Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 11 (2015) 365 – 380

Transportation Research Procedia

www.elsevier.com/locate/procedia

10th International Conference on Transport Survey Methods

Two-year worth of smart card transaction data – extracting longitudinal observations for the understanding of travel behaviour

Ka Kee Alfred Chu [a], *

[a] *Agence métropolitaine de transport, 700 de la Gauchetière ouest, 26th floor, Montréal, Québec, H3B 5M2, Canada*

Abstract

This paper proposes a sampling strategy to extract longitudinal observations from a large smart card fare validation database. Internal consistency and comparability with the population are evaluated. It is revealed that operational practices, rather than theoretical life span of the card, are the determinant factor of observation duration, sample size, and the presence of spatial and temporal bias. Using mobility and location diversity indicators, the longitudinal observations are analysed individually and aggregately for understanding travel behaviour at the day-to-day, seasonal and year-to-year levels.

Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC

Keywords: smart card; travel behaviour; longitudinal analysis; sampling; public transit

1. Introduction

Longitudinal panel surveys "collect information on the same set of variables from the same sample members at two or more points in time" (Tourangeau et al., 1997). They allow direct measurement of behavioural change down to the individual level. In the field of transportation, they are especially useful to understand trends in travel behaviour and to assess the effect of a change in service or policy. Despite their potential, longitudinal panel surveys are difficult to carry out. The administration of panel surveys involves maintaining the original sample between waves by reducing the attrition rate and refreshing the sample. In additional, traditional data collection methods that rely on memory and reporting are often restricted by data precision. Time-in-sample effect (also known as conditioning, rotation bias or panel fatigue) and seam effect are other drawbacks of panel surveys (Tourangeau et al., 1997). These severely limit the duration and the number of reliable observations that can be obtained. Because of these, few large-scale travel panel surveys with a long duration have been undertaken. A notable example has been the Puget Sound Transportation Panel Survey that lasted from 1989 to 2002 (PSRC, 2014).

Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC
doi:10.1016/j.trpro.2015.12.031

Travel data collected from passive devices can potentially overcome those shortcomings and are useful for studying multiday travel behaviour. Wolf (2000) proposes using GPS data loggers to replace travel diaries in the collection of travel data. More recently, mobility-aware and on-line technologies, such as the smartphones or wearable GPS, have been used to collect multiday travel data. Long-term data streams generated from passive devices on the same respondents can effectively be treated as panel surveys. Public transit-focused passive data streams from smart card automatic fare collection (AFC) systems are becoming increasingly common among transit agencies around the world and have been proposed as an alternative form of travel survey (Strambi et al., 2009; Chu & Chapleau 2013; Morency 2013).

This relatively-new technology passively and continuously accumulates fare validation records from all the cards over a very long timeframe. Each record contains a unique card identifier and a precise timestamp of the fare validation. When these records are organised at the card level, the culmination of validation records can potentially provide a card-based public transit panel survey, which differs significantly from traditional panel survey with its temporal and population coverage. The extended duration and the large number of the observations, which is almost impossible to match by traditional survey methods, make this data stream interesting for longitudinal analysis. In theory, longitudinal observations are collected as long as the cards are functional. It captures special events which are systematically avoided in periodic surveys in order to preserve comparability. It is not limited by hardware availability and battery life usually associated with other mobility-aware and on-line technologies. It also collects more precise data than traditional survey methods based on telephone, internet or face-to-face interview, thus avoiding the drawbacks of panel surveys. However in reality, system-specific operational practices, instead of card life, often dictate the duration of the observations. Card loss, card withdrawal and policies regarding card replacement can potentially cut short many observations and introduce unwanted biases into the data.

In sum, smart card data possess properties that align with the direction of large-scale mobility surveys (Raimond, 2009). They represent a move from periodical to continuous data collection, from cross-sectional to panel data and from single-day observations to multi-day observations of travel behaviour. As smart card data become more accessible, it is worth investigating whether longitudinal observations can be extracted from these data for travel behaviour analysis. There are numerous works on transit smart card data. Pelletier et al. (2011) provide a comprehensive review on the contribution of smart card data to transit planning: measuring transit performance (Trépanier et al., 2009), deriving the origin-destination matrices (Cui, 2006; Farzin, 2008; Munizaga et al., 2011) and understanding travel pattern. The potential for the latter has long been recognised (Bagchi & White, 2005; Utsunomiya et al., 2006; Bryan & Blythe, 2007). Some studies use a more computational approach, applying data-mining techniques to large data sets in order to detect travel patterns and classify transit users: Morency et al. (2007) use the object-oriented approach and clustering algorithm to group transit users with similar temporal travel patterns; Chakirov & Erath (2011) characterize spatial and temporal travel behaviour with statistical tools and data-mining techniques; Ma et al. (2013) apply clustering algorithms to detect and categorize travel patterns; Agard et al. (2013) propose a new distance calculation technique for a clustering algorithm which better captures the similarity of travel patterns among cardholders. Other studies follow a more analytical and descriptive approach, interpreting travel patterns from the data: Park & Kim (2008) produce transit use indicators; Lee & Hickman (2011) analyse the travel patterns of regular users over several days; Chu & Chapleau (2010) propose the use of anchor to characterize multi-day trip patterns at the individual level and look at its variation in time and space; Chu (2014) studies the long-term transit consumption and fare use patterns with a one-year dataset. The panel properties of smart card data also serve to study the change in travel behaviour following a modification of service (Mojica, 2008; Asakura et al., 2009) or the implementation of an incentive (Lathia & Capra, 2011).

This paper continues on the works from the previous ISCTSC conference (Chu & Chapleau, 2013; Morency, 2013). The objective is twofold: first, it proposes a sampling strategy to extract longitudinal observations and examines whether those observations are representative of the entire transit population. It identifies potential biases and suggests correction techniques. Second, with the observations resulting from the proposed sampling strategy, the paper investigates travel patterns at the aggregate and individual levels. System-wide travel patterns consist of the