Mathematics and Computers in Simulation* **69**(1-2): 135–150.
URL: *http://dx.doi.org/10.1016/j.matcom.2005.02.036*

Oppermann, M. (2000). Tourism Destination Loyalty, *Journal of Travel Research* **39**(1): 78–84.
URL: *http://dx.doi.org/10.1177/004728750003900110*

Ortúzar, J. D. D., Armoogum, J., Madre, J. and Potier, F. (2011). Continuous Mobility Surveys: The State of Practice, *Transport Reviews* **31**(3): 293–312.
URL: *http://dx.doi.org/10.1080/01441647.2010.510224*

Rabe-Hesketh, S. and Skrondal, A. (2013). Avoiding biased versions of Wooldridge's simple solution to the initial conditions problem, *Economics Letters* **120**(2): 346–349.
URL: *http://dx.doi.org/10.1016/j.econlet.2013.05.009*

Rindfüser, G., Mühlhans, H., Doherty, S. T. and Beackmann, K. J. (2003). Tracing the planning and execution of activities and their attributes: Design and application of a hand-held scheduling process survey, *10th International Conference on Travel Behaviour Research*, August 10-14, Lucerne, Switzerland, pp. 1–31.

Robin, T. and Bierlaire, M. (2012). Modeling investor behavior, *Journal of Choice Modelling* **5**(2): 98–130.
URL: *http://dx.doi.org/10.1016/S1755-5345(13)70054-X*

Scarpa, R. and Thiene, M. (2005). Destination Choice Models for Rock Climbing in the North eastern Alps: A Latent-Class Approach Based on Intensity of Preferences, *Land Economics* **81**(3): 426–444.

Schlich, R. (2004). *Verhaltenshomogene Gruppen in Längsschnitterhebungen*, PhD thesis, ETH Zurich.

Schwanen, T., Banister, D. and Anable, J. (2012). Rethinking habits and their role in behaviour change: the case of low-carbon mobility, *Journal of Transport Geography* **24**: 522–532.
URL: *http://dx.doi.org/10.1016/j.jtrangeo.2012.06.003*

Scott, D. M. and He, S. Y. (2012). Modeling constrained destination choice for shopping: a GIS-based, time-geographic approach, *Journal of Transport Geography* **23**: 60–71.
URL: *http://dx.doi.org/10.1016/j.jtrangeo.2012.03.021*

Seddighi, H. and Theocharous, A. (2002). A model of tourism destination choice: a theoretical and empirical analysis, *Tourism Management* **23**(5): 475–487.
URL: *http://dx.doi.org/10.1016/S0261-5177(02)00012-2*

Shobeirinejad, M., Veitch, T., Smart, J. C. R., Sipe, N. and Burke, M. (2013). Destination choice decisions of retail travellers: results from discrete choice modelling in Brisbane, *Australasian Transport Research Forum (ATRF)*, Brisbane, Queensland, Australia.

Sivakumar, A. and Bhat, C. (2007). Comprehensive, Unified Framework for Analyzing Spatial Location Choice, *Transportation Research Record: Journal of the Transportation Research Board* **2003**: 103–111.
URL: *http://dx.doi.org/10.3141/2003-13*

Thiene, M. and Scarpa, R. (2009). Deriving and Testing Efficient Estimates of WTP Distributions in Destination Choice Models, *Environmental and Resource Economics* **44**(3): 379–395.
URL: *dx.doi.org/10.1007/s10640-009-9291-7*

Thøgersen, J. (2006). Understanding repetitive travel mode choices in a stable context: A panel study approach, *Transportation Research Part A: Policy and Practice* **40**(8): 621–638.
URL: *http://dx.doi.org/10.1016/j.tra.2005.11.004*

Timmermans, H. J. P. (1996). A stated choice model of sequential mode and destination choice behaviour for shopping trips, *Environment and Planning A* **28**(1): 173–184.
URL: *http://dx.doi.org/10.1068/a280173*

Timmermans, H., van der Hagen, X. and Borgers, A. (1992). Transportation systems, retail environments and pedestrian trip chaining behaviour: Modelling issues and applications, *Transportation Research Part B: Methodological* **26**(1): 45–59.
URL: *http://dx.doi.org/10.1016/0191-2615(92)90019-S*

Tinguely, L. (2015). *Exploiting pedestrian WiFi traces for destination choice modeling*, Master thesis, EPFL.
URL: *http://infoscience.epfl.ch/record/209732*

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.

Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, University of California, Berkeley.

Um, S. and Crompton, J. L. (1990). Attitude determinants in tourism destination choice, *Annals of Tourism Research* **17**(3): 432–448.
URL: *http://dx.doi.org/10.1016/0160-7383(90)90008-F*

Verplanken, B., Walker, I., Davis, A. and Jurasek, M. (2008). Context change and travel mode choice: Combining the habit discontinuity and self-activation hypotheses, *Journal of Environmental Psychology* **28**(2): 121–127.
URL: *http://dx.doi.org/10.1016/j.jenvp.2007.10.005*

Weis, C. and Axhausen, K. W. (2009). Induced travel demand: Evidence from a pseudo panel data based structural equations model, *Research in Transportation Economics* **25**(1): 8–18.
URL: *http://dx.doi.org/10.1016/j.retrec.2009.08.007*

Whynes, D. K., Reedand, G. and Newbold, P. (1996). General Practitioners' Choice of Referral Destination: A Probit Analysis, *Managerial and Decision Economics* **17**(6): 587.

Woodside, A. G. and Lysonski, S. (1989). A General Model Of Traveler Destination Choice, *Journal of Travel Research* **27**(4): 8–14.
URL: *http://dx.doi.org/10.1177/004728758902700402*

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, *Journal of Applied Econometrics* **20**(1): 39–54.
URL: *http://dx.doi.org/10.1002/jae.770*

Wu, L. (2012). *A Tourist Behavior Model System With Multi-Faceted Dependencies and Interactions*, PhD thesis, Hiroshima University.

Yang, D. and Timmermans, H. (2015). Analysis of consumer response to fuel price fluctuations applying sample selection model to GPS panel data: Dynamics in individuals' car use, *Transportation Research Part D: Transport and Environment* **38**: 67–79.
URL: *http://dx.doi.org/10.1016/j.trd.2015.04.011*

Yang, Y., Fik, T. and Zhang, J. (2013). Modeling sequential tourist flows: Where is the next destination?, *Annals of Tourism Research* **43**: 297–320.
URL: *http://dx.doi.org/10.1016/j.annals.2013.07.005*

Yoshimura, Y., Sobolevsky, S., Ratti, C., Girardin, F., Carrascal, J. P., Blat, J. and Sinatra, R. (2014). An analysis of visitors' behavior in The Louvre Museum: a study using Bluetooth data, *Environment and Planning B: Planning and Design* **41**(6): 1113–1131.
URL: *http://dx.doi.org/10.1068/b130047p*

Zhu, W. and Timmermans, H. (2011). Modeling pedestrian shopping behavior using principles of bounded rationality: model comparison and validation, *Journal of Geographical Systems* **13**(2): 101–126.
URL: *http://dx.doi.org/10.1007/s10109-010-0122-8*

Zhu, W., Timmermans, H. and Wang, D. (2006). Temporal Variation in Consumer Spatial Behavior in Shopping Streets, *Journal of Urban Planning and Development* **132**(3): 166–171.
URL: *http://dx.doi.org/10.1061/(ASCE)0733-9488(2006)132:3(166)*

A Appendix

A.1 Descriptive statistics of the WiFi traces

As described in Danalet et al. (2014), WiFi traces have been anonymized but the category of people has been collected. Table 6 shows the number of daily observations and the total number of individuals observed per category. Employees are overrepresented in the sample.

Category	Number of observations		Number of individuals	
Employees	1219		145	
Students, among which...	649		66	
Civil engineering, Bachelor, 4th semester		131		12
Computer science, Bachelor, 4th semester		87		6
Computer science, Master, 2nd semester		53		6
Mathematics, Bachelor, 2nd semester		108		13
Life science and technology, Bachelor, 2nd semester		138		11
Physics, Bachelor, 2nd semester		132		18
Total	1868		211	

Table 6: Number of observations and of individuals per categories of individuals.

The number of times each catering location is chosen is described in Table 7. The most visited catering location is L'Esplanade, very central on the campus. Le Parmentier and Le Vinci are very close and share the same kitchen; their counts of being chosen from WiFi traces are biased towards Le Parmentier, with a larger capacity and therefore a larger attractivity (see Danalet et al.; 2014). Number of visits in catering locations in the Rolex Learning Center (RLC), Le Hodler and Le Klee, are most probably underestimated due to the large attractivity of the library (see again Danalet et al.; 2014).

The walked distance to reach a catering location (Fig. 4) is computed used a weighted shortest path (Danalet et al.; 2014). It takes into account the pedestrian network and the different floors on the campus. In 478 cases, distance could not be computed (previous location to the catering destination is not properly connected to the network).

More descriptive statistics about the data used in this case study are available in Tinguely (2015).

	Count of chosen alternatives			
Catering locations	Morning	Lunch	After lunch	*Total*
Cafeteria Cafe Le Klee	1	1	2	4
Self-service BC	46	60	40	146
Other BM	11	13	22	46
Cafeteria ELA	38	38	49	125
Cafeteria INM	3	3	7	13
Cafeteria MX	39	15	30	84
Other PH	38	7	34	79
Cafeteria L'Arcadie	19	11	8	38
Self-service L'Atlantide	73	11	51	135
Restaurant Le Copernic	0	6	0	6
Self-service Le Corbusier	17	56	0	73
Cafeteria Le Giacometti	47	44	85	176
Self-service Le Parmentier	14	68	53	135
Self-service Le Vinci	1	1	0	2
Self-service L'Esplanade	104	102	206	412
Self-service L'Ornithorynque	30	69	0	99
Caravan Pizza	18	24	22	64
Caravan Kebab	13	11	30	54
Cafeteria Satellite	37	11	87	135
Self-service Le Hodler	13	22	0	35
Restaurant Table de Vallotton	0	7	0	7
Total	562	580	728	1868

Table 7: Number of time each catering location is chosen in the dataset. Morning represents visits starting before 11:30, lunch visits starting between 11:30 and 14:00, and after lunch visits starting after 14:00.

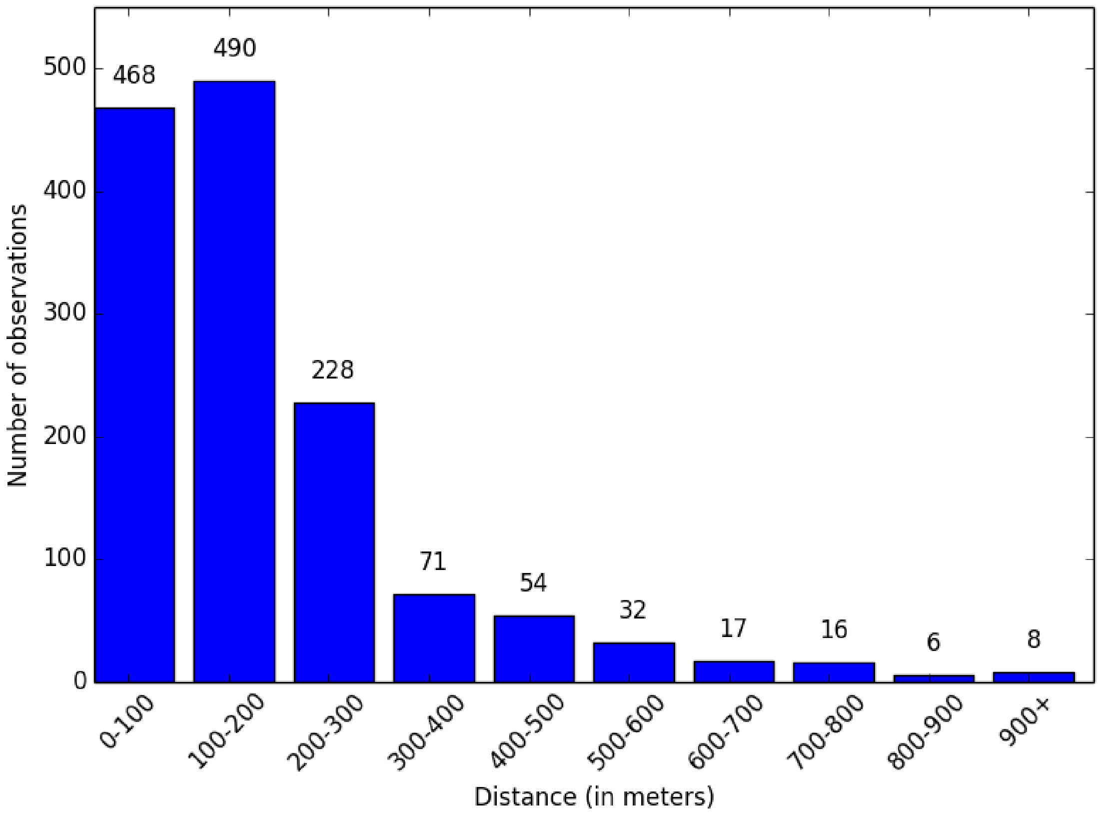

Figure 4: Walked distance to reach a catering location, in meters.

A.2 Elasticity of choice probabilities to price: detailed results

Catering locations		Static model	Dynamic model without agent effect	Dynamic model with agent effect correction	
				First choice	First choice and frequency
L'Arcadie	Employees	−1.239 89	−0.985 452	−3.395 26	−3.571 78
	Students	−2.384 84	−1.838 35	−4.578 27	−4.751 58
L'Atlantide	Employees	−1.134 13	−0.895 157	−3.180 69	−3.326 38
	Students	−2.271 22	−1.773 87	−4.444 07	−4.625
BC	Employees	−0.936 438	−0.739 586	−2.617 85	−2.7033
	Students	−1.573 02	−1.2261	−3.082 35	−3.151 99
Le Copernic	Employees	−2.353 53	−1.871 12	−6.412 68	−6.681 58
	Students	−4.515 51	−3.476 45	−8.643 27	−8.906 21
Le Corbusier	Employees	−0.929 999	−0.735 688	−2.615 05	−2.701 22
	Students	−1.549 62	−1.203 79	−2.977 14	−3.075 26
ELA	Employees	−0.684 047	−0.547 564	−1.874 49	−1.9601
	Students	−1.279 76	−0.972 102	−2.4586	−2.529 25
L'Esplanade	Employees	−0.815 134	−0.656 035	−2.194 22	−2.362 97
	Students	−1.273 79	−0.959 653	−2.356 76	−2.441
Le Giacometti	Employees	−0.693 894	−0.553 458	−1.8864	−1.968 11
	Students	−1.313 32	−1.005 51	−2.492 73	−2.587 09
Le Hodler	Employees	−1.7247	−1.376 71	−4.736 38	−4.969 36
	Students	−3.240 68	−2.477 13	−6.3035	−6.461 66
INM	Employees	−0.763 795	−0.607 341	−2.088 43	−2.179 38
	Students	−1.457 22	−1.1218	−2.796 51	−2.879 87
Kebab	Employees	−0.864 627	−0.689 636	−2.371 33	−2.463 79
	Students	−1.6376	−1.254 15	−3.1567	−3.2466
Le Klee	Employees	−0.794 574	−0.631 613	−2.175 19	−2.274 34
	Students	−1.512 89	−1.165 28	−2.910 07	−3.003 68
MX	Employees	−0.971 626	−0.767 171	−2.709 28	−2.8277
	Students	−1.625 67	−1.267 78	−3.191 63	−3.317 79
Ornithorynque	Employees	−0.844 35	−0.664 995	−2.249 17	−2.374 91
	Students	−1.670 28	−1.306 76	−3.259 05	−3.3904
Le Parmentier	Employees	−0.871 571	−0.695 141	−2.378 36	−2.494 91
	Students	−1.475 15	−1.127 95	−2.736 67	−2.8268
Pizza	Employees	−0.977 648	−0.777 131	−2.662 59	−2.793 63
	Students	−1.8736	−1.442 53	−3.584 97	−3.720 84
Sat	Employees	−0.596 256	−0.474 232	−1.639 87	−1.690 52
	Students	−1.126 53	−0.866 874	−2.1905	−2.225 46
Table de Vallotton	Employees	−3.9529	−3.140 98	−10.5511	−10.9123
	Students	−7.582 75	−5.836 44	−14.3337	−14.7175
Le Vinci	Employees	−1.024 26	−0.814 11	−2.802 12	−2.936 23
	Students	−1.708 61	−1.316 17	−3.210 97	−3.338 83

Table 8: Average sample elasticities of choice probabilities to price