



Intrapersonal variability in public transport path choice due to changes in service reliability

Ulrik Berggren^{1,2} · Carmelo D'Agostino¹ · Helena Svensson^{1,2} · Karin Brundell-Freij^{1,2,3}

Accepted: 28 July 2021 / Published online: 17 August 2021
© The Author(s) 2021

Abstract

Uncertainty during the course of a trip is regarded as a particularly disturbing phenomenon by public transport (PT) passengers and has been shown to have a substantial impact on travel behaviour. However, so far there has been little empirical evidence within contemporary research concerning the degree of mid- to long-term adaptation among PT passengers to changes in the principal cause of such uncertainty within PT transport systems: Lack of service departure reliability. While most studies to date have focussed on instantaneous or short-term behavioural responses, this paper presents results from a one-year panel dataset consisting of individuals' trip itineraries derived from smart card transactions enriched by automatic vehicle location (AVL) data. We measured long-term line route usage and departure reliability, the latter in terms of headway regularity and schedule adherence, in two panel waves. Thus, we aimed for practical indicators that may be associated with specific line routes and origin–destination PT stop pairs used for recurring trips over time. Results from logistic regression indicate a consistent significant impact of changed headway regularity and punctuality change for line routes with relatively high departure frequency. However, there is a relative indifference to long-term reliability change by a significant share of individuals that have actual trade-offs between at least two line route options, possibly indicating satisficing behaviour among PT passengers. Implications on PT path choice forecasting and route planning are discussed.

Keywords Public transport · Service reliability · Smart card data · Panel data

✉ Ulrik Berggren
Ulrik.berggren@tft.lth.se

¹ Department of Technology and Society, Division of Transport and Roads, Lund University, P.O. Box 118, 221 00 Lund, Sweden

² K2 The Swedish Knowledge Centre for Public Transport, Bruksgatan 8, 222 36 Lund, Sweden

³ WSP Advisory, Malmö, Sweden

Introduction

Irregularities and delays, or reliability deficiencies in short, are phenomena in public transport (PT) provision that have a negative impact on passengers' perceptions and time use (van Oort 2011), as well as on operating costs (Fadaei and Cats 2016; Watkins et al. 2020). Moreover, reliability has been shown to be a key factor in the choice between PT and other modes of transport according to a wide range of studies on passengers' attitudes and perceptions, (Beirão and Sarsfield Cabral 2007; dell'Olio et al. 2011). In fact, travel time uncertainty, being the opposite of reliability, has been found to have a higher disutility than nominal travel time (Bates, 2001). Unfortunately, and despite of the huge potential implications on passenger satisfaction, patronage and operations, this quite common but problematic aspect of public transport operations is usually not taken into account in appraisal methods that use path choice algorithms from state-of-the practice macroscopic transport models (Carrel et al. 2013). Despite an urgent need for empirically grounded perspectives on the behavioural consequences of reliability changes in order to facilitate appropriate policy measures in terms of PT network design, research regarding the long-term¹ behavioural effects (Carrel et al. 2013; Le et al. 2020) and the potential subsequent impact on the formation of habits (Evans 2008; Gärling and Kay 2003; Ouellette and Wood 1998; Verplanken et al. 1998) that can result from perceptions of (un)reliability, is scarce. Apart from small-scale surveys by Carrel et al. (2013) and Le et al. (2020), previous research into longitudinal change in PT use has mostly covered how short-term travel patterns (Goulet-Langlois et al. 2018; Trépanier et al. 2007) and long-term mode choice (Chatterjee 2010; Thøgersen 2006) may be impacted by exogenous factors such as changes in transport provision. The reason for the relative scarcity of research into medium- to long-term (at least one year) behavioural effects from PT reliability changes may be attributed to two fundamental challenges in obtaining empirical evidence for making robust conclusions regarding this issue. First, the sole measure of reliability has been a matter of continuous debate over at least the last 50 years, during which time innumerable attempts have been made to identify a good measure (Bagherianet al. 2016; Currie et al. 2011; Diab et al. 2015; Gittens and Shalaby 2015; Jenelius 2018; Noland and Polak 2002; Uniman et al. 2010). Second, detailed data regarding the behaviour of PT passengers have at least historically been difficult to capture, particularly behavioural responses to “vague” events such as fluctuations in reliability. Stated choice experiments have shed light on some phenomena that relate to reliability, such as passengers' approaches to risk (Li and Hensher 2012) and other attributes in a path choice context (Noland and Polak 2002). However, the inherent methodological uncertainty related to the design and wording of these studies, particularly when studying individuals' perceptions of the likelihood of travel time variations, as well as their reliance on stated intentions instead of longitudinally revealed behaviour, makes expressing strong general claims about sensitivity in terms of perceptions as well as responses, somewhat hazardous (Bates et al. 2001; Le et al. 2020). However, with the growing availability of large-scale disaggregate datasets covering supply-side automatic vehicle location (AVL) data, demand-side automatic fare collection (AFC) data, particularly “smart” travel cards (Kurauchi and Schmöcker, 2017), and automatic passenger counters (APC), the amount of research into different aspects of detailed passenger responses to changes in PT provision

¹ We use the definitions adopted by Carrel et al (2013); short-term adaptation occurs during a trip whereas long-term adaptation is influenced by past experience and may therefore occur during the trip-planning phase.

has grown considerably. However, as identified by, for example, Carrel et al. (2013), the process of medium- to long-term (from one year to the next) strategic passenger response to PT service reliability changes has been poorly researched empirically, although some successful simulation approaches (An et al. 2014) may be noted.

In this paper, we explore PT passengers' long-term revealed *sensitivity*, in terms of their propensity to change travel paths over time in response to (observed) change(s) in travel conditions—in our case, PT service departure reliability, utilising a fixed effects panel approach. Travel path change, as expressed in terms of the relative frequency of line route² usage, was surveyed in two panel waves using AFC smart card transaction data, while AVL and timetable data formed the basis for two departure reliability indicators that measure *headway regularity* and departure schedule adherence (*punctuality*), respectively. Thus, we have broken down the concept of PT service departure reliability into these two observable components. According to previous research, punctuality may be particularly influential when passengers are expected to adapt to scheduled departure times (cf. exposition presented by Carrel et al. (2013)). As succinctly summarised by Rudnicki (1997), from an operational perspective, it may be defined as any deviation from timetable adherence, measured in absolute or relative terms. Individuals who wish to use PT line routes that operate with long headways perceive lack of punctuality as particularly negative [when they adhere to schedule-based waiting behaviour, as discussed by, for instance, Liu et al. (2010)]. Usually, and in practice, a tolerance threshold is set for when a lack of punctuality is perceived as onerous by a majority of passengers, and the performance of the transport system is benchmarked according to the proportion of departures that exceed this deviation threshold value. *Regularity*, on the other hand, may be more associated with single (or multiple) line routes that run with a short (combined) headway. In such cases, the *distribution of departures* may be perceived as being more critical to tentative passengers than the strict timetable adherence of each departure time (Rudnicki 1997).

Long-term sensitivity in this case should be understood as being the degree to which individual passengers, over a measurement period of at least months, marginally change their relative usage rate of a *path* (expressed as a line route) in a given origin–destination pair in response to a (marginal) change in *service departure reliability* of that path. There should be at least two different alternative line routes and a non-zero difference in reliability for each of these line routes during the measurement period. Due to earlier findings of behavioural differences across trip purposes (Berggren et al. 2019), we controlled for day type (weekday versus weekend day) to check for systematic differences in long-term sensitivity to reliability changes assuming a different average composition of trip purposes across day types.

Thus, the overarching research hypothesis we wished to explore was that:

- (a) Passengers perceive a gradual change in their marginal utility of a certain path as the reliability of that path changes over time.

Therefore,

- (b) when a path is subject to an increase/decrease in departure reliability, this is reflected in the behaviour of passengers insofar that the proportional use of this path increases/decreases, all else being equal.

² The term *line route* is used throughout this paper to denote a unique sequence of stops or stations serviced by a specific PT line using a numerical identifier (“line number”).

This second sub-hypothesis, measurable in contrast to the first, is conditional upon the existence of a considered choice set (potential trade-off) and that the alternatives of this choice set remain or improve differently in standard when relative reliability changes between paths. Thus, a generalisation of the research question may be formulated as to whether or not it is possible to link changes in the relative usage of a certain path to the relative change in departure reliability compared to other paths in the choice set of an individual passenger, when changes in other attribute values such as departure frequency, travel time variability and travel time are controlled for.

In a conventional discrete choice framework, there are latent and random utilities assigned to each choice option as a function of the attributes of each and every potential choice option in a considered choice set. It should be noted that, in contrast to that conventional framework, we studied only the relative usage of (a subset of) paths that were chosen at least once. Therefore, in our analysis the observation of increased (decreased) relative usage over time (boolean) is the independent variable, and relationships are represented by a binary logistic regression approach. In the generalised linear regression approach thus applied, the binary “choice” was between the options (1) To change (increase or decrease) the use of a specific path (line route) or (2) To not change the use of this path, for each observed OD pair. A latent and random utility was attributed to option (1) as a function of how the path-specific attribute values had changed, while the utility of option (2) was assumed to be constant with random variation. Since the dependent variable is dichotomous (change versus no change), this binary logistic regression may be regarded as a natural reformulation of the choice process.

The rest of the paper is organised as follows. First, in section “Research background”, we outline previous research regarding the measurement of reliability and demand. Next, section “Service reliability and path choice” describes the theoretical framework regarding behaviour and PT service reliability, along with our rationale behind the choice of reliability measures and analytical approach. Section “Data and research setting” outlines the data, design, and analysis of our case study in Scania, Sweden, followed by sections “Results” and “Discussion”, in which we present and discuss the results from the case study. Finally, section “Conclusion” contains conclusions drawn from the case study, along with research ideas for the future regarding the study of behavioural responses to reliability changes in PT.

Research background

Reliability and choice in PT

The importance of how uncertainty and risk affect individual preferences in choice processes has been a topic of research for the last 40 years or so, ever since Kahneman and Tversky showed the existence of non-linear, non-reciprocal response patterns in such circumstances (Kahneman and Tversky 1979; Tversky and Kahneman 1981). This phenomenon has also been shown within the context of travel time uncertainty at path choice; although primarily for car travel and using stated preference (SP) surveys (Abdel-Aty et al. 1995; Avineri and Prashker 2004). Here, a number of framing-related issues unfortunately cloud the validity of the presented results to some extent. Noland and Polak (2002) further discuss these framing dilemmas in terms of individuals’ sensitivity, or perceptual limits, vis-a-vis travel time uncertainty (i.e., what deviations from schedules are actually

perceived as delays and how do passengers respond?). According to the authors, significant issues concern, on the one hand, how these dilemmas are presented in terms of trade-offs between probabilities and absolute magnitudes of different delays in SP surveys and, on the other, how they depend on context that, in the PT realm, essentially includes network structure and service frequency. Moreover, the magnitude of travel time deviation has been empirically shown (Yap et al. 2017) to have non-linear effects on path choice. Also, as Heiner (1983) highlights, successive learning from experience will itself affect the adaptability to uncertainty—ultimately to minimise mental stress and disutility. Here, the amplitude (occurrences of major disruptions) may have a more adverse effect on path choice than standard deviations (common, minor departure perturbations)—a notion³ supported by both prospect theory (Kahneman and Tversky 1979) and empirical results (Diab et al. 2015; Yap et al. 2017). An empirical account of the relationship between perceived reliability and path choice is presented by Carrel et al. (2013), in their survey of PT passengers, in which they find that departure regularity is the most important path-specific feature. They also mention the gradual change in passenger behaviour with respect to headway length and departure reliability, in which an interval may be discerned for PT passengers' adaptation behaviour with respect to service departure frequencies F , between chiefly stochastic behaviour at “short” headways ($H = 1/F$) and mostly deterministic behaviour at “long” headways. However, the exact figure for this interval appears to be dependent on service reliability and information, with empirical results ranging from five minutes for reliable services (Frestad Nygaard and Tørset 2016; Luethi et al. 2007) to up to 30 min for less reliable services (Ingvardson et al. 2018). As noted by Furth and Muller (2006) and Trompet et al. (2011), passenger behaviour with respect to PT service reliability is also conditional on expected waiting times, which themselves are a function of both scheduled headway and its variability.

A number of attempts have been made to identify theoretically appealing and operationally feasible measures of reliability in PT. Currie et al. (2011) conducted a qualitative evaluation of ten different measures based on international expert judgement and found that excess wait times for high-frequency services and customer journey time delays were best suited in terms of availability and information value—the latter measure conditional on the availability of AVL data. However, the most comprehensive evaluation of reliability measure approaches to our knowledge is provided by Gittens and Shalaby (2015) in their account of 20 different measures in which both theoretical and empirical underpinnings were presented. They arrived at the conclusion that a useful reliability index should include both in-vehicle and wait time variation, and that it should take the service context, i.e., headway distributions, into account. Other researchers have proposed adjusted or completely new measures to cater to advances in data capture methods.

Jenelius (2018) used AVL and APC data to extract perceived travel times during periods of congestion and crowded conditions, based on disaggregate boarding and alighting figures on a stop and trip level. Bagherian et al. (2016) used scheduled timetables and AFC data to calculate the buffer times needed by passengers to compensate for reliability deficiencies during daily PT journeys. Dixit et al. (2019) used AFC and AVL data to calculate the behavioural impacts of both common, minor, deviations and large-scale disruptions to the provision of PT, also taking composite multimodal trips into account. PT passenger path choices made under uncertainty, as the topic is described by Noland and Polak (2002),

³ Referred to by some authors as the “peak-end rule”.

are explicitly considered in the study by Yap et al. (2017). In their algorithm, they investigate the impact on multimodal PT trips from both common, small-scale variability in headways due to driver behaviour and varying passenger loads, and large-scale disruptions caused by, for example, technical failures or road closures.

Trip patterns from smart card data

The inherent difficulty in obtaining disaggregate data describing trip patterns made by PT passengers comprises two principal challenges when dealing with AFC systems that only register boarding events: (a) The inference of alighting locations, regardless of whether they are points for transfers or final destinations, and (b) the generation of complete origin–destination (OD) trip itineraries in order to obtain OD matrices. Critical factors of the latter include the degree of multi-leg PT trips in the system and the amount of location information provided by the AFC system. The first component of difficulty is usually approached by using some kind of mirroring of the boarding profile (Navick and Furth 2002; Trépanier et al. 2007), with slightly different approaches for potential transfers and other activity destinations respectively. Trépanier et al. (2007) reported an 80 per cent success rate in peak hours for their approach to infer alighting locations (bus stops). Barry et al. (2009) used a schedule-based shortest path algorithm to infer trip paths, thereby generating complete OD trips when no explicit location information was available for the intermediate trip points. Exact information on locations is sometimes available in contemporary AFC systems, but when it is lacking, AVL data have usually been merged with travel card validations in order to obtain full trip itineraries, one example of which has been reported by Farzin (2008). Different approaches have been reported regarding how to distinguish between and measure transfers, among other kinds of activities, in trip itineraries from travel cards. Different kinds of distance cut-offs have been utilised (Munizaga and Palma 2012; Zhao and Rahbee 2007), as have scheduled headways of line routes boarded after transfer (Gordon 2012; Wang et al. 2011). Seaborn et al. (2009) take bus in-vehicle times into account in their calculation of transfer time thresholds from travel card transactions at bus boardings and station gate passings while Chu and Chapleau (2008) include a maximum transfer walking distance to distinguish transfers from activities. Finally, clustering of PT stops and stations has been introduced by Goulet-Langlois et al. (2016) in their approach to infer activities based on likely activity locations.

Service reliability and path choice

This section outlines the rationale behind our choice of PT service reliability indicators and the theory and practice behind our approach to gauge passenger responses to changes in reliability, as evidenced by individual passengers' path choice.

In line with previous research regarding PT passengers' response to uncertainty variation, we have set out to explore the long-term impact on path adjustment from passengers' assumed perception of uncertainty during trips (Diab et al. 2015). Here we rely on the assumption that this perception of the predictability of the PT system may be sub-divided into (a) time-table adherence and (b) headway variability. Both these indicators of reliability are associated with the individual passenger's desire to minimise the amount of non-useful time spent travelling, including time components such as

waiting and in-vehicle travelling—components which have a certain degree of variation or fuzziness which should be specifically associated with negative perceptions of travel time uncertainty.

Similar to An et al. (2014), we propose an index to measure headway regularity that is normalised by headway durations, but only to represent the headway variation as initially proposed by Osuna and Newell (1972) in their formulation of a waiting time function. This is to be able to distinguish the relative impact on path choice from schedule adherence and headway variability. Thus, our regularity index takes the coefficient of headway variation CV_h as a starting point (Eq. 1):

$$CV_h = \left[\frac{\text{var}(H_r)}{E^2(H_s)} \right]^{1/2} \quad (1)$$

where H_r is the actual, and H_s the scheduled, service headway of interest. In our application, H_s was derived from the number of scheduled departures per hour, and the value of this line route-based measure per two-week analysis period was obtained by accounting for the relative weight of the number of trips recorded by hour and analysis period in the farecard dataset. H_r , on the other hand, was derived directly from the AVL data regarding each PT line route.

To standardise the values of the regularity index to the interval (0,1] by taking scheduled headway into account and thus enable a comparison with measures standardised in the same numeric interval such as punctuality, we propose an adjusted measure, formulated in Eq. (2):

$$RI_h = 1 / \left(1 + \frac{\text{var}(H_r)}{E^2(H_s)} \right) \quad (2)$$

The mathematical properties of this index reflect an intuitive sense of behavioural responsiveness, since, as previously mentioned, the headways of PT line routes should adhere to those time spans within which passengers intending to travel by these line routes are expected to arrive randomly at the first boarding point. Given this restraint, as variation in actual headway approaches zero, RI_h may be approximated by one (for “very reliable services”), while, as the variance in actual headway moves toward infinity (for “very unreliable services”), RI_h approaches zero asymptotically. As noted by Furth and Muller (2006), CV_h is conditional on the headway itself, in which the variation of each departure time (schedule adherence) becomes more important for long headways, while the interval between successive departures dominates at short headways as the numerator of Eq. (1) increases in relation to the denominator.

In order to gauge punctuality in an operationalised way using a binary measure, it is propitious to define a threshold value of departure deviation from schedule, above which a departure may be regarded as non-punctual (Currie et al. 2011). To find such a value, we tested both the quite common five-minute threshold value for lateness, as referred to by Currie et al. (2011), and a value that could also measure even smaller deviations. Due to its higher explanatory power, we decided to use two minutes as the final threshold value for the punctuality threshold value. The reason to not set it at zero is to partially exclude errors of measurement and rounding without causing unnecessary data loss. The chosen punctuality index may thus be formulated according to Eq. (3):

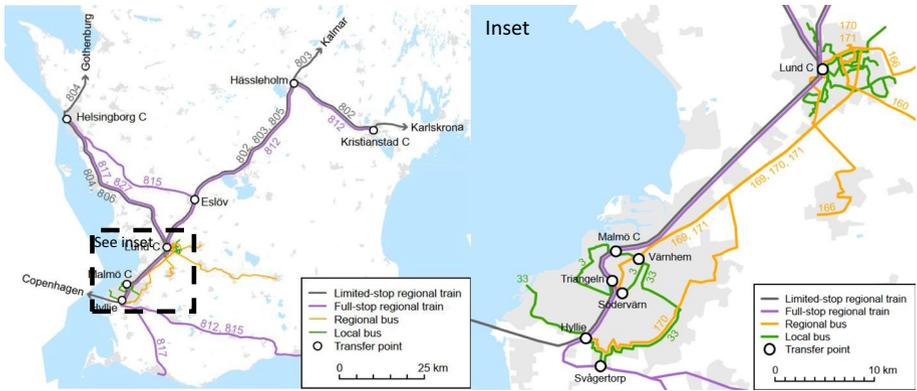


Fig. 1 Map depicting the geographical context, extent and line types of the PT subnetwork of 20 line routes used in this study (Note that local routes 731–750, running in the city of Lund, are not specifically indicated due to spatial limitations)

$$P_L = \sum_{v=1}^{V_L} \sum_{s=1}^{S_L} I_{s,v} / N_{S_L, V_L} \tag{3}$$

where $\begin{cases} I_{s,v} = 1 \text{ if an actual departure} \\ \text{deviates at least 2 minutes} \\ \text{from the schedule, and} \\ I_{s,v} = 0 \text{ otherwise} \end{cases}$, P_L is the punctuality index (proportion of non-punc-

tual departures) for line route L , s is a stop or station in the set of stops or stations S_L for line L , v_L is a vehicle run of line route L and N_{S_L} is the total number of stops or stations along line route L .

As a measure of in-vehicle time variation, we introduce the OD-based reliability buffer time RBT (Gittens and Shalaby 2015; Uniman et al. 2010) defined as the difference between the 95 percentile and the median travel time (Eq. 4) per OD pair:

$$RBT_{OD,L} = TT_{95perc,OD,L,\{i\}} - TT_{median,OD,L,\{i\}} \tag{4}$$

The travel time for each passenger trip (TT_i) and OD was calculated based on the departure and arrival time attributes of the temporally and spatially matched PT vehicle trip from AVL data, and the quantiles $TT_{95perc,OD,L}$ and $TT_{median,OD,L}$ were calculated based on the sets of trips $\{i\}$ for each day type (weekday or weekend day), OD pair OD and line route L .

For each OD pair in the AFC dataset, the line route(s) registered for fare card boarding transactions per card ID were assumed as being the “choice set” for each passenger, in which each option comprised (a sequence of) line routes between origin and destination stops or stations. In this setup, the reliability measures were associated with the line route level only, regardless of date or running direction. The rationale here was that passengers may adapt their behaviour already in their trip planning rather than taking account of specific path-based reliability properties en-route during a trip (which they had less chance to adapt to). Thus, the intention was to capture general perceptions of service reliability per line route rather than the specific reliability for each vehicular trip.

Table 1 Characteristics of the line routes included in the study. U—local city route, S/R—suburban or regional bus route, T—train route

Line route ID	Line type	Line route length (km)	Number of stops served	P_L 2016 ^a	P_L 2017 ^b	RI 2016 ⁷	RI 2017 ⁸	Minimum headway (minutes)
3	U	11	31	0.37	0.38	0.89	0.80	6
33	U	19	43	0.41	0.50	0.94	0.84	10
160	S/R	36	20	0.46	0.52	0.99	0.94	15
166	S/R	24	39	0.48	0.45	0.99	0.95	10
169	S/R	24	15	0.47	0.50	0.95	0.95	10
170	S/R	26	20	0.48	0.47	0.99	0.99	20
171	S/R	21	12	0.51	0.41	0.88	0.94	5
731	U	8	33	0.33	0.29	0.99	0.98	15
733	U	11	39	0.39	0.41	0.91	0.91	7
736	U	9	29	0.41	0.29	0.98	0.96	10
750	U	5	9	0.14	0.27	0.90	0.97	10
802	T	289	27	0.49	0.56	0.99	0.96	60
803	T	326	17	0.38	0.49	0.99	0.97	60
804	T	299	16	0.38	0.50	0.99	0.99	60
805	T	79	5	0.28	0.39	1.00	1.00	60
806	T	65	5 or 6	0.40	0.56	1.00	0.99	37
812	T	114	15	0.43	0.50	1.00	0.99	60
815	T	173	28	0.38	0.53	1.00	0.99	30
817	T	98	17	0.16	0.33	1.00	1.00	30
827	T	58	4	0.30	0.52	1.00	0.99	60

The RI_h regularity index is given by Eq. (2) while the P_L proportion of non-punctual departures is given by Eq. (3)

^aBetween 14 and 27 of November 2016

^bBetween 13 and 26 of November 2017

Data and research setting

A selection of 20 PT line routes (bus and regional trains) was made for our analysis, covering principal travel demand patterns within the commuting region of south-west Scania in southern Sweden, particularly within and between the cities of Malmö and Lund (Fig. 1). To enable a panel analysis, two two-week time frames separated by one year were selected for data extraction from AVL and AFC databases, from November 2016 and 2017, respectively. The main rationale as to the durations and spacing of the panel waves was to gather sufficiently large datasets for each wave to enable an analysis of changes in long-term travel strategies due to changes in marginal service reliability without getting into calculation issues due to the sheer size of the datasets. A similar approach was applied by Le et al. (2020) in their analysis of PT usage changes due to changed service satisfaction.

We assumed an inflection point between short and long headways to 12 min, reflecting the midpoint of the interval [10,15] minutes mentioned in the Transit Capacity and Quality of Service Manual (Board, National Academies of Sciences, and Medicine 2013), and above which the majority of PT passengers start consulting timetables and scheduling

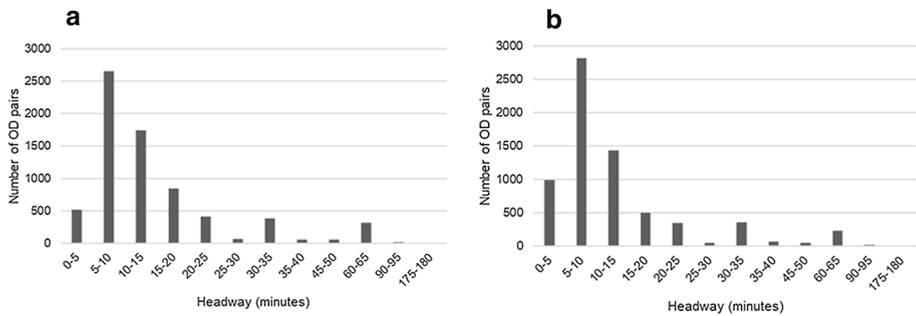


Fig. 2 Distribution of the number of observed (travelled) OD-pairs per scheduled headway in the time frames (panel waves) of 2016 (left) and 2017 (right), respectively, from the AFC dataset

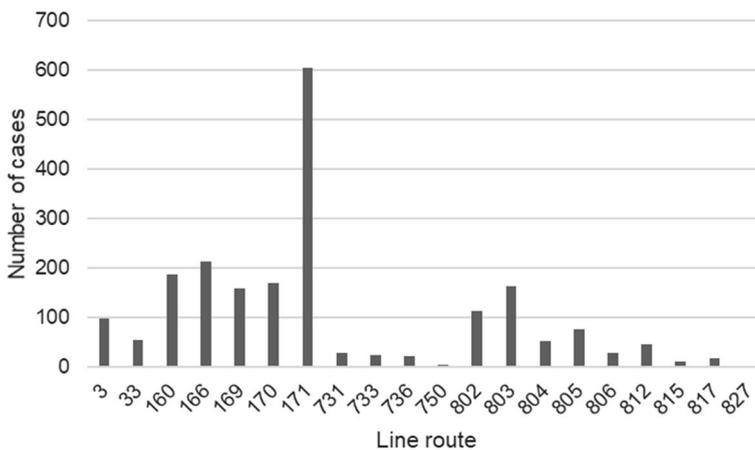


Fig. 3 Distribution of analysis cases (CardID*line route*OD pair) across line routes in the panel data set

starts to become the most important factor. Thus, shorter headways imply random arrivals at origin boarding points among PT passengers, and one would expect a stronger focus on headway regularity in order to minimise the expected wait times.

Properties of the PT network under study

The selected line routes of the case study comprise six urban bus routes, five suburban/regional and nine regional train routes with characteristics indicated in Table 1. In the autumn of 2017, substantial road works were carried out (construction of a new light rail line in the city of Lund) affecting the punctuality and/or regularity on line routes 160, 166, 169, 731, 733, 736 and 750.

The scope and location of the case study, and thus the selection of line routes, has been determined by three principal criteria: (a) There should be enough demand (card transactions) during at least two measurement periods one year apart to be able to obtain a sufficiently large dataset to enable panel analysis; (b) there should be a sufficient number of path options available between each origin and destination pair (reasonably representing

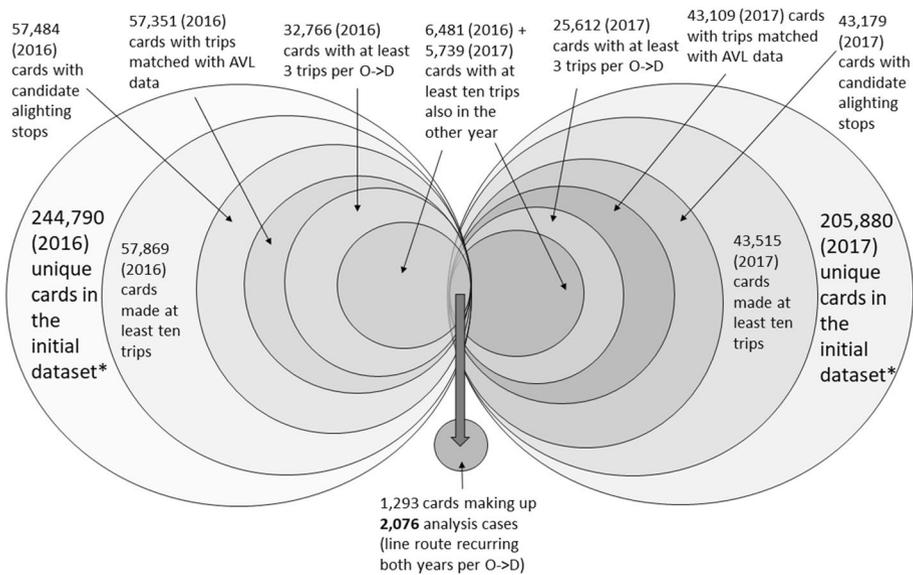


Fig. 4 The data refinement procedure of analysis cases to be included in the panel data set. One analysis case comprises the attributes of Line route (L), $RI_{L,2016}$, $RI_{L,2017}$, $P_{L,2016}$, $P_{L,2017}$, CardID, Number of trips in 2016 $_{L,(OD)}$; Number of trips in 2017 $_{L,(OD)}$, Trip proportion $_{L,(OD),2016}$, Trip proportion $_{L,(OD),2017}$, Origin stop and Destination stop. *Unique cards in each wave, but the same card may be found in both waves

the travel patterns of all PT travellers) and (c) the AFC and AVL data should be readily available for analysis.

The distribution of headways across line routes used by travel card holders in the respective datasets is quite heterogeneous (Fig. 2)—reflecting both the configuration of the timetables and ridership across line routes, the latter as indicated in Fig. 3.

Data, its enrichment and refinement

Smart card transaction data covering trips made on the selected PT line routes of Table 1 were retrieved from the AFC system of the regional PT provider Skånetrafiken, covering two weeks in November 2016 and 2017, respectively. The data comprised time stamps, stop names, line route names, as well as card identifiers for each boarding event on the selected routes during the two-week time frames. In addition, data from stationary ticket vending machines (TVM) were used. However, the line route used for the actual trip was not specified in the TVM data but were inferred from subsequent on-board validations.

In order to enable analysis of the longitudinal behavioural effects of changes in service attributes such as reliability, trip types that were subject to the impact of changes in line route properties had to be obtained. Thus, it was necessary to identify *recurring* trips primarily involving commuting of any kind (Chu and Lomone 2016) that occurred during both panel waves. In our study, this was operationalised based on two main criteria (the first criterion also applied by Chu and Lomone): (a) The same card was used to make at least ten trips in both time frames (panel waves), and (b) origin–destination pairs and line routes recurred at least three times per time frame (panel wave).

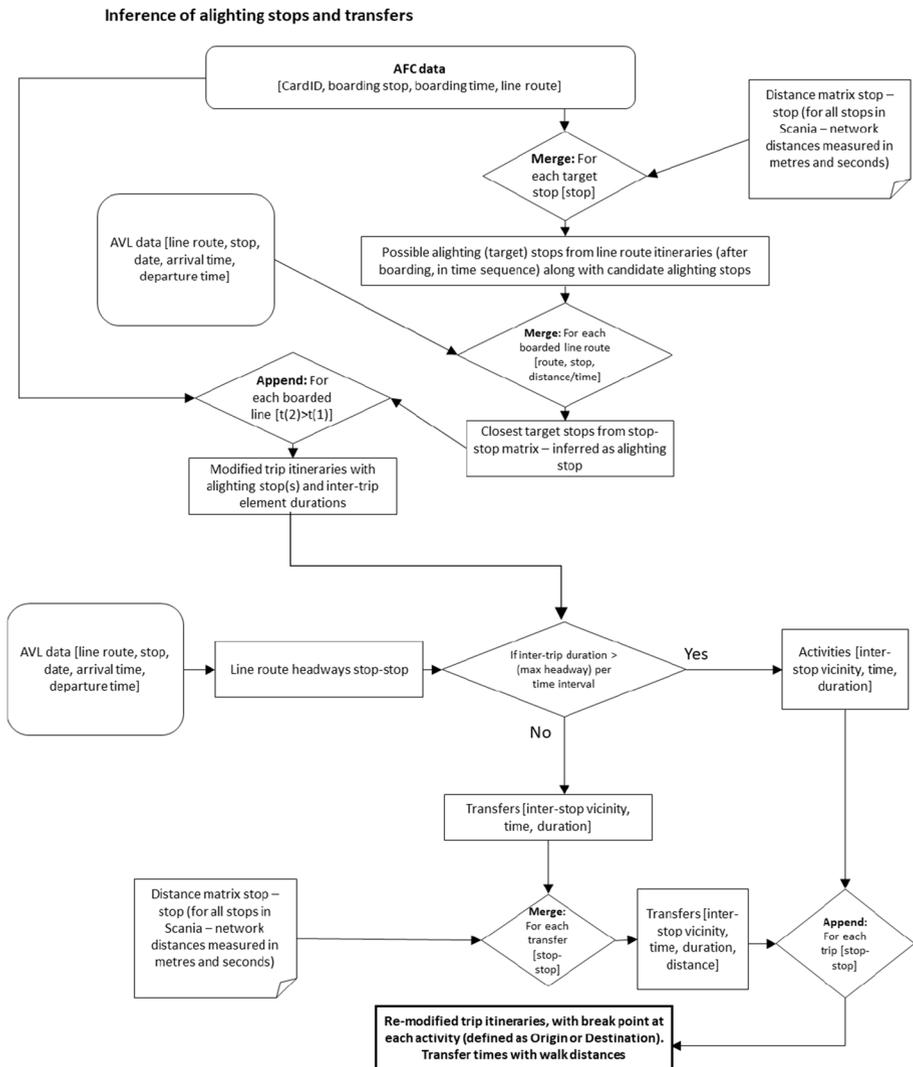


Fig. 5 Schematic diagram of OD matrix generation process

As shown in Fig. 4, transaction data from a total of 244,790 cards were collected from the 2016 wave while the corresponding figure for the 2017 wave was 205,880 cards, of which 58,346 cards were identified in both waves from their ID.⁴ The grand total of transactions was 1,588,833 in the 2016 wave and 1,246,416 in the 2017 wave. However, a total of only 10,230 cards remained after filtering out the cards that had been validated less than ten times in each wave, out of the 58,346 cards. This figure was further reduced in the identification of recurring trips (per line route and OD pair), along with some minor data

⁴ The cards used in the models are assigned random ID:s that are consistent across panel waves but are not directly traceable to the specific card numbers used in the AFC system.

loss in the trip itinerary generation process, leaving 1293 cards and 9211 and 9617 transactions per time frame, respectively. The final dataset for regression analysis was made up of relative frequencies, wave by wave, for each case of travel card ID, line route ID, and OD pair—i.e., the revealed choice of each travel card holder to choose a particular line route (included in their revealed personal choice set) for each OD trip during each panel wave.

In addition to the AFC data, reliability data, in the form of scheduled and actual departure times at all stops, were obtained from the AVL system for the same line routes as the card transaction data (cf. Table 1) and merged with the dataset of analysis cases on line route ID and time frame identifiers (the merging procedure is outlined schematically in Fig. 5), including separate identifiers for weekdays and weekends, respectively. The raw AVL data consisted of scheduled and actual stop arrival and departure times of all vehicular trips, during the two panel waves, of the same 20 line routes as the AFC dataset for all stops and stations serviced by them for passenger boarding and alighting. If data regarding actual departure times were missing, the scheduled times were used instead.

Thus, two primary datasets were prepared for analysis—one with relative line route frequencies by panel wave included regardless of the day type and one in which each frequency was associated with either a weekday or a weekend day. Both data preparation and processing, as well as regression analysis, were performed using SAS software on an ordinary PC laptop with an i7 1.80 GHz CPU and 16 GB installed RAM.

Generation of OD matrices

The PT provider for Scania (Skånetrafiken), whose PT network is used in this study, uses tap-on only as fare validation for bus trips, and on-board validation by itinerant staff for train trips. Ticket types include pre-paid periodical or rechargeable cards used for pre-payment of single-trip tickets. All purchase of single-trip train tickets must be performed in advance, either at TVMs or from the bus driver on a preceding bus trip leg.

Alighting stops had to be inferred in order to generate OD trips for subsequent behavioural studies of passenger line route usage. This was accomplished based on a series of steps, roughly outlined in the upper part of Fig. 5 and largely inspired by the work of Gordon (2012). For train trips for which the fare was pre-paid using TVMs, the line route was inferred by sorting all transactions by travel cardID and time stamp.

In the lower part of Fig. 5, the inference of full origin stop—destination stop trips using activity detection is outlined (further described below).

Like Gordon (2012) and a number of other authors, we use subsequent boarding stops as candidates for alighting stops for a particular leg of a trip (mirroring). However, in order to allow for transfers that not only use one stop point, but also transfers between stop points in close proximity to each other, we include other candidate stops for possible alighting events within a network-congruent walking distance (according to a distance matrix) of four kilometres. This cut-off distance value is based on the walking tolerances that were surveyed and presented in Berggren et al. (2019) in the same temporal and spatial contexts as the travel card data obtained for the panel analysis. From this array of possible alighting stops per trip, one of them was selected that is on the same line route as the boarding stop of that particular trip in the AFC data. Since this matching is based on both spatial (boarding stop and line route) and temporal criteria (the most likely departure according to the minimisation problem expressed in Eq. (5)), each passenger (travel card ID) trip has a corresponding vehicle movement determined from the AVL data.

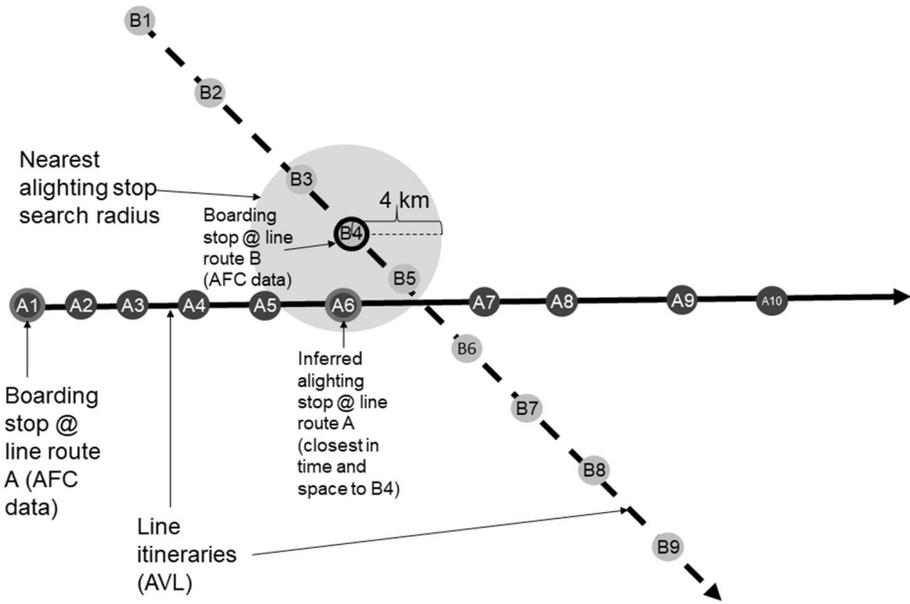


Fig. 6 Schematic representation of the inference of alighting stops from transfers with walking links [adapted from Gordon (2012)]. A passenger is recorded, in the form of a transaction in the AFC system, as boarding at stop A1 of line route A and subsequently as boarding on stop B4 of line route B. The closest stop of route A, A6, which is serviced by the service run of A recorded as previously being boarded and which is within the search radius of 4 kms, is then inferred as the alighting stop

$$t_{diff,opt} = \min_i |t_{boarding,AFC} - t_{dep,AVL,i}| \tag{5}$$

where $t_{diff,opt}$ is the optimal time difference, $t_{boarding,AFC}$ is the time of boarding according to the AFC dataset, and $t_{dep,AVL,i}$ is the departure time relevant to trip i .

Figure 6 illustrates a specific spatial case of transfer, or activity, inference when different stops are used for alighting and subsequent boarding. In our case, the headway in the peak period of the line route used for the post-intermission boarding was considered in order to discern inter-trip activities from transfers between PT services. Thus, if an intermission (“time gap”) between an alighting and a boarding event—occurring on the same day and meeting the spatial constraints mentioned above—lasted longer than the maximum weekday headway of the line route used for the subsequent boarding, it was regarded as an activity and the alighting stop was the trip destination—otherwise the intermission was regarded as a transfer event that was included in the trip.

A random sample of time gaps is shown in Fig. 7 and its cumulative distribution (CDF) in Fig. 8. As indicated in the scatterplot of the first of these diagrams, a large portion of the gaps were less than 60 min. Although the median (and mode) headway in Fig. 7 is as low as 14 min, most gaps (95 per cent) were classified as activities since time gaps also included date changes, leaving just five percent of them being transfers. This is also attributable to the relatively small selection of line routes in the data, making possible transfers between routes somewhat limited. Consequently, the mean number of trips (trip legs) per travel card ID and day was as low as 1.7 during both panel waves. The second “peak” in Fig. 7 is located around the nine-hour mark (400–650 min), suggesting that these time gaps

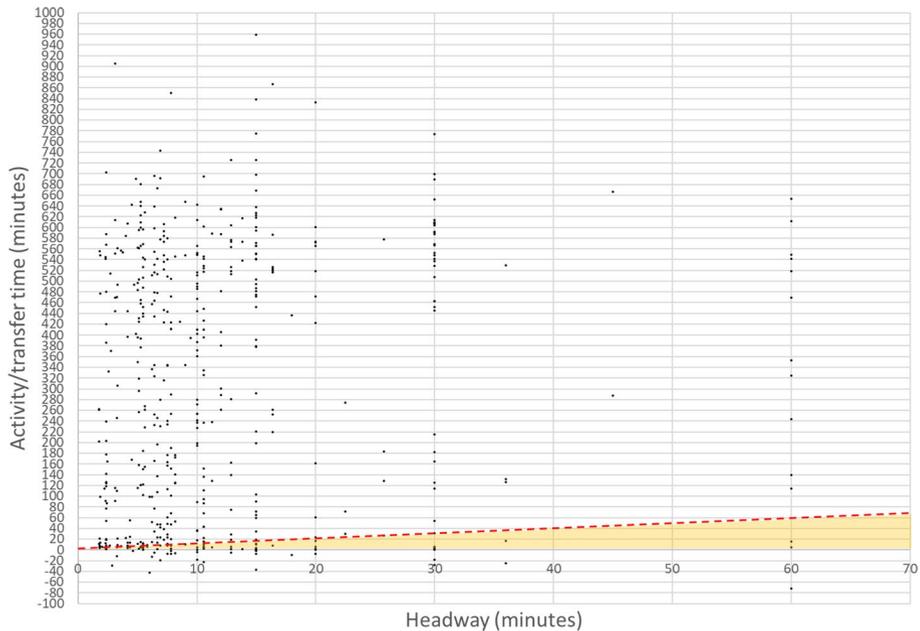


Fig. 7 Time gaps (10 per cent random sample of all weekday observations) between trip elements in minutes, plotted against headways (also in minutes) of the line route used for a subsequent trip, on days with multiple trip elements. The red line and the shaded area below it indicate gaps categorised as transfers

correspond to work or education-related activities. Negative time gaps are due to inconsistencies between boarding time stamps and previous trip alighting stop arrival times that originate from the process of inferring alighting stops or stations in which the observed trips in the travel card data were matched with the service trips in the AVL data.

To control for the use of different stop points for the same ultimate origin or destination (e.g., home or work), street network-congruent walking time matrices for neighbouring stops within 600 m were applied to increase the flexibility of OD pair definitions (inspired by the work of Goulet-Langlois et al. (2016) and with the cut-off distance value based on the empirical findings of Berggren et al. (2019)⁵). Thus, only stops within the cut-off network walking distance to neighbouring stops and present on at least one occasion in the card transaction data were included in order to rule out irrelevant stops.

Identifying recurring trips and matching across time frames (panel waves)

In order to be able to study the panel effect of changes in reliability (punctuality and regularity), it is crucial to obtain trip types that are subject to the impact of changes in line route properties. Thus, only trips fulfilling this criterion during *both measurement periods* were included in the analysis. The rationale here is to control for changes in line route usage that

⁵ The same time frames of 2016 and 2017 and location (south-west Scania, Sweden) were used by Berggren et al. (2019) as the study presented here. However, the recorded trip data from that study were based on a smartphone-based travel survey.

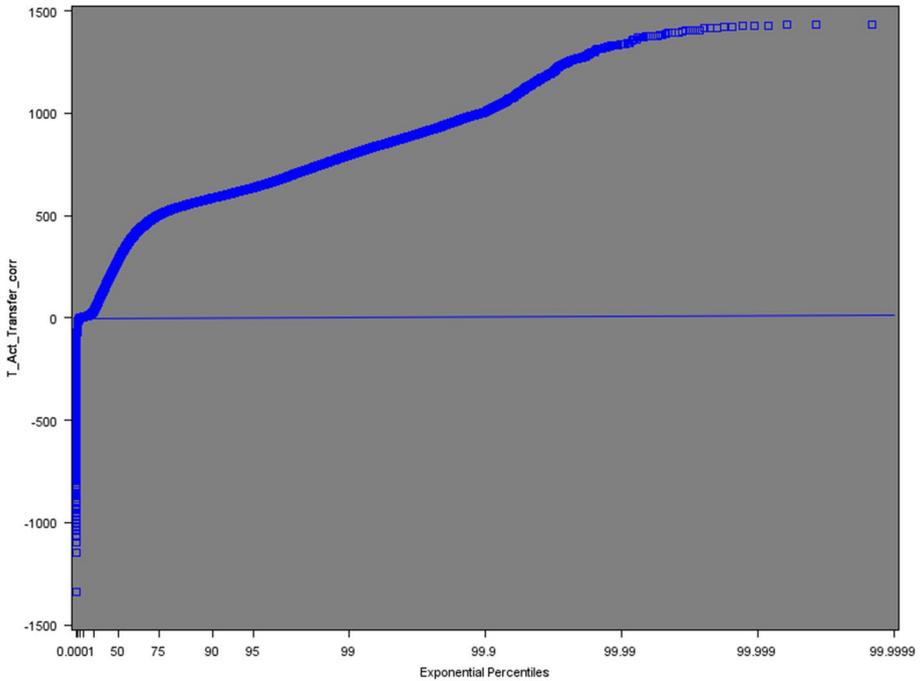


Fig. 8 Distribution (cdf) of time gaps, in minutes, between trip elements, on days with multiple trip elements

Table 2 Simple example of stop clustering structure with three travel cards, two stop clusters (A and C) and six stops

CardID	Stop cluster C —parent stop	Cluster member stop S_C	Walk distance (m) to “parent” cluster stop
1	A	A	0
1	A	B	100
1	C	C	0
1	C	D	150
2	A	A	0
2	A	B	100
2	C	C	0
2	C	D	150
2	C	E	200
3	A	A	0
3	A	B	100
3	A	F	120

are due to exogenous causes, such as a change of residence or employment location. In order to maximise the success rate of the matching of specific trips (travel card IDs using a specific line route in a specific OD pair) across the two panel waves, a procedure of stop clustering was applied, similar to the method applied by Goulet-Langlois et al. (2016). This hierarchical clustering principle is illustrated in Table 2. In this example, card no. 1 only

uses four stops according to the full AFC data set, while card no. 2 uses five stops and card no. 3 a total of three stops. The stops in each set of stops identified as being used for boarding by each card holder were used as candidate stops in the matching of trips per OD stop pair, relaxing the requirement for an exact origin stop-destination stop pair match into an origin cluster-destination cluster match criterium, each cluster involving extended matching options compared to just one origin stop-destination stop pair per trip. Cut-off values for maximum network distances between the stops of each cluster were set to 0.72 and 0.78 km for 2016 and 2017 data, respectively. These values were double the median access/egress walking distances obtained in the survey presented in Berggren et al. (2019).

As mentioned initially in this section, a total of 1,293 cards met the criteria regarding recurrence and cross-wave matching. However, since each individual may use more than one line route recurrently, the total number of cases amounted to 2,076 (cf. Fig. 4). Of these, a change in the relative usage of a line route was found in 437 cases (21 per cent of the total).

In order to study the possibility of differences in behaviour depending on whether the trip was made on a weekday or on a weekend, a day type dummy was introduced based on the date on the timestamp of each boarding event in the transaction (AFC) data. Here, the total number of cases was larger since the trip frequencies per OD pair and card ID were disaggregated based on an additional variable (day type) and in all possible day type combinations across the two measurement periods (2016 weekday vs 2017 weekday, 2016 weekend vs 2017 weekend, 2016 weekday vs 2017 weekend and 2016 weekend vs 2017 weekday). The resulting dataset contained 2,418 cases, of which 452 included a line route change between the 2016 and 2017 measurement periods.

Reliability measures and other covariates

Reliability data was derived from an AVL database covering actual and scheduled departure and arrival times for all stops along the same subset of PT line routes that was identified in the smart card dataset. From the data, regularity and punctuality indices, RI and P , were calculated for each line route, day type (weekday or weekend day), and panel wave time frame using Eqs. 2 and 3, respectively.

To check model robustness in relation to other common path choice-determining trip attributes, the potential covariates of mean OD line route *travel time* and ridership-weighted mean PT *service frequency*⁶ per line route was also appended to each case in the analysis dataset. In addition, *travel time variability* was controlled for as measured by reliability buffer time using the definition specified in Eq. (4) (RBT, Gittens and Shalaby 2015; Uniman et al. 2010).

In this context it should be noted that the set of trips per line route and OD pair used to calculate the travel time distributions was obtained from the 57,351 (2016) and 43,109 (2017) cards present in the third filtering step of AFC data (section "Data, its enrichment and refinement", Fig. 4), meaning that there might be trips included that were made by cards not being present among the recurring trips in both panel waves

⁶ For each line route, day type, and panel wave time frame, the departure frequency of each hour was weighted by the number of boardings according to the AFC data for that hour, line route, time frame, and day type, meaning service frequencies in peak periods were weighted higher than off-peak service frequencies. Only AFC data for the 2016 panel wave were used for this purpose to ensure that the variable was only a function of headway and not of ridership.

(the final filtering step in Fig. 4). However, since all calculations were made per OD pair and line route, only travel time data for relevant recurring OD pairs and line routes were used.

There was a total of 232 cases with a cross-wave increase of line route usage in the data, which constitutes 11 per cent of all cases. The number of cases with decreased line route usage across waves amounted to 8.6 per cent or 179 cases. Thus, a large majority, or 80 per cent of all cases, did not include any change of line route usage. Descriptive statistics of the covariates used in model estimations are displayed in Table 3. As indicated, the mean headway regularity RI_L has generally improved but the mean timetable adherence P_L has decreased across panel waves. However, the median values have the opposite signs for each measure, indicating outlier values with large leverage impacts on the data (e. g. one line route had a punctuality loss of 105 per cent i.e., the proportion of delayed departures increased from 22.6 to 46.4 per cent, while another gained 25 per cent). This outlier effect was also evident in travel times, regarding both mean values and the buffer time required to allow for travel time variability (RBT). Thus, median values were used in order to obtain the most representative view of the data. A total of 1,254 cases displayed increased RBT while 776 cases entailed a decreased need for a travel time buffer.

Viewing these statistics on a line route level, only five out of 20 routes showed an increase in punctuality, while the remaining routes were more delayed in the second than in the first panel wave. A similar picture was evident for headway regularity, in which five line routes improved while 13 had decreased regularity (two routes remained unchanged in this regard). For the ridership-weighted mean service frequency covariate,

Table 3 Descriptive summary statistics for regressors/independent covariates used in the four sub models, based on the total dataset of analysis cases (travel card ID*line route*OD)

Variable	N	Mean	SD	Median	Minimum	Maximum
ΔP_L	2059	-0.04217	0.18971	0.01461	-1.11095	0.2509
ΔRI_L	2059	0.04114	0.08967	-0.00272	-0.09145	0.1623
DiffSupply, WD	2059	0.01962	0.04577	0.00766	-0.0353	0.36576
DiffTravelTime, WD	2059	-0.15891	1.84069	-0.00897	-24.7406	2.91163
Diff RBT, weekday	2059	0.67488	2.28841	0.08845	-1	23.61667

All variables are measured as fractions

Table 4 Pearson’s correlation coefficients for the independent variables/regressors in the total dataset of analysis cases (cardID*line route*OD)

Variable	ΔP_L	ΔRI_L	DiffSupply, WD	DiffTravelTime, WD	Diff RBT, weekday
ΔP_L	1	0.583025	0.129708	-0.03792	-0.19857
ΔRI_L	0.583025	1	-0.037256	0.02627	-0.21205
DiffSupply, WD	0.129708	-0.037256	1	0.004177	-0.01802
DiffTravelTime, WD	-0.03792	0.02627	0.004177	1	0.028379
Diff RBT, weekday	-0.19857	-0.21205	-0.01802	0.028379	1

Significant correlations, on a 0.05 significance level, are indicated in bold

only two line routes had decreased levels while all others had increase departure frequencies across panel waves.

At the significance level of 0.05, the inter-variable correlation between the two measures of reliability was found to be significantly different from zero. Consequently, schedule adherence P and headway regularity RI ranged from 0.58 for Pearson's correlation (Table 4) to 0.69 for Spearman's correlation (not shown). PT supply frequency (the difference in ridership-weighted maximum departures per hour) correlated with P and RI , respectively, only according to the Spearman test, with a coefficient of 0.48. The other variables were not significantly correlated.

Modelling approach

We have used a fixed effects and logistic regression approach to analyse our panel data regarding the degree and magnitude of impact of changed service reliability—punctuality and regularity—on the individual passenger's long-term trade-off between line routes in the given OD pairs. More specifically, we have regressed the likelihood of changing the relative path used for each OD pair and line route, given changes in reliability, as well as other potential covariates, including service frequency, travel time variability, and travel time, across panel waves. In a sense, it is a dual model approach since increased and reduced line route usage were analysed in separate models. These were not of a pure (discrete) choice variant, in that we did not keep track of the complete choice set for each individual passenger (travel card ID) in the models. Instead, inherent to the structure of the panel data was that competing line routes were *not* taken into account, only the impact of the changing attributes of the line route registered at least once by each cardholder was, while the effects of changing attributes of other line routes couldn't be traced. In the panel dataset, the relative frequency of the usage of each line route, travel card ID, and OD stop pair were matched across the two panel waves—each comprising a two-week period for November 2016 and 2017 respectively, along with the line route reliability attributes of each panel wave.

A (Boolean) response variable indicated whether the analysis case implied a change in the relative line route usage frequency—positive, negative or not at all. The regressions were based on a linear logit transformation of the probability p of a certain outcome, as a function of *observed covariates* x_i as shown in Eq. (6):

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \alpha + \sum_k \beta_k \cdot x_k \quad (6)$$

where α is the intercept, β_k is the k :th independent coefficient and x_k is the k :th explanatory variable.

Estimates of the regression parameters α and β_k were given by maximum likelihood estimation.

A total of fourteen different models were tested, as indicated in Table 6. Two subsets of data were evaluated depending on increased or decreased line route ridership and across weekday covariates, respectively (models with acronyms ending with a “+” or “-”, respectively, in Table 6). In order to model positive changes in line route usage (increased ridership proportions), all cases of an increased line route ridership proportion were set to “1” and the remaining cases to “0” for the negative change in line usage (decreased ridership proportions); all cases with a decreased ridership proportion were set to “1” and all

other cases to “0”. The explanatory variables were normalised to a full fraction scale of change, in which for example a 50-per cent decrease in travel time or reliability index value were both set to -0.5. Using the first model pair of M1, we analysed line route switching behaviour in relation to both schedule adherence (P) and headway regularity (RI), as well as the auxiliary covariates of travel time, RBT, and service supply. Using models M2—M3, on the other hand, we were able to analyse the same dependent variable as in M1 but in relation to one reliability measure at a time (Table 6).

Since the reliability measures were found to be correlated, we had each reliability index interact with a dummy variable, indicating cases with short maximum weighted headways⁷ (at or below twelve minutes) in order to find the potential impact of short versus long headways (above twelve minutes). The rationale is to be able to discern the potentially different approaches used by passengers to minimise wait times and scheduling effort for short and long headways, respectively. To establish whether changed service reliability in high frequency line routes, i.e. routes with a maximum twelve-minute headway, imposed a different behavioural response than low frequency routes, a test was conducted using a dummy for analysis cases involving high-frequency line routes. In the models M4–M5, the reliability measures interacted with this dummy.

The influence of day type (weekdays versus weekends) on long-term change responses was tested in a data subset in which the day type for each relative line route usage was considered using a dummy variable for weekend days, which interacted with each reliability measure (M6–M7). In an additional model, not shown in Table 6 due to lack of space, we tested the difference in propensity to change line route usage across day types and panel waves regardless of the direction of change and without any additional covariates regarding service reliability, etc.

The fit of the different models was measured using McFadden’s *pseudo* R^2 (McFadden 1973) and the receiver operating characteristic (ROC). The former indicates the respective model’s ability to explain the variation in the response variable, and the latter the sensitivity vs. fall-out rate regarding the models’ ability to detect responses (Hand and Till 2001). *Pseudo* $R^2 < 0.05$ indicates low fit, *Pseudo* $R^2 > 0.20$ indicates a very good fit, and *Pseudo* $R^2 > 0.40$ is rarely observed (Andreß et al. 2013).

To check the validity of our research hypotheses regarding PT service reliability and passengers’ long-term trade-off between line routes, we evaluated the logistic regression coefficients for each independent variable and calculated a measure of sensitivity for the impact of each independent variable on the propensity to change line route—both regarding increased and decreased proportional ridership. The sensitivity S , evaluated at each regressor β_k , indicates the marginal change in the probability of changing the proportional use of a line route with respect to a marginal change of that specific covariate. The effect of intercept α is subtracted according to Eq. (7) to remove the effects of unexplained variation on the dependent variable.

$$S = \frac{e^{(\alpha+\beta_k)}}{1 + e^{(\alpha+\beta_k)}} - \frac{e^\alpha}{1 + e^\alpha} \quad (7)$$

⁷ The selection condition for an analysis case is valid only if at least one of the time frames contains a line route with a certain ridership-weighted headway for that case.

Results

The estimation results of the fourteen feasible models, presented in Table 6, do support some parts of our research hypothesis—but not in all possible trip conditions and not for all reliability indicators applied. In this section, we first present the model fit and outcome from each model, followed by a joint interpretation and discussion regarding the general picture that the results convey.

The model with by far the best fit was model M1-, which resulted in a McFadden's *pseudo R*² equal to 0.154 and an area under the ROC curve of 0.81, while model M6+ had the worst fit with a McFadden's *pseudo R*² of 0.04 and an area under the ROC curve of 0.57. Since we assumed that there could be a certain degree of collinearity between a few of the auxiliary reliability measures (RI and *P* had a collinearity *R*² value of 0.3 and RBT and RI had a *R*² value of 0.05), the results of the multi-variate regression models were somewhat difficult to interpret. Thus, in our analysis, we focused on the simple models of M2 through M7 to isolate the variables and thus implicitly accepted a lower goodness of fit.

Strategic choice behaviour regardless of line route headway

When regressing the probability of an increased line route usage rate against both line route regularity and punctuality, as in model M1+ (Fig. 9), only regularity emerged as a significant driver towards a higher propensity to choose these line routes, while punctuality appeared to work the opposite way (Table 6). In terms of relative impact, the sensitivity (Table 5) to the regularity *RI*_L entailed an increased long-term choice propensity of 0.89 units for each unit of increased regularity, while the corresponding (and reverse) impact of punctuality was ten times smaller.

Both reliability measures negatively impacted on decreased line route usage (Table 6) according to model M1- (Fig. 10), which appears reasonable since the increased reliability of a line route should counteract a passenger's propensity to switch to a different route.

Also, increased RBT positively impacted decreased line route ridership, which appear to be reasonable given the potential disutility of a longer travel time buffer to compensate for the increased travel time uncertainty captured by the RBT measure. However, increased RBT also affected increased line route usage positively, which is less intuitive.

The two other covariates of M1, Travel time change and Departure frequency change had a less significant and at least in part unintuitive impact on line route switching

Table 5 Sensitivity to the marginal probability of changing (increasing or reducing, respectively) the proportionate line route ridership in relation to each independent weekday variable in model M1

Regressor	Sensitivity value for increased use (M1+)	Sensitivity value for decreased use (M1-)
Change in <i>RI</i> _L	0.90	-0.05
Change in <i>P</i> _L	-0.08	-0.05
Change in travel time	0.01	-0.001
Change in service frequency	-0.08	0.59
Change in RBT	0.005	0.003

behaviour. Thus, increased departure frequency in an OD entailed a reduced propensity to increase line route usage, and increased travel time entailed a reduced propensity to decrease line route usage. No significant effect on increased line route usage was found from increased OD travel time.

Sensitivities (Table 5) specify the propensities at around equal, with a five percentage-unit decreased likelihood of a negative change in the line route usage for each unit of reliability increase.

Models M2 and M3 indicate the impact on the line route switching behaviour of each reliability measure separately. Comparing M2+ with M1+, regularity in itself had minimal explanatory power regarding increased line route usage compared to when it was accompanied by punctuality, travel time, and service supply attributes. In M2–, however, regularity had a stronger effect than when it was combined with the other line route attributes of model M1–. However, when punctuality in model M3+ was correspondingly analysed, it displayed a counter-intuitive negative impact on increased line route usage, but a somewhat stronger positive (decreased incidence of decreased line route usage) impact in model M3–.

Strategic choice behaviour and headway categories

The systematic impact of regularity on line route switching behaviour was most pronounced for line routes that had at least five departures per hour (HW1), as indicated by the signs and values of the parameters for *RI* in model M4 in Table 6. Accordingly, there was a positive impact of *RI* on increased line route usage (M4+) and a negative impact on decreased line route usage (M4–). However, for line routes that had less than five departures per hour (HW2), there were no such indications of a consistent impact since *RI* had a negative influence on both increased and decreased line route usage. A corresponding pattern was evident in model M5, in which the impact of punctuality *P* on line route usage was analysed (Table 6).

Influence of day type

Finally, the last models of Table 6, M6 and M7, analysed the effects of day type—week-day or weekend day—on the propensity to change the proportional use of line routes in a positive or negative direction. The results did not indicate a significantly different outcome regarding how reliability affected long-term line route choice depending on day type.

However, from the results of the additional day type model (not present in Table 6), in which neither the direction of change regarding line route usage nor reliability differences across panel waves and day types were taken into account, there was evidence of a difference in line route switching behaviour between weekdays and weekends. In this model, and depending on the direction of measurement, a significant parameter value of 0.76 for line route usage change between weekdays and weekend days and, correspondingly, values of -1.6 to -1.8 for differences between weekend days and weekdays, indicated a more pronounced switching behaviour during weekends than during weekdays. This difference across day types in “sticking” to a specific line route or path is in line with previous research by Kim et al. (2017), who also showed that PT passengers are more prone to sticking to specific paths on weekdays than on weekends. However, it should be noted that their analysis covered a period of nine consecutive weekdays and two consecutive weekends, while our data spans one year.

Table 6 Results from the logistic regression models (M1 to M7) in terms of parameter estimates. Significance at the 0.05 level is indicated in bold

Regressor type	M1+	M1-	M2+	M2-	M3+	M3-	M4+	M4-	M5+	M5-	M6+	M6-	M7+	M7-
Reliability														
Intercept	-2.39	-2.90	-1.99	-2.24	-2.26	-2.81	-2.21	-2.14	-2.77	-2.90	-2.07	-2.28	-2.27	-2.86
Change in regularity index	6.28	-7.03	0.32	-10.10	-	-	-	-	-	-	-0.13	-10.22	-	-
Change in punctuality index	-3.26	-3.91	-	-	-2.28	-4.10	-	-	-	-	-	-	-2.25	-4.20
Change in regularity index, HW1	-	-	-	-	-	-	1.57	-9.86	-	-	-	-	-	-
Change in regularity index, HW2	-	-	-	-	-	-	-8.60	-5.06	-	-	-	-	-	-
Change in punctuality index, HW1	-	-	-	-	-	-	-	-	4.98	-2.57	-	-	-	-
Change in punctuality index, HW2	-	-	-	-	-	-	-	-	-4.16	-4.44	-	-	-	-
Change in regularity index, WE	-	-	-	-	-	-	-	-	-	-	204.6	42.98	-	-
Change in punctuality index, WE	-	-	-	-	-	-	-	-	-	-	-	-	-204.9	-14.43
Travel time														
Travel time change	0.12	-0.03	-	-	-	-	-	-	-	-	-	-	-	-
Change in RBT	0.06	0.05	-	-	-	-	-	-	-	-	-	-	-	-
Service supply														
Change in service frequency	-19.26	3.49	-	-	-	-	-	-	-	-	-	-	-	-
McFadden's pseudo R ²	0.15	0.00008	0.07	0.04	0.11	0.01	0.07	0.07	0.11	0.007	0.08	0.04	0.12	
Number of observations	2,059	2,076	2,076	2,076	2,076	2,076	2,076	2,076	2,076	2,125	2,125	2,125	2,125	

HW1, HW2—variables interacted with dummies for line routes with headways at or below or above twelve minutes, respectively. WE—variable interacted with the dummy for weekend days

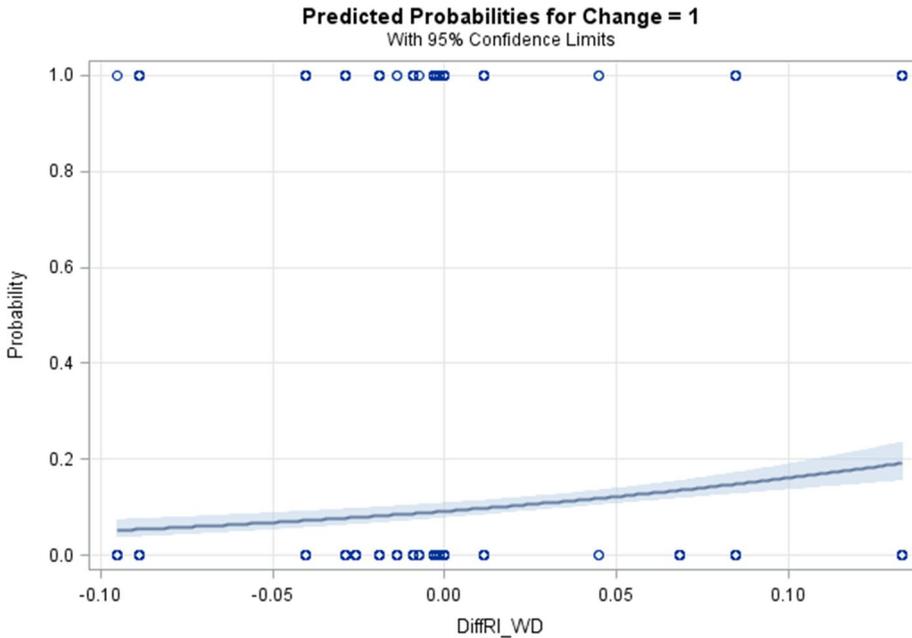


Fig. 9 Visualisation of model M1+ by which the probabilities of positive ridership change are regressed against the weekday RI_L difference (DiffRI_WD), with the other regressors (displayed in \backslash^* MERGEFORMAT Table 3) fixed at their median values: Change in P_L = -- 0.04, Change in travel time = -- 0.01, Change in RBT = 0.08 and Change in service frequency = 0.008

To summarise our findings, the regression analyses indicated that headway regularity appeared to have the most significant impact on line route switching behaviour regardless of service frequency, in which the greatest influence of improvements on regularity is to attract more trips, whereas the repelling effect of a decline in regularity is less articulated. For punctuality, on the other hand, the clear and positive impact was restricted to line routes with at least five departures per hour. Thus, our data give some support to somewhat different kinds of adaptation behaviour to changes in reliability depending on headway, although only for punctuality. The reader should note that the data itself indicated a general flux of passengers from low-frequency train routes to high-frequency bus routes across our panel, illustrated in Table 7. Since the train routes in our AVL dataset experienced a general decline in both regularity and punctuality during this period, while the most popular bus routes displayed improved reliability figures, this shift of passengers is perfectly analogous to an expected passenger response to reliability change. It also sheds some explanatory light on the somewhat ambiguous impact of the reliability of low-frequent line routes on changes in their usage.

Discussion

We applied average line route-specific departure reliability metrics by day type in order to analyse marginal changes in the relative passenger demand per line route. The main rationale was to get a generalised measure that targeted the general impression that passengers

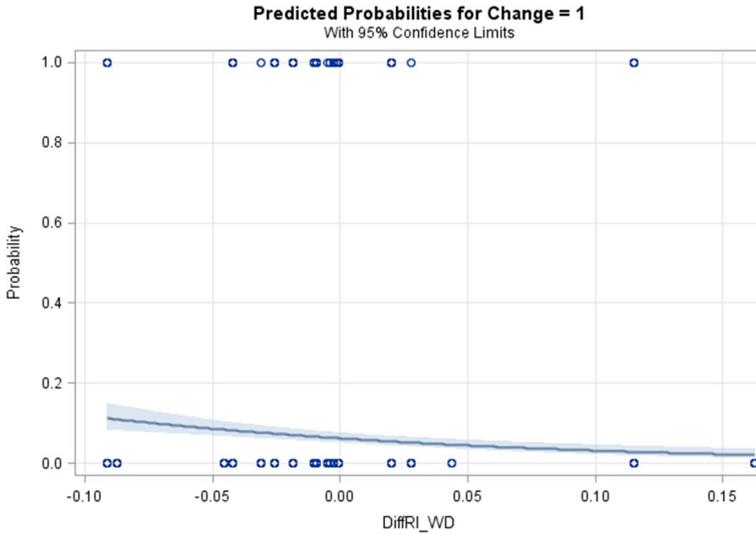


Fig. 10 Visualisation of model **M1**- by which the probabilities of **negative** ridership change are regressed against the weekday R_{I_L} difference (DiffRI_WD), with the other regressors (as displayed in \backslash^* MERGEFORMAT Table 3) fixed at their median values: Change in $P_L = -0.04$, Change in travel time = -0.01 , Change in RBT = 0.08 and Change in service frequency = 0.008

Table 7 Mean values of relative line route usage frequencies per panel wave, for cases with changed relative frequencies across panel waves, subdivided by headway group

	Year 1 2016	Year 2 2017	Net change	
HW1 (≤ 12 min)	100% bus trip cases		Increased:	10.32%
	Mean relative frequency:		Decreased:	2.02%
	63.12%	87.40%	Difference:	8.30%
HW2 (> 12 min)	80% train trip cases		Increased:	14.27%
	Mean relative frequency:		Decreased:	15.86%
	61.26%	58.08%	Difference:	-1.59%

The distribution of changes across panel waves (years) are indicated in the rightmost column. HW1—line route headways at and below 12 min, HW2—line route headways above 12 min. The blue arrow indicates the general trend of line route usage change, which was from low-frequency to high-frequency line routes across panel waves

are expected to have of the service reliability of the line routes that they use on a recurring basis. Thus, the exact reliability of specific vehicular trips or on particular days or time periods was not of interest here, due to the increasing difficulties by which to associate each passengers’ line route perception to these disaggregate metrics. It would be of interest for future research efforts to look into whether passengers make trade-offs across line routes differently depending on e.g., travel direction and time of day, but in this paper, we restricted the analysis to focus only on day type specific trade off patterns.

In this study, indicators of revealed choice constituted the target of measurement. However, as noted, the explanatory power is rather low, and the model suffers from multicollinearity between some of the independent variables. These phenomena are not uncommon in models of human behaviour relying on the kind of empirical mobility data used in this

paper but may be possible to manage more directly in an experimental setting such as that applied when recording stated choices of selected respondents in a controlled environment.

As our results suggest, the positive impact of improved service departure reliability (and negative impact from decreased reliability) on long-term line route choice probabilities is confined to cases where (a) departure reliability is being measured using line route headway regularity only, regardless of service frequency, or (b) where regularity and punctuality are only regarded for line routes with an hourly service frequency of at least five departures. To some extent, these results challenge the consensus among researchers and practitioners regarding the importance of schedule adherence for low-frequency services. However, we believe that these results may be attributed to properties of the data, particularly the phenomenon of a general decline in punctuality among less frequent line routes described in the last paragraph of Chapter 5 (and indicated in Table 7). In addition, there may be other factors at play that were not captured by the applied model framework, such as factors related to social choice as well as pure randomness. Nevertheless, the relationships between a relative change of these two indicators of departure reliability and travel demand warrant further research efforts to bring more clarity to the issue.

We believe that a number of behavioural phenomena may be at play in the data. Primarily, the relatively few cases that actually included a line route change across panel waves (21 per cent of all cases, 54 per cent of cases in which there were multiple distinct alternative line routes for the particular case OD) may indicate a kind of endowment effect (Kahneman et al. 1991), or satisficing behaviour, i.e., a reluctance on the part of a passenger to change away from a route with which (s)he is sufficiently content (Kahneman and Krueger 2006; Kaufman 1990). On the other hand, people who actually responded to reliability changes across years may display some aspect of risk-averse behaviour, although the net impact from reliability change on behaviour was relatively weak in our data. There is also the issue of perception, or sensitivity, to both absolute reliability and changes in reliability (Friman et al. 2001), in which we have focused on common deviations from the timetable—punctuality and regularity—and have left the potential behavioural effects of large-scale disruptions outside the scope of our analysis. However, as also described by Yap et al. (2017), the long-term effects of what Friman et al. call “remembered frequency of negative critical incidents” may also have substantial effects on long-term PT path choice. Our analysis has focused on what Carrel et al. (2013) call the trip planning phase; in that we have analysed long-term behavioural change. However, it is not possible for us to truly disentangle these effects from the short-term decisions that are made based on events occurring en route. Having said that, no large-scale interruptions were reported by the transport provider, nor by the media during the two periods of our data collection.

A potential cause for confounding effects is the modification to the PT route network that occurred between the panel waves, and which affected six of the line routes included in our study. Our approach to mitigating these effects included the clustering of origin and destination stops based on walking distance. However, there may be other effects on behaviour from this event, the mechanisms of which have not been fully captured by our models.

Another phenomenon that may have a potentially counteracting effect on behavioural response to reliability deficiencies is the occurrence of bus “overtakings”, as reported by Barabino et al. (2015). This feature of bus operations, typically, entails a re-ordering of bus trajectories which may result in severe delays being recorded on some vehicle trips in the AVL system, but which are not perceived as serious by passengers if their total travel time is not negatively affected compared to their expected travel time, including the levels of actual and expected total trip time predictability. Fortunately, this phenomenon tends to occur quite infrequently in our AVL data when comparing scheduled and actual vehicle

running times. A total of 2.5 per cent of all stop or station departure times, regardless of the day type or line route, included a vehicular trip that had been subject to re-ordering of the actual running sequence of buses—and this applies to both data collection periods.

Conclusion

In the study described in this paper, we applied a panel approach in order to study the long-term behavioural impact of changes on the level of PT service departure reliability. We used a panel of revealed trip data from tap-on travel card transactions of an AFC system in south-west Scania, Sweden, and combined them with true vehicle trajectories from AVL data covering the same line routes and time frames. The common challenge of obtaining full OD trips with data lacking tap-off information was approached by means of an alighting and transfer stop inference procedure. Despite the unavoidable uncertainties that such an inference procedure imposed on the observations, the results appear reasonable given the additional measurement noise. The main rationale behind the chosen panel approach was to enable the controlling for factors other than line route-based changes in service reliability. By using a stop clustering procedure and keeping both origin and destination stop clusters constant for each individual passenger (card identifier) between the panel waves, we were able to reduce the impact of other causes of path change, such as changes in the individual passenger's locations of origin or destination (related to, for example, change of residential or employment location). Other line route-based attributes such as travel times and service frequency were controlled for in the model formulation in order to elicit the relative impact of departure reliability on long-term path choice.

Our research hypothesis regarding the behavioural impact of departure reliability change in PT is rather bold, yet straightforward. Although we found that passengers in general tend to shift from marginally unreliable to marginally more reliable line routes, this appears not to include all passengers (which could be assumed from a literal interpretation of our research hypothesis), nor all service frequencies. In fact, the majority of cases we studied showed an indifferent reaction on the part of the passenger to changes in reliability. However, the behavioural response foreseen in the second part of the research hypothesis is clearly supported by our findings. Thus, results suggest a positive impact on long-term choice probabilities from an increased level of service reliability on the line route level, but a consistent impact may only be found for only one of the measures applied in this paper; headway regularity, and only for cases with at least a certain minimum level of service frequency. In general terms, the main takeaway of this study would be that service reliability has a significant impact on the strategic path choice of PT passengers according to the panel data based evidence presented in this paper, although with certain limitations.

The results presented in this paper should have important implications for modelling and forecasting the demand for public transport, and ultimately on PT network design. This would primarily concern the inclusion of line route reliability—particularly headway regularity—in path choice algorithms and—eventually—mode choice utility functions. In addition to the well-researched measure of RBT, penalties or substitution rates attributable to reliability will have to be estimated and tested in order to accurately model passengers' response to reliability changes in PT. These would constitute a couple of directions for future research that are within immediate reach. However, the subject of behavioural long-term responses to changes in PT service reliability is surprisingly uncharted terrain in transportation research, in which most studies thus far have focused on short-term effects.

In addition, the way in which different passenger groups perceive and respond to irregularities and delays in PT constitutes another area of great interest, although there are few examples of large-scale studies of this kind of disaggregate revealed behaviour to date.

Acknowledgements The research presented in this paper was made possible thanks to two sources of funding: The Swedish Transport Administration (grant no. TRV 2016/10730) and the EU Interreg programme Öresund-Kattegat-Skagerrak, specifically the EU project: NYPS 202015553. The authors wish to express their warm gratitude to these organisations. In addition, we are most grateful to the public transport provider in the region of Scania (Skånetrafiken) for their kind contribution of travel card transaction data as well as real-time vehicle location data. Two anonymous reviewers contributed significantly in the improvement of this paper. This research was performed in collaboration with K2—The Swedish Knowledge Centre for Public Transport.

Authors contributions UB: Conceptualisation, Data curation, Formal analysis, Methodology, Writing—original draft; CD: Conceptualisation, Methodology, Software, Supervision, Validation, Writing—review and editing; HS: Funding acquisition, Project administration, Supervision, Validation, Writing—review and editing; KBF: Methodology, Supervision, Validation, Writing—review and editing.

Funding Open access funding provided by Lund University. The Swedish Transport Administration (grant no. TRV 2016/10730). EU Interreg programme Öresund-Kattegat-Skagerrak, EU project no. NYPS 202015553.

Declarations

Conflicts of interest Authors declare that they have no conflict of interest.

Availability of data and material All data, in anonymised form if appropriate, available from the corresponding author upon request.

Code availability All software code (sas/sql) available from the corresponding author upon request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdel-Aty, M., Kitamura, R., Jovanis, P.P.: Investigating effect of travel time variability on route choice using repeated-measurement stated preference data. *Transp. Res. Rec.* **1493**, 39–45 (1995)
- An, J., Liu, Y., Yang, X.: Measuring route-level passenger perceived transit service reliability with an agent-based simulation approach. *Transp. Res. Record J. Transp. Res. Board* **2415**, 48–58 (2014). <https://doi.org/10.3141/2415-05>
- Andreß, H.-J., Golsch, K., Schmidt, A.W.: *Applied Panel Data Analysis For Economic and Social Surveys*. Springer Verlag, Berlin (2013)
- Avineri, E., Prashker, J.N.: Violations of expected utility theory in route-choice stated preferences certainty effect and inflation of small probabilities. *Transp. Res. Record* **1894**, 222–229 (2004)
- Bagherian, M.C., van Oded, O., Niels, H.M.: Measuring passenger travel time reliability using smart card data. In: Paper presented at the TRISTAN IX: Triennial Symposium on Transportation Analysis, Oranjestad, Aruba (2016)

- Barabino, B., Di Francesco, M., Mozzoni, S.: Rethinking bus punctuality by integrating automatic vehicle location data and passenger patterns. *Transp. Res. Part A Policy Pract.* **75**, 84–95 (2015). <https://doi.org/10.1016/j.tra.2015.03.012>
- Barry, J.J., Freimer, R., Slavin, H.: Use of entry-only automatic fare collection data to estimate linked transit trips in New York City. *Transp. Res. Record J. Transp. Res. Board* **2112**(1), 53–61 (2009). <https://doi.org/10.3141/2112-07>
- BatesJohnPeter, J.P.J.C.A.: The valuation of reliability for personal travel. *Transp. Res. Part E Logist. Transp. Rev.* **37**, 191–229 (2001)
- Beirão, G., Sarsfield Cabral, J.A.: Understanding attitudes towards public transport and private car: a qualitative study. *Transp. Policy* **14**(6), 478–489 (2007). <https://doi.org/10.1016/j.tranpol.2007.04.009>
- Berggren, U., Johnsson, C., Svensson, H., Wretstrand, A.: Exploring waiting times in public transport through a semi-automated dedicated smartphone app survey. *Travel Behaviour and Society* **15**, 1–14 (2019). <https://doi.org/10.1016/j.tbs.2018.11.002>
- Board, T.R., National Academies of Sciences, Engineering and Medicine: *Transit Capacity and Quality of Service Manual*, 3rd edn. The National Academies Press, Washington (2013)
- Carrel, A., Halvorsen, A., Walker, J.: Passengers' perception of and behavioral adaptation to unreliability in public transportation. *Transp. Res. Record J. Transp. Res. Board* **2351**, 153–162 (2013). <https://doi.org/10.3141/2351-17>
- Chatterjee, K.: Modelling the dynamics of bus use in a changing travel environment using panel data. *Transportation* **38**(3), 487–509 (2010). <https://doi.org/10.1007/s11116-010-9312-y>
- Chu, K.K.A., Chapleau, R.: Enriching archived smart card transaction data for transit demand modeling. *Transp. Res. Record J. Transp. Res. Board* **2063**(1), 63–72 (2008). <https://doi.org/10.3141/2063-08>
- Chu, K.K.A., Lomone, A.: Reproducing Longitudinal in-vehicle traveler experience and the impact of a service reduction with public transit smart card data. *Transp. Res. Record J. Transp. Res. Board* **2541**, 81–89 (2016). <https://doi.org/10.3141/2541-10>
- Currie, G. Douglas, N., Kearns, I.: An assessment of alternative bus reliability indicators. In: Paper Presented at the Australasian Transport Research Forum, Perth, Australia (2011). 09/28–09/30/2011
- dell'Olio, L., Ibeas, A., Cecin, P.: The quality of service desired by public transport users. *Transp. Policy* **18**(1), 217–227 (2011). <https://doi.org/10.1016/j.tranpol.2010.08.005>
- Diab, E.I., Badami, M.G., El-Geneidy, A.M.: Bus Transit service reliability and improvement strategies: integrating the perspectives of passengers and transit agencies in North America. *Transp. Rev.* **35**(3), 292–328 (2015). <https://doi.org/10.1080/01441647.2015.1005034>
- Dixit, M., Brands, T., van Oort, N., Cats, O., Hoogendoorn, S.: Passenger travel time reliability for multimodal public transport journeys. *Transp. Res. Record J. Transp. Res. Board* **2673**(2), 149–160 (2019). <https://doi.org/10.1177/0361198118825459>
- Evans, J.S.: Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* **59**, 255–278 (2008). <https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Fadaei, M., Cats, O.: Evaluating the impacts and benefits of public transport design and operational measures. *Transp. Policy* **48**, 105–116 (2016). <https://doi.org/10.1016/j.tranpol.2016.02.015>
- Farzin, J.M.: Constructing an automated bus origin-destination matrix using farecard and global positioning system data in São Paulo, Brazil. *Transp. Res. Record J. Transp. Res. Board* **2072**(1), 30–37 (2008). <https://doi.org/10.3141/2072-04>
- Frestad Nygaard, M., Tørset, T.: Waiting time strategy for public transport passengers. In: Paper Presented at the Trafikdage, Aalborg, Denmark (2016)
- Friman, M., Edvardsson, B., Gärling, T.: Frequency of negative critical incidents and satisfaction with public transport services. *J. Retail. Consum. Serv.* **8**, 95–104 (2001). [https://doi.org/10.1016/S0969-6989\(00\)00003-5](https://doi.org/10.1016/S0969-6989(00)00003-5)
- Furth, P.G., Muller, T.H.J.: Service reliability and hidden waiting time—insights from automatic vehicle location data. *Transp. Res. Rec.* **1955**, 79–87 (2006)
- Gittens, A., Shalaby, A.: Evaluation of bus reliability measures and development of a new composite indicator. *Transp. Res. Record J. Transp. Res. Board* **2533**(1), 91–99 (2015). <https://doi.org/10.3141/2533-10>
- Gordon, J.B.: *Intermodal Passenger Flows on London's Public Transport Network Automated Inference of Full Passenger Journeys Using Fare-Transaction and Vehicle-Location Data*. (PhD). Massachusetts Institute of Technology, Boston (2012)
- Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, J.: Inferring patterns in the multi-week activity sequences of public transport users. *Transp. Res. Part c Emerg. Technol.* **64**, 1–16 (2016). <https://doi.org/10.1016/j.trc.2015.12.012>

- Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, Z., Zhao, J.: Measuring regularity of individual travel patterns. *IEEE Trans. Intell. Transp. Syst.* **19**(5), 1583–1592 (2018). <https://doi.org/10.1109/tits.2017.2728704>
- Gärling, T.A., Kay, W.: Introduction: habitual travel choice. *Transportation* **30**, 1–11 (2003)
- Hand, D.J., Till, R.J.: A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **45**, 171–186 (2001)
- Heiner, R.A.: The origin of predictable behavior. *Am. Econ. Rev.* **73**(4), 560–595 (1983). Retrieved from <https://www.jstor.org/stable/1816559>
- Ingvardson, J.B., Nielsen, O.A., Raveau, S., Nielsen, B.F.: Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: a smart card data analysis. *Transp. Res. Part c Emerg. Technol.* **90**, 292–306 (2018). <https://doi.org/10.1016/j.trc.2018.03.006>
- Jenelius, E.: Experienced public transport service reliability—integrating travel time and travel conditions. In: Paper Presented at the 7th International Symposium on Transport Network Reliability (INSTR), Sydney, Australia (2018). 04/09/2018
- Kahneman, D., Knetsch, J.L., Thaler, R.H.: Anomalies: the endowment effect, loss aversion, and status quo bias. *J. Econ. Persp.* **5**(1), 193–206 (1991). <https://doi.org/10.1257/jep.5.1.193>
- Kahneman, D., Krueger, A.B.: Developments in the measurement of subjective well-being. *J. Econ. Persp.* **20**(1), 3–24 (2006)
- Kahneman, D., Tversky, A.: prospecttheory: an analysis of decision under risk. *Econometrica*, **47**(2), 263–292 (1979). Retrieved from <https://www.jstor.org/stable/1914185>
- Kaufman, B.E.: A new theory of satisficing. *The Journal of Behavioral Economics* **19**(1), 35–51 (1990)
- Kim, J., Corcoran, J., Papamanolis, M.: Route choice stickiness of public transport passengers: measuring habitual bus ridership behaviour using smart card data. *Transp. Res. Part c Emerg. Technol.* **83**, 146–164 (2017). <https://doi.org/10.1016/j.trc.2017.08.005>
- Kurauchi, F., Schmöcker, J.D.E.: *Public Transport Planning with Smart Card Data* (Kurauchi F., Schmöcker J.-D., eds.). CRC Press, Boca Raton (2017).
- Le, H.T.K., Carrel, A.L., Li, M.: How much dissatisfaction is too much for transit? Linking transit user satisfaction and loyalty using panel data. *Travel Behav. Soc.* **20**, 144–154 (2020). <https://doi.org/10.1016/j.tbs.2020.03.007>
- Li, Z., Hensher, D.A.: Behavioural implications of preferences, risk attitudes and beliefs in modelling risky travel choice with travel time variability. *Transportation* **40**(3), 505–523 (2012). <https://doi.org/10.1007/s11116-012-9445-2>
- Liu, Y., Bunker, J., Ferreira, L.: Transit users' route-choice modelling in transit assignment: a review. *Transp. Rev.* **30**(6), 753–769 (2010). Doi: <https://doi.org/10.1080/01441641003744261>. Retrieved from <https://trid.trb.org/view/1082792>
- Luethi, M.W., Weidmann, U., Nash, A.: Passenger arrival rates at public transport stations. In: Paper Presented at the Transportation Research Board 86th Annual Meeting No. 07–063., Washington (2007)
- McFadden, D.: Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (ed.) *Frontiers in Econometrics*, pp. 105–142. Academic Press, New York (1973)
- Munizaga, M.A., Palma, C.: Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. *Transp. Res. Part c Emerg. Technol.* **24**, 9–18 (2012). <https://doi.org/10.1016/j.trc.2012.01.007>
- Navick, D.S., Furth, P.G.: estimating passenger miles, origin-destination patterns', and loads with location-stamped farebox data. *Transp. Res. Rec.* **1799**(02–2466), 107–113 (2002)
- Noland, R.B., Polak, J.W.: Travel time variability: a review of theoretical and empirical issues. *Transp. Rev.* **22**(1), 39–54 (2002). <https://doi.org/10.1080/01441640010022456>
- Osuna, E.E., Newell, G.F.: control strategies for an idealized public transportation system. *Transp. Sci.*, **6**(1), 52–72 (1972). Retrieved from <http://www.jstor.com/stable/25767635>
- Ouellette, J.A., Wood, W.: Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. *Psychol. Bull.* **124**(1), 54–74 (1998). <https://doi.org/10.1037/0033-2909.124.1.54>
- Rudnicki, A.: Measures of regularity and punctuality in public transport operation. *IFAC Proc. Vol.* **30**(8), 661–666 (1997). [https://doi.org/10.1016/s1474-6670\(17\)43896-1](https://doi.org/10.1016/s1474-6670(17)43896-1)
- Seaborn, C., Attanucci, J., Wilson, N.H.M.: Analyzing multimodal public transport journeys in london with smart card fare payment data. *Transp. Res. Record J. Transp. Res. Board* **2121**(1), 55–62 (2009). <https://doi.org/10.3141/2121-06>
- Thøgersen, J.: Understanding repetitive travel mode choices in a stable context: a panel study approach. *Transp. Res. Part a Policy Pract.* **40**(8), 621–638 (2006). <https://doi.org/10.1016/j.tra.2005.11.004>

- Trépanier, M., Tranchant, N., Chapleau, R.: Individual trip destination estimation in a transit smart card automated fare collection system. *J. Intell. Transp. Syst.* **11**(1), 1–14 (2007). <https://doi.org/10.1080/15472450601122256>
- Trompet, M., Liu, X., Graham, D.J.: Development of key performance indicator to compare regularity of service between urban bus operators. *Transp. Res. Record J. Transp. Res. Board* **2216**(1), 33–41 (2011). <https://doi.org/10.3141/2216-04>
- Tversky, A., Kahneman, D.: The framing of decisions and the psychology of choice. *Science* **211**(4481), 453–458 (1981). <https://doi.org/10.1126/science.7455683>
- Uniman, D.L., Attanucci, J., Mishalani, R.G., Wilson, N.H.M.: Service reliability measurement using automated fare card data. *Transp. Res. Record J. Transp. Res. Board* **2143**(1), 92–99 (2010). <https://doi.org/10.3141/2143-12>
- van Oort, N.: Service reliability and urban public transport design. (PhD). TU Delft, Delft. (T2011/2) (2011)
- Wang, W., Attanucci, J.P., Wilson, N.H.M.: Bus passenger origin-destination estimation and related analyses using automated data collection systems. *J. Public Transp.* **14**(4), 131–150 (2011)
- Verplanken, B., Aarts, H., van Knippenberg, A., Moonen, A.: Habit versus planned behaviour: a field experiment. *Br. J. Soc. Psychol.* **37**, 111–128 (1998)
- Watkins, K., Queen, C., Berrebi, S.: Minutes matter: a bus transit service reliability guidebook (2020)
- Yap, M.D., Cats, O., van Oort, N., Hoogendoorn, S.P.: A robust transfer inference algorithm for public transport journeys during disruptions. *Transp. Res. Procedia* **27**, 1042–1049 (2017). <https://doi.org/10.1016/j.trpro.2017.12.099>
- Zhao, J., Rahbee, A.: Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Comput. Aid. Civil Infrastruct. Eng.* **22**, 376–387 (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ulrik Berggren is a PhD candidate of the Faculty of Engineering at Lund University. He has a master degree in civil engineering from Luleå university of technology and has worked as a consultant in transport planning for a decade before returning to academia. His main research interest regards how to understand and forecast passenger behaviour in public transport systems.

Carmelo D'Agostino is a Post-Doctoral Fellow at Lund University. His research focuses on the quantitative measurement and analysis of behavioural aspects of traffic safety, particularly for vulnerable road users.

Helena Svensson is a Lecturer at Lund University and education manager at K2, the Swedish National Knowledge Centre for Public transport, responsible for the Bachelor programme in transport and road and rail and her research interests are mobility, accessibility and preferences among older people and people with disabilities.

Karin Brundell Freij is a Professor (part time, adjunct) at Lund University, affiliated to K2, the Swedish National Knowledge Centre for Public transport. She is also a senior consultant with WSP Sweden. Over the last 30 years, she has published papers in a variety of journals on a wide range of issues within the fields of Transport Analysis, Transport Modelling and Transport Policy.