Travel Behaviour and Society 11 (2018) 31–42

Contents lists available at ScienceDirect

Travel Behaviour and Society

journal homepage: www.elsevier.com/locate/tbs

Closer to the total? Long-distance travel of French mobile phone users

Maxim Janzen[a,*], Maarten Vanhoof[b,c], Zbigniew Smoreda[b], Kay W. Axhausen[a]

[a] IVT, ETH Zurich, Stefano-Franscini-Platz 5, CH-8093 Zurich, Switzerland
[b] Orange Labs, 44 avenue de la Republique, CS 50010, FR-92326 Chatillon Cedex, France
[c] Open Lab, School of Computing, Newcastle University, Urban Sciences Building, Wellington St, Newcastle upon Tyne NE4 5TC, UK

ARTICLE INFO

Keywords:
Long-distance travel demand
Mobile phone data
Travel surveys
Soft refuse

ABSTRACT

Analyzing long-distance travel demand has become increasingly relevant because the share of traffic induced by journeys related to remote activities which are not part of daily life is growing. In today's mobile world, such journeys are responsible for almost 50 percent of all traffic. Traditionally, surveys have been used to gather data needed to analyze travel demand. Due to the high response burden and memory issues, respondents are known to underreport their number of long-distance journeys. The question of the actual number of long-distance journeys therefore remains unanswered without additional data sources. This paper is the first to quantify the underreporting of long-distance tour frequencies in travel diaries. We took a sample of mobile phone billing data covering five months and compared the observed long-distance travel with the results of a national travel survey covering the same period and the same country. The comparison shows that most of the estimates of the number of missing tours by researchers have thus been too low. Our work suggests that the actual number of long-distance journeys is twice as high as that reported in surveys. Two different causes of underreporting were identified. Firstly, soft refusers travelled long distances but reported no long-distance tours. Secondly, respondents underestimated their number of long-distance tours. Consequently, there is a need to use alternative data sources in order to gain better estimates of long-distance travel demand.

1. Introduction

Analyzing long-distance travel behavior has become more important in recent years because the contribution of long-distance journeys to overall traffic is continuously growing. Therefore, the impact on planning urban areas, highways, railroads etc. is becoming greater. Long-distance travel is usually defined as trips which take place outside of a person's environment. However, the definition of a person's environment varies in the literature. It can be defined either spatially, temporally, purpose-based or a combination of these three. This paper utilizes the spatial definition, meaning that all trips within a certain distance of a person's home are considered to be daily life travel. All trips beyond the distance threshold are considered to be a long-distance journeys. Temporal definitions might characterize all overnight stays as long-distance trips. Purpose-based definitions utilize the purpose of a trip to decide whether it is a long-distance journey. In order to develop tools which are able to provide reliable predictions, one needs data sources that describe the current state of long-distance travel demand.

Data collection methods in the field of travel demand research have been investigated in the past (Axhausen et al., 2002; Armoogum and Madre, 2002; Bonnel et al., 2009; Zmud et al., 2013; Richardson et al., 1995; Arentze et al., 2000; Draijer et al., 2000). The most frequently used data sources are surveys. In the case of long-distance travel, the number of available surveys is limited (the main sources are national travel surveys). However, all long-distance travel surveys involve similar problems. Due to the high response burden, surveys tend to have a low number of respondents. Furthermore, it is known that the number of journeys reported in such surveys is too low (Madre et al., 2007; Armoogum and Madre, 2002). Both factors limit the explanatory power of the studies and leave the question of the quality of the results unanswered (Kuhnimhof and Last, 2009).

To overcome these limitations alternative data sources are needed. We propose in this paper to use mobile phone billing data in order to obtain better estimates of long-distance travel demand. The advantage is the large number of people that can be tracked without having being asked to spend a lot of effort on a survey. We analyzed five months of mobile phone billing data covering one third of the total French population. The data was provided by Orange™ France. After reconstructing long-distance journeys from the data, we were able to quantify the error reported by the French National travel survey. The main analysis is split in two parts. Firstly, we quantify the number of persons that do not travel long distances at all. This analysis will show

* Corresponding author.
E-mail address: maxim.janzen@ivt.baug.ethz.ch (M. Janzen).

https://doi.org/10.1016/j.tbs.2017.12.001
Received 29 March 2017; Received in revised form 31 October 2017; Accepted 5 December 2017

that there are more non-travellers among survey respondents than among the Orange customers. Secondly, we quantify the number of long-distance tours that are done by the mobile persons. It will be shown that mobile Orange customers travel significantly more than survey respondents. Both results indicate that the number of tours was heavily underreported in the survey. The aim of this paper to confirm the assumed underestimation of long-distance tours and demonstrate that there is a need of alternative data sources.

This paper is structured as follows: After a literature review we describe in detail the mobile phone data made available for our studies as well as the French national travel survey. In section four, our tour reconstruction methodology is described. Afterwards, we present outcomes and comparisons. We then offer a discussion and a conclusion.

2. Previous work

Data collection has always been an important issue in the field of travel demand research. Different methods of data collection have been investigated in the past (Axhausen et al., 2002; Armoogum and Madre, 2002). The data sources used have mostly been various forms of surveys to suit the diverse requirements of the researchers (Dillman, 2000).

In the case of long-distance travel, the number of recent surveys is limited. For Europe, *Mobidrive* studies are available (Zimmermann et al., 2001; Axhausen et al., 2002; Chalasani and Axhausen, 2004). Each of these studies encompasses a six-week period, which is usually not sufficient for a deep analysis of long-distance travel behavior. Other sources are national travel surveys like the French (Armoogum et al., 2008), British (Department for Transport, 2016) or Austrian (BMVI, 2012) ones. An additional longitudinal perspective is provided by the INVERMO study from Germany(Chlond et al., 2006). Several European studies have been combined for an analysis of long-distance travel demand in Europe (Frick and Grimm, 2014). A similar approach led to a nationwide model for the United States (Outwater et al., 2015a; Outwater et al., 2015b; Bradley et al., 2015).

An overview of available studies of annual long-distance travel rates can be found in Table 1, which reports the study area and year. Variations in the definition of long-distance travel are also reported, which include the distance-threshold used, the destinations included in the analysis and whether single-day tours were excluded from the set of long-distance journeys. Finally, the main indicator, the annual number of long-distance tours, reported in the studies are presented. The values that had to be extrapolated are marked. The studies included are: the California Statewide Household Travel Survey (CSHTS) (Bierce and Kurth, 2014; Cambridge Systematics Inc., 2013), an ifmo study (Frick and Grimm, 2014; Kuhnimhof et al., 2014), the INVERMO project (Zumkeller et al., 2005; Chlond et al., 2006), the Knowledge Base for Intermodal Passenger Travel in Europe (KITE) (Frei et al., 2010), the DATELINE study (Neumann, 2003), the French national travel survey (ENTD) (Armoogum et al., 2008), the Microcensus Switzerland (MCS) (Swiss Federal Statistical Office (BFS), 2010) a Eurostat report (Weckström-Eno, 1999), Methods for European Surveys of Travel Behaviour (MEST) (Axhausen and Youssefzadeh, 1999), and US National Transportation Statistics (US NTS) (Bureau of Transportation Statistics, 2016). All of these studies surveyed 8–12 weeks of long-distance travel and estimated annual tour rates. A correction factor is incorporated in most of the tour rates. The ifmo study reports a higher value than the other studies due to several reasons. Firstly, it is one of the most recent studies and it is known that the amount of long-distance journeys is growing. Secondly, it is combining several studies to get a full picture and, in particular, it estimates 5.0 everydays long-distance tours (e.g. commuting) which is more than in any other study.

Other long-distance travel studies have been performed with a special emphasis on tourism. Guidelines for tourism studies (Harris et al., 1994) and preferred analysis methods (Crouch, 1994) have been presented in the past. Many tourism studies have been performed, including the Travel Market Switzerland study (Bieger and Lässer, 2008) and the Net Traveler Survey (Schonland and Williams, 1996). Almost all of them focus on tourism activities within a single country. A summary of international studies can be found in Lennon, 2003 or the Eurostat database (Eurostat, 2016). However, the results of tourism surveys are limited due to the known issue of unobserved tourism (De Cantis et al., 2015).

Due to the high response burden that is usually associated with long-distance surveys (Axhausen et al., 2015; Axhausen and Weis, 2010), it can be expected that the number of long-distance trips is usually underreported. This is due to non-responding frequent travellers as well as travellers claiming not to travel while answering other questions, or so-called soft refusers (Madre et al., 2007). Furthermore, there is a memory effect. Respondents tend to forget tours, which happened some time before the survey (Smith and Wood, 1977; Bradburn et al., 1987; Tourangeau, 1999). Additionally, the vehicle miles travelled are usually heavily underestimated as shown by Wolf et al. (2003). Consequently, there is a need for survey weighting and expanding (Bar-Gera et al., 2009). Assumptions about underreporting long-distance tour rates in surveys led researchers to introduce correction factors in several studies (Cambridge Systematics Inc., 2013; Armoogum et al., 2008). In the case of tourist surveys, a weight correcting for the response bias is essential (Leeworthy et al., 2001). A correction factor is the only method currently available to account for underreporting. Assumptions about the inaccuracy of long-distance travel surveys are supported by evidence that two surveys of the same scope can suggest non-consistent travel behavior (Perdue and Botkin, 1988).

In order to estimate the level of underreporting in surveys, one

Table 1
Annual long-distance tour frequencies: Other studies (* based on own extrapolation).

Study	Year	Area	Destination	Long-dist. definition	Exclude single-day	Annual tours per capita
DATELINE (Neumann, 2003)	2001–02	Europe	international	75 km	No	2.7
DATELINE (Neumann, 2003)	2001–02	France	international	75 km	No	3.8
ENTD (Armoogum et al., 2008)	2007–08	France	France	80 km	No	5.1
MEST (Axhausen and Youssefzadeh, 1999)	1997–98	France	international	100 km	No	*7.4
MCS (Swiss Federal Statistical Office (BFS), 2010)	2010	Switzerland	international	100 km	No	*7.8
MEST (Axhausen and Youssefzadeh, 1999)	1997–98	Europe	domestic	100 km	No	*7.9
KITE (Frei et al., 2010)	2008–09	Switzerland	international	100 km	Yes	8.2
KITE (Frei et al., 2010)	2008–09	Portugal	international	100 km	Yes	8.2
CSHTS (Bierce and Kurth, 2014; Cambridge Systematics Inc., 2013)	2012	California	state-wide	50 miles	No	8.2
Eurostat (Weckström-Eno, 1999)	1999	France	international	100 km	No	8.5
INVERMO (Chlond et al., 2006)	2001–03	Germany	international	100 km	No	8.8
MEST (Axhausen and Youssefzadeh, 1999)	1997–98	Europe	international	100 km	No	*8.9
KITE (Frei et al., 2010)	2008–09	Czech Rep.	international	100 km	Yes	9.0
US NTS (Bureau of Transportation Statistics, 2016)	2001	USA	international	50 miles	No	*9.4
ifmo (Frick and Grimm, 2014; Kuhnimhof et al., 2014)	2011	Germany	international	100 km	No	15.9