Probabilistic choice set formation incorporating activity spaces into the context of mode and destination choice modelling

Panagiotis Tsoleridis[a,*], Charisma F. Choudhury[a], Stephane Hess[a]

[a] *University of Leeds, University road, Leeds, UK, LS2 9JT*

Abstract

Understanding the constraints that individuals face during their spatial choices is important from a policy perspective. Such constraints, however, are often overlooked in the choice set generation process during model development. In order to address that gap, the current study proposes a probabilistic choice set formation based on Manski's framework assuming that the actual choice set of an individual is latent (unobserved). Though latent class models with heterogeneous choice sets have been used previously in the context of mode and route choice, their application in the context of spatial choices have been hindered due to the inherently large choice sets making the problem computationally intractable. To address this issue, we propose to computationally simplify the problem by utilising the geography-derived notions of Activity Spaces to delineate a range of potential choice sets per individual helping us to capture both issues of spatial awareness and time-space constraints. In order to account for the latent nature of the true choice set, we propose a Latent Class Choice Modelling (LCCM) framework to allocate the individuals probabilistically into the different resulting choice sets, with each class having a different choice set and a different set of parameters. Thus the LCCM is able to capture heterogeneity in the choice sets and in the sensitivities, at the same time. The proposed LCCM framework is empirically tested on joint mode and shopping destination choices captured through a GPS smartphone application. It is compared to a base MNL model estimated on the global choice set, an LCCM capturing heterogeneity only in the sensitivities and a LCCM with latent consideration choice sets, similarly to the proposed model, but with generic parameters across classes. Our proposed specification is able to outperform all of the remaining models, while also providing insights on the factors affecting individuals to be constrained in their location choices across space hinting to cases of spatial cognition, the importance of the home and workplace geography and the individual's socioeconomic status. Such insights can be important for developing more behaviourally realistic models that can be used by planners and policy makers to formulate more effective measures that better relate to the underlying population. Furthermore, the analysis provides insights into the discrepancies that can emerge by accounting for latent consideration sets in willingness-to-pay measures and demand elasticities, which could have significant implications in the effectiveness of policy measures.

Tags: probabilistic choice set formation, latent class choice models, individual constraints, activity spaces, GPS trip diaries

*Corresponding Author

Email addresses: P.Tsoleridis@leeds.ac.uk (Panagiotis Tsoleridis), C.F.Choudhury@leeds.ac.uk (Charisma F. Choudhury), S.Hess@leeds.ac.uk (Stephane Hess)

1. Introduction

Individuals face constraints during their decision making process. In the context of daily mobility choices, those constraints can refer to either time-space constraints and/or limited spatial awareness, among others. Time-space constraints can arise from the need to participate in specific activities in specific locations and for specific durations, i.e. travel to workplace and stay there for a duration of 8 hours, which limits the ability of the individuals to equally evaluate every possible location for shopping and other discretionary activities before, during or after working hours. On the other hand, spatial awareness or the lack of it could be caused as a result of the regularity of daily mobility patterns, which reinforces travelling within the already familiar geographical space.

Despite being often overlooked, it is important from a policy perspective to uncover those latent constraints from the data for the purpose of proposing more effective policy measures suitable for addressing the needs and intricacies of the underlying population. Furthermore, there is value in understanding the influence of various sociodemographic characteristics in the formation of those constraints or phrase it in a different way, identifying what could be the most likely characteristics of the individuals facing each specific constraint. An example of the derived value from such an analysis, could be the identification of instances of social exclusion (Schönfelder and Axhausen, 2003) or the potential presence of very restrictive space-time constraints that prevent individuals from exploring a wider range of alternatives for their daily choices.

Discrete choice modelling (DCM) has been an important tool for policy making since the seminal work of McFadden (1973) and the development of the Multinomial Logit (MNL) model for understanding individual mobility behaviour. The MNL model has a strong grounding on microeconomic theory postulating that individuals will choose the alternative that maximises their utility for a specific choice task among a set of alternatives, also known as the choice set. Traditionally, a DCM is estimated under the assumption that individuals will consider and equally evaluate all alternatives in the choice set. There have been attempts, however, aiming to relax that assumption acknowledging the fact that not only all alternatives might not be available to all individuals, such as making a car trip for individuals with no car access in their household, but they might also not be considered by everyone, as well, such as making a large distance trip by walking or choosing to go for shopping in a store that is unknown to the individual. Assigning an unrealistic choice set for model estimation is yet another case of model misspecification as noted in Williams and Ortuzar (1982), who emphasised the potential adverse effects arising from that, such as biased estimates and incorrect choice probabilities. Gaudry and Dagenais (1979) demonstrated that accounting for *captive* decision makers (individuals who choose only a specific alternative) will have a significant impact on the estimated market shares. Li et al. (2015) have also showed the potential biased welfare measures that could be caused as a result from a choice set misspecification on simulated data.

Several studies mainly originating from consumer behaviour research have suggested that decision makers utilise heuristic processes to segment the alternatives into a range of cognitive subsets, unobserved to the analyst. According to Punj and Srinivasan (1989), the global choice set, i.e. all alternatives of the case study, such as transport modes for a mode choice model or activity locations for a destination choice model, can be decomposed into two different subsets, an *awareness* and a *consideration* set (Howard and Sheth, 1969; Wright and Barbour, 1977; Pagliara and Timmermans, 2009; Capurso et al., 2019). Awareness set is a subset of the global choice set, which includes the alternatives the individuals are aware of due to various reasons, such as past experience, familiarity, word of mouth etc. The consideration set is a subset of the awareness set and it is the final set of feasible alternatives that the individuals actually evaluate during their decision making process. In a spatial context, the respective terms of spatial information fields and spatial usage fields have been proposed by Potter (1979). As such, spatial information fields contain the activity locations the individual is aware of (awareness set), regardless if she has ever travelled to those places, while spatial usage fields are a subset of the former containing the locations that have been visited and are actively considered by the individual for conducting her activities (consideration set) (Timmermans et al., 1982). Therefore, there have been studies in the literature suggesting that choice models should be estimated using the consideration set instead of the common practice of simply using the global choice set.

Several approaches have been proposed over the years for defining alternative availability/consideration. A usual approach is the availability to be exogenously defined using deterministic thresholds based on the analyst's assumptions and the observations in the data, e.g. walking is not considered for trips of distances above the maximum observed walking distance in the data (Calastri et al., 2019; Hasnine et al., 2018;

Tsoleridis et al., 2022). A tour-based approach has also been proposed to account for feasible constraints in terms of mode availability, such as the need for a driver to return her car back home at the end of the tour (Tsoleridis et al., 2022). Contrary to that, a behaviourally richer framework was proposed by Manski (1977) suggesting that the probability of individual n choosing alternative i from a consideration choice set C can be decomposed into a probability of choice set C being the actual choice set considered and a probability of choosing alternative i conditional on belonging in the consideration set C from a set of S possible non-empty choice sets, as shown in *Equation 1*.

$$P_{in}(C) = \sum_{s=1}^{S} P_n(i|C_s)P(C_s) \tag{1}$$

Manski's framework requires a complete enumeration of all possible combinations of alternatives forming potential S non-empty choice sets. That number of combinations increases rapidly with the number of alternatives J, as $2^J - 1$, making this approach computationally infeasible for choice contexts with a large number of alternatives, such as in a spatial choice model. In fact, several implementations of Manski's probabilistic choice set formation have been performed, but the vast majority of them is limited in a mode choice context (Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995; Calastri et al., 2019; Capurso et al., 2019), which generally offers tractable and well-defined choice sets. To simplify the problem, deterministic availability of alternatives can be initially defined in terms of their feasibility. Examples include, defining car as unavailable for individuals with no car in their household in the context of mode choice or excluding routes involving detours above a certain threshold from the habitual route in the context of route choice. Probabilistic availability can thus be incorporated only to previously defined feasible alternatives to account for the uncertainty of the analyst, such as the maximum distance for walking to be considered as an alternative (Swait and Ben-Akiva, 1986) or the spatio-temporal constraints associated with the route choice (Kaplan and Prato, 2012).

One of the first operational implementations of Manski's framework was the logit captivity model of Swait and Ben-Akiva (1986), who proposed a probabilistic choice set generation framework to account for captive decision makers on mode choice for commuting trips. That specification presents a simplified version of a probabilistic choice set generation model in which the number of non empty choice sets is restricted to choice sets of only one alternative (captive individuals) and choice sets including all feasible alternatives (individuals free to choose). That study also laid the foundations for the more general Independent Availability Logit (IAL) model proposed in Swait and Ben-Akiva (1987) assuming that the inclusion of one alternative in the choice set is independent of the remaining alternatives, a necessary assumption to make the specification computationally tractable.

Several attempts were made over the years to approximate Manski's framework, while also relaxing its computational complexity, such as the Implicit Availability/Perception model of Cascetta and Papola (2001) and the Constrained Multinomial Logit (CMNL) model Martinez et al. (2009) incorporating additional terms in the utility function to capture latent constraints. Specifically, the CMNL model includes penalties in the utility function for alternatives exceeding certain attribute thresholds. Despite the authors suggesting that their approach is an approximation of Manski's model, Bierlaire et al. (2010) highlighted some limitations of the CMNL as its inability to produce unbiased estimates on simulated data, contrary to the IAL specification of Swait and Ben-Akiva (1987). Similar conclusions were derived from the study of Li et al. (2015), as well, regarding the comparison of CMNL and IAL models. In the same study, the authors also showed that models capable of uncovering taste heterogeneity, such as MNL models with sociodemographic interactions, mixed Logit or Latent Class Choice Models, will also produce biased estimates and welfare measures, since the choice set formation process is confounded in the sensitivity heterogeneity that is being captured. More specifically, the authors demonstrated -using simulated data- that the IAL specification is the only one capable of producing unbiased estimates in the presence of a latent price threshold.

Consideration of alternatives was traditionally understood by asking additional information to the decision makers (Ben-Akiva and Boccara, 1995; Kaplan and Prato, 2012). A prominent example in the literature is the study of Ben-Akiva and Boccara (1995), in which answers to relevant questions were used as indicators for defining latent variables regarding the consideration of alternatives. Nonetheless, with passively or even semi-passively collected Revealed Preference (RP) data, such as a GPS-based trip diary, only the observed

choice is known without having any additional information on the non-chosen alternatives and the reasons for not choosing them. Hence only assumptions can be made about the extend of the considered choice set and the non-chosen alternatives belonging to that. A relevant example can be found in the study of Calastri et al. (2019), where the authors utilised a GPS dataset to specify a Latent Class Choice Model for mode choice with a range of classes adhering to specific combinations of mode alternatives

Contrary to the examples presented so far, a spatial choice model and more specifically a destination choice model of discretionary activities, e.g. shopping, is characterised by large choice sets, which could be comprised of traffic analysis zones or geographic zones (e.g. Middle Super Output Areas - MSOAs), general shopping areas, shopping malls or even specific parcels and stores depending on the level of spatial granularity offered by the utilised dataset and the level of detail required by the analyst. Sampling of alternatives has been proposed as a method suitable for reducing the computational complexity for choice models with large choice sets (Guevara and M. Ben-Akiva, 2013a; Guevara and M. Ben-Akiva, 2013b; Tsoleridis et al., 2021). Besides the computational advantages, however, sampling of alternatives does not account for the latent consideration of alternatives during the decision making process. Therefore, in order to account for the potential presence of latent choice set formation mechanisms in a spatial context, certain simplifications might be necessary to make the problem computationally tractable. To the best of our knowledge, only the study of Thill and Horowitz (1997b) has attempted to develop a probabilistic choice set formation specification for a spatial choice model first using simulated data and then applied on the real-world context of shopping destinations for home-based trips (Thill and Horowitz, 1997a). Their approach relies on the notion that individuals have to make destination choices, while being subject to latent time constraints, therefore the set of destinations under consideration will depend on each individual's time budget. The simplification applied in that study, relative to the IAL specification of Swait and Ben-Akiva (1987), is that individuals are probabilistically allocated into a finite number of exogenously defined time thresholds as concentric circles from their home locations within which reachable destinations form the respective consideration sets. Their specification which is effectively an LCCM specification –although not stated as such– is able to outperform a base unconstrained MNL model. Nonetheless, it is also worth mentioning that the specification of Thill and Horowitz (1997a) does not account for differences in sensitivities across classes, thus having the possibility of confounding the presence of latent choice set formation constraints with individual unobserved heterogeneity. Furthermore, no additional covariates were used in the allocation of individuals to the specified time thresholds limiting our ability to link latent constraints with specific sociodemographic attributes that could lead to better informed policy measures.

In the current study, we follow a similar approach to Thill and Horowitz (1997a), but we differ our utilised proxy measures of latent constraints by considering the coexistence of spatial awareness/cognition and space-time constraints. The geography-derived notions of Activity Spaces (Hagerstrand, 1970) in the form of detour ellipses (Justen et al., 2013; Leite Mariante et al., 2018; Tsoleridis et al., 2021) and standard deviational ellipses (Brown and Moore, 1970; Horton and Reynolds, 1970; Horton and Reynolds, 1971; Yuill, 1971; Schönfelder and Axhausen, 2003; Schönfelder and Axhausen, 2004; Schönfelder and Axhausen, 2010; Manley, 2016) are utilised to define proxy measures of trip-specific space-time and individual-specific spatial awareness/cognition constraints, respectively. Our proposed specification, is compared against a base MNL model estimated using a choice set of feasible alternatives based on logical checks and deterministic exogenous thresholds and a base LCCM specification estimated again using the same choice set and capturing heterogeneity on the sensitivities across classes. Contrary to that base LCCM, our approach is capturing heterogeneity both in the sensitivities and the consideration choice sets at the same time. More specifically, our proposed specification includes three classes each with its own choice set. The choice set of *class a* includes alternatives within the estimated detour ellipses adhering to trip-specific space-time constraints. The choice set of *class b* includes alternatives, within both the estimated detour ellipses as before, but also within the individual-specific standard deviational ellipses thus capturing additional spatial cognitive constraints of the individuals. Finally, *class c* has the same choice set as the base MNL and the base LCCM models including all feasible alternatives, thus representing individuals that do not face any of the two aforementioned latent constraints. In addition, each of the three classes has a range of class-specific parameters to avoid confounding unobserved heterogeneity among individuals with the presence of latent choice set formation mechanisms, while also sociodemographic attributes are used to assist the allocation of individuals to each class, thus addressing some of the limitations identified in the work of Thill and Horowitz (1997a). An equivalent LCCM specification with different choice sets, but with generic parameters across classes and only

constants in the class allocation, is also utilised for comparison purposes and to highlight the discrepancies in the two approaches.

According to the description above, the proposed specification is a similar case to the logit captivity model of Swait and Ben-Akiva (1986), where individuals are either captive to their space-time constraints, their spatial awareness or are free to choose form all the range of available and feasible alternatives being more inclined to explore the space around them. The proposed approach is empirically tested on a joint mode-destination choice model of shopping activities utilising a 2-week GPS trip diary. The model's objective is to jointly capture individual behaviour regarding the choice of an intermediate shopping location, as well as the modes of the two trip legs travelling to and from that location. The results suggest that our proposed LCCM framework is capable of uncovering significant latent constraints, which also leads to significant model fit improvements compared to the base LCCM specification. More importantly, however, a specification like that is able to shed light into the types of individuals that are more likely to face certain latent constraints during their daily decision making process, thus deriving valuable insights for more effective policy measures able to account for those constraints.

The remainder of the paper is as follows. In the second chapter, a description of the different forms of activity spaces is performed. In the third chapter, the methodological frameworks of the proposed model specifications are thoroughly explained, while in the following chapter the data used in the practical application is described. In the fifth chapter, the modelling outputs and their interpretation are highlighted. Finally, in the last chapter the conclusions and limitations of the study are summarised and recommendations for future research are suggested.

2. Forms of Activity Spaces

The notions of Activity Spaces (AS) originate from the fields of time-space (Hagerstrand, 1970) and behavioural geography (Brown and Moore, 1970; Horton and Reynolds, 1970; Yuill, 1971) and they have been studied extensively in different heterogeneous research areas since their inception with an emphasis on understanding when and where activity participation occurs and identify potential reachable opportunities given the remaining time budget (Schönfelder and Axhausen, 2004; Schönfelder, 2006; Schönfelder and Axhausen, 2010; Kamruzzaman and Hine, 2012; Lam et al., 2018). They are mainly used as a measure of describing the spatial distribution of visited locations and they incorporate a notion of individual spatial awareness (Manley, 2016) by providing invaluable information about the exposure to specific locations and activities that individuals might perform based on their usual mobility patterns and their time-space constraints. Due to the vast range of studies and application domains, there are several different forms of AS proposed in the literature depending on the aspect under examination in each case and the level of analysis. In a systematic review, Smith et al. (2019) summarised the different AS forms, which among others (such as convex hulls, daily path areas, kernel densities, interpolation etc.), include the following:

- Ellipses formed around two fixed points of a specific trip chain, labelled here as *Detour Ellipses (DEs)*
- Ellipses formed around the observed trips of an individual during a survey period, most commonly known as *Standard Deviational Ellipses (SDEs)*

In order to produce generalised results, representative of the sample in our dataset, we refrained from using the observed Activity Spaces of both forms, Detour Ellipses and Standard Deviational Ellipses, and used estimated Activity Spaces instead as proxy measures of latent constraints entering our behavioural specifications. To accomplish that, we estimate a range of continuous regression models using specific structural components of the Activity Space as dependent variables.

2.1. Detour Ellipse

The first type of ellipse utilised in this study aims to capture the reachable destinations based on trip and individual characteristics and the location of the following activity. Several approaches have been proposed in the literature to define what is commonly known as Potential Path Areas (PPAs) based on Hagerstrand (1970) work on time-space geography. Those approaches typically take into account the available net time between the fixed activity locations and an average travel speed (Kamruzzaman and Hine, 2012) or real network travel

times based on the time of day in more advanced cases (Miller, 1991) to identify the intermediate locations that are potentially reachable within those time budgets (Lam et al., 2018). A slightly different approach is taken in the current study, where we propose the use of Detour Ellipses between fixed locations to capture the intermediate reachable areas. A Detour Ellipse (DE) is a type of an ellipse formed around two fixed locations, the foci of the ellipse, which are also referred to as *'pegs'*. More specifically, a DE is based on the Detour Factor (DF) defined as the ratio of the sum of the distances between O(previous origin)-S(shopping destination) and S(shopping destination)-D(next destination) over the distance between O-D, as defined in *Equation 2* (Justen et al., 2013). In other words, a DF measures the deviation that an individual is willing to make to reach an intermediate shopping location S between the O-D (Leite Mariante et al., 2018) and it serves as a measure of spatial dependence among destinations in a trip/activity chain. It is also clear that $DF \geq 1$ should always hold in cases where O and D are different.

$$DF = \frac{l_{OS} + l_{SD}}{l_{OD}} \tag{2}$$

Previous studies have used fixed DFs for intermediate locations to be considered along observed O-D paths (Cascetta and Papola, 2009). Such an example is presented in the study of Newsome et al. (1998), who defined DEs based on the furthest visited intermediate location between fixed home and work locations. That approach, however, fails to take into account the influence of the total OD distance on the resulting DF, as it can be easily understood that longer OD paths will lead to smaller DFs due to the presence of time constraints for reaching both the intermediate and the following location D. That means that the longer the OD path the smaller the time budget available to the individuals to deviate further away from that path. That relation between DF and OD distance has been taken into consideration in Justen et al. (2013), although their approach is limited by the fact that only average values per DF percentile are considered, while trip-specific and sociodemographic attributes that could have an impact on the formation of DEs have not been taken into account.

In the current study, we make a distinction between trip chains with an intermediate shopping location S between an initial origin O and a following destination D, referred to as OSD, and simple tours in which individuals travel for shopping to a location S and then return back to their origin O, referred to as OSO trip chains in the remainder of the paper. The detour factors of OSD trip chains are always greater than 1.0 and they are exactly equal to 1.0 only in the extreme case of choosing a shopping location S directly on top of the OD path. Therefore, the estimated detour factors from that model will need to adhere to that restriction. In order to accomplish that, the dependent variable, in that case the observed detour factors, are transformed accordingly as $y_i - 1$ and are assumed to follow a log-normal distribution to ensure the estimation of strictly positive values, as shown in *Equation 3*.

$$log(y_i - 1) = \Sigma b_{x_i} x_i + \sigma \tag{3}$$

where x_i is a vector of mode-specific, trip-specific (including the straight OD distance), locational and sociodemographic explanatory variables and b_{x_i} are the respective parameters to be estimated. The disturbance term for the log-transformed DF is assumed to follow a normal distribution with $N(0, \sigma)$, where σ is the standard deviation that is estimated alongside the rest of the parameters

For OSO trip chains, a different modelling approach was developed by using the straight distance to the shopping location S as the dependent variable and use it to define the estimated radius of a circle from the respective O. In that case as well, in order to ensure only positive estimated values, the dependent variable is assumed to follow a log-normal distribution. As a result, the modelling formulation of *Equation 4* is proposed, where y_i is the observed straight OS distance and as previously x_i is a vector of mode-specific, trip-specific, locational and sociodemographic explanatory variables and b_{x_i} is a vector of parameters to be estimated.

$$log(y_i) = \Sigma b_{x_i} x_i \tag{4}$$

2.2. Standard Deviational Ellipse

The second type of Activity Space aims to capture the spatial awareness of the participants. The Standard Deviational Ellipse (SDE) is proposed for that purpose, which originates from the fields of behavioural and

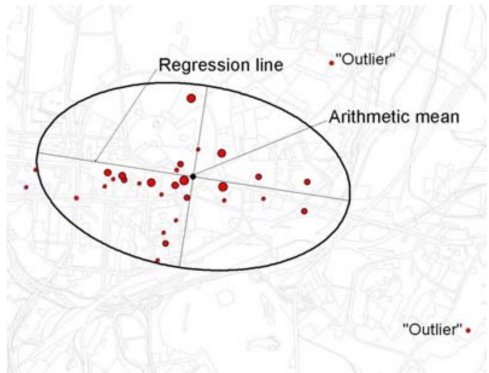

Figure 1: Weighted standard deviational ellipse around observed/visited destinations (Schönfelder, 2003)

social geography (Brown and Moore, 1970; Horton and Reynolds, 1970; Yuill, 1971). The SDE captures the spatial dispersion of the visited locations (observed latitude/longitude coordinates) of an individual and has been proposed as a measure of capturing the exposure to opportunities as a consequence of daily activities (Horton and Reynolds, 1971). In that sense, a SDE provides additional information on the individual awareness of certain destinations, that the Detour Ellipse and other forms of Potential Path Areas are not able to provide. Activity spaces formed by SDEs are considered a subset of a larger latent *awareness space* (Brown and Moore, 1970; Patterson and Farber, 2015) or *spatial information field* (Potter, 1979) suggesting that individuals would likely possess spatial knowledge that far exceeds the SDE formed around the observed destinations due to word of mouth and the usually limited durations of surveys.

An SDE is generally considered the two-dimensional equivalence of a standard 95% confidence interval. The mathematical process of defining an SDE, as described in Yuill (1971), involves the calculation of the covariance matrix of the latitude/longitude coordinates and the calculation of the rotation matrix leading to the final definition of the ellipse's perimeter (Tsoleridis et al., 2021). In *(Figure 1)* the main components of an SDE are depicted, such as the ellipse's major axis indicating the axis of major dispersion, which can also be considered as the regression line of latitude/longitude coordinates, the ellipse's orientation capturing the slope of the regression and the sign of the correlation sign among the coordinates and finally the arithmetic mean or else the centroid of the ellipse capturing the centre of gravity of the individual's usual movements (Schönfelder, 2003). Destinations that are outside of an SDE are considered *outliers* relative to the usual movement areas of an individual.

Several measures can be extracted from a SDE that describe the mobility patterns of an individual, such as its shape (minor to major axis ratio), size (area, number of polygons located within etc.), orientation and eccentricity (Yuill, 1971). An ellipse is a generalised form of a circle with one axis (major axis) more elongated than the other (minor axis). In the case of a very small minor axis close to 0, the ellipse resembles a straight line, while in the case of equal axes the ellipse takes the form of a circle. The ratio of minor/major axis, $\frac{b}{a}$, can provide an initial indication of time constraints or spatial exploration. For instance, an ellipse with a small $\frac{b}{a}$ can characterise a person with significant time constraints that does not have the freedom or the willingness to roam around, and the opposite can be said for an ellipse resembling with $\frac{b}{a}$ close to 1.0 resembling a circle.

Temporal factors and individual sociodemographic characteristics can also be taken into account, such as examining weekday/weekend differences on the stability of SDEs among full-time and part-time workers (Srivastava and Schoenfelder, 2003; Smith et al., 2019). Time also plays an important role in the evolution of the Activity Spaces providing opportunities to the individuals for more spatial search and exposure to the surrounding area, thus increasing their spatial cognition. SDEs might stabilise after a certain period of time, but significant longitudinal data will be required in order to capture that. Schönfelder (2006) used longitudinal data from traditional trip diaries to study the impact of time on the formation of activity spaces (SDEs). He concludes that, in general, datasets of longer durations (around 6 weeks) than the common ones (1-2 weeks) are necessary for SDEs to reach stability. That finding, however, could be different for emerging data sources, such as GPS trip diaries, which generally offer a more comprehensive depiction of individual

Table 1: Literature timeline of Activity Spaces and Choice Modelling dealing with choice set formation

Year	Activity Spaces	Choice Modelling
1970	Hagerstrand; Brown and Moore; Horton and Reynolds	
1971	Horton and Reynolds; Yuill	
1979	Potter	Gaudry and Dagenais
1982		Williams and Ortuzar
1986		Swait and Ben-Akiva
1987		Swait and Ben-Akiva
1991	Miller	
1995		Ben-Akiva and Boccara
1997		Thill and Horowitz (a); Thill and Horowitz (b)
1998	Newsome et al.	
2003	Schönfelder; Srivastava and Schönfelder	
2004	Schönfelder and Axhausen	
2006	Schönfelder	
2009	Cascetta and Papola	
2010	Schönfelder and Axhausen	
2013	Justen et al.	
2015	Patterson and Farber	Li et al.
2018	Leite Mariante et al.	
2019	Smith et al.	Calastri et al.
2021	Tsoleridis et al.	

mobility behaviour by capturing a higher number of smaller trips during the day. A general timeline of the evolution of Activity Space literature and the key papers of Choice Modelling mainly dealing with the subject of choice set formation is presented in *Table 1* along with the studies where those two were merged together.

In order to reconstruct estimated SDEs per individual we need to estimate a range of models referring to several important structural components of an SDE (see *Figure 1*). Specifically, we need to estimate the distance $dist_{hc}$ and the angle θ_{hc} between the home location h and the ellipse centroid c, the orientation θ_{sde} and the total area of the ellipse A_{sde}, as well as the shape of the ellipse defined as the ratio between the small and large axis ($\frac{b_{sde}}{a_{sde}}$). Therefore, five continuous models have been developed, in total, to assist with the reconstruction of estimated activity spaces. From the estimated distance $dist_{hc}$ and angle θ_{hc}, the centroid of the estimated confidence ellipse per individual can be defined. Following that, we are using the estimated minor/major axis ratio and the estimated area to define the shape of the ellipse and finally the estimated orientation of the ellipse to complete its mapping over space.

The models for $dist_{hc}$ and A_{sde} are specified as linear regression models with log-normally distributed dependent variables similarly to the *OSO* model previously described in *Equation 4*. To estimate the models referring to the angles, i.e. for the angle between home-centroid θ_{hc} and the ellipse orientation θ_{sde}, the predicted values had to be restricted between $0^o - 360^o$. In order to achieve that, the dependent variables were scaled between 0-1 by dividing them with 360. Then the scaled angles were used as the dependent variables in beta regression models. Beta regression is a type of model suitable for modelling continuous variables with values between 0 and 1 (Ferrari and Cribari-Neto, 2004). A similar approach was followed for the specification of the minor/major axis ratio $\frac{b_{sde}}{a_{sde}}$ model, where the dependent variables should also take values only between 0-1. The mean μ of the dependent variable y can be calculated as $\mu = E(y) = \Sigma b_{x_i} x_i$ and the variance as $Var(y) = \frac{\mu(1-\mu)}{1+\phi}$, where ϕ is a precision parameter with a larger precision parameter leading to smaller variance (Ferrari and Cribari-Neto, 2004).

3. Methodology

The MNL model has been the main workhorse of DCM and has been applied in numerous studies of behavioural modelling in the fields of transport, environment and health, among others (McFadden, 1973; Domencich and McFadden, 1975; McFadden, 1978; Ben-Akiva and Lerman, 1985; McFadden, 2000). According to that specification, an individual n will choose among a set of alternatives J, the alternative i that provides the highest utility for a specific choice task t. The utility U_{int} is a latent construct comprised of a systematic part V_{int} and a disturbance term ϵ_{int} as $U_{int} = V_{int} + \epsilon_{int}$, assuming an additive disturbance term. Different distributional assumptions regarding the disturbance term will yield different specifications, with the assumption of a Type-I Extreme Value distributed ϵ_{int} leading to an MNL model.

In an MNL model, individual heterogeneity can be captured by specifying interaction terms with sociodemographic attributes as shifts from the base level of specific parameters. More advanced specifications are necessary, however, to account for unobserved heterogeneity, such as mixed Logit (McFadden and Train, 2000) and LCCM (Kamakura and Russell, 1989). In the mixed Logit model, heterogeneity is captured in a continuous manner by assuming that sensitivities across individuals follow a certain distribution with normal, uniform or log-normal distributions being the most commonly used ones. The mixed Logit has been widely used for the purpose of capturing heterogeneity in the sample and it is considered the most flexible logit specification (McFadden and Train, 2000; McFadden, 2000). Mixed Logit models usually require the use of simulation to estimate the parameters of the specified distributions and a large number of draws is usually needed to reach a certain level of estimation stability, which significantly increases the estimation times. Furthermore, that means that the estimated parameters largely depend on the analyst's distributional assumptions, which can have adverse effects in the case of behaviourally inaccurate ones. On the other hand, LCCMs capture heterogeneity by assuming that individuals in the sample can be probabilistically allocated into a discrete and finite number of latent classes based on their sociodemographics and their choice behaviour.

Three consideration sets are constructed per individual and choice task after the creation of the mode-specific Detour Ellipses and individual-specific Standard Deviational Ellipses. Individuals can belong to each consideration set with a positive probability, thus acknowledging the latent nature of the true choice set. The probability of belonging to a consideration set C is modelled as a multinomial logit, as shown in *Equation 5*. The hypothesis behind this specification, is that the individuals in the sample are likely to be subject to either time-space constraints for the specific trip chain (class a) or to be constrained due to issues of spatial awareness (class b) or alternatively to be free to consider alternatives from the global choice set (class c) during their choice of mode and shopping location. As such, the proposed LCCM specification allocates individuals probabilistically into three classes, with each class having a different choice set. The choice set of *class a* includes mode-destination alternatives (feasible in terms of mode) that are located within the estimated mode-specific Detour Ellipses, therefore capturing individuals who are *captive* to their time-space constraints. The choice set of *class b* includes alternatives within the merged area of the estimated Detour and Standard Deviational Ellipses, thus capturing individuals *captive* to their spatial awareness. Finally, the choice set of *class c* includes all feasible alternatives in the global choice set per trip.

$$\pi_C = \frac{e^{\delta_c + \gamma_c x_n}}{\sum_{r=1}^{C} e^{\delta_r + \gamma_r x_n}} \tag{5}$$

According to this framework, the probability of choosing an alternative i is calculated based on the logit function if alternative i belongs to consideration set C and it is 0 otherwise, as shown in *Equation 6*. Finally, the unconditional likelihood of observing a sequence of choices for individual n is calculated by weighting the class-specific conditional probabilities for alternative i with the class membership probabilities $\pi_n(C)$ across G potential classes, i.e. non-empty choice sets, which in that case is limited to 3, as shown in *Equation 7*. The coefficients of both model components are jointly estimated by using Maximum Likelihood Estimation.

$$P_{i|C} = \begin{cases} \frac{e^{V_i}}{\sum_j^J e^{V_j}} & \text{, if } i \in C \\ 0 & \text{, otherwise} \end{cases} \tag{6}$$

$$P_{ni} = \sum_{C \in G} \pi_n(C) P_{ni|C} \tag{7}$$

Further complexity can be added to the proposed specification in order to capture additional heterogeneity within each class by specifying a 2-stage LCCM similar to the study of Song et al. (2019). Nonetheless, that is out of the scope of the current and it is left as an idea for potential future research.

The utility function of the behavioural model follows the size variable specification proposed in Daly (1982) and specifically the one implemented in a joint mode and destination choice model presented in the study of Kristoffersson et al. (2018) combining mode preferences with shopping destination attraction. According to that, the systematic utility V_{md} for mode m and destination d presented in *Equation 3* is comprised of several parts referring to mode- and destination-specific Level-of-Service (LOS) variables, locational variables capturing the quality of each destination and variables capturing its size, also known as size variables. That size variable specification was proposed to account for the utility of the elemental destinations within the aggregated destination alternatives.

$$V_{md} = \sum_{r \in R} b_r x_{rmd} + \sum_{q \in Q} b_q y_{qd} + \phi log(S_d) \tag{8}$$

The first component includes mode- and destination-specific variables that best describe the trip to destination d with mode m, such as travel time and cost for motorised modes and distance for active travel, as well as ASCs capturing inherent preferences for specific modes/destinations and sociodemographic interactions. With this, x_{rmd} is the r-th LOS variable for mode m and destination d. The second component captures the impact (positive or negative) that certain characteristics could have on the utility of a specific destination, such as available parking space for car users, where y_{qd} is the q-th quality variable for destination d. The final component in *Equation* aims to capture the attraction or the *"size"* of a destination d, S_d and is specified as a composite logarithmic term as shown in the following *Equation* :

$$S_d = a_{1d} + \sum_{r>1} exp(\gamma_r) a_{rd} \tag{9}$$

where a_{1d} is the attraction attribute used as a base with a γ parameter normalised to 1.0, a_{rd} are the additional attraction attributes of destination d relative to the base attribute, and γ_r are the parameters to be estimated capturing the effect of those attributes on the attraction of the target destination. The γ_r parameters are constrained to be positive by using an exponential transform.

The log-size parameter ϕ is usually kept fixed to 1.0 ensuring that a change in the size of a destination will affect proportionately its utility. Therefore, the choice probabilities will not be affected by the zoning discretisation that usually forms the destination alternatives. Kristoffersson et al. (2018), however, proposed a freely estimated ϕ, which can lead to estimated values different than 1.0, leading to a behavioural interpretation on the formation of destination alternatives. Specifically, in the case of $\phi < 1$, the authors suggest that the model captures significant correlation among the utilities of the elemental alternatives within each aggregate destination alternative.

4. Data

For the practical implementation of the proposed approach, a 2-week GPS-based trip diary captured by a smartphone application is utilised. The GPS trip diary was collected as part of the research project "DECISIONS" conducted by the Institute for Transport Studies, University of Leeds, between October 2016-March 2017. Besides the utilised GPS trip diary, the "DECISIONS" project aimed to capture a range of different aspects of individual behaviour, such as in-home and out-of-home activity participation, energy appliance usage and the effect of social networks. The GPS trip diary captured the daily trips over the survey period with the use of a smartphone application, which was tracking the traces of the participants.

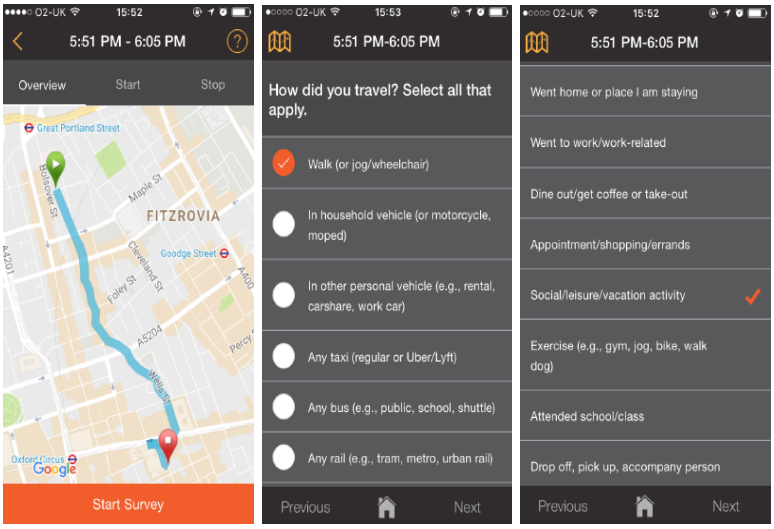

Figure 2: User interface of smartphone application used for the trip diary

After the end of each trip, the participants had the chance to correct the information of the logged trip and provide additional information regarding the chosen mode and the type of the activity at the destination (trip purpose). A depiction of the application's interface is presented in *Figure 2*. A background household survey was also conducted in order to capture several important sociodemographic attributes of the participants, such as their age, their household composition, the availability of mobility tools (e.g. private vehicles and PT season ticket ownership) and their personal and household income, among others. A more detailed description of the data collected during the "DECISIONS" project and its various sub-modules is provided in Calastri et al. (2020).

The initial GPS trip diary included trips captured throughout the UK, but the vast majority of them were in the region of Yorkshire and more specifically in and around the city of Leeds. Therefore, the final dataset used for the subsequent analysis included only individuals residing within the local authority of Leeds. As previously described, the purpose of the analysis is to understand where the individuals are more likely to travel in order to cover their daily shopping needs and how to travel there and to the following activity location, thus acknowledging the interrelations between mode and destination and among the locations of consecutive activities. As a result, only the shopping trips and their following trips were selected for the subsequent analysis resulting in a final dataset of 1,541 shopping-following trip chains performed by 270 unique individuals. The analysis is conducted at the trip chain level, 66% of which are *OSD* trip chains and the remaining 34% are *OSO* trip chains. Shopping trips are comprised of three subcategories, namely grocery (82%), clothes (12.7%), and other types of shopping (5.3%), mainly for durables. The vast majority of following trips were trips going home (61.5%), while there was a small percentage of 9.3% of a consecutive shopping trip to a different shopping destination. The alternative modes of transport included car, public transport (PT) – as a combination of bus and rail – and walking.

The high spatial resolution of the GPS traces, despite the benefits, it also provides additional challenges for their analysis compared to traditional data sources. Therefore, because each GPS trace is a unique pair of latitude and longitude coordinates, a clustering approach had to be developed to identify unique activity locations. Hierarchical Agglomerative Clustering (HAC) was utilised for that purpose as it does not require any a priori assumptions about the number of the required clusters. HAC, however, required a specific distance threshold to be defined to assign points within that threshold in the same cluster. A distance threshold of 200m was chosen to ensure a small distance difference (approximately 100m) of points allocated in the same cluster of an activity location per individual. The implementation of HAC helped us to define unique home and work locations based on the purposes of trips going to those destinations in order to further define daily tours and sub-tours. In cases, where trip purposes were not enough to define home and work locations, additional information on time of day and activity duration was used, such as assigning a work location to a cluster if an individual spends the majority of working hours (09:00-17:00) there. The home

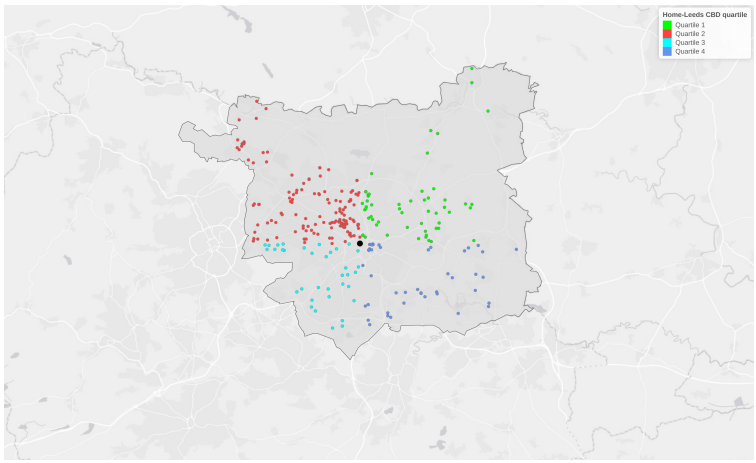

Figure 3: Home locations segmented in quartiles relative to their position from Leeds CBD (black circle)

locations of the individuals were also segmented according to their position relative to Leeds Central Business District (CBD) into four quartiles, as depicted in *Figure 3*. Out of the 270 individuals, 55 (20.4%) live in quartile 1, 136 (50.4%) in quartile 2, 35 (12.9%) in quartile 3 and 35 (16.3%) in quartile 4.

In order to take advantage of the high spatial resolution provided by the GPS data, the definition of shopping destination alternatives was not limited to the usual UK geographical boundaries, but were defined at a more granular level by clustering the observed elemental shopping destinations. HAC was implemented again with a distance threshold of 800 metres among the shopping trip destinations. The cluster centroids defined as the mean of the latitude/longitude coordinates of the points in each cluster were then used to replace the original destination points of each shopping trip belonging to the cluster. The main goal of choosing an appropriate distance threshold was to ensure a small average distance difference between the original destination points of a cluster and its centroid. After trying different distance thresholds between 500m-1,000m, a 800m distance threshold was selected resulting in small average distance differences of around 4-5 minutes of walking (assuming a 5 km/h average walking speed). As a final step, a 400m buffer was defined around each cluster centroid to create the aggregate shopping areas used as destination alternatives in the analysis. This process resulted in the definition of 176 general shopping areas around the region of Yorkshire, capturing 76% of the retail polygons located within the Local Authority of Leeds, as defined in OpenStreetMaps (OSM).

Further steps were necessary in order to enrich the initial dataset with additional information important for behavioural modelling. Initially, the dataset contained only the self-reported travel times/distances for the chosen modes, however, the values of the unchosen mode alternatives were also required to properly define the attributes for all alternatives. For that reason, the Bing maps route API[1] was utilised to obtain the travel times and distances for all the modes (car, bus/rail, walking) and for trips starting from each initial origin to each shopping cluster and from each shopping cluster to each following destination. For consistency reasons, the travel times/distances of the chosen mode alternatives were recalculated as well to ensure that the data used for estimation would come from the same data generating process. The total number of queries passed to the API was 1,627,296 (1,541 trips $\times$ 176 shopping destinations $\times$ 3 modes $\times$ 2 for the current and the subsequent trip). After that data collection stage, deterministic mode availability was assigned based on logical feasibility checks of the results obtained from the API, such as cases of short distance PT trips for which the API returned only walking segments, or in specific cases where car was the chosen mode and the participant had to return it back home. For that latter case, special attention was given to the stated size of the party that participated in the trip in order to understand whether the participant of the survey was the actual driver. As such, if the individual was the only person in a car trip, then she was assigned as the car driver and all the remaining modes would become unavailable only in the case where the following trip was

[1]Details can be found here: https://docs.microsoft.com/en-us/bingmaps/rest-services/routes/

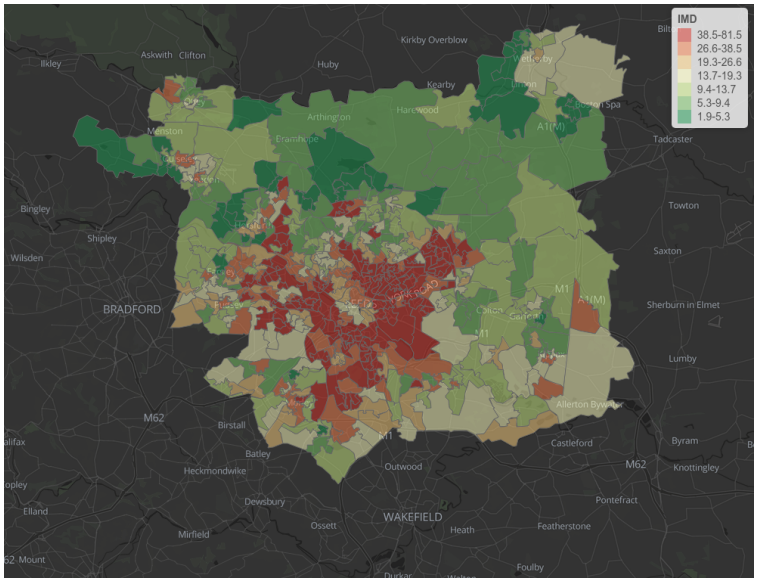

Figure 4: Index of Multiple Deprivation for 2015

to return back home. For other trip purposes for the following trip, it is assumed that the individual is free to consider all the available modes. On the contrary, if there were more than 1 people participating in a car trip, then we could not safely assume that the individual was the driver and all the modes would remain available for the following trip, as well.

Information on travel cost was also missing both for chosen and unchosen modes. Car travel cost was computed using the UK's official Transport Appraisal Guidance (WEBTag) specifications for fuel and operating costs (Department for Transport, 2014). Parking cost was also calculated for trips with destinations in central areas/high streets across the region of Yorkshire based on information on hourly or fixed parking costs provided by the respective Local Authorities. Fuel, operating and parking costs were then aggregated to form a total car travel cost used for estimation. For PT, an average distance-based fare was used for bus and rail and a total PT cost was calculated per trip based on the distance of the leg performed by bus or rail. A discount was also applied for trips made by season ticket holders.

Additional locational data were acquired from the Office for National Statistics (ONS) and OpenStreetMaps (OSM) to be used in the modelling specifications as attraction variables for specific shopping destinations. Specifically, the total areas of retail, grocery and durable shopping parcels within each aggregated shopping destination was calculated from OSM together with the total parking areas and the locations of the most popular retailers per shopping type in the UK market. The population around shopping locations was also extracted from the Office of National Statistics (ONS) together with Indices for Multiple Deprivation for 2015[2] (*Figure 4*), as well as average house prices and percentages of white residents for 2016-2017 in order to capture instances of spatial inequalities.

The different types of shopping stores among the elemental shopping destinations within an aggregate destination alternative was also acquired using the OSM categorisation. The purpose of that was to capture the impact of shopping store variability using Shannon's entropy (Shannon, 1948; Whittaker, 1949), H_d, measuring the percentage of the area covered by a specific store type $t \in T$ inside a shopping destination d from a total number of N different store types as shown in *Equation 10*. Shannon's entropy is used to examine whether an increased variability in store types makes a shopping destination more likely to be chosen, since that would enable the completion of different shopping activities within the same trip.

[2]Details can be found here: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015

$$H_d = -\frac{\sum_{t=1}^{T} (p_t \ln (p_t))}{\ln N} \tag{10}$$

In order to capture agglomeration effects and the impact of neighbouring shopping destinations on the attraction of a target shopping destination, the same information on the aforementioned locational variables was extracted for additional buffers between 400-1,000m, 1,000-2,000m and 2,000-5,000m from each cluster centroid, similar to the study of Kristoffersson et al. (2018).

Finally, the impact of the location of the intermediate shopping destination S, in relation to the straight distance between O and D, was captured by calculating the angles between *OS-OD* and *SD-OD*. It is assumed that shopping destinations, which require a significant deviation from the straight OD path would be less likely to be chosen compared to others that are in the same direction.

5. Results

In the following, the modelling outputs are going to described starting first with the auxiliary specifications for OSD and OSO trip chains and then the five models for the structural components of the SDE –namely models for the distance and the angle between home locations and the SDE centroid, its orientation, its minor/major axis ratio and its area– before moving on to the main outputs of the proposed LCCM specification.

5.1. Detour Ellipse outputs

The estimates of the Detour Factor and the distance model for the OSD and OSO trip chains, respectively, described in the following were used to re-create estimated Detour Ellipses for both cases (OSD and OSO trip chains). The resulting trip- and mode-specific estimated Detour Ellipses, on average, contain 19.5 shopping locations out of the 176 identified locations within the study area (11.1%).

The Detour Factor model for the OSD trip chains achieves a correlation of 0.72 between observed and estimated detour factors. The estimated parameters of the corresponding model are presented in *Table 2* along with the robust t-ratios to account for the multiple observations per individual (Daly and Hess, 2010). According to the model, the individuals are willing to take a shorter detour to reach a shopping destination, when the destination for the next activity is further away indicating the time-space constraints the individuals are subject to. Choosing walking for the first shopping trip or also travelling by mode combinations of car-PT, car-walking will result in smaller detours than the base mode combination of car-car. Nonetheless, the opposite is true when travelling by PT in both trip legs with individuals choosing longer detours. Longer detours are also predicted for clothes and durables shopping, when the following activity is for education purposes or the shopping trip is part of a tour with education as its main purpose, for male individuals and for mostly older ages (above 76 years old). Furthermore, individuals with a household income between £40,000-£50,000 will also tend to take longer detours, while the higher the average income around the individuals' home locations the longer the detours they are willing to choose, as well. Longer detours were also estimated for individuals living in households of mostly above 3 members, those employed in Healthcare, Education, Academia or other types of occupations, when the trip occurs during morning weekend hours and when the are more available parking areas around the shopping location. Contrary to that, smaller detours are estimated for individuals living together with more than four employed individuals possibly due to the increased time requirements to accommodate everyone's daily mobility needs, for part-time workers most likely due to increased housekeeping responsibilities, for shopping trips during night or early morning periods during weekdays and for individuals of the smallest household income band (below £10,000).

Moving on to the model for the estimated distances of the OSO trip chains, a correlation of 0.67 was achieved between observed and predicted distances. The modelling outputs are presented in *Table 3*, where it can be seen that mode combinations of car-walking and walking-walking lead to smaller straight distances to reach a shopping destination compared to the base mode combination of car-car. Shopping for durables will lead to larger distances mainly due to the more specialised type of stores, which can be located further away from home locations. Living in households of either 1 or 4 members will lead to smaller distances, while that is also the case for male individuals, when the shopping trip is part of a tour with work or education as its main purpose, when the following activity is for social, other purposes or to return back to work

Table 2: Modelling outputs of the DF model for *O-S-D* trip chains

Parameters	**Estimates**	**Rob. t-ratios**
Constant	-4.7480	-3.62
O-D straigth distance (km) (log)	-1.3221	-20.57
Car-PT	-0.4283	-1.46
Car-Walking	-2.1897	-6.60
PT-PT	0.9615	2.31
Walking-Car	-1.4913	-4.24
Walking-PT	-0.6923	-2.59
Walking-Walking	-1.7347	-7.34
Shopping: Clothes - Other	0.5791	3.22
Household size: 3-4 members	0.4599	3.05
Household size: 5 members	-0.9965	-1.46
Household size: 6 members	1.1187	2.58
Other employed household members > 4	-1.8039	-8.44
Part time workers	-0.4286	-2.54
Occupation: Healthcare	1.0691	3.77
Occupation: Education	0.2591	1.46
Occupation: Academia	0.4369	1.74
Occupation: Students	0.6211	1.42
Occupation: Other	0.5894	2.81
Time of day: Weekday night	-0.4682	-1.65
Time of day: Weekday morning	-1.1338	-1.43
Time of day: Weekend morning	0.6974	3.21
Following trip purpose: Groceries	0.3666	1.78
Following trip purpose: Education	2.5978	5.21
Age 25-29	0.3771	1.61
Age 60-65	0.4492	1.50
Age ≥ 76	1.2577	2.96
Parking areas 400m around shopping destination in $1,000m^2$ *for car trips*	0.0186	6.21
Shannon's entropy 400m around shopping destination (log)	0.3833	2.24
Total passengers in shopping trip >1	0.2520	1.73
Houshold income <£10,000	-0.6036	-1.53
Houshold income £40,000-£50,000	0.4140	2.06
Male	0.2854	1.86
Population in 400m around home location in 1,000 people	0.4275	1.56
Average annual income in 400m around home location in £1,000 (log)	1.0374	2.82
Trip part of a tour with main purpose: Education	0.6589	1.44
Sigma	1.9616	33.05

with the latter mostly referring to short shopping trips from work locations during lunch break. Longer distances can be expected for trips during the interpeak or during morning hours for retired people and for engineers, community/social workers, students and those working on management. The model was also able to uncover that younger and older individuals, specifically those between 25-29 and above 76 years old, and those who have not disclosed their income tend to choose longer distances. In addition, participating in shopping activities of longer durations (which is considered as exogenous in that case) will make individuals to also travel for longer distances. Considering shopping duration as endogenous within a discrete-continuous framework, however, would be a more accurate approach, leading to better capturing the trade-offs between shopping duration and distance travelled, which could be an interesting direction for future research. Finally, decreasing marginal utilities/disutilities have been uncovered for a range of locational variables specified in a logarithmic function. Specifically, shopping locations with more retail and parking areas will cause individuals to travel for longer distances to visit them. Individuals living in areas with a higher average income tend to choose longer distances. Finally, an increased shopping store variability (Shannon's entropy) will lead to smaller distances.

5.2. Standard Deviational Ellipse outputs

In the following, the estimates of the five SDE-related sub-models are detailed. Those estimates were used to re-create estimated individual-specific SDEs, which on average contain 73.2 (41.6%) of the shopping locations in the study area.

5.2.1. Home-SDE centroid distance outputs

The estimated distances between home and SDE centroids achieved a correlation of 0.48 with the observed distances. The outputs of the corresponding model are presented in *Table 4*. According to those, individuals living in areas other than quartile 3 tend to have an SDE centroid, i.e. the centre of gravity of their usual movements, further away from their home locations. The distance between home and Leeds CBD also has an impact on the home-SDE centroid distance with larger distances leading also to larger home-SDE centroid distances for quartiles 2 and 4. On the contrary, distances between work locations and Leeds CBD will lead to smaller home-SDE centroid distances for quartiles 2 and 3. The CBD of York also was found to have an impact on the estimated distances. Specifically, larger home-York CBD distances will lead to larger home-SDE centroid distances for individuals in quartile 3, but to smaller distances or those living in quartile 1. An interesting insight also arises by investigating the impact of the distance between home and work locations, where larger distances will lead to larger home-SDE centroid distances, as well, since the individuals are in generally required to travel further away from their home due to the distant work location. Regarding the remaining parameters, individuals of household income between £30,000-50,000 and £75,000-100,000 have higher home-SDE centroid distances, while that also holds for elderly individuals (above 66 years old), for unemployed people and for those working in the sales or maintenance industry, social workers and academics. Smaller distances, however, are expected for married individuals and younger people (between 25-29 years old).

5.2.2. Home-SDE centroid angle outputs

The model regarding the angle between home locations and SDE centroids achieved the highest level of accuracy of the five SDE-related sub-models with a correlation of 0.78. The model outputs are presented in *Table 5*. That model, together with the previous one for home-SDE centroid distance, is important for the identification of the coordinates of the estimated SDE centroid, since a point can be defined by its distance and its angle from another known point, in that case the home locations. Nonetheless, it is difficult to extract any meaningful behavioural interpretation from its estimated parameters. It is important to note that angles were measured from the home location to the SDE centroid, since the directionality matters for the size of the calculated angles. The main interpretation that can be extracted from the estimates is that on average the CBD of Leeds acts a strong attraction to the centre of gravity of individuals' usual areas of movement, as captured by the SDE centroid, regardless of the geographic quartiles they reside. Specifically, individuals living in quartile 3 have the smallest home-SDE centroid angle with an average angle of 87.5^o. Individuals from quartile 4 and 1 have larger home-SDE centroid angles with average values of 150.9^o and 195.7^o, respectively. Finally, quartile 2 has the largest angles with an average value of 280.8^o.

Table 3: Modelling outputs of the distance (km) model for *OSO* trip chains

Parameters	**Estimates**	**Rob. t-ratios**
Constant	-1.8424	-3.52
Car-Walking	-1.3592	-2.04
Walking-Walking	-1.3846	-13.83
Shopping: Other	0.4366	2.97
Household size: 1 member	-0.1666	-2.29
Household size: 4 members	-0.2578	-2.06
Occupation: Engineering/Community/ Social/Management/Student/Other	0.2336	3.54
Occupation: Retired	0.4607	2.33
Time of day: Interpeak	0.1588	1.87
Time of day: Weekday night	-0.2577	-2.04
Time of day: Weekday morning	1.1395	6.99
Time of day: Weekend morning	0.1459	1.59
Following trip purpose: Other	-0.4919	-1.36
Following trip purpose: Social/Leisure	-0.6376	-3.27
Following trip purpose: Return to Work	-0.1687	-1.45
Age 25-29	0.4385	4.99
Age 30-39	0.1261	1.69
Age 50-59	0.1626	1.71
Age ≥ 76	0.2502	1.94
Shannon's entropy 400m around shopping destination (log)	-0.1079	-1.51
Retail areas 400m around shopping destination in $1,000m^2$ *(log)*	0.1196	4.05
Household income £20,000-£30,000	-0.1305	-1.28
Household income £75,000-£100,000	-0.1872	-1.29
Household income- Not disclosed	0.4409	3.37
Parking areas 400m around shopping destination in $1,000m^2$ *for car trips (log)*	0.0456	1.38
Gender: Male	-0.1104	-1.64
Total passengers in shopping trip >1	0.0851	1.36
Duration of shopping activity	0.1953	3.53
Marital status: Widowed	-0.6593	-2.95
Trip part of a tour with main purpose: Education/Work	-0.7579	-2.40
Average annual income in 400m around home location in £1,000 (log)	0.5275	3.25
Sigma	0.6150	23.43

Table 4: Modelling outputs for the distance of the Standard Deviational Ellipse centroid from the home location ($dist_{hc}$)

Parameters	**Estimates**	**t-ratios**
Constant	-12.0510	-1.51
Home location in quartile 2, 4	12.0179	1.51
Home location in quartile 1	18.0662	2.15
Home-Leeds CBD distance in km (quartile 4)	0.1871	1.23
Home-Leeds CBD distance in km (quartile 2)	0.0432	2.39
Work-Leeds CBD distance in km (quartile 2, 3)	-0.1351	-2.63
Household income £30,000-50,000	0.1382	1.55
Household income £75,000-100,000	0.2617	1.58
Occupation: Sales, Maintenance	0.6715	2.51
Occupation: Social	0.2706	1.71
Occupation: Academics, Research	0.3332	1.84
Home-Work distance in km (quartile 1)	0.2583	2.96
Home-Work distance in km (quartiles 2, 3)	0.3372	5.31
Home-Work distance in km (quartile 4)	0.4320	3.13
Married	-0.2224	-2.38
Age 25-29	-0.3202	-2.01
Age >66	0.5669	2.37
Home-York CBD distance in km (quartile 1)	-1.6410	-2.28
Home-York CBD distance in km (quartile 3)	3.3845	1.55
Unemployed (quartile 4)	0.7089	1.91
Sigma	0.6753	23.24

All the above are also captured in the model from the estimated parameters for quartiles 1, 2 and 4 using quartile 3 as the base. It is also interesting to note that angles in quartile 2 are decreasing as home and work locations are further away from Leeds CBD and the distance between home-work locations is increasing. In addition, the neighbouring cities of Wakefield and Sheffield located in the south of Leeds also have an influence on the position of the SDE centroid relative to the home locations in quartile 2 with those increasing as the home-Wakefield CBD distance also increases and are decreasing as the home-Sheffield CBD distance increases. Larger distances between home and work locations also can cause an increase in angles in quartile 3. Population has an influence on the home-SDE centroid angles in quartile 3 with a larger population around home locations leading to increased angles. From the remaining parameters, individuals working in maintenance occupations and those aged between 30-39 or 50-65 years old have smaller angles, while individuals of lower income, engineers, students and those working in the food and serving industry have larger home-SDE centroid angles, all else held equal.

5.2.3. SDE orientation outputs

The modelling specification for the SDE orientation achieved a correlation of 0.43 between observed and predicted values. The estimated parameters are depicted in *Table 6*. Similarly to the previously described home-SDE centroid angles, this model also captures the impact of the Leeds CBD to the orientation of the SDE. The values of the SDE orientation range from 0° to 180°, since those angles in that case are measured from the SDE centroid to the top end of the ellipse. Because of that, SDEs for individuals residing in quartiles 1 and 3 have smaller values largely between $0^o - 90^o$ and average values of 79.2^o and 71.1^o, respectively, while individuals living in quartiles 2 and 4 have SDEs with orientations between $90^o - 180^o$ and average values of 109.0^o and 111.5^o, respectively. That is also captured in the model, with larger orientation angles being predicted for quartile 2 and smaller for quartile 1 (although statistically significant only at the 80% confidence level) relative to the base quartile 3. Orientations in quartile 4 are larger than those in quartile 3, but they were omitted from the final model since they were not statistically significant even at the 50% confidence level. Regarding the remaining parameters, the larger the population around the home location and the distance between home-Leeds CBD the larger the angle orientation of the SDE. Individuals with personal income of £10,000-30,000 and £40,000-75,000 have larger orientations, while the opposite is true for

Table 5: Modelling outputs for the angle between the Standard Deviational Ellipse centroid and the home location (θ_{hc})

Parameters	**Estimates**	**t-ratios**
Constant	-4.1798	-4.11
Home in quartile 1	3.3350	3.40
Home in quartile 2	40.6517	2.61
Home in quartile 4	2.9321	2.99
Home-Leeds CBD distance in km (linear) (quartile 2)	-1.3490	-2.76
Home-Leeds CBD distance in km (log) (quartile 2)	1.4916	3.50
Work-Leeds CBD distance in km (log) (quartile 2)	-0.2239	-2.41
Population 400m around home location in 1,000 people (log) (quartile 3)	4.1139	2.13
Personal income <£10,000	0.9963	4.20
Occupation: Engineering	0.7541	3.27
Occupation: Maintenance	-1.2733	-1.72
Occupation: Food and serving	1.0490	1.46
Occupation: Business, Education	0.2439	1.68
Student	1.1086	3.94
Home-Work distance in km (log) (quartile 2)	-0.3182	-2.53
Age 30-39 and 50-59	-0.3822	-2.87
Age 60-65	-0.9361	-2.93
Home-Work angle (quartile 2)	0.7091	2.48
Home-Work angle (quartile 3)	1.3336	2.35
Home-Wakefield CBD distance in km (quartile 2)	2.4829	2.59
Home-Sheffield CBD distance in km (quartile 2)	-1.4075	-2.50
Precision ϕ	3.3457	12.28

higher personal incomes between £75,000-100,000, unemployed people living in quartile 3 and for those living in more deprived areas in quartile 2.

5.2.4. SDE minor/major axis ratio outputs

The estimated model for the minor/major axis ratio ($\frac{b_{sde}}{a_{sde}}$) of the SDEs achieved a correlation of 0.45 between observed and predicted values and its outputs are presented in *Table 7*. Overall, individuals located at quartile 4 have larger $\frac{b_{sde}}{a_{sde}}$ ratios compared to the rest and especially compared to those residing in quartile 2. Those living in quartile 2 also tend to have larger ratios as the variability of shopping store types around their home locations increases. An interesting finding can be observed for the effect of income on the estimated ratios, with an almost monotonically decreasing $\frac{b_{sde}}{a_{sde}}$ ratio as the personal income increases. In addition, as the size of the household increases the $\frac{b_{sde}}{a_{sde}}$ ratio decreases possibly due to increased family commitments, while smaller ratios are also expected for people with no car availability, unmarried, working in Education or in arts, media and the sports industry and being between 18-24 or 40-49 years old.

5.2.5. SDE area outputs

The model regarding the SDE area achieved a correlation of 0.42 between the observed and estimated areas and its outputs are presented in *Table 8*. In general, individuals residing in quartile 1 have a larger SDE area. Furthermore, distance between home and Leeds CBD is a significant factor influencing the area with higher instances also leading to larger areas. The same also holds for distances between the individuals' work locations and the Leeds CBD, as well as distances between home-work locations (for quartiles 2 and 4), which also positively affects the SDE area with larger distances leading to larger SDE areas, i.e. the visited locations of the individual are more dispersed in space. Male individuals, students, and low income individuals have lower SDE areas, while the opposite is true for retired, divorced and unemployed individuals residing in quartiles 1 and 4. Higher population in buffers of 400m around home locations will lead to larger areas for homes in quartiles 1 and 4, but smaller for quartile 4. Finally, a more diverse set of land uses around

Table 6: Modelling outputs for the orientation of the Standard Deviational Ellipse (θ_{sde})

Parameters	**Estimates**	**t-ratios**
Constant	0.4372	1.07
Home located in quartile 2	1.7330	3.22
Home located in quartile 1	-0.5826	-1.35
Home-Leeds CBD distance in km (log) (quartile 4)	0.4272	3.69
Population 400m around home in 1,000 people (log) (quartile 1)	1.3999	1.92
Personal income £10,000-30,000	0.2429	1.79
Personal income £40,000-75,000	0.2962	1.32
Personal income £75,000-100,000	-1.0363	-1.66
Unemployed (quartile 3)	-1.0155	-1.60
Average IMD 400m around home (log) (quartile 2)	-0.3718	-2.20
Precision ϕ	2.5125	13.3

Table 7: Modelling outputs for the minor/major axis ratio of the Standard Deviational Ellipse ($\frac{b_{sde}}{a_{sde}}$)

Parameters	**Estimates**	**t-ratios**
Constant	0.2799	1.45
Home located in quartile 4	0.3888	1.55
Home located in quartile 2	-0.4586	-1.20
Home located in quartile 1	0.1548	0.91
Average Shannon's entropy 400m around home (quartile 2)	1.3516	1.72
Angle between home location and Leeds CBD (quartile 2)	-0.0045	-1.49
Personal income £10,000-20,000	-0.3262	-2.18
Personal income £20,000-30,000	-0.5208	-3.52
Personal income £30,000-40,000	-0.2023	-1.33
Personal income £40,000-50,000	-0.6026	-2.84
Personal income £75,000-100,000	-1.0159	-2.11
Occupation: Food and serving	1.2547	2.20
Occupation: Art, sports and media	-1.0400	-1.63
Occupation: Education	-0.3602	-2.41
Household size	-0.0777	-2.50
No car ownership	-0.3759	-3.00
Home-Work distance in km (log) (quartile 4)	-0.2513	-2.27
Non married	-0.3119	-2.80
Age 18-24	-0.2264	-1.46
Age 40-49	-0.2259	-1.79
Precision ϕ	5.9588	12.43

Table 8: Modelling outputs for the area of the Standard Deviational Ellipse (A_{sde})

Parameters	**Estimates**	**t-ratios**
Constant	1.9131	3.26
Home-Leeds CBD (quartile 1)	2.7535	3.02
Home-Leeds CBD distance (log) (quartile 1)	0.4817	1.83
Home-Leeds CBD distance (log) (quartile 2)	0.4843	2.12
Home-Leeds CBD distance (log) (quartile 3)	1.0470	3.72
Work-Leeds CBD distance (log)	0.1603	2.34
Population 400m around home location in 1,000 people (log) (quartile 1)	2.0909	3.79
Population 400m around home location in 1,000 people (log) (quartile 3)	1.8691	1.49
Population 400m around home location in 1,000 people (log) (quartile 4)	-1.2987	-1.42
Average Shannon's entropy 400m around home (quartile 1)	-2.0843	-1.57
Average Shannon's entropy 400m around home (quartile 4)	4.1394	3.06
Personal income £10,000-20,000	-0.3074	-1.58
Student	-0.9935	-3.62
Retired	0.6626	1.83
Number of cars	0.4861	3.82
Home-Work distance (log) (quartile 2)	0.4179	2.91
Home-Work distance (log) (quartile 4)	0.7289	3.98
Divorced	1.0171	3.52
Male	-0.4708	-3.01
Unemployed (quartile 1)	0.8565	1.59
Unemployed (quartile 4)	2.2312	3.48
Sigma	1.1722	23.24

home locations (as captured by Shannon's entropy) will lead to lower SDE areas for homes in quartile 2, but the opposite is true for quartile 4 leading to larger areas.

It is worth highlighting at this point that the aforementioned models of structural ellipsoidal SDE components were estimated on small sample sizes of 270 individuals. As a result, we opted to be more lenient on assessing the statistical significance of each estimated parameters through their t-ratios. Furthermore, the correlation between most of the respective observed and estimated values lies around 0.45 with only the model for home-SDE centroid angle achieving a higher correlation of 0.78. That signifies that there are other unobserved factors influencing those structural SDE elements and more specifically the SDE orientation, the minor/major axis ratio and the SDE area, which the small sample size and the limited duration of the survey (2 weeks) did not allow us to capture. Therefore, caution is advised when using those estimates for prediction and more research is required to uncover those latent factors, which should be a subject for future studies.

5.3. Latent Class Choice Model outputs

Moving on to the estimation of the behavioural models, it is worth mentioning that they were estimated on choice sets comprised of 1584 alternatives (3 modes for first shopping trip x 176 shopping locations x 3 modes for following trip). The general specification of the models presented includes a range of Alternative Specific Constants (ASCs) as shifts from the base alternative, which in this case is assumed to be travelling by car for both trip legs to destination 1, which refers to the central shopping mall in the city of Leeds (car-car-dest 1). As expected in a case of a model including $J = 1584$ alternatives, it is not possible to estimate $J - 1$ ASCs for numerical and computational reasons. As a result, we opted to group alternatives based on their general geographical area relative to Leeds Central Business District (CBD), with that grouping comprising of 9 categories, namely Leeds CBD, north-east-south-west of Leeds and north-east-south-west of the remaining region of Yorkshire. The shifts from the base ASC, include interaction effects for specific modes and for specific areas, separately for individuals with and without car ownership in their household, for students and for married individuals. The Level-of-Service (LOS) parameters include travel times for the mechanised modes of car and Public Transport (PT), distance for walking and travel cost for car and

PT. Base parameters for the three LOS attributes specifically for the first shopping trip were specified using a Box-Cox transformation for the purpose of capturing the presence of non-linearities in the sensitivities (Box and Cox, 1964). Using that approach, the LOS attributes x are entering the Utility function as $\beta\frac{x^{\lambda}-1}{\lambda}$, with β being the base parameter for a specific LOS attribute and λ being an additional estimated parameter capturing the degree of non-linearity. An estimated $\lambda == 1$ indicates a linear specification as the Box-Cox transformation effectively collapses to βx, a $\lambda < 0$ indicates the presence of decreasing marginal disutilities, while in the case of $\lambda = 0$ the Box-Cox specification takes a logarithmic form $\beta log(x)$. Finally, the flexibility of a Box-Cox transformation allows for the case of $\lambda > 1$ indicating the presence of increasing marginal disutilities, such as instances of increasing discomfort or time restrictions for the remaining time budget as travel time and/or cost increases. Shifts from those base LOS parameters are also specified for the following trip, for specific types of shopping activities, types of tours, trip chains and times of day. Furthermore, interactions with continuous measures of shopping duration are specified for time and walking and of personal income for cost. The remaining specified parameters refer to a measure capturing the angular deviation from the straight OD path that is required to reach the intermediate shopping location, a range of locational variables aiming to capture the preference of car drivers for parking spaces, the preference of individuals living in richer places to go shopping in poorer areas and vice versa, as well as a similar measure for individuals living in less racially diverse to go shopping to more racially diverse neighbourhoods. In that study, the quartiles of the distribution of house prices around the home location are used to characterise an area as rich (quartile 4) or poor (quartile 1) and the same is done for racial diversity using the distribution of the percentage of white residents to characterise an area as less (quartile 4) or more racially diverse (quartile 1. The additional attraction due to the presence of major retailers per shopping type is also captured, while finally a range of size variables referring to the population around the shopping location, the total retail area per shopping type and the diversity of shopping store types are also included in a composite logarithmic term as described in *Section 3*.

The fit statistics of the behavioural models are presented in *Table 9*. Besides the proposed PCS-LCCM specification, which is able to capture heterogeneity in the choice sets and in the sensitivities, three additional models are presented, a base MNL model, *MNL-base*, using the global choice set of feasible alternatives, a base LCCM specification, *LCCM-base*, using the same choice set as MNL-base and capturing unobserved heterogeneity in the sensitivities and a simplified probabilistic choice set formation LCCM, *PCS-LCCM-generic*, using the same choice set structure across classes as the proposed PCS-LCCM, but with generic parameters across classes and only constants in the class allocation. PCS-LCCM-generic is similar to the specification proposed in the study of Thill and Horowitz (1997a) but using different proxy measures for latent constraints. The inclusion of the PCS-LCCM-generic has the purpose of capturing the impact of confounding unobserved heterogeneity in choice set formation with heterogeneity in sensitivities. That model still provides significant model fit improvements from the MNL-base model with 33.54 LL units with just two additional parameters, namely the two constants in the class allocation model, as depicted in *Table 9*. The remaining two models that are able to capture unobserved individual heterogeneity in the sensitivities, namely LCCM-base and PCS-LCCM, provide further improvements over the MNL-base model by 117.9 and 148.1 LL units, respectively, with 52 additional parameters. The proposed PCS-LCCM is also able to outperform the LCCM-base model by 30.18 LL units with the same number of parameters, although a direct comparison of the log-likelihoods of the two models is not a valid approach, since those two are not nested specifications. Nonetheless, the improvements in model fit become evident by examining and comparing the AIC and BIC statistics (Ben-Akiva and Swait, 1986), both of which are improved for the PCS-LCCM compared to the LCCM-base. The proposed PCS-LCCM model is also able to outperform the PCS-LCCM-generic model by a significant margin, namely by 114.51 LL units with 50 additional parameters.

More important, however, are the differences in the estimated shares of the latent classes, with PCS-LCCM-generic allocating individuals by 28.2% to *class a*, 2.9% to *class b* and 68.9% to *class c*. Contrary to that, the PCS-LCCM model estimates a much larger share for *class b*, specifically 29.3% of the sample, and also smaller shares for *class a* and *c*, namely 24.8% and 45.9%, respectively. That serves as an indication of the necessity to capture heterogeneity both in the choice sets and in the sensitivities across the estimated classes, which is an aspect missing in Thill and Horowitz (1997a).

According to the estimated ASCs presented in *Table 10*, destination 1 (base) is more likely to be chosen when travelling by a PT or walking due to the parking restrictions in Leeds CBD and due to the increasing

Table 9: Fit statistics of the modelling specifications

Fit statistics	**MNL-base**	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Log-likelihood (model)	-4,106.186	-3,988.274	-4,072.650	-3,958.099
AIC	8,316.37	8,184.55	8,253.3	8,124.2
BIC	8,594.06	8,739.93	8,541.67	8,679.58
Number of parameters	52	104	54	104
Number of individuals			270	
Number of observations			1,541	

promotion of more sustainable modes. The same also holds for the remaining destinations in the city centre, which are also in general less likely to be chosen compared to destination 1. Destinations in the remaining study area further away from the city centre are even less favourable, especially for modes other than car, although shopping locations in local high streets are again more likely to be visited by walking. Mode combinations other than those involving car are also more likely to be chosen for individuals with no car availability in their households. Furthermore, shopping trips including more than 1 passenger are more likely to be performed by car, at least for one of the two legs, due to its convenience. Finally, walking for both legs is more likely to be chosen by students, but less likely by married couples hinting to cases of more constrained time budgets for the latter demographic group compared to the former.

Table 10: Estimated ASCs and sociodemographic shifts of the modelling specifications

Parameters	**Estimate (Rob. t-rat. 0)** ***(Rob. t-rat. 1.0)***			
	MNL-base	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Households with car ownership (base: car-car/dest 1)				
ASC dest 1 shift Car-PT/Car-Walking	-1.6871 (-2.55)	-1.5334 (-2.25)	-1.5453 (-2.32)	-1.7506 (-2.15)
ASC dest 1 shift PT-PT	1.4219 (3.70)	1.5695 (3.86)	1.4275 (3.71)	1.6282 (4.14)
ASC dest 1 shift Walking-PT/Walking-Walking	2.5854 (9.25)	2.7180 (9.45)	2.6698 (9.40)	3.0955 (7.34)
ASC rest Leeds CBD	-2.2826 (-6.47)	-2.2197 (-6.08)	-2.2339 (-6.46)	-2.1619 (-5.84)
ASC rest Leeds CBD PT-Car/Walking-Car/ PT-PT/PT-walking	1.7345 (4.20)	1.7530 (4.33)	1.6992 (4.27)	1.7232 (3.95)
ASC rest Leeds CBD Walking-PT	2.9668 (6.50)	2.9974 (6.49)	2.9381 (6.66)	3.2687 (5.76)
ASC rest Leeds CBD Walking-Walking	3.8365 (9.02)	3.9309 (8.98)	3.8931 (9.20)	4.0709 (8.05)
ASC rest Leeds (no CBD)	-0.6215 (-5.66)	-0.6500 (-5.59)	-0.6525 (-5.71)	-0.5632 (-4.94)
ASC rest Leeds (no CBD) Car-PT/Car-Walking	-2.8230 (-8.53)	-2.8247 (-8.97)	-2.6713 (-8.06)	-3.2396 (-7.29)
ASC rest Leeds (no CBD) PT-Car/PT-PT/ PT-walking/Walking-Car/Walking-PT	-1.2653 (-5.59)	-1.4362 (-5.83)	-1.2973 (-5.66)	-1.6280 (-4.96)
ASC rest Leeds (no CBD) Walking-Walking	0.8037 (2.88)	0.8646 (2.99)	0.8828 (3.09)	0.7185 (1.50)
ASC rest Yorkshire (no Leeds) Car-PT/Car-Walking/ PT-Car/PT-PT/PT-Walking/Walking-Car/Walking-PT	-1.7449 (-5.79)	-1.8331 (-5.76)	-1.6678 (-5.54)	-1.9882 (-5.47)
Shifts for households with no car ownership				
Car-PT/Car-Walking/Walking-PT/Walking-Walking	2.4892 (7.18)	2.2060 (7.11)	2.5350 (7.53)	2.7200 (7.01)
PT-PT	4.3207 (10.94)	4.1018 (10.27)	4.2962 (11.37)	4.7209 (9.61)
PT-Walking	3.1890 (7.06)	2.4761 (5.97)	3.2227 (7.29)	3.4159 (4.66)
Shifts for central areas outside Leeds city centre				
Walking-PT/Walking-Walking	2.0958 (3.41)	2.0963 (3.12)	2.1850 (3.66)	2.2892 (3.74)
Shifts for trips with more than 1 passenger				
PT for first/following trip	-1.4224 (-5.57)	-1.3454 (-5.22)	-1.4396 (-5.68)	-1.6195 (-5.33)
Walking for first/following trip	-0.5303 (-3.94)	-0.5217 (-4.25)	-0.5185 (-3.96)	-0.6501 (-4.46)
Shifts for students				
Walking-Walking	1.2361 (3.60)	1.2182 (3.32)	1.2759 (3.79)	1.7160 (3.28)
Shifts for married individuals				
Walking-Walking	-0.5279 (-2.10)	-0.7629 (-3.19)	-0.5077 (-2.01)	-0.6256 (-2.14)

Regarding the LOS variables of travel time, travel distance and travel cost, depicted in *Table 11*, statistically significant non-linearities were found for PT time, walking distance and travel cost, while only linear sensitivities were found for car time. In general, decreasing travel time and walking distance sensitivities were found as the shopping duration increases, while decreasing cost sensitivities were found as personal

income increases, but not for all classes for the LCCM specifications. Finally, travel time for motorised modes and walking distance sensitivities were slightly higher for the following trip relative to the first shopping trip.

More specifically for PCS-LCCM, *class a* representing individuals with significant space-time constraints has the highest travel time for car/PT and walking distance sensitivity for the first trip to the shopping destination. At the same time, those individuals are half as sensitive for travelling to the following activity location by a motorised mode and equally sensitive for walking there. That indicates that individuals in *class a* are more likely to choose a shopping destination closer to their initial origin to cover their shopping needs before travelling by car/PT to their next activity. Also worth noting is that the Box-Cox λ for walking distance of *class a* is above 1.0 indicating increasing marginal disutilities as distance increases. In contrast, *class b*, with individuals restricted within their usual area of movement, shows the smallest travel time and walking distance sensitivities for the first trip and the largest for the following one, meaning that they are more likely to choose a shopping destination closer to the destination of the following activity. Contrary to those, *class c* has time and walking distance sensitivities more in line with the estimates of the MNL-base model. All three classes show a decreasing marginal time disutility as the duration of the shopping activity increases. The same does not hold, however, across classes for the sensitivity to walking distance, with class a being the exception showing an increasing marginal disutility. Regarding cost sensitivities, *classes a* and *b* have similar base cost parameters, but *class a* also shows an increasing marginal cost disutility as captured by the Box-Cox λ and a decreasing one as income increases, while *class b* has a linear marginal cost disutility and an increasing one as income decreases. Finally, individuals in *class c* have a much lower and in fact not statistically significant linear base cost sensitivity, but one that increases proportionately to income.

Table 11: Estimated LOS parameters of the modelling specifications

Parameters	**Estimate (Rob. t-rat. 0)** ***(Rob. t-rat. 1.0)***			
	MNL-base	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Households with car ownership (base: car-car/dest 1)				
ASC dest 1 shift Car-PT/Car-Walking	-1.6871 (-2.55)	-1.5334 (-2.25)	-1.5453 (-2.32)	-1.7506 (-2.15)
ASC dest 1 shift PT-PT	1.4219 (3.70)	1.5695 (3.86)	1.4275 (3.71)	1.6282 (4.14)
ASC dest 1 shift Walking-PT/Walking-Walking	2.5854 (9.25)	2.7180 (9.45)	2.6698 (9.40)	3.0955 (7.34)
ASC rest Leeds CBD	-2.2826 (-6.47)	-2.2197 (-6.08)	-2.2339 (-6.46)	-2.1619 (-5.84)
ASC rest Leeds CBD PT-Car/Walking-Car/ PT-PT/PT-walking	1.7345 (4.20)	1.7530 (4.33)	1.6992 (4.27)	1.7232 (3.95)
ASC rest Leeds CBD Walking-PT	2.9668 (6.50)	2.9974 (6.49)	2.9381 (6.66)	3.2687 (5.76)
ASC rest Leeds CBD Walking-Walking	3.8365 (9.02)	3.9309 (8.98)	3.8931 (9.20)	4.0709 (8.05)
ASC rest Leeds (no CBD)	-0.6215 (-5.66)	-0.6500 (-5.59)	-0.6525 (-5.71)	-0.5632 (-4.94)
ASC rest Leeds (no CBD) Car-PT/Car-Walking	-2.8230 (-8.53)	-2.8247 (-8.97)	-2.6713 (-8.06)	-3.2396 (-7.29)
ASC rest Leeds (no CBD) PT-Car/PT-PT/ PT-walking/Walking-Car/Walking-PT	-1.2653 (-5.59)	-1.4362 (-5.83)	-1.2973 (-5.66)	-1.6280 (-4.96)
ASC rest Leeds (no CBD) Walking-Walking	0.8037 (2.88)	0.8646 (2.99)	0.8828 (3.09)	0.7185 (1.50)
ASC rest Yorkshire (no Leeds) Car-PT/Car-Walking/ PT-Car/PT-PT/PT-Walking/Walking-Car/Walking-PT	-1.7449 (-5.79)	-1.8331 (-5.76)	-1.6678 (-5.54)	-1.9882 (-5.47)
Shifts for households with no car ownership				
Car-PT/Car-Walking/Walking-PT/Walking-Walking	2.4892 (7.18)	2.2060 (7.11)	2.5350 (7.53)	2.7200 (7.01)
PT-PT	4.3207 (10.94)	4.1018 (10.27)	4.2962 (11.37)	4.7209 (9.61)
PT-Walking	3.1890 (7.06)	2.4761 (5.97)	3.2227 (7.29)	3.4159 (4.66)
Shifts for central areas outside Leeds city centre				
Walking-PT/Walking-Walking	2.0958 (3.41)	2.0963 (3.12)	2.1850 (3.66)	2.2892 (3.74)
Shifts for trips with more than 1 passenger				
PT for first/following trip	-1.4224 (-5.57)	-1.3454 (-5.22)	-1.4396 (-5.68)	-1.6195 (-5.33)
Walking for first/following trip	-0.5303 (-3.94)	-0.5217 (-4.25)	-0.5185 (-3.96)	-0.6501 (-4.46)
Shifts for students				
Walking-Walking	1.2361 (3.60)	1.2182 (3.32)	1.2759 (3.79)	1.7160 (3.28)
Shifts for married individuals				
Walking-Walking	-0.5279 (-2.10)	-0.7629 (-3.19)	-0.5077 (-2.01)	-0.6256 (-2.14)
LOS variables				
Travel time car, PT for first trip (mins)	-0.0996 (-11.46)	–	-0.0876 (-10.28)	–
Travel time car, PT for first trip (mins) (class a)	–	-0.0796 (-776.23)	–	-0.2033 (-4.24)
Travel time car, PT for first trip (mins) (class b)	–	-0.1217 (-3.98)	–	-0.0233 (-0.33)
Travel time car, PT for first trip (mins) (class c)	–	-0.1025 (-10.20)	–	-0.0989 (-6.60)
Travel time shift for clothes shopping	0.0371 (3.87)	–	0.0350 (3.75)	–
Travel time shift for clothes shopping (class a)	–	0.0321 (19.86)	–	-0.0752 (-1.07)
Travel time shift for clothes shopping (class b)	–	0.0188 (0.37)	–	0.0242 (0.36)
Travel time shift for clothes shopping (class c)	–	0.0310 (2.28)	–	0.0211 (0.96)
Travel time shift for OSO trip chains	0.0154 (2.44)	–	0.0139 (2.26)	–
Travel time shift for OSO trip chains (class a)	–	0.0542 (130.85)	–	0.1906 (2.24)
Travel time shift for OSO trip chains (class b)	–	-0.0066 (-0.30)	–	-0.0011 (-0.13)
Travel time shift for OSO trip chains (class c)	–	0.0109 (1.49)	–	0.0069 (0.67)
Travel time shift for HWH tours	-0.0477 (-4.55)	–	-0.0438 -(4.37)	–
Travel time shift for HWH tours (class a)	–	-0.0384 (-0.96)	–	0.0155 (0.25)

Continued on next page

Table 11 – continued from previous page

Parameters	Estimate (Rob. t-rat. 0) *(Rob. t-rat. 1.0)*			
	MNL-base	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Travel time shift for HWH tours (class b)	–	-0.1006 (-1.77)	–	-0.0080 (-0.28)
Travel time shift for HWH tours (class c)	–	-0.0339 (-2.74)	–	-0.0639 (-3.57)
Travel time shift for morning/weekend night	-0.0513 (-3.45)	–	-0.0456 (-2.83)	–
Travel time shift for morning/weekend night (class a)	–	-0.0855 (-1.59)	–	-0.0961 (-1.22)
Travel time shift for morning/weekend night (class b)	–	0.0104 (0.28)	–	-0.0019 (-0.09)
Travel time shift for morning/weekend night (class c)	–	-0.0844 (-3.35)	–	-0.0755 (-1.69)
Travel time multiplier for following trip	1.2386 *(2.94)*	–	1.2813 *(3.14)*	–
Travel time multiplier for following trip (class a)	–	1.4275 *(1.09)*	–	0.4945 *(-4.51)*
Travel time multiplier for following trip (class b)	–	2.0494 *(2.21)*	–	7.7936 *(0.26)*
Travel time multiplier for following trip (class c)	–	1.0417 *(0.44)*	–	1.0793 *(0.33)*
Shopping duration-travel time elasticity	-0.3261 (-10.85)	–	-0.3540 (-10.71)	–
Shopping duration-travel time elasticity (class a)	–	-0.3522 (-3.68)	–	-0.3005 (-2.48)
Shopping duration-travel time elasticity (class b)	–	-0.2807 (-3.04)	–	-0.4121 (-1.72)
Shopping duration-travel time elasticity (class c)	–	-0.3573 (-9.15)	–	-0.3384 (-5.98)
Box-Cox λ PT travel time	0.7652 *(-9.31)*	–	0.7748 *(-8.60)*	–
Box-Cox λ PT travel time (class a)	–	1.3010 *(6.57)*	–	0.8088 *(-1.54)*
Box-Cox λ PT travel time (class b)	–	0.6012 *(-6.86)*	–	0.7093 *(-2.87)*
Box-Cox λ PT travel time (class c)	–	0.8025 *(-6.36)*	–	0.7856 *(-5.86)*
Walking distance for first trip (km)	-1.5930 (-12.60)	–	-1.4470 (-11.67)	–
Walking distance for first trip (km) (class a)	–	-7.4632 (-4.02)	–	-6.5966 (-3.36)
Walking distance for first trip (km) (class b)	–	-1.9114 (-2.80)	–	-0.1913 (-0.15)
Walking distance for first trip (km) (class c)	–	-1.3102 (-8.66)	–	-1.4650 (-5.42)
Walking distance shift for OSO trip chains	0.1981 (1.80)	–	0.2026 (1.88)	–
Walking distance shift for OSO trip chains (class a)	–	-1.2952 (-0.65)	–	2.8216 (1.81)
Walking distance shift for OSO trip chains (class b)	–	0.1782 (0.18)	–	-0.1261 (-0.36)
Walking distance shift for OSO trip chains (class c)	–	0.1886 (1.22)	–	0.2416 (1.18)
Walking distance multiplier for following trip	1.2272 *(2.41)*	–	1.2660 *(2.58)*	–
Walking distance multiplier for following trip (class a)	–	0.1448 *(-8.04)*	–	1.2761 *(1.27)*
Walking distance multiplier for following trip (class b)	–	1.0851 *(0.41)*	–	22.2251 *(0.13)*
Walking distance multiplier for following trip (class c)	–	1.4818 *(2.63)*	–	0.7755 *(-0.83)*
Box-Cox λ walking distance	0.7887 *(-4.08)*	–	0.8219 *(-3.21)*	–
Box-Cox λ walking distance (class a)	–	1.7529 *(1.44)*	–	1.8287 *(1.76)*
Box-Cox λ walking distance (class b)	–	0.8440 *(-0.89)*	–	2.0684 *(0.29)*
Box-Cox λ walking distance (class c)	–	0.8353 *(-1.98)*	–	0.9832 *(-0.12)*
Shopping duration-walking distance elasticity	-0.1443 (-4.32)	–	-0.1593 (-4.45)	–
Shopping duration-walking distance elasticity (class a)	–	0.1608 (1.13)	–	0.1591 (1.89)
Shopping duration-walking distance elasticity (class b)	–	-0.3807 (-2.13)	–	-0.1580 (-0.91)
Shopping duration-walking distance elasticity (class c)	–	-0.1398 (-3.06)	–	-0.2710 (-4.08)
Travel cost (£)	-0.5887 (-8.44)	–	-0.5165 (-7.46)	–
Travel cost (£) (class a)	–	-2.7637 (-8.71)	–	-0.9448 (-2.38)
Travel cost (£) (class b)	–	-0.6751 (-2.40)	–	-1.1802 (-1.66)
Travel cost (£) (class c)	–	-0.4275 (-5.33)	–	-0.1618 (-1.06)
Box-Cox λ travel cost	0.5900 *(-7.43)*	–	0.6568 *(-6.02)*	–
Box-Cox λ travel cost (class a)	–	0.8878 *(-0.68)*	–	1.2473 *(1.83)*
Box-Cox λ travel cost (class b)	–	1.0110 *(0.03)*	–	1.0049 *(0.04)*
Box-Cox λ travel cost (class c)	–	0.7106 *(-3.93)*	–	0.7888 *(-0.81)*
Cost-Personal income elasticity	-0.2862 (-2.91)	–	-0.3112 (-2.72)	–
Cost-Personal income elasticity (class a)	–	0.1863 (1.92)	–	-0.5483 (-2.54)
Cost-Personal income elasticity (class b)	–	0.7297 (1.11)	–	0.5531 (1.21)
Cost-Personal income elasticity (class c)	–	-0.5207 (-3.03)	–	-1.0933 (-1.54)

The model is also able to uncover interesting insights that could hint to instances of spatial and economic inequality in the area of Leeds, as shown in *Table 12*. According to the model, all else held equal, individuals living in more affluent areas (highest percentile of house prices) are less likely to visit shopping destinations located in poorer areas (lowest percentile of house prices). In contrast, individuals living in poorer areas do not show any preference difference in visiting richer or equally poor areas to cover their shopping needs. A potential interpretation of the above, relevant for policy makers, could be that a pound earned in the most affluent areas is more likely to be spent, hence distributed, in similarly affluent areas, while a pound earned in the least affluent areas is more likely to be equally distributed across space. Therefore, in the long run, wealth accumulation would favour more the already wealthy areas in Leeds compared to the rest leading to increased spatial inequalities. In a similar way, individuals living in areas with a higher percentage of white residents are less likely to shop in areas at the lowest percentile of white residents. Nonetheless, the same does not hold for individuals living in those areas at the lowest percentile as they are more willing to visit shopping destinations located in areas at the second and third percentile than in areas similar to theirs. As in every other problem in the urban context, there is of course a circular causality to disentangle here, as well (Bettencourt, 2021). According to that, all of the above, could be the result of the agglomeration of higher quality elemental shopping stores or better urban environment in general in richer or white dominated areas,

which leads to a reinforcing feedback loop favouring specific areas over others. That is exactly the problem that policy makers could try to alleviate by breaking that loop with the implementation of proper policy measures to provide relevant investment incentives in less affluent areas, as well.

Most of the remaining parameters, i.e. angular deviation, locational parameters and size variables, were found to be generic across classes and their estimates are similar with those of the MNL-base and LCCM-base models. The only exception is the parameter capturing how likely is for individuals living in richer areas to go shopping to poorer areas, which has been allowed to differ across classes. According to the estimated PCS-LCCM, individuals in *class a* who live in richer areas have a dispreference to travel to poorer areas for covering their shopping needs, but the opposite is true for individuals in *class b*. Nonetheless, both of those parameters are not statistically significant. Residents of richer areas of *class c*, however, show an even higher and statistically significant dispreference for shopping in areas with lower house prices.

A Box-Cox transformation of parking areas captured significantly positive, but also decreasing sensitivities as the area of parking increases. The presence of major retail attractions per shopping category (clothes, grocery, other) significantly increases the likelihood of visiting the shopping destination for trips of the respective shopping category. With regard to the direction of travel, shopping destinations located in places where the angular deviation between OS and OD is greater than 90° are less likely to be chosen compared to others, conforming to our initial assumptions.

The estimated multiplier ϕ of the logarithm of the composite size variable is significantly lower than 1.0 in all of the models presented. According to Kristoffersson et al. (2018), that hints to instances of significant unobserved correlation among the elemental alternatives within the aggregate shopping destinations used in the choice set. An increased cumulative retail floor area of grocery, clothes and durable stores in a destination acts as a more significant attractor for trips of the respective shopping category than population that was used as the base size variable. Furthermore, the cumulative floor area of grocery stores and an increased store type variability in neighbouring destinations in medium distances (1000-2000 m) will add to the attraction of the shopping destination, when the subsequent trip is also for shopping.

Table 12: Estimated locational parameters of the modelling specifications

Parameters	**Estimate (Rob. t-rat. 0)** ***(Rob. t-rat. 1.0)***			
	MNL-base	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Direction of travel				
Presence of angle>90° between O-S and O-D	-0.2621 (-2.22)	-0.2968 (-2.55)	-0.1920 (-1.58)	-0.2424 (-1.87)
Locational variables				
Parking areas (400m buffer)	0.1036 (4.00)	0.1091 (4.34)	0.1057 (4.05)	0.1151 (4.20)
Box-Cox λ for parking areas (400m buffer)	0.4219 *(-7.86)*	0.4189 *(-8.39)*	0.4237 *(-7.94)*	0.4085 *(-8.44)*
Living in rich areas - shopping in poor areas	-0.6817 (-2.51)	–	-0.7166 (-2.64)	–
Living in rich areas - shopping in poor areas (class a)	–	0.1737 (0.31)	–	-0.8626 (-0.85)
Living in rich areas - shopping in poor areas (class b)	–	0.1674 (0.33)	–	0.5875 (0.70)
Living in rich areas - shopping in poor areas (class c)	–	-1.0518 (-2.71)	–	-2.2892 (-2.62)
Living in areas with high % of whites (quart.4) - shopping in low % whites (quart.1)	-0.3732 (-1.72)	-0.3971 (-1.90)	-0.3664 (-1.68)	-0.2679 (-1.10)
Living in areas with low % of whites (quart.1) - shopping in high % whites (quart.4)	0.2889 (1.07)	0.2830 (0.93)	0.3011 (1.06)	0.4794 (1.61)
Living in areas with low % of whites (quart.1) - shopping in medium % whites (quart.2-3)	0.5348 (2.72)	0.5826 (3.05)	0.5576 (2.76)	0.6755 (2.99)
Major clothes shopping retailers (400m buffer)	1.3288 (6.03)	1.3515 (6.09)	1.3211 (5.94)	1.3672 (5.87)
Major grocery retailers (400m buffer)	0.4576 (4.51)	0.4657 (4.40)	0.4469 (4.33)	0.4180 (3.84)
Major durables retailers (400m buffer)	2.0015 (2.51)	1.9768 (2.59)	2.2064 (2.64)	1.9689 (2.74)
Size variables				
Natural logarithm multiplier ϕ	0.6638 *(-3.88)*	0.6698 *(-3.59)*	0.6594 *(-3.93)*	0.6730 *(-3.71)*
Population (400m buffer) (base)	1.0000 (–)	1.0000 (–)	1.0000 (–)	1.0000 (–)
Retail areas for clothes (400m buffer) (log.)	0.4515 (0.95)	0.4260 (0.86)	0.4833 (1.00)	0.5543 (1.04)
Retail areas for groceries (400m buffer) (log.)	0.8961 (2.24)	0.8995 (2.15)	0.9113 (2.27)	1.0641 (2.53)
Retail areas for durables (400m buffer) (log.)	0.3905 (0.56)	0.2969 (0.42)	0.3035 (0.44)	0.2744 (0.38)
Shopping store variability when following trip purpose is shopping (1000-2000m buffer) (log.)	2.1171 (1.75)	1.8938 (1.38)	2.0269 (1.58)	1.9265 (1.35)
Retail areas for groceries when following trip purpose is shopping (1000-2000m buffer) (log.)	-0.8628 (-1.08)	-0.9231 (-1.18)	-0.7967 (-1.02)	-0.6079 (-0.80)

As already mentioned, according to the class allocation model presented in *Table 13*, 24.8% of the sample is allocated to *class a*, therefore it is subject to space-time constraints, 29.3% to *class b* being subject to spatial cognition limitations and the remaining 45.9% is allocated to *class c*, labelled as the *explorers*, since they are willing to move in areas not visited before or at least not captured systematically during the survey period. With regard to the behavioural profiling of the estimated classes, *class a* is more likely to include individuals of higher household income above £65,000 (30.0% - average household income = £54,500), who

are in a possession of at least one car (27.3% - average number of cars = 1.0), but also living in higher than average deprived areas (30.6% living in areas with IMD>30 - average IMD = 23.7), live in a household of a size above 4 people (31.3% - average household size = 2.0) and be of younger age below 30 years old (27.9%) or older age above 60 years old (33.6% - average age = 39.5). According to the remaining sociodemographic attributes not used in the class allocation model, individuals in *class a* are more likely to be employed in social (28.8%) and legal occupations (42.9%) or be retired (33.9%), cohabiting with their partner (28.5%) and have at least three more employed adults in their household (28.0%). From that profile analysis, it can be concluded that those individuals could face space-time constraints mainly due to their family requirements and the need to accommodate the daily needs of the household members or due to their age. The highly deprived immediate neighbourhood that they reside might not provide the necessary incentives either in terms of infrastructure or general urban environment for them to wander in the space around them and explore new opportunities and amenities in neighbouring areas. As a result, they are more constrained to their pre-defined schedules and time budget constraints. Especially, for the case of the elderly individuals it is important to provide the necessary conditions to avoid cases of social exclusion.

With regard to *class b*, it is also more likely to include individuals of higher household income above £65,000 (39.2% - average household income = £54,000), who live in a household of a size above 4 people (38.1% - average household size = 2.0), but those also have a very low percentage of uniquely visited locations (44.6% have below 20% of uniquely visited locations). According to their remaining sociodemographic attributes, individuals in *class b* are more likely to be employed in the media/sports industry (45.1%) and technical (31.0%) and sales-related (31.6%) occupations, be in a possession of a season ticket for PT (34.6%) and live in a household with at least four more employed adults (38.1%). The behavioural profiling of *class b* indicates that those individuals are more likely to be bounded to their usual areas of movement due to the need to accommodate a significant amount of daily household needs, especially for the case of households with many employed individuals and probably not in a possession of a sufficient amount of private vehicles to accommodate all those individual needs. Contrary to *class a*, however, that increased number of employed individuals in *class b* might also provide the necessary incentives and be the reason behind the increased area of movement of those individuals, who are at least not bounded to their space-time constraints in a similar way as individuals of *class a*. The increased family commitments, however, do constrain their time budgets to not roam outside their familiar space to explore new opportunities and visit new locations. In order to understand the impact of the increased family commitments of individuals allocated into *classes a* and *b*, the ratio of personal to household income is calculated. According to that, the personal income of individuals allocated into *classes a* and *b* is more likely to be only a small percentage of the total household income with 29.3% and 33.3% allocated to *class a* and *b*, respectively, having a personal income of less than 20% of the total household income. Those allocation probabilities to *classes a* and *b* drop almost monotonically as their personal income ratio increases. That means that individuals in those classes are not the top earners in their family, hence are more likely to be involved with in-home activities to support the needs of the household.

Finally, *class c* is more likely to include younger individuals below 30 years old (48.3% - average age = 40.5) of lower household income below £35,000 (55.6% - average household income = £46,000) with no car ownership (54.5%) being alone in their household (67.0%) and visiting unique locations more frequently (51.0% have above 60% of uniquely visited locations). In addition, individuals in *class c* are more likely to be students (49.7%), working in maintenance and repairing occupations (61.8%) or be unemployed (55.1%) and be single with regard to their marital status (49.4%). That behavioural profiling hints to individuals free of the increased family commitments captured in *classes a* and *b* and hence able and willing to explore the space beyond their usual areas of movement.

Table 13: Estimated parameters of the class allocation models

Parameters	**Estimate (Rob. t-rat. 0)** ***(Rob. t-rat. 1.0)***			
	MNL-base	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Constant class a	–	-2.6019 (-1.94)	-0.8920 (-3.97)	-1.9282 (-1.93)
Constant class b	–	-1.9054 (-1.01)	-3.1578 (-0.85)	-0.7224 (-0.65)
Household income class a	–	-0.0326 (-1.74)	–	0.0165 (1.98)
IMD 400m around home class a	–	0.0355 (2.02)	–	0.0281 (2.32)
No car ownership class a	–	-1.6226 (-2.22)	–	-1.2401 (-2.62)
Population 400m around home class a	–	1.0254 (1.73)	–	-0.1348 (-0.31)
Low unique visits class a	–	1.1357 (1.59)	–	0.6834 (1.25)
Household size above 4 members class a	–	0.4062 (0.25)	–	0.9635 (0.70)
Age below 25 years old class a	–	-0.8398 (-0.82)	–	0.6862 (1.20)
Age above 60 years old class a	–	-0.4860 (-0.61)	–	0.6982 (1.24)
Household income class b	–	0.0016 (0.16)	–	0.0095 (1.28)
IMD 400m around home class b	–	0.0102 (0.49)	–	0.0088 (0.63)
No car ownership class b	–	0.9518 (1.23)	–	-0.2809 (-0.36)
Population 400m around home class b	–	-0.1769 (-0.15)	–	-0.3375 (-0.50)
Low unique visits class b	–	1.8876 (2.74)	–	1.4264 (1.66)

Continued on next page

Table 14: Comparison of Values of Travel Time estimates across models ($£/hr$)

Parameters	**MNL-base**	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Shopping trip				
Car	11.38	15.46	11.98	28.13
PT IVT	7.29	9.65	7.51	14.86
Following trip				
Car	13.17	18.52	14.46	30.61
PT IVT	11.01	15.18	11.68	21.63

Table 13 – continued from previous page

Parameters	**Estimate (Rob. t-rat. 0)** ***(Rob. t-rat. 1.0)***			
	MNL-base	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Household size above 4 members class b	–	2.1574 (1.42)	–	1.4468 (1.28)
Age below 25 years old class b	–	-0.3860 (-0.34)	–	-0.9790 (-0.71)
Age above 60 years old class b	–	-2.6021 (-0.45)	–	0.0576 (0.06)

Besides the improvements in model fit and the interesting behavioural insights derived from the PCS-LCCM specification, it is also important to note the discrepancies across models in the trade-offs of the individuals as captured by the Values of Travel Time (VTT) estimates, depicted in *Table 14*. In general, car VTTs are higher than PT values and VTTs for the first trip leg (shopping trips) are smaller than the ones for the following trips, which also include commuting trips, among others. According to the VTTs across the models examined, the values derived from PCS-LCCM are larger than the VTTs from the remaining models indicating the impact of capturing latent constraints on the VTT estimation. Furthermore, demand elasticities for different scenarios referring to a 1% increase in cost and time for car and PT are presented in *Table 15*. From that comparison, it can be concluded that PCS-LCCM leads to generally more modest demand elasticities/cross-elasticities compared to the remaining specifications examined and especially compared to MNL-base. It is important to note the significant overestimation of demand elasticities resulted from MNL-base in all scenarios examined both for time and cost and across modes. Similar findings were also derived in the study of Basar and Bhat (2004) indicating important policy implications from choice set misspecification.

6. Conclusions

In the current study, we focused on proposing a specification able to capture space-time and spatial awareness constraints in a spatial choice context. The study illustrates that the proposed approach is able to perform better than a range of other specifications used for comparison purposes. Furthermore, the study also demonstrates the impact of capturing latent choice set formation mechanisms on the estimated VTT values and demand elasticities, which can be important measures of analysis from a policy-perspective.

The geography-derived notions of Activity Spaces have been utilised in this study to define proxy measures for capturing the aforementioned latent space-time and spatial awareness constraints. Nonetheless, other measures could also be used for that purpose and future studies should continue on that direction to provide more computationally efficient ways of defining latent consideration choice sets in a spatial choice modelling context and its inherent complexity.

As a limitation of the current study, we should acknowledge the limited duration of the survey, which could have significant impacts on the proper estimation of the Standard Deviational Ellipses. In fact we could easily assume that we would end up with larger SDEs as a result of capturing more trips, which seem as less frequent ones with the current utilised dataset. On the other hand, however, we also can not exclude the possibility that the significant spatio-temporal regularities of travel would put more weight to specific locations leading to less variance and noise and hence more compact SDEs for specific individuals than the current ones.

In any case, the current study proposes an operational implementation of Manski's framework and its IAL version suitable and applicable for a spatial choice context with the necessary simplifications to make the problem computationally tractable. The study also demonstrates that Activity Spaces can be incorporated as proxy measures of capturing latent space-time and spatial cognition constraints leading to interesting insights

Table 15: Comparison of demand elasticities across models

Parameters	**MNL-base**	**LCCM**	**PCS-LCCM-generic**	**PCS-LCCM**
Increase car cost by 1%				
Car	-0.122	-0.100	-0.105	-0.089
PT	0.340	0.287	0.300	0.300
Walking	0.180	0.150	0.154	0.121
Increase PT cost by 1%				
Car	0.059	0.056	0.057	0.061
PT	-0.549	-0.516	-0.516	-0.526
Walking	-0.016	-0.016	-0.017	-0.021
Increase car time by 1%				
Car	-0.383	-0.365	-0.348	-0.291
PT	1.187	1.182	1.068	0.938
Walking	0.548	0.517	0.500	0.424
Increase PT IVT by 1%				
Car	0.090	0.087	0.087	0.077
PT	-0.748	-0.747	-0.698	-0.637
Walking	-0.039	-0.030	-0.039	-0.025

that could inform policy making. Besides the impact on VTTs and demand elasticities, the behavioural profiling of the estimated classes of the LCCM can provide invaluable information to policy makers for the purpose of proposing measures more suitable to the constraints of the underlying population, while also being able to timely identify and prevent cases of social exclusion, such as the individuals in class a. More research efforts in that direction, however, are necessary to disentangle the inherent complexity behind the formation of latent constraints and choice set elicitation mechanisms.

Acknowledgements

The current research was funded by the Advanced Quantitative Methods (AQM) scholarship of the Economic and Social Research Council (ESRC).

References

Basar, G. and Bhat, C.R. 2004. A Parameterized Consideration Set Model for Airport Choice: An Application to the San Francisco Bay Area. *Transportation Research Part B: Methodological*. **38**(10), pp.889–904. 10.1016/j.trb.2004.01.001.

Ben-Akiva, M. and Boccara, B. 1995. Discrete choice models with latent choice sets. *International Journal of Research in Marketing*. **12**, pp.9–24. 10.1016/0167-8116(95)00002-J.

Ben-Akiva, M.E. and Swait, J. 1986. The Akaike Likelihood Ratio Index. *Transportation Science*. **20**, pp.133–136. 10.1287/trsc.20.2.133.

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

Bettencourt, L.M.A. 2021. *Introduction to Urban Science: Evidence and Theory of Cities as Complex Systems*. Cambridge, Massachusetts: MIT Press.

Bierlaire, M., Hurtubia, M. and Flötteröd, G. 2010. Analysis of Implicit Choice Set Generation Using a Constrained Multinomial Logit Model. *Transportation Research Record: Journal of the Transportation Research Board*. 10.3141/2175-11.

Box, G.E.P. and Cox, D.R. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society*. **26**(2), pp.211–252.

Brown, L.A. and Moore, E.G. 1970. The Intra-Urban Migration Process: A Perspective. *Geografiska Annaler. Series B, Human Geography.* **52**(1), pp.1–13. 10.2307/490436.

Calastri, C., Crastes dit Sourd, R. and Hess, S. 2020. We want it all: Experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning. *Transportation.* **47**, pp.175–201. 10.1007/s11116-018-9858-7.

Calastri, C., Hess, S., Choudhury, C.F., Daly, A. and Gabrielli, L. 2018. Mode choice with latent availability and consideration: Theory and a case study. *Transportation Research Part B: Methodological.*

Capurso, M., Hess, S. and Dekker, T. 2019. Modelling the role of consideration of alternatives in mode choice: An application on the Rome- Milan corridor. *Transportation Research Part A: Policy and Practice.* **129**(10). 10.1016/j.tra.2019.07.011.

Cascetta, E. and Papola, A. 2009. Dominance among alternatives in random utility models. *Transportation Research Part A: Policy and Practice.* **43**, pp.170–179. 10.1016/j.tra.2008.10.003.

Cascetta, E. and Papola, A. 2001. Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies.* **9**(4), pp.249–263.

Daly, A. 1982. Estimating Choice Models containing Attraction Variables. *Transportation Research Part B: Methodological.* **16B**(1), pp.5–15.

Daly, A. and Hess, S. 2010. Simple approaches for random utility modelling with panel data *In*: *European Transport Conference.* Glasgow.

Department for Transport 2014. *Values of Time and Vehicle Operating Costs TAG Unit 3.5.6.* Department for Transport.

Domencich, T.A. and McFadden, D. 1975. *Urban Travel Demand: A Behavioral Analysis.* Oxford, England: North-Holland Publishing Company Limited.

Ferrari, S. and Cribari-Neto, F. 2004. Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics.* **31**(7), pp.799–815. 10.1080/0266476042000214501.

Gaudry, M.J.I. and Dagenais, M.G. 1979. The Dogit Model. *Transportation Research B.* **13B**(2), pp.105–111.

Guevara, A.C. and Ben-Akiva, M. 2013a. Sampling of alternatives in Logit Mixture models. *Transportation Research Part B: Methodological.* **58**, pp.185–198. 10.1016/j.trb.2013.08.011.

Guevara, A.C. and Ben-Akiva, M. 2013b. Sampling of alternatives in Multivariate Extreme Value (MEV) models. *Transportation Research Part B: Methodological.* **48**, pp.31–52.

Hagerstrand, T. 1970. What about people in regional science? *In*: *Ninth European Congress of the Regional Science Association.*

Hasnine, M.S., Lin, T.Y., Weis, A. and Habib, K.N. 2018. Determinants of travel mode choices of post-secondary students in a large metropolitan area: The case of the city of Toronto. *Journal of Transport Geography.* **70**, pp.161–171.

Horton, F.E. and Reynolds, D.R. 1970. Action space formation: A behavioral approach to predicting urban travel behavior. *Highway Research Record.* **332**, pp.136–148.

Horton, F.E. and Reynolds, D.R. 1971. Effects of Urban Spatial Structure on Individual Behavior. *Economic Geography.* **47**(1), pp.36–48. 10.2307/143224.

Howard, J.A. and Sheth, J.N. 1969. *The theory of buyer behavior.* New York: John Wiley & Sons.

Justen, A., Martínez, F.J. and Cortés, C.E. 2013. The use of space-time constraints for the selection of discretionary activity locations. *Journal of Transport Geography.* **33**, pp.146–152.

Kamakura, W.A. and Russell, G. 1989. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research.* **26**, pp.379–390.

Kamruzzaman, M. and Hine, J. 2012. Analysis of rural activity spaces and transport disadvantage using a multi-method approach. *Transport Policy.* **19**(1), pp.105–120. 10.1016/j.tranpol.2011.09.007.

Kaplan, S. and Prato, C.G. 2012. Closing the gap between behavior and models in route choice: The role of

spatiotemporal constraints and latent traits in choice set formation. *Transportation Research Part F: Traffic Psychology and Behaviour*. **15**(1), pp.9–24. 10.1016/j.trf.2011.11.001.

Kristoffersson, I., Daly, A. and Algers, S. 2018. Modelling the attraction of travel to shopping destinations in large-scale modelling. *Transport Policy*. **68**, pp.52–62. 10.1016/j.tranpol.2018.04.013.

Lam, W.H.K., Chen, B.Y. and Miller, E.J. 2018. Evaluating Place-based Accessibility using Space-time Utility Perspective *In*: *Proceedings of the 3rd International Conference on Civil, Structural and Transportation Engineering (ICCSTE'18)*. Niagara Falls, Canada, June 10-12, 2018. 10.11159/iccste18.109.

Leite Mariante, G., Ma, T.-Y. and Van Acker, V. 2018. Modeling discretionary activity location choice using detour factors and sampling of alternatives for mixed logit models. *Journal of Transport Geography*. **72**, pp.151–165. 10.1016/j.jtrangeo.2018.09.003.

Li, L., Adamowicz, W. and Swait, J. 2015. The effect of choice set misspecification on welfare measures in random utility models. *Resource and Energy Economics*. **42**, pp.71–92. 10.1016/j.reseneeco.2015.07.001.

Manley, E. 2016. Estimating the Topological Structure of Driver Spatial Knowledge. *Applied Spatial Analysis and Policy*. **9**, pp.165–189. 10.1007/s12061-014-9115-1.

Manski, C.F. 1977. The structure of random utility models. *Theory and decision*. **8**(3), pp.229–254.

Martinez, F., Aguila, F. and Hurtubia, R. 2009. The constrained multinomial logit: A semi-compensatory choice model. *Transportation Research Part B: Methodological*. **43**, pp.365–377.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior *In*: P. Zarembka, ed. *Frontiers in Econometrics* [Online]. New York: Academic Press, pp.105–142.

McFadden, D. 2000. Disaggregate Behavioural Travel Demand's RUM Side: A 30-Year Retrospective *In*: D. Hensher, ed. *Travel Behaviour Research: The Leading Edge* [Online]. Pergamon Press: Oxford, UK, 2000, pp.17–63.

McFadden, D. 1978. Modelling the choice of residential location *In*: A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull, eds. *Spatial Interaction Theory and Planning Models*. North Holland, Amsterdam, pp.75–96.

McFadden, D. and Train, K. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics*. **15**, pp.447–470.

Miller, H.J. 1991. Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information Systems*. **5**(3), pp.287–301. 10.1080/02693799108927856.

Newsome, T.H., Walcott, W.A. and Smith, P.D. 1998. Urban activity spaces: Illustrations and application of a conceptual model for integrating the time and space dimensions. *Transportation*. **25**, pp.357–377.

Pagliara, F. and Timmermans, H.J.P. 2009. Choice set generation in spatial contexts: A review. *Transportation Letters The International Journal of Transportation Research*. **1**(3), pp.181–196. 10.3328/TL.2009.01.03.181-196.

Patterson, Z. and Farber, S. 2015. Potential Path Areas and Activity Spaces in Application: A Review. *Transport Reviews: A Transnational Transdisciplinary Journal.*, pp.679–700. 10.1080/01441647.2015.1042944.

Potter, R.B. 1979. Perception of urban retailing facilities: An analysis of consumer information fields. *Geografiska Annaler. Series B, Human Geography*. **61**(1), pp.19–29. 10.2307/490402.

Punj, G. and Srinivasan, S. 1989. Influence of Expertise and Purchase Experience on the Formation of Evoked Sets. *Advances in Consumer Research. Association for COnsumer Research (U.S.)*. **16**, pp.507–514.

Schönfelder, S. 2003. Between routines and variety seeking: The characteristics of locational choice in daily travel *In*: Lucerne. 10.3929/ethz-b-000023602.

Schönfelder, S. 2006. *Urban Rhythms: Modelling The Rhythms Of Individual Travel Behaviour*. PhD thesis, Zürich: Eidgenössischen Technischen Hochschule Zürich.

Schönfelder, S. and Axhausen, K.W. 2003. Activity spaces: Measures of social exclusion? *Transport Policy*. **10**(4), pp.273–286. 10.1016/j.tranpol.2003.07.002.

Schönfelder, S. and Axhausen, K.W. 2004. *Structure and innovation of human activity spaces*. ETH Zürich.

Schönfelder, S. and Axhausen, K.W. 2010. *Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel*. England: Ashgate Publication Company.

Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*. **27**(3), pp.379–423. 10.1002/j.1538-7305.1948.tb01338.x.

Smith, L., Foley, L. and Panter, J. 2019. Activity spaces in studies of the environment and physical activity: A review and synthesis of implications for causality. *Health & Place*. **58**, p.102113. 10.1016/j.healthplace.2019.04.003.

Song, F., Hess, S. and Dekker, T. 2019. Fancy sharing an air taxi? Uncovering the impact of variety seeking on the demand for new shared mobility services. *Business*.

Srivastava, G. and Schoenfelder, S. 2003. *On the temporal variation of human activity spaces*. Zürich: ETH Zürich.

Swait, J. and Ben-Akiva, M. 1986. Constraints on Individual Travel Behavior in a Brazilian City. *Transportation Research Record*. **1085**.

Swait, J. and Ben-Akiva, M. 1987. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*. **21**, pp.91–102.

Thill, J.-D. and Horowitz, J.L. 1997a. Modelling Non-Work Destination Choices with Choice Sets Defined by Travel-Time Constraints *In*: M. M. Fischer and A. Getis, eds. *Recent Developments in Spatial Analysis. Advances in Spatial Science* [Online]. Berlin: Springer, pp.186–208.

Thill, J.-D. and Horowitz, J.L. 1997b. Travel-time constraints on destination-choice sets. *Geographical Analysis*. **29**(2), pp.108–123. 10.1111/j.1538-4632.1997.tb00951.x.

Timmermans, H., Heijden, R.V.D. and Westerveld, H. 1982. Perception of Urban Retailing Environments: An Empirical Analysis of Consumer Information and Usage Fields. *Geoforum*. **13**(1), pp.27–37. 10.1016/0016-7185(82)90005-7.

Tsoleridis, P., Choudhury, C.F. and Hess, S. 2022. Deriving transport appraisal values from emerging revealed preference data. *Transportation Research Part A: Policy and Practice*. **165**, pp.225–245. 10.1016/j.tra.2022.08.016.

Tsoleridis, P., Choudhury, C.F. and Hess, S. 2021. Utilising activity space concepts to sampling of alternatives for mode and destination choice modelling of discretionary activities. *Journal of Choice Modelling*. **42**(1), p.100336. 10.1016/j.jocm.2021.100336.

Whittaker, R.H. 1949. Evolution and measurement of species diversity. *Taxon*. **21**(2/3), pp.213–251. 10.2307/1218190.

Williams, H.C.W.L. and Ortuzar, J.D. 1982. Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research Part B: Methodological*. **16**(3), pp.167–219. 10.1016/0191-2615(82)90024-8.

Wright, P. and Barbour, F. 1977. Phased decision strategies: Sequels to initial screening *In*: M. Starr and M. Zeleny, eds. *Multiple criteria decision making*. Amsterdam: North Holland, pp.91–109.

Yuill, R.S. 1971. The standard deviational ellipse; An updated tool for spatial description. *Geografiska Annaler. Series B, Human Geography*. **53**(1), pp.28–39.