Assessing transport related social exclusion using a Capabilities Approach to accessibility framework: A dynamic Bayesian network approach

Abstract

Accessibility is considered to be a valuable concept that can be used to generate insights on issues related to social exclusion due to limited access to transport options. Recently, researchers have attempted to link accessibility with popular theories of social justice such as Amartya Sen's Capabilities Approach (CA). Such studies have set the theoretical foundations on the way accessibility can be expressed through the CA, however, attempts to operationalise this approach remain fragmented and predominantly qualitative in nature. In this study, a novel framework of expressing accessibility at the level of an individual is proposed, based on the basic elements of the CA. In particular, dynamic Bayesian networks are used to express the causal relationship between capabilities, functionings, personal and environmental characteristics. This is done by introducing informative Dirichlet prior distributions constructed using data from traditional mobility surveys, modelling the transition probabilities with data related to place based characteristics and defining an observation model from unlabelled mobility data and places of interest (POI). We demonstrate the usefulness of the proposed framework by assessing the equality levels and their link to transport related social exclusion of different population groups in London, using unlabelled, service provider generated mobility data.

1 Introduction

The concept of accessibility has been the focus of different disciplines such as geography, urban planning and transport planning for some time. The wide adoption of the term resulted in different definitions commonly encountered throughout literature: A very early definition originates from Hansen (1959) who defined accessibility as a potential of interaction between destinations. Within transport economics Ben-Akiva (1979) defined accessibility based on the benefits provided by the interaction between transport and land use. In transport geography Geurs & Van Wee (2004) defined accessibility as the extent to which transport and land-use systems enable individuals or groups of individuals to reach activities or destinations by means of transport modes.

When the focus of the studies is the connection between accessibility and social processes causing disadvantage, such as transport related social exclusion, the term accessibility is generally viewed as a fundamental property of individuals to participate in different activities within civil society (Burns 1980, Preston & Rajé 2007) and refers to the extent to which a person is able to reach a range of destinations that facilitate different social, leisure and employment activities considered to be normal for their society (Evans 2009, Nutley 1998). This ability takes the wider urban environment characteristics into consideration, such as transport provision (buses, trains etc.) and environmental characteristics as well as individual preferences and capabilities (Farrington 2007, Kwan 2013). Related to this, Church et al. (2000) identified seven distinct factors that could reduce access to opportunities, covering aspects such as physical characteristics of an individual (eg. mobility difficulties, impairments etc.), geographical and place based characteristics, time based restrictions as well as economic and societal factors. It is important to note that these factors tend not to appear in isolation, and coexisting factors are more likely to increase the risk of transport related social exclusion.

Although the above description of accessibility overlaps with the notion of mobility, it also highlights some key concepts that tend to be overlooked by thinking only in terms of mobility. Traditionally, in transportation planning and engineering, individual mobility refers to the resources and characteristics of individuals (financial status, age, access to a car etc.) that enable a person to move from place to place (Tyler 2006). However, increased mobility does not necessarily result in increased accessibility. For example, a person can be thoroughly mobile and still experience barriers when attempting to reach an activity. Besides physical and geographical, these barriers could be of a social nature such as social discrimination or fear of crime. (Church et al. 2000, Evans 2009). In any case, the mobility component is implicitly included in the definition of accessibility as given above.

In terms of accessibility measurement, there is a rich history of different numerical approaches, depending on the geographical scale and target group of an accessibility assessment. One fundamental categorisation given by Miller (2005) is place-based measures and people-based measures. The former focuses on place based and spatial separation concepts while the latter focuses on individual accessibility/mobility patterns. While measures from both categories have been used to investigate issues of transport related social exclusion, the choice of measure can produce dramatically different results. These range from overestimating equity of access to different urban services, as is the case of place-based measures, to producing more conservative, but oversensitive results (Neutens et al. 2010).

Recently, there has been an interest in using Amartya Sen's Capabilities Approach (CA) to express accessibility using theories of social justice (Hananel & Berechman 2016). This framework can then be used for investigating equity issues in transport. Apart from providing decision makers with a framework for considering equality in transport provision, the components of the approach provide the flexibility to express complex concepts, such as accessibility, through a causal structure (Pereira et al. 2017, Hananel & Berechman 2016, Beyazit 2011). In this study, a new numerical framework is presented for evaluating individual accessibility, using the Capabilities Approach. The implementation is based on dynamic Bayesian networks, and provides both inferential and computational intelligence capabilities using unlabelled mobility data. Furthermore, this study links the discovered accessibility patterns with socio-economic qualitative attributes at an individual level.

Contrary to previous articles that have used this approach within qualitative case studies, we demonstrate its applicability to investigating equality levels in accessibility using a combination of machine generated service provider data, in particular London's Automatic Fare Collection (AFC) system, and further socio-economic data. The paper is organised as follows. In section 2, the link between accessibility and transport social exclusion is briefly discussed as well as the placement of accessibility within social justice theories. Section 3 introduces the Capabilities Approach to accessibility model and provides implementation details by specifying the elements of the Capabilities set and how these relate to the observed functionings. Finally section 4 provides the results and sections 5 and 6 provide the discussion and conclusions, respectively.

2 Research Background

2.1 Accessibility and transport related social exclusion

The link between transport disadvantage and issues such as social exclusion, well-being and discussions around issues of equity and equality has been recognised since the 1960's. Fairly recently however, this discussion has been extended to recognise the fundamental role of accessibility in such issues (Pereira et al. 2017, Lucas 2012, Casas 2007). According to a widely cited definition by Kenyon (2003), transport related social exclusion is a process by which individuals are prevented from participating in different aspects of a social life in a community. This may be because of reduced accessibility to opportunities, services and social networks or due to insufficient mobility in a society. Such a process leads to decreased levels of well-being particularly for vulnerable population groups (Currie et al. 2010).

A social exclusion approach to transport disadvantage puts the focus on the outcomes of transport deprivation (Titheridge et al. 2014), however, it is important to notice that this concept emphasizes both the causal factors that lead to such a condition and the interactions between them (Lucas 2012). Such factors include characteristics that lie with the individual, characteristics of the local area as well as wider economic societal and governance factors. The lack of available transport options or inability to use them, together with disadvantaged personal status reduces the ability of an individual to reach different opportunities, causing lack of accessibility, which is in turn manifested as social exclusion.

Along the same line of thought, Preston & Rajé (2007) argue that the effects of social exclusion are not due to a lack of social opportunities, but because of a lack of access to those opportunities. According to the authors, addressing social exclusion requires extending the knowledge of person/place interaction beyond transport geography and into the domain of social-spatial research. Approaching accessibility from this angle, Farrington & Farrington (2005) redefine the terms used to describe accessibility: Opportunities become more than locations on a map. They are potentials for achieving an individual's needs, wants, aspirations and desires. Reaching opportunities becomes more than a function of space, as an individual won't necessarily be able to participate in the activities associated with each destination (Pereira et al. 2017).

At this point, it should be noted that case studies seeking to quantify transport disadvantage or transport related social exclusion [1] rarely adopt the above described definition of accessibility in its entirety. Instead, existing accessibility indicators covering aspects of the above definition are used (Kamruzzaman et al. 2016, Pyrialakou et al. 2016). For example, Preston & Rajé (2007) used a gravity and utility based accessibility indicator to identify areas of different levels of mobility/accessibility at an aggregated level. Wu & Hine (2003) used a contour based accessibility approach to identify transport disadvantage in households living in areas with limited transport coverage. Commonly mentioned reasons for this are the lack of data availability, the convenience of using already established models as well as the need to communicate the findings in a familiar manner to policy makers.

[1]Although transport disadvantage and transport related social exclusion are different concepts, the indicators used in case studies are often identical (Kamruzzaman et al. 2016)

However, attempting to approach issues of social equity using existing accessibility measures can be problematic. Using different frameworks of social theory, Martens & Golub (2012) examined how different accessibility measures could perform in terms of equity. They argued that neither place based frameworks such as distance and infrastructure based measures, nor people based frameworks, such as space-time and utility based measures, are suited to address issues of social justice in transport. The former is mainly mobility oriented, in that it focuses only on the ability of a person to travel in space but not actually to what he/she can do with the opportunities offered at a destination. For example, there might be fully accessible buses for people with disability, but this is not of much use if the destinations lack accessible facilities. People based frameworks on the other hand, by looking only at actual travelling patterns, conceal a basic equity argument (Sen et al. 1990): a person might adjust his/her expectations to deal with the conditions at hand. For example, a disabled person might manage to get to work by travelling twice the amount of time compared to a non-disabled person, but that doesn't mean that the person should not opt for better transportation conditions. These examples highlight that an individualistic, people based approach to accessibility needs to be combined with qualitative appraisals that could aid towards a deeper understanding of inequalities.

Finally, it should also be mentioned that accessibility is only one way of quantifying transport related social exclusion, albeit the most holistic one. Other methods include structured questionnaires and basic statistical analysis (Delbosc & Currie 2011), outcome based analysis such as measurement of individual activity spaces (Schönfelder & Axhausen 2003), deprivation based measures (Noble et al. 2007), mobility based measures (Dodson et al. 2006) and structural equation models (Golob & McNally 1997).

2.2 The Capabilities Approach and accessibility

The CA was first introduced by the philosopher and economist Amartya Sen in the 1980's (Sen et al. 1990), and was originally developed as an alternative to the predominant utilitarian way of viewing notions such as quality of life and well-being in welfare economics. It's success as a theory of social justice has led to the creation of the Human Development Index by the United Nations Development Programme for the purposes of ranking countries by the level of well-being. In essence, the CA describes the ability of an individual to function given the set of practical opportunities that are available to them (Sen et al. 1990). Contrary to Rawl's egalitarian approach (Rawls 2009) where the emphasis is on the primary goods, the CA focuses on human capabilities which result from a combination of personal abilities, and the wider environment (Pereira et al. 2017).

The CA can be perceived as a normative evaluation concept, aiming at promoting public policies towards improvement of the abilities of individuals to function as opposed to just describing the problem. This allows for the relative assessment of different policy proposals and the effect that those will have on a person's well-being (Alkire 2008). As accessibility has been traditionally used as a concept that can push towards policy changes (Pirie 1981) the Capabilities Approach seem to fit in that framework. Viewing accessibility within this context encompasses not only the ability of individuals to move so that they can conduct the activities they value or have reason to value, but also includes all the policies that enable people to do so (Pereira et al. 2017). Two notions are central in this theory: capabilities and functionings:

- **Capabilities:** These refer to the practical opportunities available and are the combinations of beings and doings that a person can achieve.
- **Functionings:** These refer to the various things a person may value doing and being (Sen 2014) and are usually observed (realised) representing what an individual actually achieves.

In accessibility terms, functioning can be understood as the realisation of day-to-day activities (e.g. shopping, getting to work etc.). The practical opportunities constitute the capabilities that each person has to complete the activity. Although the capability set is not directly observable, it can be derived from a set of functioning vectors from which the person has the freedom to choose (Mitra 2006). In this reading, the Capabilities Approach can be used to capture elements of social freedom (the ability to achieve various functions and realise one's potential), welfare (the capability to achieve these functions) and equity (Hananel & Berechman 2016).

Within the CA, the notion of functioning vectors refers to all factors that shape the capabilities set. The scope of functioning vectors can be very broad and can include different elements such as an individual's characteristics (e.g. age, income, impairment etc.), characteristics of the environment (e.g. social, physical, cultural etc.) or commodities (e.g. availability of public transport modes).

2.3 The capabilities approach in transportation literature

Literature on applications of the CA in transportation is sparse, however, it clearly sets the scene on how it can be used to assess accessibility.

Hananel & Berechman (2016) argue that the first step towards translating the CA in the transportation domain is to define what is meant by capabilities. In their view, a combination of the extent of mobility and access to opportunities for individual population groups, especially the disadvantaged ones, could be considered as good candidates for capabilities. These capabilities should reflect the minimum conditions that allow the least advantaged groups to benefit from any transportation interventions. Thus, the functioning vectors may include measures such as the maximum allowable travel time, travel distance or travel expenses for all residents in the area of influence, focusing on the more disadvantaged. The authors conclude that the capabilities and functioning vectors should not be viewed independent from one another, but recognise and address the interactions between them.

In another study, Beyazit (2011) juxtaposed the core elements of the CA with concepts in transport research. In their analysis, functionings refer to the wider definition of accessibility as it has been described in the section 1. Particularly, the transport system constitutes the goods, while the provision of access to ones needs and wants is the functioning of the transport system. Travelling for leisure could be one of these functionings, as is travelling for social interaction. The capabilities then refer to the mobility element that enables people to move from one location to another physically, socially and financially, within a society and across societies. In this way, people possess a capabilities set which translates into an opportunities set of achievable functionings, from which they are free to chose. Manifestations of these choices could be the traveling mode or modes, the choice of locations, the reason to travel and the choice of travel time.

Hickman et al. (2017) interpretation of functionings and capabilities within the transport context is similar to that of the authors above. In their view, the functionings represent what a person actually does and how. The realised functioning element is represented by the actual travel behaviour and participation in activities and as such, it is easier to measure. Measurement of capabilities on the other hand is more challenging. The authors propose an individual based accessibility definition that encompasses, alongside physical accessibility, issues such as the type of available infrastructure, land use, social and cultural norms and individual characteristics. The defined capabilities set is specific to each individual and reflects the freedom to choose from different potential functionings. However, this doesn't mean that two persons with similar functionings have the same capabilities. For example, a person with higher income may choose to have a similar mobility level to a person of a lower income by choosing not to own or use a car.

This distinction between functionings and capabilities is beneficial in that it helps towards understanding why improvements in certain levels of accessibility (such as improvement in levels of public transport, new cycling infrastructure etc.) might not lead to improvement in the overall accessibility.

Pereira et al. (2017) proposes framing accessibility in terms of combined capabilities, having two separable but interacting components. This first one relates to a person's capability to access and use the transportation system, which depends on the interplay between personal and external factors. Personal factors may be individual characteristics such as physical and mental health, accumulated experience and financial resources. External factors may be the social environment as well as the transport system's design, price level information or availability. The second component refers to the more macroscopic view of accessibility which is related to the interaction between the transportation system and land-use patterns, and how this interaction acts as an enabler towards the expansion of capabilities. This includes elements of the transportation network such as network coverage and connectivity, as well as the spatial distribution of activities.

Tyler (2006) approached accessibility through the CA following a more microscopic view. In this setting, capabilities are perceived as the combination between the individual abilities of a person, and the capabilities the environment provides. To recite the author's example, the physical infrastructure might require someone to be able to step up 30cm to participate in an activity. If the person is not able to provide this capability based on the individual characteristics (eg. wheelchair user or the elderly), then participation in the activity is not possible. Therefore, there is an interaction between what an individual can offer and what the environment can provide.

Looking at the mobility component of accessibility for elderly people, Ryan et al. (2015), approached capabilities as the outcome of an individual's mobility resources. In this sense, the potential of an individual to use public transport constitutes an element of the capabilities set. Functionings are chosen by an individual from the elements of the capabilities set, which could be all the different transportation options. The definition of realised functionings as actual behaviour is in line with the previously described studies.

In a study to identify Minimum Income Standards for transport use within rural communities, (Smith et al. 2012) used the CA to place income in the wider notion of well-being. Income however, is only one of the factors that affect people's capabilities to function. As a result, the authors extended the definition of minimum income to refer to all the goods, services, opportunities and choices to participate in society.

In another study (Hickman et al. 2017), used the notions of desired and actual transportation situation to distinguish between capabilities and functionings. Particular elements of the capabilities set included, among others, proximity to transportation access points, air quality levels, levels of security, levels of enjoyment when travelling, levels of accessibility to employment, availability of transportation modes, commuting time and transport costs.

Other authors (Orr 2010), proposed defining the capabilities set by focusing on activities, both realised and potential. Once these are identified, the individual capabilities required to achieve these can be mapped out (eg. access to sufficient income). This approach is framed in terms of evaluation of different transportation interventions aiming at minimising transport disadvantage and social exclusion for elderly and disabled people. They proposed to break down an activity into individual tasks and assess each task individually. For example, the activity 'going to a shop' has a set of necessary tasks embedded, one which could be 'taking the bus'. Specific barriers can then be associated with particular tasks, such as 'fear of crime walking to the bus stop'. In contrast to the above mentioned case studies, this approach to defining and measuring the capabilities set is inherently data driven.

Table 1 below summarises the way different authors approached definitions of the capabilities set and functionings as well as input variables.

Judging from the reviewed studies, the CA has been applied to a wide range of social issues in transport, ranging from investigating the impact of specific transport interventions to evaluating transport related social exclusion. In nearly all cases, the studies were based on empirical findings within a specific geographical context while the focus was on disadvantaged groups (eg. low income people, elderly, slum dwellers etc.) and within a comparative evaluation framework. A considerable proportion of the reviewed studies were qualitative, in line with the body of literature covering social aspects of transport (Lucas & Porter 2016). The ones that were more quantitatively oriented used statistical tools such as Structural Equation Models, Principal Component Analysis, logistic regression etc. This suggests that there is currently no consensus among researchers on how to quantitatively operationalise the CA for issues related to transport and social aspects.

Authors	Scope	Factors included	Capabilities	Functionings	Methodology
Hickman et al. (2017)	Investigating transport disadvantage between different income groups in Manilla, Philippines	Proximity to transport, security, air pollution, access to employment, income etc.	eg. Travel to work and other activities, Information, Natural environment	Similar to capabilities	Qualitative interviews with focus groups, self-disclosing desired and actual levels of PT experience
Ryan et al. (2015)	Evaluating levels of interaction of PT for eldery people	Income, driving licence, population density, gender, age, difficulties in boarding a bus etc.	The extent that elderly people can use public transport for the majority of their trips	Frequency of public transport use	Logistic regressions for capabilities and functionings using likert scale responses for dependent variable and factors as independent variables
Nordbakke (2013)	Investigating mobility of older women	Social networks, access to car, physical accessibility of the build environment, security	Availability of PT, availability of activities, access to information, access to alternative transport	Mobility levels	Qualitative interviews
Smith et al. (2012)	Benchmarking transportation costs for rural communities	Income, accessibility to services, age and household composition etc.	Access and sustain of activities such as education, social participation, employment, healthcare etc.	Types and number of trips	Stratified sampling followed by qualitative interviews
Rashid et al. (2010)	Exploring transport disadvantage	Income, ethnicity, household composition etc.	Trip frequency, travel time, car dependency	Similar to capabilities	Principal component analysis followed by multi-criteria evaluation
Maciel et al. (2015)	Exploring the mobility dimension of deprivation in Sao Paulo	Income, education, housing, access to information etc.	Mobility and accessibility	Commuting to work patterns	Generation of deprivation and accessibility indices

Goodman et al. (2014)	Investigating the effects of providing free bus transport to young Londoners	Location, gender age, ethnicity, deprivation	Social participation	Free bus journeys	Qualitative interviews
Yang & Day (2016)	Effect of job relocation on travel wellbeing	Income, age, vehicle ownership, location, traffic etc.	Preferred PT mode	Used PT mode	SEM
Chikaraishi et al. (2017)	Association between individual capabilities and travel time spent	Income, years of schooling, car ownership	Leisure (consumption) and employment (production)	Mobility	PCA derived index

Table 1: Reviewed literature

In terms of the link between the Capabilities Approach and accessibility, three common themes have been identified in all reviewed studies:

- Capabilities represent the potential of an individual to reach and engage with opportunities. Realised functionings represent the observed behaviour of the above. Both of the terms are in line with the general definition of accessibility as set out in section 1.

- The focus of the CA is the individual and in this sense in line with person-based accessibility. Moreover, it takes into consideration the influence of internal and external factors that shape the individual capabilities set.

- The capabilities set is not static but in constant interaction with the components that shape it and the realised behaviour expressed by the actual functionings. The evolving nature of the capabilities set extends both spatially and temporally, in the sense that is modified based on location and time.

Moreover, the case studies emphasize the causal structure between the factors that shape the capabilities, the capabilities themselves and the functionings. This causal structure appears to be hierarchical in nature, with the functionings appearing at the bottom of the hierarchy and the factors appearing at the top.

The definition of the elements included in the capabilities set and the corresponding functionings is used interchangeably for some studies. This is not uncommon and has been identified in applications of the CA to other social aspects beyond transport (such as quality of life) (Robeyns 2005). Reasons for this can be traced in the definition of functionings as enablers to achieve the defined Capabilities, but also the close relationship between transport concepts such as mobility and accessibility (for example, mobility can be considered both a functioning (using the bus) and a capability (ability to move) (Chikaraishi 2017). In all cases, however, there is a distinction between what is measured (functionings) and hypothesis to be tested (capabilities).

On the other hand, there exists a general consensus on the factors influencing the capabilities set. This includes either focusing on the socioeconomic characteristics of an individual, the wider environment (both physical and social) or both. In line with social exclusion definition as provided by the Social Exclusion Unit (Social Exclusion Unit 2003), sociodemographic variables such as income, age and gender are all defining factors that influence accessibility and have been included in the majority of the studies. Variables of the wider social environment such as deprivation, although not explicitly accounted for, have been taken into account during the design phase of most of the reviewed studies. Physical characteristics such as distance to amenities, density of public transport etc. have also been adopted as important factors that shape the capabilities set by the majority of the studies.

Finally, in spite of the advantages of passively generated mobility data from transport service providers, namely larger samples, regular update rate, low cost and the potential for longitudinal studies (Pelletier et al. 2011, Bagchi & White 2005), none of the reviewed literature has explored their potential to extract quantifiable evidence of social exclusion and transport disadvantage. This is true within the accessibility literature in general (Anda et al. 2017) and the CA particular. This is largely due to the unlabelled nature of such datasets, requiring an additional step to infer activity types at a destination.

3 A Capabilities Approach to accessibility framework

3.1 Conceptual framework

The different concepts of the CA and the way they are linked are shown in Fig. 1 (Mitra 2006, Beyazit 2011). At an observable level, one encounters the functionings of an individual. Within an accessibility setting, this node is referring to the realised activities as well as the realised transportation modes used to reach those activities. Moving one level up the hierarchy there exist the latent set of capabilities which form the choice set of an individual. These are all the potential opportunities an individual could choose. In this setting, a realisation of a chosen element of the capabilities set leads to an observed functioning. This in turn is influenced by personal, environmental and social characteristics as well as the commodities a person has in his/her possession. All variables of this representation are expressed through stochastic quantities which aim to quantify uncertainty from incomplete knowledge about the state of variables, as is the case of capabilities, or from noisy and erroneous measurements as is the case of functionings.

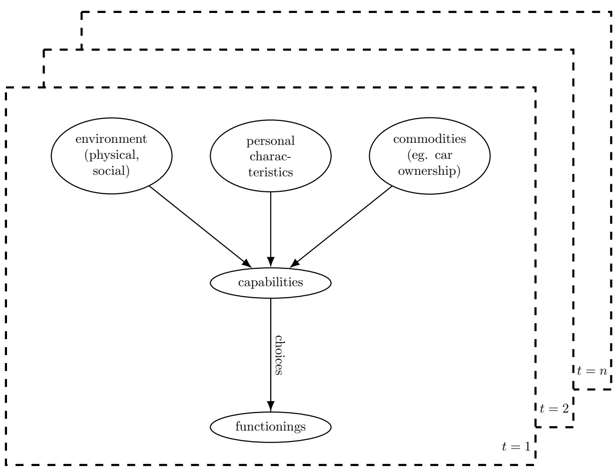


Figure 1: The CA (adopted from Mitra (2006))

The process is relevant for each individual and takes place in space and time during the act of reaching opportunities. In this setting, the capabilities set is changing depending of the characteristics of the environment that exist in each location at a particular point in time ($t = 1...n$). This representation imposes a structure on accessibility through the use of a directed graph where the nodes represent the components of the CA and the edges the relationship between them. The graph is acyclic, in the sense that no closed loops appear between the nodes. This allows information to flow from the top level to the bottom level nodes. The whole process should not be independent between subsequent time steps but should capture the dynamic evolution of capabilities in time.

In terms of mathematical implementation, the above described conceptual framework can be implemented through a Bayesian network structure, the details of which is described in subsequent sections. Bayesian networks have been successfully used within transportation research for a wide range of applications, ranging from transportation mode detection (Bantis & Haworth 2017) to travel behaviour analysis (Daziano et al. 2013). In the context of this study, advantages of using Bayesian networks can be summarised by the requirement of expressing accessibility through the causal structure of CA at an individual level, while at the same time providing inferential abilities from unlabelled mobility data. In addition, through the use of the posterior quantities for the model's nodes, Bayesian networks can represent uncertainty as a function of the different configurations of the states of all other variables in the model. Other approaches commonly used in the literature to represent casual relationships, such as SEMs, become unsuitable in the context of unlabelled mobility data. This is because SEMs do not provide inferential capabilities to extract semantic information from low level data.

3.2 Data

For this study, individual mobility data from London's AFC system (referred to as Oyster card) were used to infer the potential activity types an individual is likely to perform as well as the transportation modes used. In the 8-week sample provided (late October - mid December 2013), the individual trajectories represent the locations of public transport access points an individual used throughout their trips, as well as the public transportation modes used (Bus, Rail, Tram). The potential activities at each location were represented using Ordnance Survey's Points of Interest (POI) dataset (Ordnance Survey 2012), bounded by a 20 minute walking distance isochrone area at each location. From the 10-fold classification scheme defined by OS (Ordnance Survey 2012), four were considered representative for non-workplace activities (*Accommodation, eating and drinking, Outdoors and recreation, Education and health, Retail*) and one for employment activities (*Commercial services*).

Personal socio-demographic characteristics for each individual were obtained from a travel diary survey (London Travel Demand Survey, LTDS). LTDS is a continuous household survey aimed at probing London's public transport customers' sociodemographic background and patterns of transport, with a geographic coverage extending up to outer Greater London but within the M25 boundary.

During the 2011/2012 LTDS survey, respondents were asked if they were willing to provide their Oyster card unique ID for Transport for London (TfL) to undertake further analysis of their travel. Since then, the relevant data has been stored by TfL's Customer Experience department from mid-June 2011 to March 2014. The Oyster card sample provided for this study, overlapped with that of the LTDS sample for the period of October/mid-December 2013. The trajectories were linked to sociodemographic characteristics by means of a the Oyster card unique ID present in both datasets.

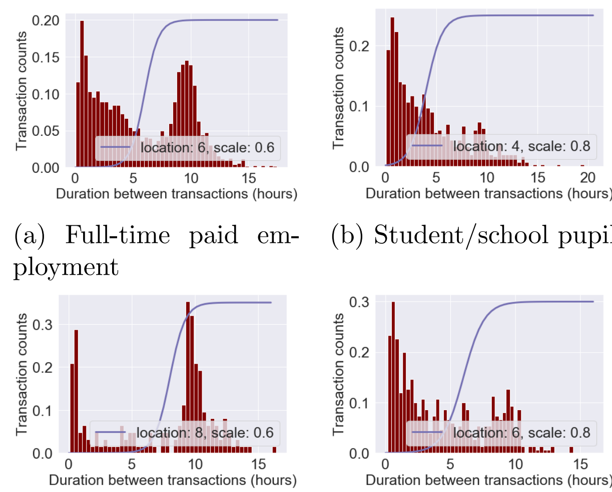
Differentiation between employment and non-employment activity types within an individual's trajectory was done by utilising the information contained in the combined LTDS/Oyster card dataset and duration of stay[2]. In particular, the distributions of duration of stay for individual daily journeys were plotted and examined for each employment status. The probability of an activity belonging to *Employment* was then calculated using the probability density function of a logistic random variable.

$$f(x;\mu,\sigma) = \frac{e^{-\frac{\mu-x}{\sigma}}}{\sigma(1+e^{-\frac{\mu-x}{\sigma}})^2} \tag{1}$$

where x is the duration in hours, μ the location parameter and σ the standard deviation (scale). The parameters μ and σ were adjusted to reflect different working assumptions as assessed empirically by the duration of stay distribution of Figure 10.

Equation 1 was then used to weight the OS activity type vector corresponding to *Employment/Education* activity types.

[2]In the context of this study, this was defined as the duration between individual trip segments, using only the records that are less likely to belong to an interchange trip



(a) Full-time paid employment

(b) Student/school pupil

(c) Full-time self-employment

(d) Part-time paid employment

Figure 2: Distribution of duration between transactions for different employment types, with logistic cumulative distribution functions (CDF) overlayed. Note that the histograms were normalised and the CDFs were scaled accordingly.

In addition to the above described data, a number of other data sources were used to shape assumptions about the influence of personal characteristics and external environment in an individual's ability to reach activities using the public transport. Table 2 provides a summary of the data used in this study.

Table 2: Description of data used in this study

Dataset	Source	Description
Oyster card	TfL	AFC individual trajectories between PT access points. Used as the primary mobility dataset.
LTDS	TfL	Travel diary survey of mobility habits. Used to shape assumptions about an individual and for qualitative analysis.
RODS	TfL	Origin Destination survey using the PT. Used to shape assumptions about an individual's activity preferences.
OS POI	OS	Points of Interest dataset. Used in activity type inference.
IMD	National Statistics	Index of Multiple Deprivation. Used to shape assumptions about the influence of external environment.
Traffic density	Department for transport	Traffic volume form traffic counters aggregated at LSOA level. Used to shape assumptions about the influence of external environment
Crime rates	London metropolitan police	Reported crimes aggregated at LSOA level. Used to shape assumptions about the influence of external environment.
Proportion of green areas	OpenStreetMap	Fraction of green areas computed as a proportion of the total LSOA area. Used to shape assumptions about the influence of external environment.
PTAL	TfL	Public transport accessibility levels (PTAL) is a combined measure of different aspects of public transport quality of service (such as reliability, waiting time, availability). Used to shape assumptions about the influence of external environment.
Trip duration	TfL	Trip duration calculated per trip segment using the Oyster card dataset. Used to shape assumptions about the influence of external environment.

Using this information, the framework of comparison for this case study is based on three population groups: individuals having annual household income below £15,000, individuals > 60 years of age and an unconstrained (base) population group. The choice of those population groups was based on two factors: the sample size of each group and past research providing evidence of population groups with significantly different accessibility levels compared to the majority of population (Páez et al. 2010, Hickman et al. 2017, Kamruzzaman et al. 2016, Titheridge et al. 2009). Deviations from equality will be assessed by analysing the elements of the capabilities set using the Theil's index. The results of this task can then identify gaps between the defined capability nodes that contribute to different levels of mobility using public transport and access to activity types.

Figure 3 below shows the geographic distribution of visited places for each population group.

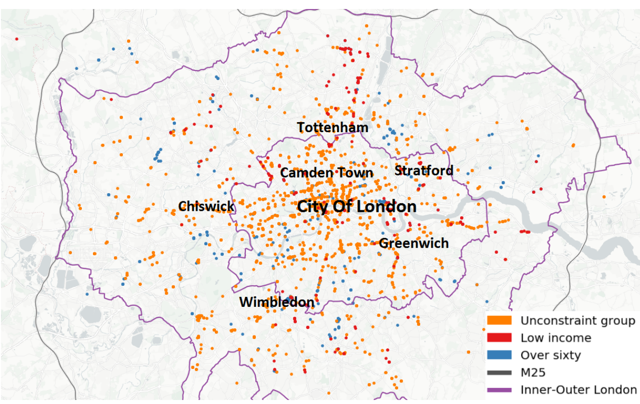


Figure 3: Visited places per population group.

As it can be seen, the majority of visited locations for the unconstrained and > 60 years old population groups is concentrated within the boundaries of Inner London in general, and around the area of City of London in particular. This is not surprising since the majority of employment opportunities is located in this area. On the other hand, the geographical distribution of the low income population group appears to span radially from Inner London, with a significant concentration around Tottenham area.

3.3 Model implementation

In the specification of this case study, two distinct but interacting components of an individual's act of reaching opportunities are included:

- Ability to interact with the public transportation modes available
- Ability to interact with the destination/opportunities available

Following the graphical representation of Figure 1, a dynamic Bayesian network was defined, the nodes of which reflect the structure of Figure 1. The sections below describe the individual components of the model.

3.3.1 Defining the Capabilities Set

By definition, the set of capabilities should be constructed in a way that reflects an individual's choice to to realise their desired goals, as well as the potential opportunities an individual has to make those choices. According to Hananel & Berechman (2016), an evaluation of the capabilities set should start by explicitly stating what these are. In the context of accessibility and adopting the definition of capabilities from Tyler (2006), these are framed around:

- the ability to engage with available opportunities and
- the ability to use the public transport to do so.

The first bullet point is related to the probability distribution of activity types bounded by the isochrone polygon, while the second is related to the probability distribution of using the different public transport modes at each access point in a trajectory.

In particular, this case study investigates the following elements:

- Potential accessibility to activities
- Potential mobility
- Potential accessibility and potential mobility dynamics

Potential accessibility to activities This element of the capabilities set describes the potential range of activity types that are reachable from a public transport access point. In the model specification of this case study, the set of potential activities is represented as a sequence of latent (unobserved) stochastic variables which are inferred by the propensity to perform each activity category based on personal characteristics and the number of defined POI types within reach from the public transport access point.

The effect of personal characteristics to the likelihood of reaching an activity type was captured through a Dirichlet distribution, through with the concentration parameter vector α_z. This represents the degree of prior belief that an individual is likely to be performing one activity type over the other. For example, it might be that prior studies point that arrival time between 11:00-12:00 pm and age group < 21 years old can be used to determine education over employment activity. This assumption can be represented by setting $\alpha_{education} > \alpha_{employment}$.

Smaller ($0 < \alpha_z < 1$) values of α_z express less uncertainty in the preference of an activity type over the other. On the other hand, larger values ($\alpha_z > 1$) express more uncertainty about the preference of an individual for an activity type. In this study, the calculation of the shape of the prior was based on rolling origin destination survey (RODS) data. In particular, a multinomial regression model was fitted on the complete RODS dataset, and the predicted probabilities for each activity type were generated for each individual based on age, sex, disability status and arrival time. These were then used to construct the concentration parameter vector α_z. The resulting predicted probabilities were then multiplied with Gamma distributed random variables with shape and rate parameters of the Gamma distribution a = b = 1 to ensure that the concentration parameters follow an exponential distribution with rate proportional to the RODS predicted probabilities (Bantis & Haworth 2019).

Potential mobility This element of capabilities set describes the potential of public transport use from the modes that are available. Similarly to potential accessibility, potential mobility is represented by a latent stochastic quantity that is inferred using the propensity of public transport use given an individual's sociodemographic characteristics and the distribution of transport modes from the Oyster card data.

Similarly to section 3.3.1, potential mobility was modelled using a categorical random variable over the Oyster card transportation mode types. This time, the propensity of an individual to use one mode over another was modelled using a multinomial regression on the LTDS dataset. In this case, the personal characteristics determining the choice of transportation mode were age, income, possession of travel pass, disability, car license, sex and ethnic group. The predicted probabilities were then recovered and used to shape the prior belief of using one mode over the others through the Dirichlet concentration parameters.

Potential accessibility and potential mobility dynamics In the context of this study, the dynamic evolution of the capabilities sets was captured using a set of transition matrices.

The underlying assumption that is made in this modelling step is that characteristics of the environment have a varying effect on the ability of an individual to reach an activity. For example, the levels of deprivation change from location to location, and this is expected to influence the choice of performing an activity type at a particular location. Similarly, the existence of more transportation options (expressed as increased levels of transport accessibility) are expected to influence the choice of transportation modes.

This assumption was represented by modelling the transition between subsequent transportation modes and activity types using a set of Multinomial logistic regressions on external covariates. The transition sequences for the inferred activities/transportation modes specified per each category were constructed as follows:
where $y_i = argmax(z_i)$ in the case of activities, and $y_i = m_i$ in the case of transportation mode.

Essentially, the algorithm generates a transition dataset from one category to another by looping through the trajectory locations and identifying if a

Algorithm 1 Construction of transition sequence

```
procedure CONSTRUCT TRANSITION SEQUENCE(input = c^{1...κ})
    for i in 1 : N do
        for κ in K do
            if y_i = κ then
                append y_{i−1} to c^κ
```

transition between location n and location n-1 is related to activity types/transportation mode k. For example, consider a trajectory with transportation modes $bus_1, bus_2, rail_3, bus_4$. In this case, the row of the transition matrix corresponding to bus related transitions will be inferred using the sequence $bus, rail$ as there is one bus/bus related transition (from $n = 1$ to $n = 2$) and one $bus/rail$ related transition (from $n = 2$ to $n = 3$)

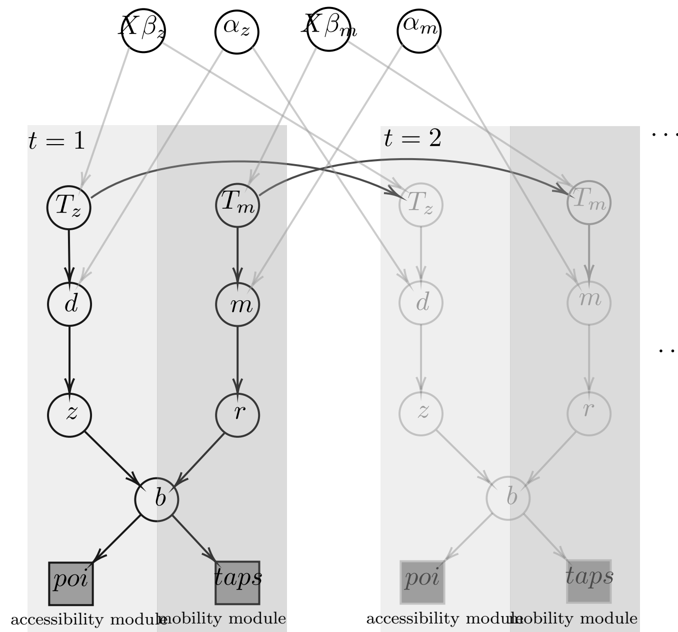
The external covariates used were *IMD, proportion of green spaces, traffic density, crime rate* for the transitions between activity types and *trip duration, PTAL* for transitioning between different transportation modes (see table **??**). These separate row regressions were organised in a row stochastic transition matrix and used in the calculation of the likelihood of potential accessibility/mobility. This resulted in two square row stochastic matrices, a $5x5$ matrix for the 5 activity categories T_z and a $3x3$ matrix for the transportation modes T_m.

3.3.2 Bringing it all together: Defining the structure of the model using Bayesian networks

The CA to accessibility (CAA) model consists of two distinct but intertwined modules: 1) activity detection and modelling and; 2) mobility modelling. These are combined using a switch variable that activates the relevant module depending on whether an individual is using public transport or performing an activity. Figure 4 illustrates a graphical representation of the joint model.

Formally, the model is defined in equation 2:

Figure 4: Graphical representation of CAA model.



$$\beta \sim Normal(0, 10^{-3}) \tag{2}$$
$$s^{1\dots\kappa} = \frac{exp(\alpha + \beta X_n)}{1 + \sum_{\kappa=1}^{K-1} exp(\alpha + \beta_\kappa X_n)}$$
$$c^{1\dots\kappa} \sim Categorical(s^{1\dots\kappa})$$
$$T = \begin{bmatrix} s^{\kappa=1} \\ s^{\kappa=2} \\ \vdots \\ s^{\kappa=K} \end{bmatrix}$$
$$P(d_n|\alpha_z, T_z) \sim Dir(\alpha_z, T_{row=argmax(d_{n-1})})$$
$$P(z_n|d_n) \sim Mult(poi_n, d_n)$$

The mobility part of the inference is similar to the above, with the exception that this time the observation vector of transportation modes is modelled through a categorical distribution $P(r_n|m_n) \sim Cat(m_n)$ with $P(m_n|\alpha_m) \sim Dir(\alpha_m)$. In this case, the notion of functionings is more straightforward as the Oyster card 'taps' are direct observations on the actual choice of transportation mode made by the individual. The node b is a stochastic variable acting as a switch that controls which module is activated for inference (accessibility or mobility). It is assumed to follow a Bernoulli distribution $P(b) \sim Bernoulli(p)$, the probability of which is determined by the duration of stay relative to the cutoff value determined from the 95th percentile of the distribution of interchange times for bus and rail services (in the case of rail services, this was 15 minutes while for buses this was 36 minutes). For example, if the duration of stay between two subsequent bus trips is more than 36 minutes, then it is more likely that an activity is carried out at the stop (as opposed to being an interchange stop).

Finally, at the very bottom of the hierarchy of Figure 4, the square nodes represent the observed mobility and POI data used to infer the parent nodes. Table 3 below summarises the notation of the model:

Table 3: Description of variables of Figure 4

Variable	Description
poi	Count of POIs per activity within the isochrone boundary
z	Multinomial probability distribution of activities.
r	Categorical probability distribution of transportation modes.
d	Dirichlet distribution on z.
m	Dirichlet distribution on r.
$\alpha_{z,m}$	RODS/LTDS derived concentration parameter vectors for activity types/transportation modes.
$T_{z,m}$	Transition matrices for activity types and transportation modes.
$\beta_{z,m}$	Prior on external covariates X $\beta \sim Normal(0, 10^{-3})$.
b	Bernoulli variable switching between modules during inference.

Inference on this the model was performed using Markov Chain Monte Carlo (MCMC) methods. A total of 20,000 sampling iterations were used, discarding the first 1,000 as non representative of the posterior quantities. The starting values of the stochastic variables were sampled from the prior distributions.

3.4 A Thiel index based assessment framework

Within the proposed framework of the CAA model, two components of an individual's ability to reach opportunities were identified and quantified, given personal characteristics and external factors: potential accessibility to different activity types using the public transport, and potential mobility of using the different transportation modes. The first one is related to the concept of equality of opportunities, while the second is related to issues of transport disadvantage. The next step of the analysis is to explore the relationship of the components between individuals using the posterior quantities as a basis of comparison.

Within the wider accessibility literature, the Theil index has been proposed as theoretically capable of quantifying accessibility related equity issues (Van Wee & Geurs 2011) and has been applied as an equity evaluation tool for different case studies (Delafontaine et al. 2011, López et al. 2008).

The Theil index quantifies the actual entropy relative to the maximum entropy of the data and practically is a measure of difference between complete randomness and uncertainty and the observed state of the dataset configuration (equation 3):

$$S_{Theil} = \sum_{i=0}^{N} \left(\frac{x_i}{N\bar{x}} ln \frac{N\bar{x}}{x_i} \right) \tag{3}$$

$$S_{max} = lnN$$

$$T = S_{max} - S_{Theil}$$

where x is a vector of non-negative elements, S_{Theil} is the observed entropy and S_{max} is the theoretical maximum entropy of the dataset.

It is interesting to observe that the above formulation is similar to Kullback-Leibler (KL) divergence (or relative entropy) if the vector x is a valid discrete probability distribution, as is the case for the posterior distributions of model 4, and S_{max} is the maximum entropy defined by the cardinality of the event set. KL-divergence is a commonly used probability divergence measure used to compare probability distributions within the context of applications in information theory (Cohen & Kempermann 1998) and statistics (Pardo 2005).

By definition, $T \geq 0$, with 0 meaning that the distribution is identical to the uniform distribution (the observed entropy is equal to the maximum) and higher values signify increased deviation from the uniform case and thus increased inequality. It is important to note that the index is invariant under state switching in the set. For example two individuals, one using the bus 90% and the remaining modes 10% of the time, and a second individual using the rail 90% and the remaining modes 10% of the time, will be assigned the same Theil value. This doesn't take into consideration which transportation mode is more favourable under a given circumstance. From this perspective, arguments related to equity are not possible by assessing the output of the index alone, and some qualitative discussion of the results is needed. This is also true for the weighted version of the index, as the weighting scheme needs to be decided to reflect equity considerations. Moreover, the uniform level of equality specified by maximum entropy represents a theoretical case that links to egalitarian approaches under the idea of equality of opportunity. However, it is legitimate to expect a certain level of inequality to exist, provided that it is caused by an individual's own choices and not unfavourable circumstances such as having low income (Pereira et al. 2017).

For the purposes of identifying individuals that experience a relative disadvantage, the posterior distributions are compared and contrasted using the Theil index against the state of complete equality characterised by maximum entropy. Since the Oyster card dataset doesn't provide any information related to an individual's preferences or desires, Theil values will be assessed under the assumption that any significant deviations of the individual Theil values from the group population mean could be attributed to particularities of the group (eg. low income, age), treating individual preferences as random fluctuations in the Theil values within the group.

4 Results

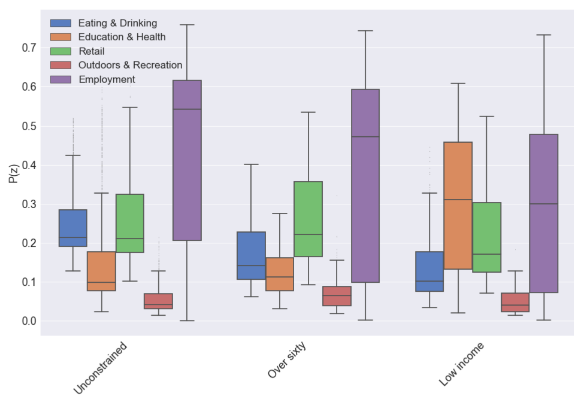
In this section the posterior distributions for activity types and transportation modes are presented for the three population groups: individuals > 60 years old, low income individuals ($<$ £15000 yearly income) and an unconstrained population group. The overarching goal is to emphasize the different accessibility patterns that indicate transport related social exclusion faced by individuals belonging to different population groups, compared with the unconstrained population group. For this task, a popular equality index is used (Theil index) to quantify equality levels between the three population groups (Van Wee & Geurs 2011, Delafontaine et al. 2011, López et al. 2008).

4.1 Distributions of activity types

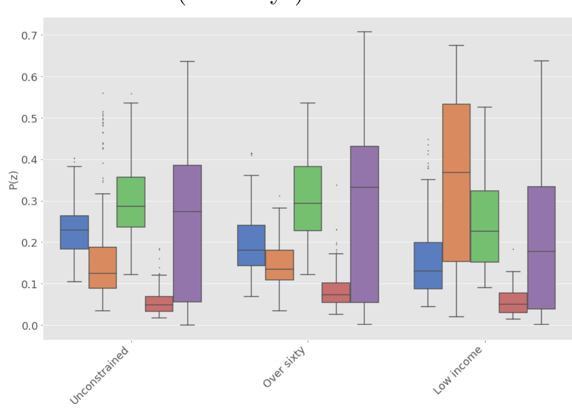
The posterior distribution of activity types corresponds to the latent d node, expressing the posterior distributions of activity types given the individual's sociodemographic characteristics, duration of stay and number of reachable POIs from the alighting point. Appendix A shows the posterior quantities of $P(d)$ for the trajectories of all users in the target groups.

4.1.1 Posterior results

The posterior quantities for the individuals in the low income group and > 60 years old group were similar to the unconstrained group for all categories (Figures 13,12,11). For the low income group one notable difference is the shorter tails of the daily distributions for the majority of the individuals for the *Employment* activity (Figure 12e). Empirically, this could signify reduced flexibility in using public transport to reach this activity compared to the unconstrained population group. Moreover, the probabilities of *Eating and Drinking* (Figure 12a) and *Retail* (Figure 12c) activity types is significantly lower throughout the day, remaining below the threshold for random probability allocation for the specified number of activity types (< 0.2). Contrary to the rest of the population groups, for the > 60 years old group the *Education and Health* (Figure 13b) category is characterised by a gradual increase over the later hours of the day for the majority of the individuals. This could be attributed to health related activities as opposed to *Education* related activities. Again, the *Employment* (Figure 13e) activity type seems to dominate the daily trajectory of this group for the early hours of the day. This is not surprising considering the fact that the majority of the individuals in this group were below the UK national pension age (63 years for women and 65 years for men). Nevertheless, a general shift of this activity type to slightly later hours of the day can be observed compared to the rest of the groups, reflecting some flexibility in using public transport to access employment. Figure 5 shows aggregated boxplots of posterior distributions for the different activity types, for all individuals in the target groups for both weekdays and weekends. As it can be seen, the general pattern of activity distribution remains with the exception of employment activity which is considerably lower in the weekends.



(a) Aggregated activity distribution boxplots for all individuals (weekdays)



(b) Aggregated activity distribution boxplots for all individuals (weekends)

Figure 5: Aggregated activity type boxplots for the three population groups.

4.1.2 Assessing equality levels

Looking at the distribution of Theil values ($T-values$) for the three population groups (Figure 6), the low income group has the largest mean compared to the rest of the groups, signalling overall increased inequality levels (at the .05 significance level, one way ANOVA, see Table 4). The equality assumption made here is that, throughout an individual's trajectory, all defined activity types should be equally reachable by an individual regardless of factors such as age, income etc., and thus the distribution of these activity types should approach the uniform distribution ($T = 0$). Increasing deviation from this can be considered deviation from equality.

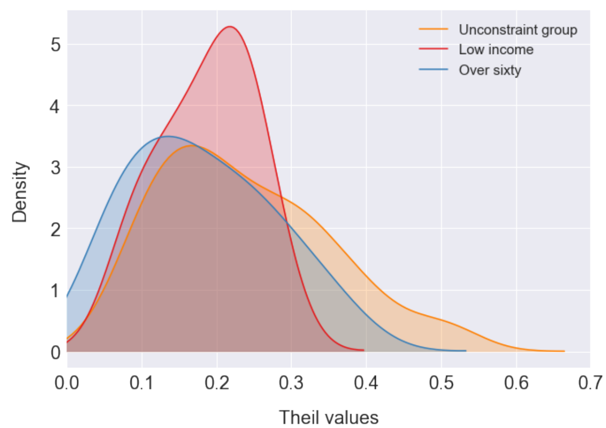


Figure 6: Density plots of Theil indices for the three population groups

Table 4: Desciptive statistics and one way ANOVA for the Theil indices

group	**count**	**mean**	**std.**	**75 perc.**
Unconstrained	181	0.18	0.11	0.31
Over sixty	30	0.12	0.09	0.25
Low income	13	0.21	0.06	0.23

(a) Descriptive statistics

	sum. sq.	**df**	**F**	**p-value**
group	0.105	2.0	4.42	0.013
Residual	2.62	221.0	NA	NA

(b) One way ANOVA

The distribution of Theil indices for the > 60 and unconstrained population groups are similar, however, the tail of the unconstrained population group is considerably longer compared to both remaining groups. Under closer examination, these outliers (> 75% percentile, $T > 0.5$) are characterised by high *Employment* probabilities at the expense of the rest of activity types. The demographic status of these individuals is composed of a mix of ethnicities, while the place of residence is outer London in most cases. All of the individuals in this sample are full time permanent employed in central London, with activity patterns being limited to <5 unique locations.

The outliers (> 75% percentile $T > 0.23$) of the low income group on the other hand, are characterised by individuals that are part-time workers and students, again residing in outer London. Their ethnic background is a mix of Asian/Arab/Black or Black British - African and Black or Black British - Caribbean with age spanning from 21 to 38 years old. The household characteristics are lone parents or couples with children. Compared to the outliers of the unconstrained group, the number of unique locations visited is greater. However, the mean distance between these locations (9.5km) is much smaller compared to the unconstrained group (20.4km). This pattern could be explained by the relatively high rates of travelling by bus and provides evidence of a reduced space where activities can take place compared to the unconstrained group. In the absence of access to individual preference mechanisms, it is difficult to make assertions as to whether this pattern is due to genuine individual choices or whether is related to higher risk of social exclusion. However, given that in London the price of a single bus journey is nearly half the price of rail and taking into consideration the sociodemographic profile of these individuals, it is likely that the observed pattern is due to necessity.

Finally, looking at the demographic characteristics of the outliers of the > 60 population group (> 75% percentile, $T > 0.24$), the ethnic backgrounds are mainly White - English/Welsh/Scottish/Northern Irish/Other White residing in outer Greater London with place of employment in Greater London area, inside the M25 motorway. All of the individuals in this percentile were employed full time with annual income ranging between £25,000-100,000. Similarly to the unconstrained population group, these individuals have high *Employment* activity type probabilities (~ 0.5). However, the probabilities for the rest of the activity types appear to be more balanced. This population subgroup has the greatest number of unique visited locations compared to the unconstrained and low income groups. However, contrary to the low income group, the mean distance between these locations is slightly larger (10km), a fact which could be explained by the higher rate of travelling by rail for activity *Eating and Drinking*.

4.2 Distribution of transportation modes

The next posterior quantity of interest is the distribution of transportation modes for each individual in the population groups. This corresponds to the latent m node of the model 4 and relates to the mobility element of the Capabilities set. Similarly to the d node, the results for the unconstrained, low income and > 60 population groups are presented in Appendix B. The segmentation per activity type was made by taking the one with the highest probability from each individual activity distribution at visited location.

4.2.1 Posterior quantities

Compared to the unconstrained population group, for the low income group the probabilities of using the rail to reach activity type *Eating and Drinking* are significantly lower, with bus being the predominant transport mode for this activity type (Figures 15a, 15b,15c). The posterior probabilities for the *Education and Health* (Figure 15d,15e,15f) activity type are slightly higher for using the bus compared to rail services, and the same holds for the *Retail* (Figure 15g,15h,15i) activity type. In terms of the overall shape of the distributions, *Retail* seems to follow the trend observed with the unconstrained population group, coinciding with retail shops' most popular shopping times. Looking at the *Employment* (Figure 15m,15n,15o) activity type, the shape of posterior distributions for bus, rail and tram services is significantly wider throughout the day, compared to the unconstrained sample, characterised by two peaks, in the morning and early afternoon.

This pattern most likely reflects the varying schedule of part-time workers. Finally, for this population group, only one individual was attributed with reaching *Outdoors and Recreation* (Figure 15j, 15k,15l) activity type with higher probability of using the rail services.

For the > 60 population group, the probabilities of using the bus and using rail to reach activity type *Eating and Drinking* seem to complement each other, with the probabilities of bus use being higher in the morning/afternoon and rail being higher in the afternoon/evening hours (Figure 16a, 16b). Overall, using the bus versus rail is similar for *Education and Health*, with rail services appearing to have a slightly shifted distribution mode toward the afternoon hours (Figure 16d, 16e). Using the different transportation modes to reach *Retail* (Figure 16g, 16h, 16i) appears to be similar with the rest of population groups. However, the distribution of transport modes used throughout the day appears to be wider for a significant number of individuals. Moreover, compared to the rest of focus groups, more people are found to be using the bus to reach *Outdoors and Recreation* activities.

Figures 7 and 8 below show aggregated boxplots of the transportation modes posterior distributions for all in individuals in the three population groups, categorised by weekdays and weekends. As it can be seen, the overall use of public transport for activity *Employment* is generally lower in the weekends for the unconstrained and > 60 group, particularly for using the bus. For the low income group, using rail for reaching activity *Eating and Drinking* is lower in the weekends compared to weekdays. On the other hand, weekdays dominate the use of public transport to reach *Education and Health* for the > 60 population group.

4.2.2 Assessing equality levels

The equality assumption made here is similar to potential accessibility: all transportation modes should be equally available regardless of any personal or place based characteristics. It is important to note that this assumption is useful only in the context of benchmarking the individual Theil values, as it is well known that the public transportation network is designed so that each mode complements the other. Moreover, as in the case

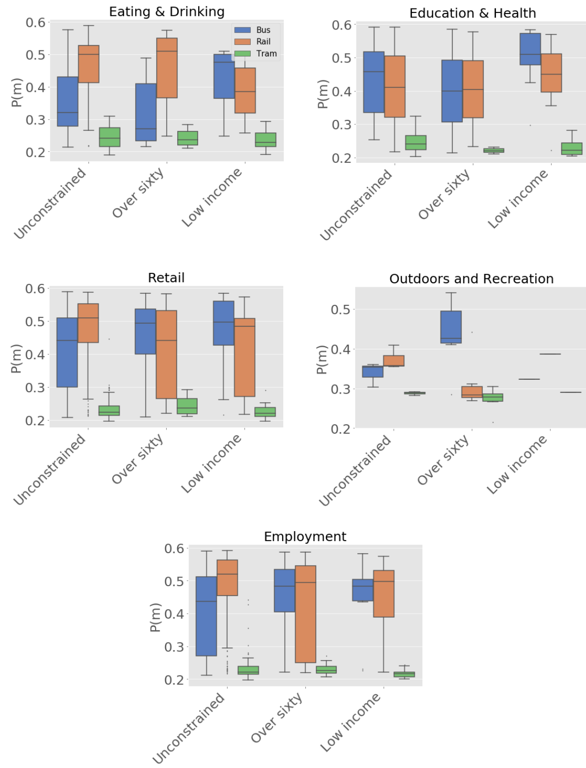


Figure 7: Aggregated transportation mode boxplots (weekdays)

of *Tram* services in London, some transportation modes are operate on a local scale only, so by default are not readily available to the general population. Nevertheless, by evaluating the individual Theil indices in a relative way, it is possible to identify cases where the use of a transport mode is not possible due to factors beyond the control of an individual (such as their sociodemographic status), a fact which could relate to transport disadvantage.

Figure 9 below shows density plots of Theil indices for the posterior transportation mode distributions for each population group:

Similar to Section 3.3.1, a one way ANOVA test was performed which resulted in failure to reject the null hypothesis, concluding that the distributions belong to the same population (Table 5). However, this result could be an artifact of the lower cardinality of the transportation mode set, particularly considering the very low use of *Tram* services, resulting in small differences in Theil values.

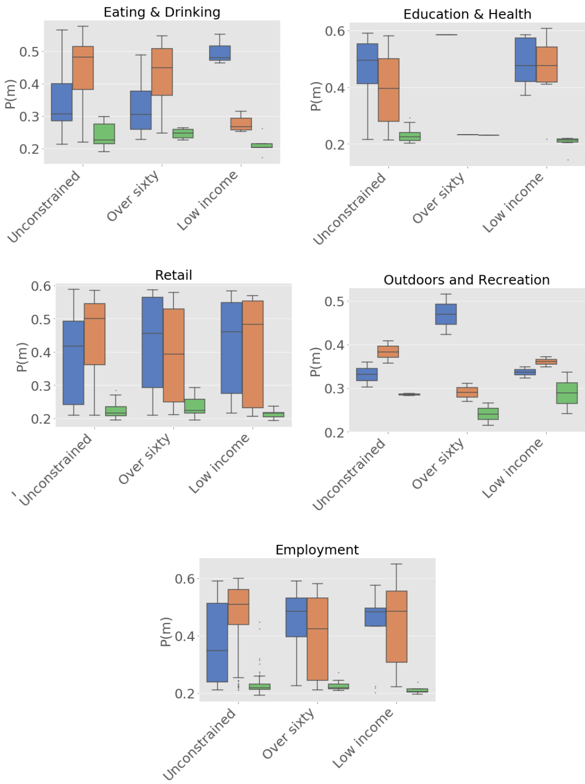


Figure 8: Aggregated transportation mode box-plots (weekends)

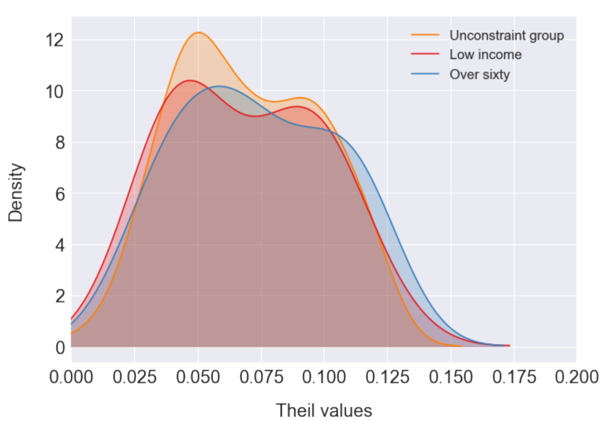


Figure 9: Density plots of mobility Theil indices for the three population groups

Exploring the Theil values distributions qualitatively, one notices a bimodality in all three population groups, meaning that, for those individuals, the use of one transportation mode dominates over

Table 5: Descriptive statistics and one way ANOVA for the Theil indices of transportation modes

group	count	mean	std.	75 perc.
Unconstrained	181	0.07	0.03	0.1
Over sixty	30	0.073	0.029	0.11
Low income	13	0.06	0.02	0.09

(a) Descriptive statistics

	sum. sq.	df	F	p-value
group	0.00038	2.0	0.23	0.79
Residual	0.18	221.0	NA	NA

(b) One way ANOVA

all others. It is interesting to observe that, in contrast to the unconstrained population group, the second mode of the low income group is attributed to very high probabilities of *Bus* use. Examining the outliers (> 75% percentile, $T > 0.1$) of the low income distribution, one notices that the majority of individuals in this set are a subset of the low income outliers of Section 3.3.1. This fact provides further evidence of the potential for social exclusion for these individuals.

4.3 Activity and mobility dynamics

In this section, the posterior results of transition matrices T_m and T_z are presented. Intuitively, these matrices capture the transition dynamics for the accessibility and mobility modules of model shown in Figure 4 taking into consideration the effects of external factors as individuals transition from one transportation mode/activity to another during the trajectory. It is important to note that, contrary to T_m where the transportation mode states are inferred using the observed Oyster card modes, T_z captures the transition dynamics of inferred activity types. The results for the unconstrained, low income and > 60 population groups are presented per activity type in appendix C.

4.3.1 Activity type transitions posterior results

Figure 17 in appendix C presents the posterior distributions for each element of the activity type transition matrix T_z for all individuals in the unconstrained population sample. As it can be seen, the transition patterns from activity type *Employment* to all other activities vary significantly between individuals, ranging from $0.2 < P(T_z) < 0.8$, with an overall mean probability of ≈ 0.4 for the transition from all other types to *Employment* making this the dominant sequence in this group. Looking at the transition between *Education and Health* and *Education and Health*, individuals seem to be divided into two clusters, one with relatively low probability $P(T_{11}) < 0.2$ and one with probabilities $P(T_{11}) > 0.2$, a behaviour which could be attributed to the students/pupils in the sample. Relatively high probabilities for many individuals are also observed between transitions *Retail/Retail*, *Eating and Drinking/Retail*.

Results for the low income population group are shown in Figure 18. The transition patterns are very similar with the unconstrained population group, however in this case, the probabilities of transitioning from *Employment* to all other activity types is lower on sample population level $P(T_z) < 0.2$.

Finally, Figure 19 presents the results for the over sixty years old population group. Again, the results here are very similar to the rest of the target groups, with the dominant transition sequences being between *Employment* and the rest of activity types.

4.3.2 Transportation mode transitions posterior results

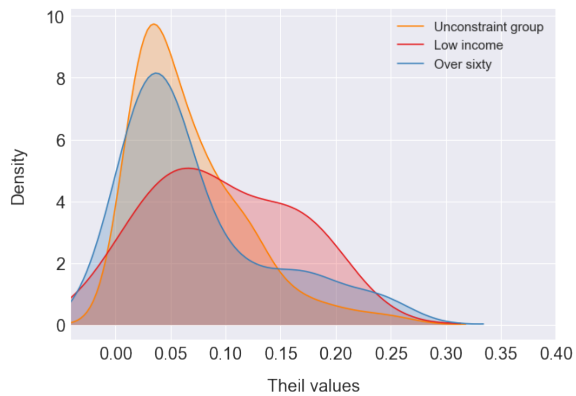
Looking at the unconstrained population group (Figure 20), there is a clear tendency to persistently transition from rail services to rail services ($T_{m_{11}}$) with a population level probability of $P(T_{m_{11}}) \approx 0.65$, a behaviour which could largely be attributed to commuting to employment activities. The transition probabilities from bus to rail are also relatively high in the sample ($P(T_{m_{01}}) \approx 0.45$), in par with transition from bus to bus ($P(T_{m_{00}}) \approx 0.475$). The transition from rail to bus on the other hand ($P(T_{m_{10}}) \approx 0.305$) is relatively low comparatively, indicating that, on average, individuals of this sample seem not to prefer finishing their journey on the bus if rail was the prior choice. As expected, due to the lack of tram transactions in the sample but also due to the limited coverage of tram services, on average the transition probabilities between the rest of transportation modes and tram is relatively small. This is confirmed by the uniform allocation of probabilities between tram and the rest of modes. This pattern is similar to the over sixty and low income population groups.

Transition probability patterns for the over sixty population group (Figure 21) are different, providing evidence that, on average, there is increased likelihood of using the bus persistently throughout a trajectory ($P(Tm_{00}) \approx 0.73$). The inverse is true for transitioning from bus to rail ($P(Tm_{01}) \approx 0.24$). Transitioning from rail to all other modes appear to be less clustered ($P(Tm_{11}) \approx 0.40, P(Tm_{10}) \approx 0.45$), indicating perhaps the less frequent use of rail services in this target group.

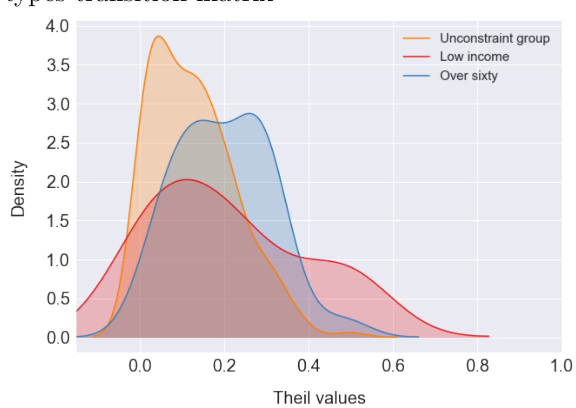
Finally, the low income population group (Figure 22) provides evidence of a broader transition probabilities spread amongst individuals for the bus services, with a tendency to prefer using the bus throughout the trajectory ($P(Tm_{00}) \approx 0.55$) compared to transitioning from bus to rail ($P(Tm_{01}) \approx 0.41$). The overall pattern of rail use is similar to the over sixty population group, showing a uniform distribution of transitions between rail/bus and rail/rail ($P(Tm_{10}) \approx 0.46, P(Tm_{11}) \approx 0.43$).

4.3.3 Assessing equality levels

This element of the capabilities set is aiming to quantify the dynamic component between different activity types/transportation modes through the use of external factor informed transition matrices. Regarding activity types, the underlying assumption being made is that an individual is less likely to be socially excluded if they maintain a uniform level of interaction with the available activities, as this translates to more frequent trips per activity type which is thought to map to increased levels of social involvement (Schönfelder & Axhausen 2003). A similar rationale holds for the interaction with public transport modes as expressed through the mobility transition matrix, in that increased levels of transition between modes could translate to an expansion of the set of activities within reach. Figures 10a and 10b below show the distribution of Theil values for T_z, T_m for the three population groups.



(a) Density plots of Theil indices for the activity types transition matrix



(b) Density plots of Theil indices for the mobility transition matrix

Figure 10: Density plots of Theil indices for the activity types transition matrix.

The ANOVA test for the three population groups (Table 6) failed to reject the null hypothesis (same distributions), however for the low income group there are some outliers that seem to have increased Theil values.

Similarly to 3.3.1, the employment status of outliers ($> 75\%$ percentile, $T > 0.2$) of the low income group are a mixture of student, part-time and full time workers, residing in outer Greater London. Not surprisingly, these individuals are characterised by increased probabilities of transitions related to *Health and Education* (for the student and part-time employed individual) and increased

Table 6: Descriptive statistics and one way ANOVA for the Theil indices of activity transition matrices

group	**count**	**mean**	**std.**	**75 perc.**
Unconstrained	181	0.06	0.05	0.094
Over sixty	30	0.072	0.066	0.10
Low income	13	0.097	0.06	0.15

(a) Descriptive statistics

	sum. sq.	**df**	**F**	**p-value**
group	0.007	2.0	1.21	0.298
Residual	0.55	221.0	NA	NA

(b) One way ANOVA

transition probabilities related to *Employment* for the full time workers. The latter is the same for nearly all outliers of the unconstrained and over sixty population group.

The ANOVA test for the distribution of individual transportation mode transition (Table 7) matrices has rejected the null hypothesis ($F-value = 8.908, p = 0.0002$) stating that the Theil distributions are different. Looking at the mean Theil values for all three population groups in Figure 10b, the over sixty and low income groups seem to have similar inequality levels ($\bar{T} = 0.21$ for low income, $\bar{T} = 0.20$ for over sixty). However, for the individuals with Theil values belonging to the tails of the distribution, the levels of inequality seem to be particularly high. The qualitative profile of those individuals is similar to the ones of activity transition matrices (mixture of employment statuses and residing in Outer London) with high transition probabilities of using a particular mode (*Bus* or *Rail*).

5 Discussion

The link between social exclusion and transport disadvantage is a complex one that has been approached in different ways in the literature. Having as a starting point the Capabilities Approach,

Table 7: Descriptive statistics and one way ANOVA for the Theil indices of transportation mode transition matrices

group	**count**	**mean**	**std.**	**75 perc.**
Unconstrained	181	0.12	0.09	0.19
Over sixty	30	0.20	0.10	0.27
Low income	13	0.21	0.17	0.29

(a) Descriptive statistics

	sum. sq.	**df**	**F**	**p-value**
group	0.17	2.0	8.908	0.0002
Residual	2.22	221.0	NA	NA

(b) One way ANOVA

this study proposed an accessibility framework to quantify the links between limited access to opportunities and reduced access to public transport using unlabelled mobility data.

The general trend for distribution of activity types throughout the day was found to be similar in the three focus population groups, a fact which is not surprising given that activity types were imputed, and not observed, from secondary data. However, including assumptions about the nature of the sociodemographic background of individuals allowed shaping of the head and tails of these distributions, particularly for the *Employment* activity type.

The low income population group was found to have smaller probabilities of activity type *Eating and Drinking*, which could be linked to the reduction of the capability to use public transport to reaching entertainment related activities. Although it is hard to make assertions due to the small sample size, it is nevertheless interesting to observe that the activity type that has the lowest probability range is the one that is the most elastic, compared to *Employment* or *Education and Health* for example. On the other hand, the probabilities of activity type *Education and Health* are higher for the low income group. Considering that a large number of individuals in this population group are students, this fact comes as no surprise.

For all population groups, activity type *Outdoors and Recreation* was found to have the lowest probability compared to the rest of the activity types. Besides the limited number of POIs in this category type, the prior assumptions of participating in this activity type were weak relative to the rest of activity types (the mean value was ≈ 0.13 for all individuals in the Oyster sample). This finding, together with the fact that *Outdoors and Recreation* had the highest accuracy of prediction (following *Employment*) makes the assertion of absence of such activity types from the dataset plausible, as opposed to being an artefact of the modelling process.

A further breakdown of results can be made by examining the distributions of transportation mode use in relation to each activity type. For the unconstrained population group, this reveals increased diversity in the probability distributions among individuals on the relative use of public transport, particularly for *Eating and Drinking* and *Education and Health* activity types. This could be an indication of the varying capability levels experienced by different people when reaching these activities. Overall, for this population group, the probability of using *Rail* services to reach the different activity types is higher compared to the rest of the modes. The exception is *Education and Health* where *Bus* seems to be considerably higher (≈ 0.42 for *Bus* and ≈ 0.36 for *Rail*). Assuming that these activities predominantly map to individuals that are either students or people that reach for medical care, decreased levels of *Rail* use compared to *Bus* provides evidence of reduced capability of using the *Rail* services by those individuals.

The use of *Bus* services is also the predominant mode of transport for the low income group for all activity types, a finding that is in line with existing evidence (Transport for London 2011). This should hardly come as a surprise, as cost is a significant barrier to transport in London, particularly for rail services. Other reasons mentioned in the literature for increased bus use are the possession of bus/rail cards for the low income groups. However, none of the individuals in the group reported possessing one in the LTDS. The predominant use of bus for people in this group could also be the main reason for the observed geographical pattern, which is characterised by a tendency to avoid inner London. This fact, combined with reduced participation in activities as determined by the increased Theil index for this group, provides evidence of transport disadvantage compared to the rest of the groups.

Furthermore, the distribution of transportation mode use for the *Employment* activity type has been found to have a distinct multimodal shape throughout the day, characterised by relatively high probabilities in the morning and afternoon. This pattern could be explained by the mixture of employment statuses of the individuals in this group: full-time employed, part-time employed and students.

Compared to the rest of the groups, individuals in the > 60 sample have wider distributions of public transport use throughout the day, spanning a temporal window between morning and early evening. It is difficult to interpret this shape, as from a data driven perspective, this group had the least number of transactions on average (≈ 30 transactions per individual) compared to the rest of the groups (≈ 32 for low income and ≈ 38 for the unconstrained group) which contributes to increased uncertainty of estimates. This fact coincides with evidence of non-travel (people who do not make trips) for this population group (Transport for London 2011) in London. From this perspective, it is difficult to make assertions of increased capability of using the public transport network throughout the day compared to the unconstrained population group.

The general pattern of transition probabilities between activity types was found similar in the three population groups. One notable exception was the relatively low transition probabilities of *Employment* for the low income group. Empirically, this pattern could be attributed to the nature of working status of the individuals in the sample, half of which were students, 25% full-time and 25% part-time employed. Moreover, this population group was found to have a larger distribution mean, providing evidence of less uniform transitions between activities. Indeed examining the outliers of the Theil distribution, one notices increased transition probabilities for either education or employment related activities, depending on the individual's employment status.

Looking at the results of the mobility transition matrix, a number of interesting transportation habits are revealed. For a significant number of individuals belonging in the unconstrained population group, the use of *Rail* services seem to be persisting throughout their trajectory, characterised by high *Rail/Rail* and low *Rail/Bus* transition probabilities. On the other hand, the inverse seems to be true for the over sixty and low income population groups, with high *Bus/Bus* and low *Bus/Rail* probabilities. This transition pattern is less uniform for the low income group, judging from the increased Theil values. Although it is difficult to make assumptions on the drivers behind this modal split pattern, it seems that, besides factors commonly mentioned in the literature such as egress and waiting time (**?**), sociodemographic factors (such as employment status and income) also play a role in the modal split habits of individuals. In each case, the lack of sensitivity in switching between different public transport modes could be regarded as a reduced capability of using the public transportation network that can result to transport disadvantage.

In terms of individual levels of equality as determined using the Theil index, the distribution of individual Theil values for the low income group is characterised by a statistically significant larger mean compared to the rest of the population groups. This provides evidence of increased levels of inequality experienced by this group, as the range of activities that are being reached is narrower.

Examining the sociodemographic variables of outliers of this distribution (individuals belonging over the 75% of the Thiel distribution, a total of 10 individual) using the matched Oyster card and LTDS IDs, further probing of the personal characteristics contributing to exclusion from activities and access to transportation modes can be deduced: 90% belong to black, Asian and minority ethnic backgrounds, 70% are women, all of them report income earned below £15,000 and all of them reside in outer Greater London. Moreover, the labour profile for these individuals is more unstable, with 8/10 people being either part-time employed or students.

In contrast, the sociodemographic profile of the Thiel outliers of the unconstrained population group (16 individuals) consisted of 44% belonging to to black, Asian and minority ethnic backgrounds, 62% women, all of them earning £25,000 or more and 67% residing in outer Greater London areas with 15/16 being full time employed. The sociodemographic profile of the > 60 population group is similar to the unconstrained group (11 individuals) with 63% women, all of them earning £25,000 or more and 80% residing in outer Greater London areas and 9/11 full time employed. The ethnic background however of the outliers in this group is different with 10% belonging to black, Asian and minority ethnic backgrounds and the rest are White/British white. Judging from the above profiles, it is clear that the low income group is characterised by most of the risk factors that could result in social exclusion.

6 Conclusions

This study proposed a novel approach to evaluating individual accessibility by framing the modelling methodology through the CA. Following the main concepts of the CA and the way they are related, the different components that shape an individual's ability to reach opportunities are explicitly modelled in a probabilistic way through the notions of latent capabilities and observed functionings. The potential of the proposed methodological framework to evaluate individual based transport based social exclusion was assessed through a case study using London's transport data. It was found that the proposed framework could identify individuals that exhibit high risk of social exclusion by comparing the distributions of the capabilities sets. Limitations of this approach can be summarised to the nature of passively generated mobility data. Since such data were generated by the service provider for reasons other than the one used in this study, any generalisations to population level characteristics should be made keeping this important consideration in mind. For example, population groups that are thought to present high risk of transport related social exclusion such as the unemployed, disabled and retired were not represented in the sample. It would be of great value if the analysis was repeated with these groups, as it would demonstrate the degree of robustness of the proposed methodological framework.

Related to the above, it is important to note that while this study represented an individual's potential accessibility to activities and potential mobility using public transport through the Bayesian network structure, access to an individual's actual "wants" and "desires" behind their choices remains out of reach and can only be uncovered through extensive qualitative studies. For example, people with lower income may choose to eat and drink out less due to lack of resources. Transport accessibility may be a factor, but its effect might be exaggerated by the lack of access to the drivers behind the choices made by those individuals. Nevertheless, the current structure of the proposed model could be used to identify deviations from the average equality levels so that further investigation can be undertaken. Future direction will be steered towards a qualitative validation of the findings.

Furthermore, the varying sample size of Oyster card data for the different population groups has an impact on the geographic representativeness of the study area. While for the unconstrained and > 60 sample the visited locations appear to be uniform throughout the study area, the visited locations of low income group appear to span radially from outer to inner London (see Figure 3). Although a positive correlation exists between the index of deprivation and the visiting locations of the low income group (OLS slope 0.013 compared to a negative correlation for the unconstrained group with OLS slope -0.044 and the > 60 group with OLS slope -0.001) which intuitively is what would one expect, it is difficult to make any firm assertions regarding the geographic representativeness of this population group.

With regards to the temporal extent of the study, the available dataset did not allow any deeper evaluation on the way individuals adjust their activity/travel behaviour in the face of an event that could impact accessibility. Such an event can be related to personal characteristics (such as a change in employment status) or can be infrastructure related (for example, an introduction of a new public transport connection. A future direction could involve using an extended time span together with information on significant events to assess whether an adaption of behaviour is represented in the evolution of mobility/accessibility nodes of the model.

In light of the ever increasing trend of urbanisation, accessibility is likely to be a major problem for future cities, as current infrastructure will be stressed to accommodate the needs of an increasing urban population. With the levels of inequality in transport likely to increase as a result of competition for resources, policy makers will need more information on the causes of transport related social exclusion. To that extend, new technologies combined with big data that provide interpretable results could provide evidence to promote equity.

References

Alkire, S. (2008), 'Using the capability approach: Prospective and evaluative analyses", *The Capability Approach: Concepts, Measures and Applications, Cambridge University Press, Cambridge* pp. 26–50.

Anda, C., Erath, A. & Fourie, P. J. (2017), 'Transport modelling in the age of big data', *International Journal of Urban Sciences* **21**(sup1), 19–42.

Bagchi, M. & White, P. R. (2005), 'The potential of public transport smart card data', *Transport Policy* **12**(5), 464–474.

Bantis, T. & Haworth, J. (2017), 'Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics', *Transportation Research Part C: Emerging Technologies* **80**, 286–309.

Bantis, T. & Haworth, J. (2019), 'Non-employment activity type imputation from points of interest and mobility data at an individual level: How accurate can we get?', *ISPRS International Journal of Geo-Information* **8**(12), 560.

Ben-Akiva, M. (1979), 'Disaggregate travel and mobility choice models and measures of accessibility', *Behavioural travel modelling* .

Beyazit, E. (2011), 'Evaluating social justice in transport: lessons to be learned from the capability approach', *Transport reviews* **31**(1), 117–134.

Burns, L. D. (1980), 'Transportation, temporal, and spatial components of accessibility'.

Casas, I. (2007), 'Social exclusion and the disabled: An accessibility approach', *The Professional Geographer* **59**(4), 463–477.

Chikaraishi, M. (2017), Mobility of the elderly, *in* 'Life-oriented behavioral research for urban policy', Springer, pp. 267–291.

Chikaraishi, M., Jana, A., Bardhan, R., Varghese, V. & Fujiwara, A. (2017), 'A framework to analyze capability and travel in formal and informal urban settings: a case from mumbai', *Journal of transport geography* **65**, 101–110.

Church, A., Frost, M. & Sullivan, K. (2000), 'Transport and social exclusion in london', *Transport Policy* **7**(3), 195–205.

Cohen, J. E. & Kempermann, J. (1998), *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population*, Springer Science & Business Media.

Currie, G., Richardson, T., Smyth, P., Vella-Brodrick, D., Hine, J., Lucas, K., Stanley, J., Morris, J., Kinnear, R. & Stanley, J. (2010), 'Investigating links between transport disadvantage, social exclusion and well-being in melbourne–updated results', *Research in transportation economics* **29**(1), 287–295.

Daziano, R. A., Miranda-Moreno, L. & Heydari, S. (2013), 'Computational bayesian statistics in transportation modeling: from road safety analysis to discrete choice', *Transport reviews* **33**(5), 570–592.

Delafontaine, M., Neutens, T., Schwanen, T. & Van de Weghe, N. (2011), 'The impact of opening hours on the equity of individual space–time accessibility', *Computers, environment and urban systems* **35**(4), 276–288.

Delbosc, A. & Currie, G. (2011), 'The spatial context of transport disadvantage, social exclusion and well-being', *Journal of Transport Geography* **19**(6), 1130–1137.

Dodson, J., Buchanan, N., Gleeson, B. & Sipe, N. (2006), 'Investigating the social dimensions of transport disadvantage—i. towards new concepts and methods', *Urban Policy and Research* **24**(4), 433–453.

Evans, G. (2009), 'Accessibility, urban design and the whole journey environment', *Built environment* **35**(3), 366–385.

Farrington, J. & Farrington, C. (2005), 'Rural accessibility, social inclusion and social justice: towards conceptualisation', *Journal of Transport geography* **13**(1), 1–12.

Farrington, J. H. (2007), 'The new narrative of accessibility: its potential contribution to discourses in (transport) geography', *Journal of Transport Geography* **15**(5), 319–330.

Geurs, K. T. & Van Wee, B. (2004), 'Accessibility evaluation of land-use and transport strategies: review and research directions', *Journal of Transport geography* **12**(2), 127–140.

Golob, T. F. & McNally, M. G. (1997), 'A model of activity participation and travel interactions between household heads', *Transportation Research Part B: Methodological* **31**(3), 177–194.

Goodman, A., Jones, A., Roberts, H., Steinbach, R. & Green, J. (2014), "we can all just get on a bus and go': Rethinking independent mobility in the context of the universal provision of free bus travel to young londoners', *Mobilities* **9**(2), 275–293.

Hananel, R. & Berechman, J. (2016), 'Justice and transportation decision-making: The capabilities approach', *Transport Policy* **49**, 78–85.

Hansen, W. G. (1959), 'How accessibility shapes land use', *Journal of the American Institute of planners* **25**(2), 73–76.

Hickman, R., Cao, M., Mella Lira, B., Fillone, A. & Bienvenido Biona, J. (2017), 'Understanding capabilities, functionings and travel in high and low income neighbourhoods in manila', *Social Inclusion* **5**(4), 161–174.

Kamruzzaman, M., Yigitcanlar, T., Yang, J. & Mohamed, M. A. (2016), 'Measures of transport-related social exclusion: A critical review of the literature', *Sustainability* **8**(7), 696.

Kenyon, S. (2003), Understanding social exclusion and social inclusion, *in* 'Proceedings of the Institution of Civil Engineers-Municipal Engineer', Vol. 156, Thomas Telford Ltd, pp. 97–104.

Kwan, M.-P. (2013), 'Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility: Space–time integration in geography and giscience', *Annals of the Association of American Geographers* **103**(5), 1078–1086.

López, E., Gutiérrez, J. & Gómez, G. (2008), 'Measuring regional cohesion effects of large-scale transport infrastructure investments: an accessibility approach', *European Planning Studies* **16**(2), 277–301.

Lucas, K. (2012), 'Transport and social exclusion: Where are we now?', *Transport policy* **20**, 105–113.

Lucas, K. & Porter, G. (2016), 'Mobilities and livelihoods in urban development contexts: introduction.', *Journal of transport geography.* **55**, 129–131.

Maciel, V., Kuwahara, M., Fronzaglia, M., Scarano, P. & Muramatsu, R. (2015), Accessibility and well-being: Measuring urban (im)mobility as deprivation, *in* '2015 Human Development Capabilities Association Conference'.

Martens, K. & Golub, A. (2012), '11. a justice-theoretic exploration of accessibility measures', *Accessibility analysis and transport planning: Challenges for Europe and North America* **195**.

Miller, H. J. (2005), Place-based versus people-based accessibility, *in* 'Access to destinations', Emerald Group Publishing Limited, pp. 63–89.

Mitra, S. (2006), 'The capability approach and disability', *Journal of disability policy studies* **16**(4), 236–247.

Neutens, T., Schwanen, T., Witlox, F. & De Maeyer, P. (2010), 'Equity of urban service delivery: a comparison of different accessibility measures', *Environment and Planning A* **42**(7), 1613–1635.

Noble, M., McLennan, D., Wilkinson, K., Whitworth, A., Exley, S., Barnes, H., Dibben, C., McLennan, D. et al. (2007), 'The english indices of deprivation 2007'.

Nordbakke, S. (2013), 'Capabilities for mobility among urban older women: barriers, strategies and options', *Journal of transport geography* **26**, 166–174.

Nutley, S. (1998), 'Rural areas: the accessibility problem', *Modern Transport Geography, 2nd rev. ed., Wiley and sons, Chichester* pp. 185–215.

Ordnance Survey (2012), 'Points of interest classification scheme'.

Orr, S. (2010), Evaluation of transport accessibility for elderly and disabled people: a proposal for an activity-based quality of life approach, *in* 'European Transport Conference, 2010Association for European Transport'.

Páez, A., Gertes Mercado, R., Farber, S., Morency, C. & Roorda, M. (2010), 'Relative accessibility deprivation indicators for urban settings: definitions and application to food deserts in montreal', *Urban Studies* **47**(7), 1415–1438.

Pardo, L. (2005), *Statistical inference based on divergence measures*, CRC Press.

Pelletier, M.-P., Trépanier, M. & Morency, C. (2011), 'Smart card data use in public transit: A literature review', *Transportation Research Part C: Emerging Technologies* **19**(4), 557–568.

Pereira, R. H., Schwanen, T. & Banister, D. (2017), 'Distributive justice and equity in transportation', *Transport Reviews* **37**(2), 170–191.

Pirie, G. (1981), 'The possibility and potential of public policy on accessibility', *Transportation Research Part A: General* **15**(5), 377–381.

Preston, J. & Rajé, F. (2007), 'Accessibility, mobility and transport-related social exclusion', *Journal of Transport Geography* **15**(3), 151–160.

Pyrialakou, V. D., Gkritza, K. & Fricker, J. D. (2016), 'Accessibility, mobility, and realized travel behavior: Assessing transport disadvantage from a policy perspective', *Journal of transport geography* **51**, 252–269.

Rashid, K., Yigitcanlar, T. & Bunker, J. M. (2010), 'Minimising transport disadvantage to support knowledge city formation: applying the capability approach to select indicators', *Melbourne 2010 Knowledge Cities World Summit: 3rd Knowledge Cities World Summit.* .

Rawls, J. (2009), *A theory of justice*, Harvard university press.

Robeyns, I. (2005), 'Selecting capabilities for quality of life measurement', *Social indicators research* **74**(1), 191–215.

Ryan, J., Wretstrand, A. & Schmidt, S. M. (2015), 'Exploring public transport as an element of older persons' mobility: A capability approach perspective', *Journal of transport geography* **48**, 105–114.

Schönfelder, S. & Axhausen, K. W. (2003), 'Activity spaces: measures of social exclusion?', *Transport policy* **10**(4), 273–286.

Sen, A. (2014), 'Development as freedom (1999)', *Roberts, JT, Hite, AB & Chorev, N. The Globalization and Development Reader: Perspectives on Development and Global Change* **2**, 525–547.

Sen, A. et al. (1990), 'Development as capability expansion', *Human development and the international development strategy for the 1990s* **1**.

Smith, N., Hirsch, D. & Davis, A. (2012), 'Accessibility and capability: the minimum transport needs and costs of rural households', *Journal of Transport Geography* **21**, 93–101.

Social Exclusion Unit (2003), 'Making the connections: final report on transport and social exclusion', *http://webarchive. nationalarchives. gov. uk* .

Titheridge, H., Achuthan, K., Mackett, R. & Solomon, J. (2009), 'Assessing the extent of transport social exclusion among the elderly'.

Titheridge, H., Mackett, R., Christie, N., Oviedo Hernández, D. & Ye, R. (2014), 'Transport and poverty: a review of the evidence'.

Transport for London (2011), 'Travel in london,supplementary report:london travel demandsurvey (ltds)'. Online; Accessed 29/03/2016.
URL: *http://content.tfl.gov.uk/london-travel-demand-survey.pdf*

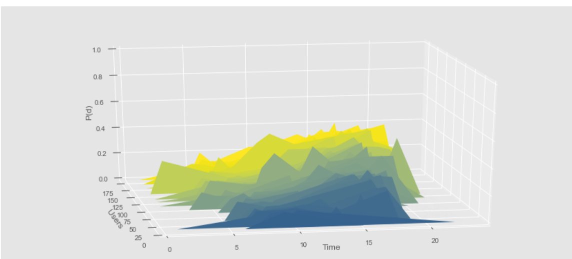
Tyler, N. (2006), 'Capabilities and radicalism: Engineering accessibility in the 21st century', *Transportation planning and technology* **29**(5), 331–358.

Van Wee, B. & Geurs, K. (2011), 'Discussing equity and social exclusion in accessibility evaluations', *EJTIR* **11**(4), 350–367.

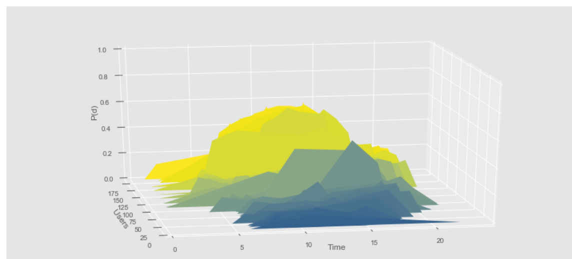
Wu, B. M. & Hine, J. P. (2003), 'A ptal approach to measuring changes in bus service accessibility', *Transport Policy* **10**(4), 307–320.

Yang, X. & Day, J. (2016), Operationalizing the capabilities approach for urban policy evaluation: The travel welfare impacts of government job resettlement, *in* 'Geography Research Forum', Vol. 35, pp. 113–137.

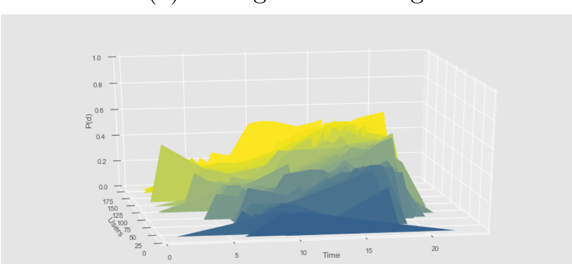
A Posterior activity distributions for the individuals of the unconstrained, > 60 and low income population groups



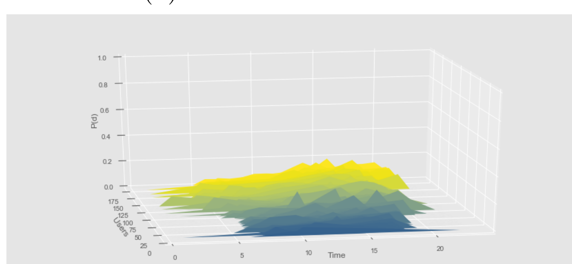
(a) Eating and drinking



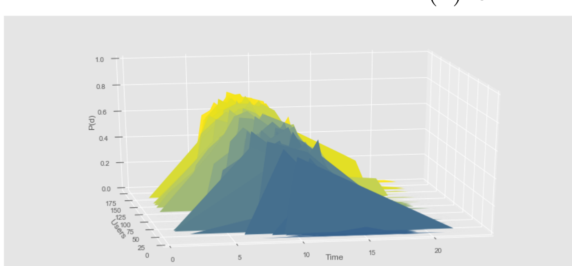
(b) Education and health



(c) Retail

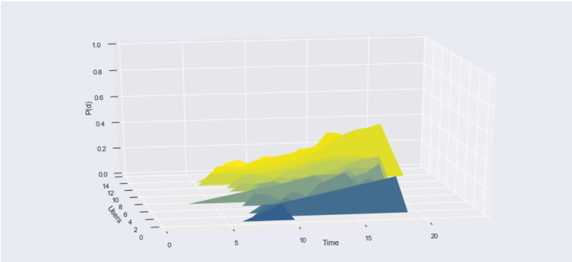


(d) Outdoors and Recreation

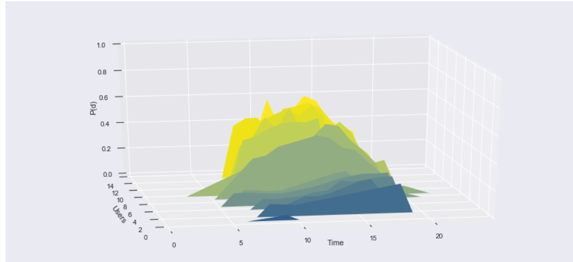


(e) Employment

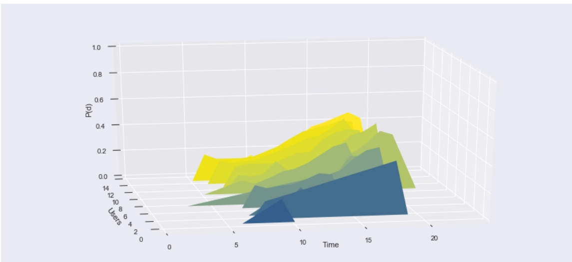
Figure 11: Posterior distributions of activity types for the unconstrained population sample



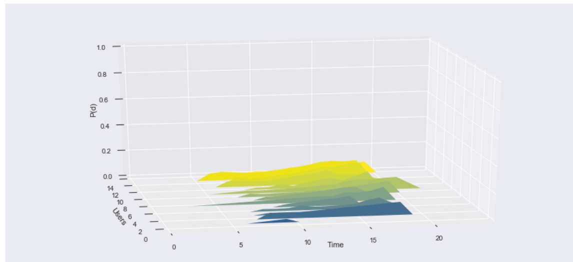
(a) Eating and drinking



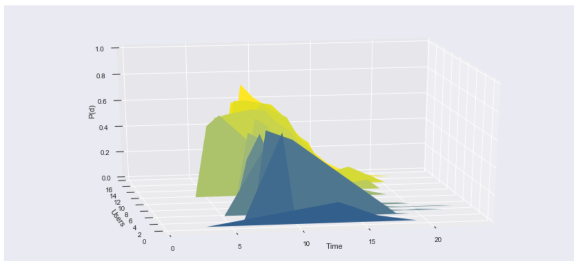
(b) Education and health



(c) Retail

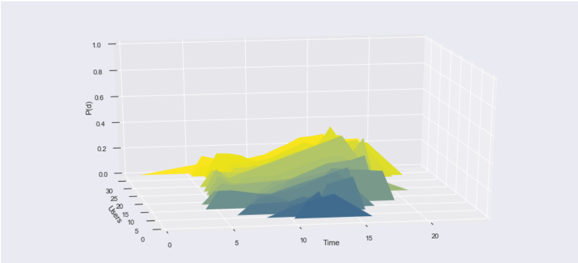


(d) Outdoors and Recreation

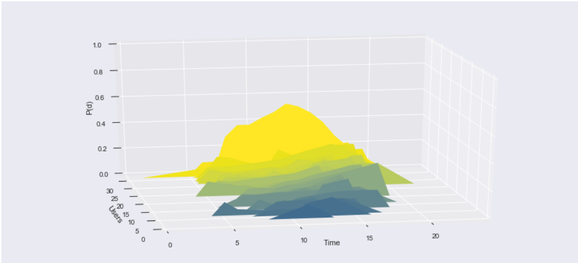


(e) Employment

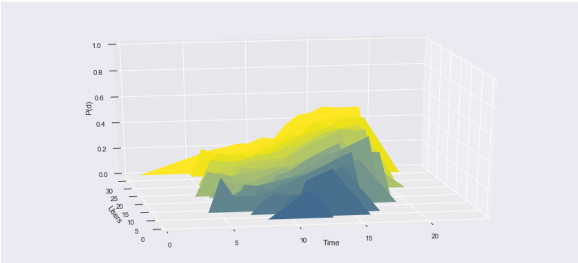
Figure 12: Posterior distributions of activity types for the low income population sample



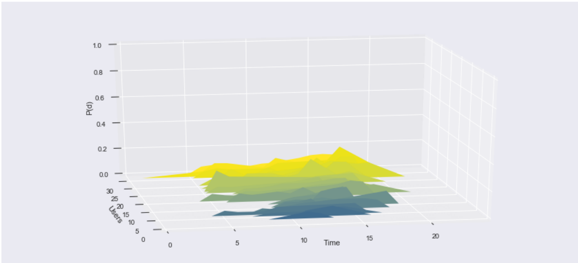
(a) Eating and drinking



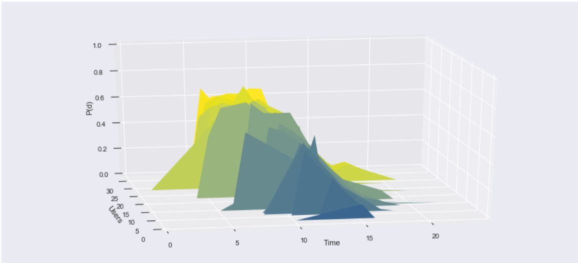
(b) Education and health



(c) Retail



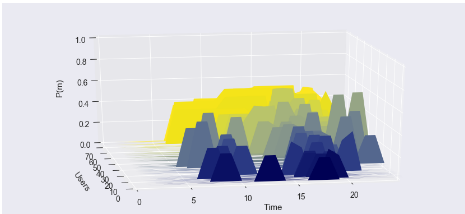
(d) Outdoors and Recreation



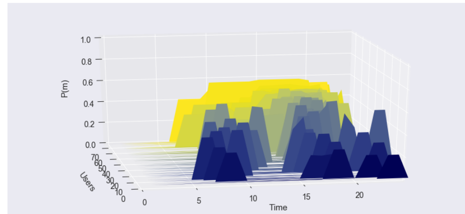
(e) Employment

Figure 13: Posterior distributions of activity types for the > 60 years old population sample

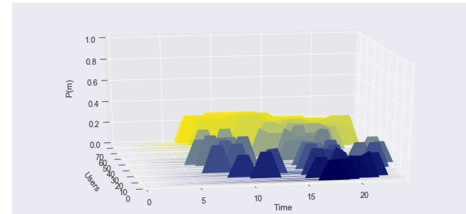
B Posterior transportation mode distributions for the individuals of the unconstrained, > 60 and low income population groups



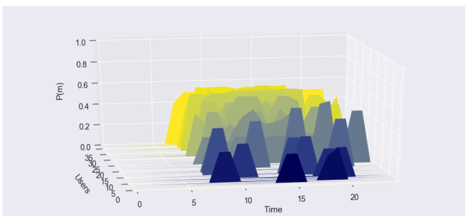
(a) Posterior means Bus/Eating and Drinking



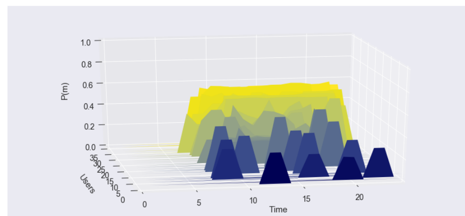
(b) Posterior means Rail/Eating and Drinking



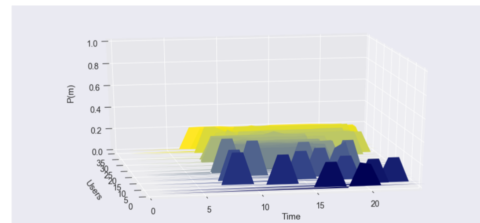
(c) Posterior means Tram/Eating and Drinking



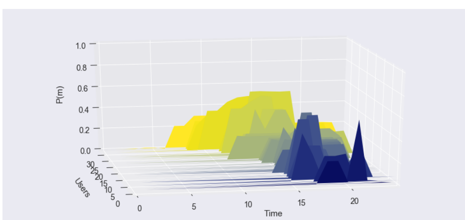
(d) Posterior means Bus/Education and Health



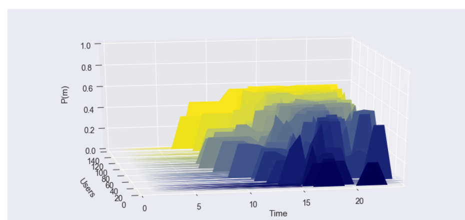
(e) Posterior means Rail/Education and Health



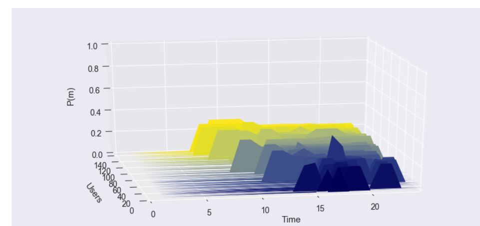
(f) Posterior means Tram/Education and Health



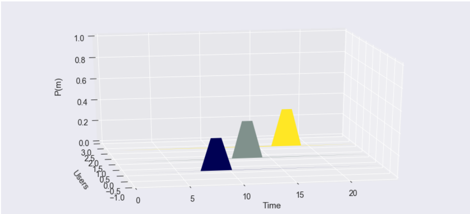
(g) Posterior means Bus/Retail



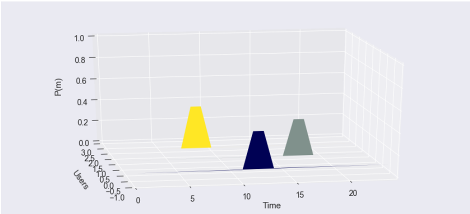
(h) Posterior means Rail/Retail



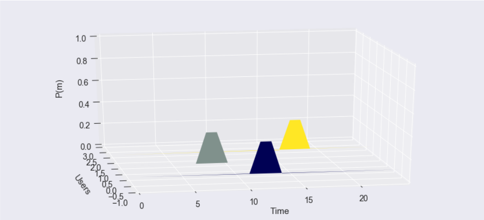
(i) Posterior means Tram/Retail



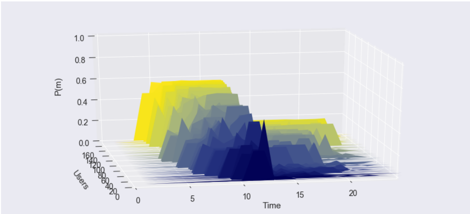
(j) Posterior means Bus/Outdoors and Recreation



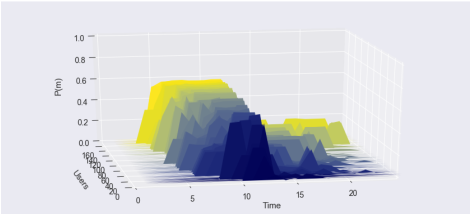
(k) Posterior means Rail/Outdoors and Recreation



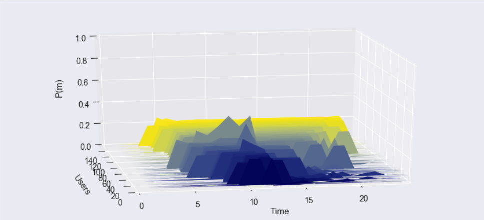
(l) Posterior means Tram/Outdoors and Recreation



(m) Posterior means Bus/Employment

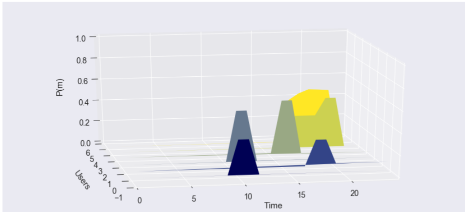


(n) Posterior means Rail/Employment

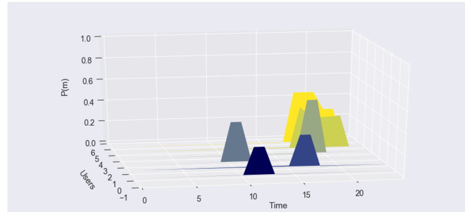


(o) Posterior means Tram/Employment

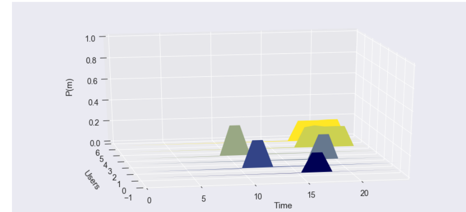
Figure 14: Posterior means for the unconstrained population group



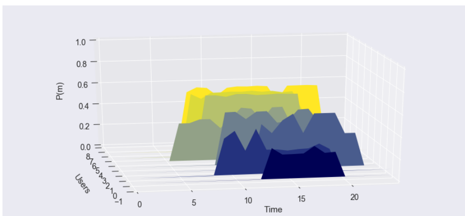
(a) Posterior means Bus/Eating and Drinking



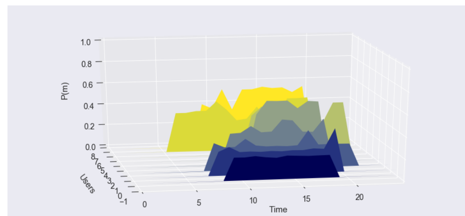
(b) Posterior means Rail/Eating and Drinking



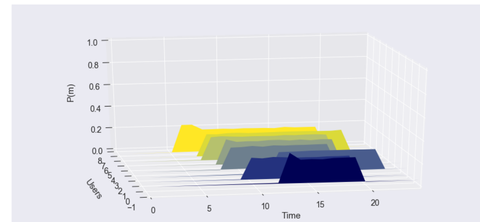
(c) Posterior means Tram/Eating and Drinking



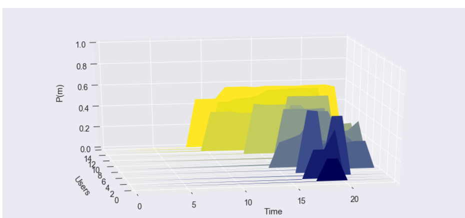
(d) Posterior means Bus/Education and Health



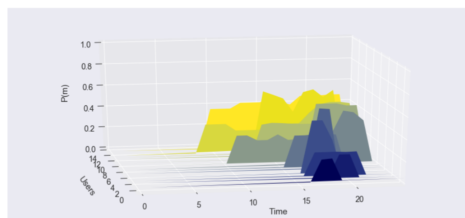
(e) Posterior means Rail/Education and Health



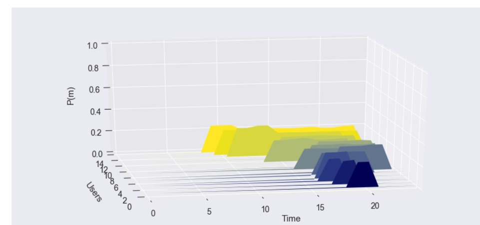
(f) Posterior means Tram/Education and Health



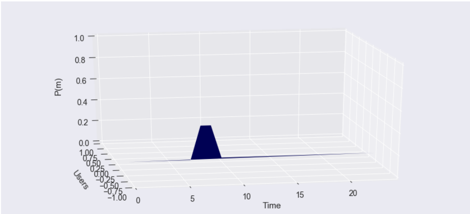
(g) Posterior means Bus/Retail



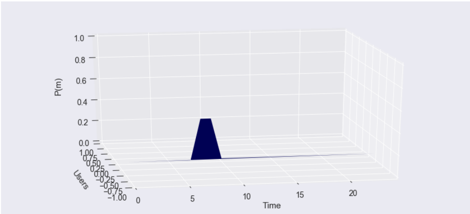
(h) Posterior means Rail/Retail



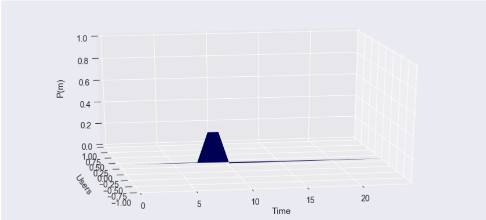
(i) Posterior means Tram/Retail



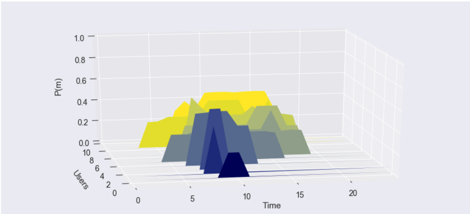
(j) Posterior means Bus/Outdoors and Recreation



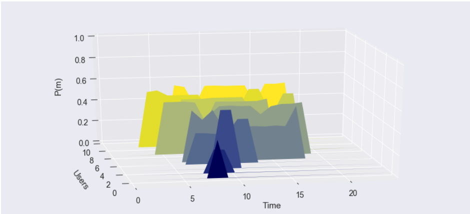
(k) Posterior means Rail/Outdoors and Recreation



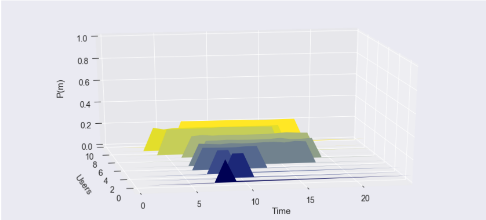
(l) Posterior means Tram/Outdoors and Recreation



(m) Posterior means Bus/Employment

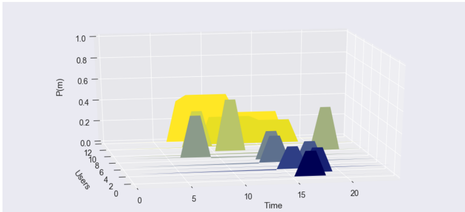


(n) Posterior means Rail/Employment

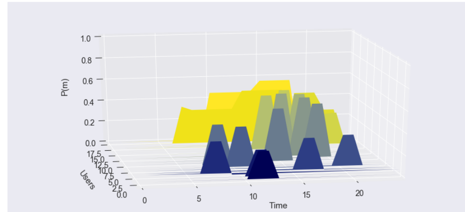


(o) Posterior means Tram/Employment

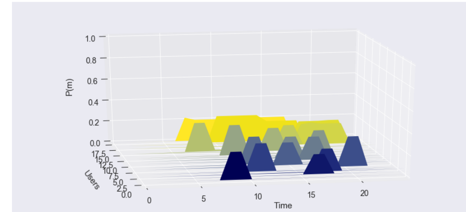
Figure 15: Posterior means for the low income population group



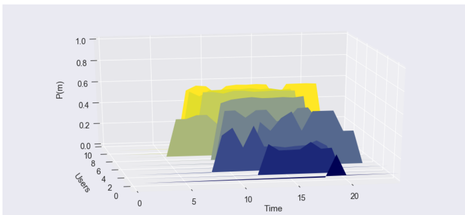
(a) Posterior means Bus/Eating and Drinking



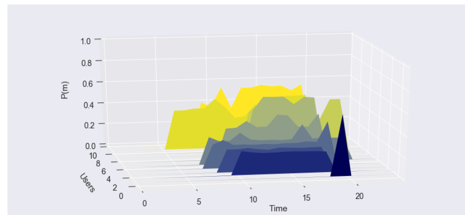
(b) Posterior means Rail/Eating and Drinking



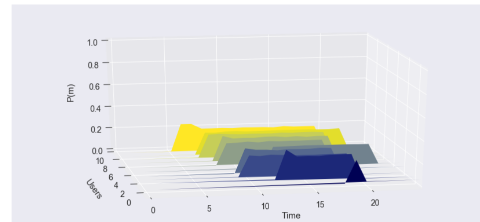
(c) Posterior means Tram/Eating and Drinking



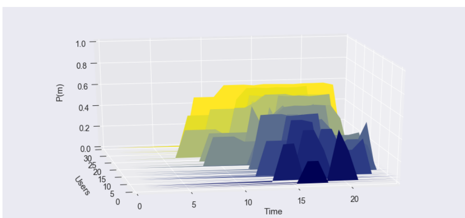
(d) Posterior means Bus/Education and Health



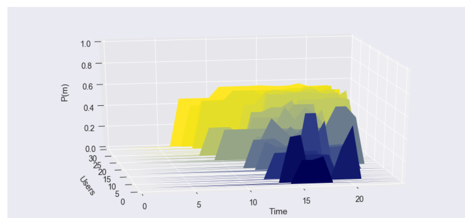
(e) Posterior means Rail/Education and Health



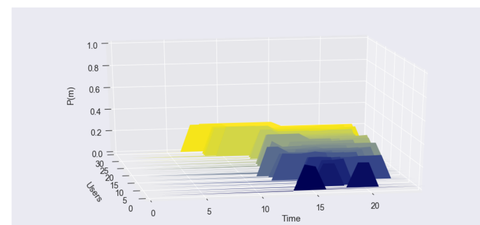
(f) Posterior means Tram/Education and Health



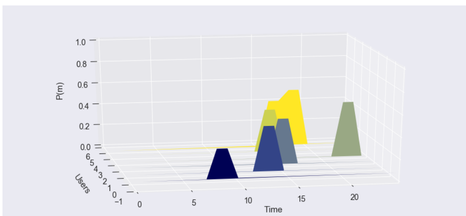
(g) Posterior means Bus/Retail



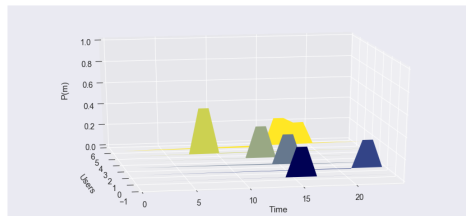
(h) Posterior means Rail/Retail



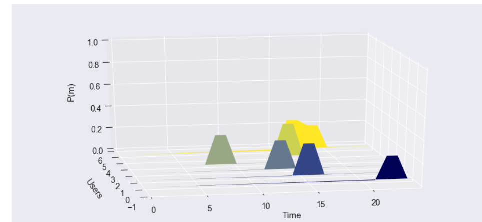
(i) Posterior means Tram/Retail



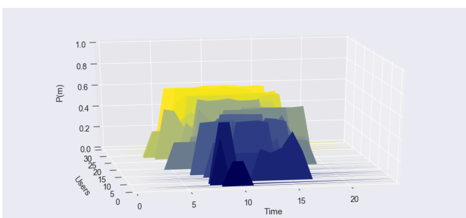
(j) Posterior means Bus/Outdoors and Recreation



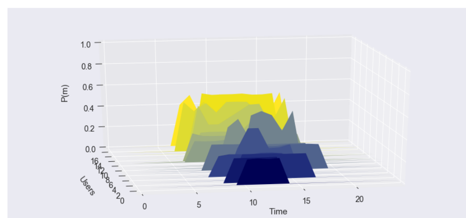
(k) Posterior means Rail/Outdoors and Recreation



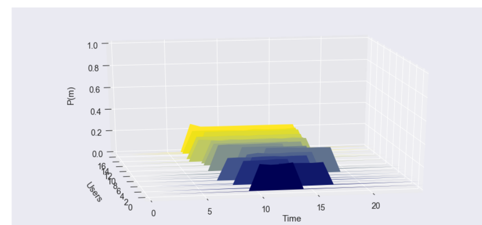
(l) Posterior means Tram/Outdoors and Recreation



(m) Posterior means Bus/Employment



(n) Posterior means Rail/Employment



(o) Posterior means Tram/Employment

Figure 16: Posterior means for the ¿ 60 population group

C Posterior activity and mobility dynamics distributions for the individuals of the unconstrained, over sixty and low income population groups

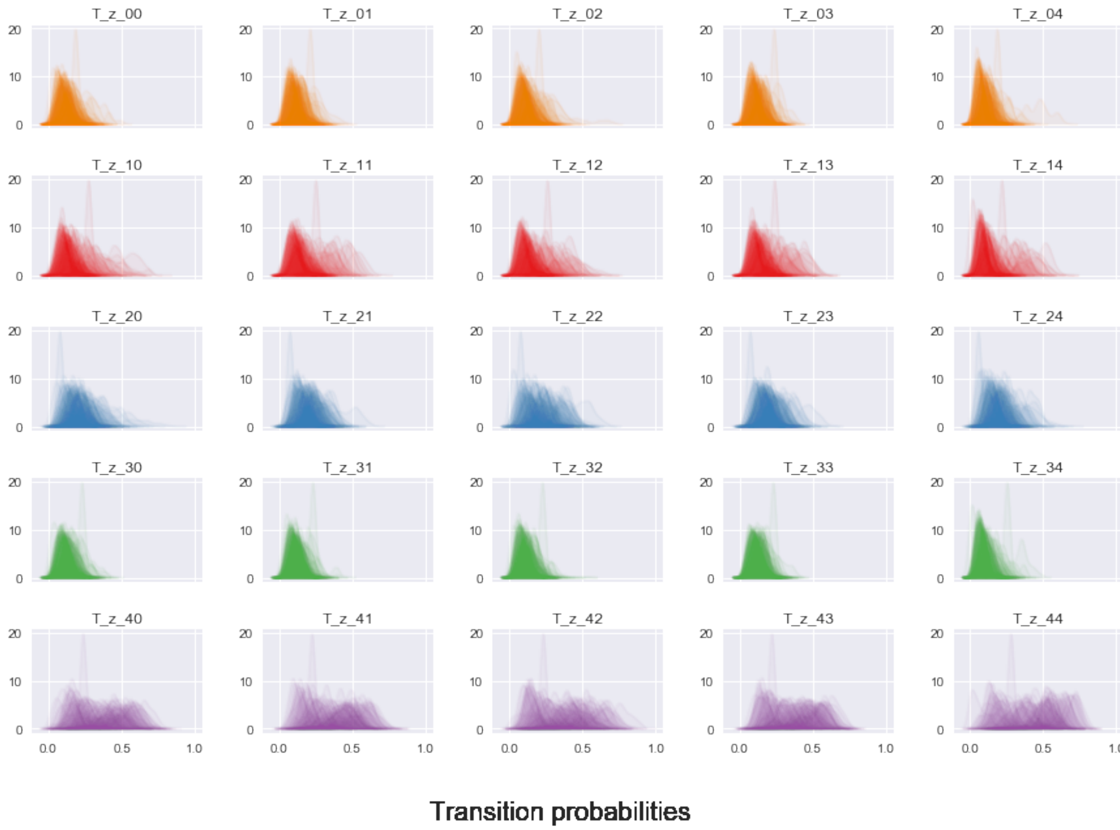


Figure 17: Posterior densities for T_z

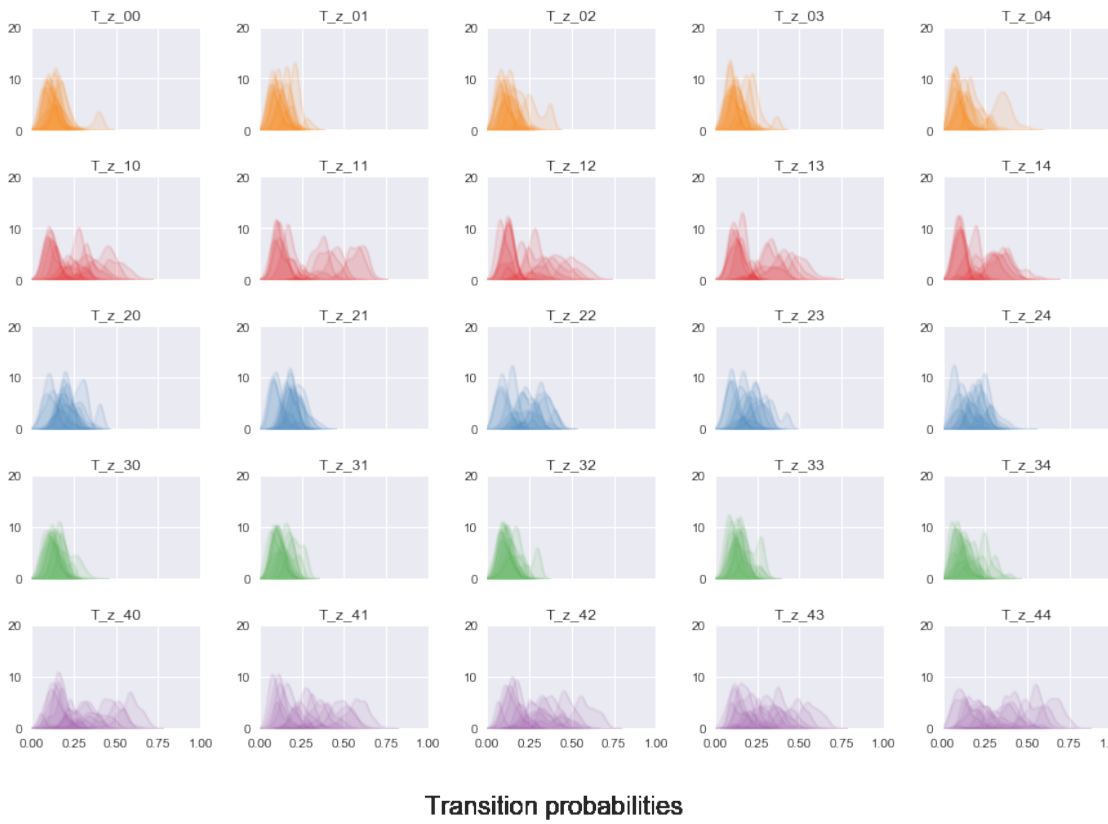


Figure 18: Posterior densities for T_z for the low income population group

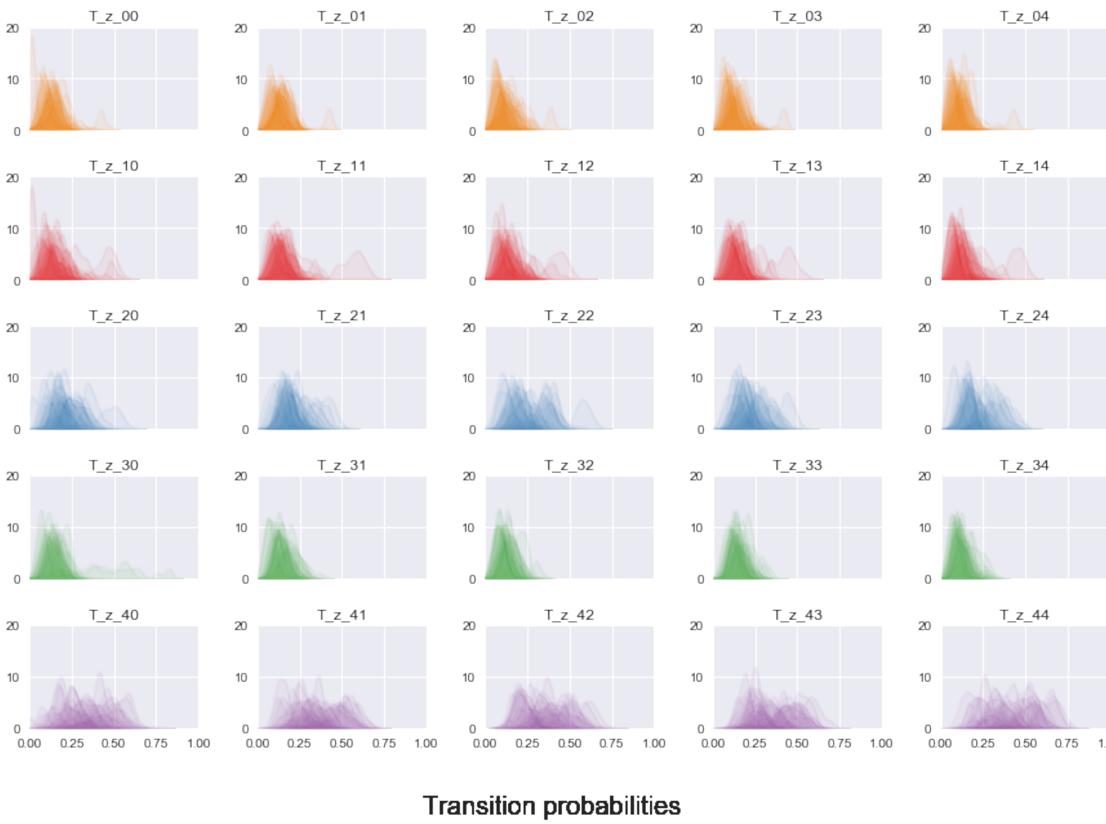


Figure 19: Posterior densities for T_z for the over sixty population group

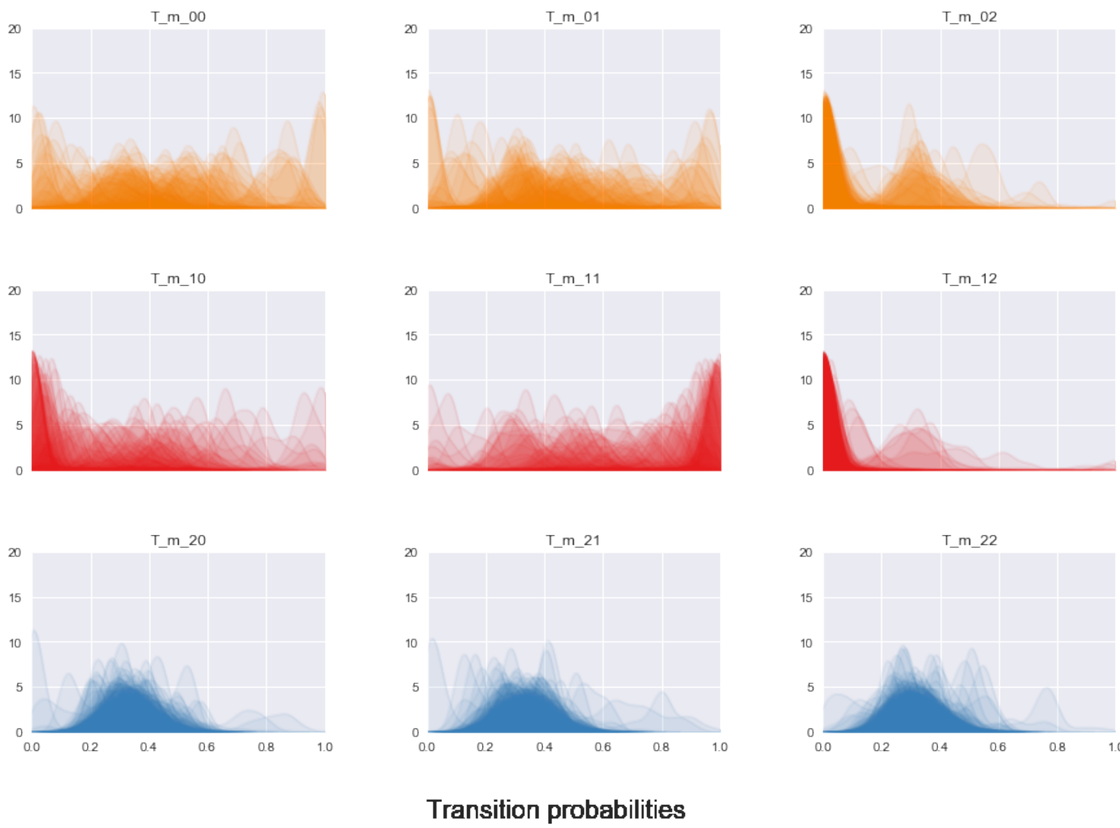


Figure 20: Posterior densities for T_m for the unconstrained population group

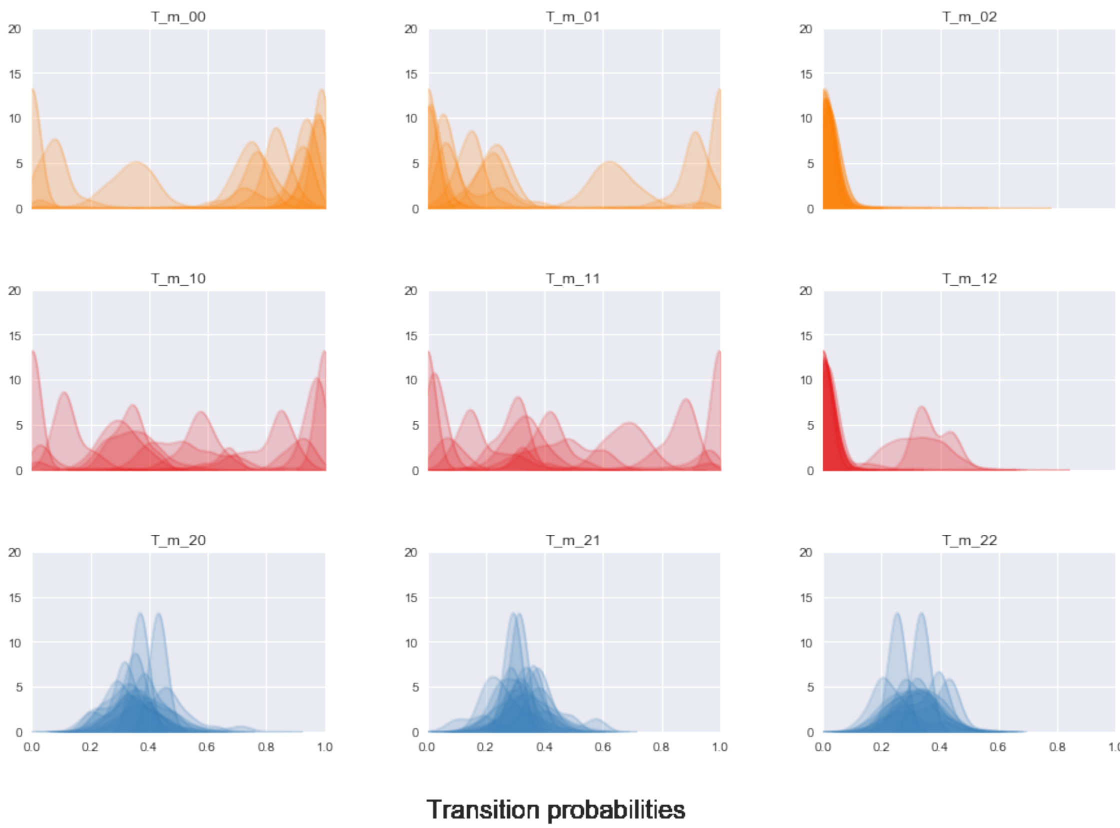


Figure 21: Posterior densities for T_m for the over sixty population group

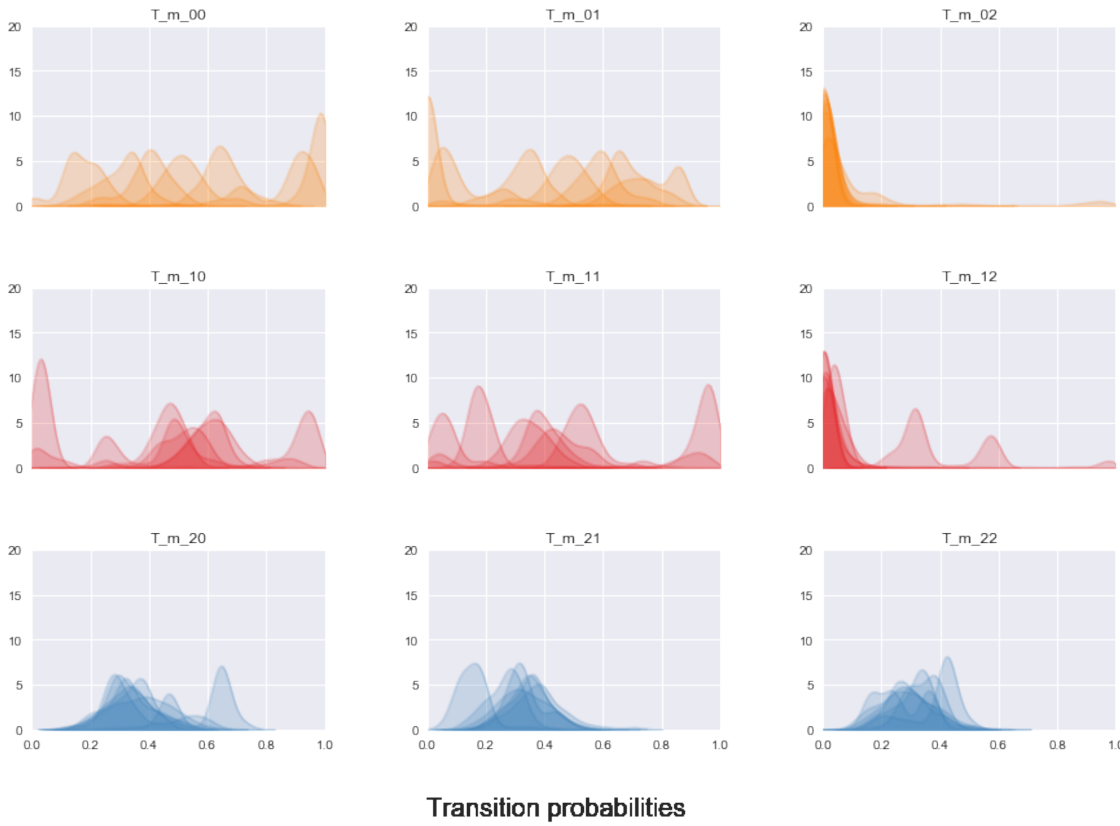


Figure 22: Posterior densities for T_m for the low income population group