Estimation of a route choice model for urban public transport using smart card data

Ľudmila Jánošíková, Jiří Slavík & Michal Koháň

Department of Transportation Networks, Faculty of Management Science and Informatics, University of Žilina, Univerzitná 1, 01026 Žilina, Slovak Republic

Published online: 11 Jul 2014.

To cite this article: Ľudmila Jánošíková, Jiří Slavík & Michal Koháň (2014): Estimation of a route choice model for urban public transport using smart card data, Transportation Planning and Technology, DOI: 10.1080/03081060.2014.935570

To link to this article: http://dx.doi.org/10.1080/03081060.2014.935570

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Estimation of a route choice model for urban public transport using smart card data

Ľudmila Jánošiková*, Jiří Slavík and Michal Kohání

Department of Transportation Networks, Faculty of Management Science and Informatics, University of Žilina, Univerzitná 1, 01026 Žilina, Slovak Republic

(Received 17 April 2012; accepted 30 April 2014)

This paper describes a logit model of route choice for urban public transport and explains how the archived data from a smart card-based fare payment system can be used for the choice set generation and model estimation. It demonstrates the feasibility and simplicity of applying a trip-chaining method to infer passenger journeys from smart card transactions data. Not only origins and destinations of passenger journeys can be inferred but also the interchanges between the segments of a linked journey can be recognised. The attributes of the corresponding routes, such as in-vehicle travel time, transfer walking time and to get from alighting stop to trip destination, the need to change, and the time headway of the first transportation line, can be determined by the combination of smart card data with other data sources, such as a street map and timetable. The smart card data represent a large volume of revealed preference data that allows travellers’ behaviour to be modelled with higher accuracy than by using traditional survey data. A multinomial route choice model is proposed and estimated by the maximum likelihood method, using urban public transport in Žilina, the Slovak Republic, as a case study.

Keywords: urban public transport; route choice model; smart card data; revealed preference data; Žilina

Introduction

The phenomenon that frequently occurs in urban public transport is that passengers have several possibilities to get to their trip destinations. They can often choose not only a bus, metro, trolleybus or tramline but also a stop at which to get on, off or change lines. In many cases, they have several direct lines at their disposal, as well as the ability to choose to use a combination of lines and to change transport modes (e.g., from metro to bus), or same mode transfers. Since a line is characterised by its route, the choices of stops and lines mentioned above are equivalent to the choice of a route in a line network. Route choice determines the distribution of passenger flows over the links of the line network. The link flow volumes denote how many people want to travel through a given link within a unit of time. In other words, they express the demand for the transportation service. The supply of the transportation service is the number of passenger places on the transportation vehicles running through the given link within a unit of time. The supply and demand ratio is one of the important indicators of the quality of an urban public transportation system. Evaluating this ratio on particular links of the transportation

*Corresponding author. Email: Ludmila.Janosikova@fri.uniza.sk

© 2014 Taylor & Francis
network enables us to identify whether the line routes and frequencies are properly
designed and meeting the public’s needs. If not, if there are links on which the offer
substantially exceeds demand, or vice versa, the offer is lower than demand, then the
line system should be reorganised. The optimization of urban public transport lines is
addressed, for example, by Fan and Machemehl (2004), Borndörfer, Grötschel, and
Pfetsch (2007), Michaelis and Schöbel (2009), Jánošíková, Blatň, and Teichmann
(2010) and Álvarez et al. (2010).

Link flows can be detected by a transportation survey at chosen links or nodes of the
transportation network. However, this method has two big disadvantages: first, it is
expensive and time-consuming and, second, it only detects the current state of the
network – it does not permit predictions of how flows will change with changes in the
public transportation system, e.g., with changes of fare, timetables, line routes and
frequencies.

Another possible method to detect link loads is to use a mathematical model of
passenger assignment. This refers to the manner in which a given aggregate level of
origin–destination (OD) travel demand is assigned to feasible routes for a particular OD
pair. If we know how many people choose every feasible route, then the volume of the
passenger flow on each link can be computed as the sum of the numbers of passengers on
those routes that pass through the given link.

The passenger assignment model captures passengers’ sensitivity to the attributes of
the routes that influence their choice, e.g., in-vehicle travel time and number of transfers.
Route attributes are exogenous variables in the model. They can be evaluated if the
routes, travelling times and frequencies of the lines are known. The attributes are
multiplied by the parameters indicating how the attributes affect the choice.

A number of route choice models can be used. In principle, the models differ in the
level of the knowledge attributed to a passenger, i.e., how much the passenger knows
about the service. In the simplest case, it is supposed that passengers only know which
lines operate between origin and destination stops, they do not know either the timetable
or travel time, or time does not matter to them. Passengers use the first vehicle that arrives
at the origin stop and travel to the destination stop. Another model assumes that
passengers come to the stop randomly but they know travel time and always use the
shortest route. However, both these models are rather far from real passengers’ behaviour.
Commuters travelling daily to work or to school usually have a good overview of the
transportation service, they know the timetables as well as travel times, and they adapt
their arrivals at a stop and route choices to this knowledge. Also irregular travellers are
supposed to be well informed, if the timetables are published. If we assume that
passengers know the values of the route attributes, we can legitimately expect that they
will choose the most convenient route. This reflection coincides with the assumption
made by dell’Olio, Ibeas, and Ruisánchez (2011) when optimising bus size and headway
in transit networks. However, we can never predict a traveller’s decision with certainty.
Therefore, a more suitable approach to modelling passengers’ behaviour is to make use of
discrete choice models (Ben-Akiva and Lerman 1985; Ben-Akiva and Bierlaire 1999;
Koppelman and Bhat 2006; Train 2009). These models are intended to analyse and
predict the decision-making behaviour of a group of individuals in the situation when
they have to choose one alternative from the finite set of mutually exclusive alternatives.

The route choice problem within the frame of one transportation mode is precisely
defined by Ben-Akiva and Bierlaire (1999). Given a transportation network composed
of nodes, links, origins and destinations, and given an origin o, a destination d and a
transportation mode $m$, what is the chosen route between $o$ and $d$ on modem? Besides the specification of the problem, the authors describe the alternative discrete choice model forms that can be used to solve the problem, such as the multinomial logit, C-logit and path-size logit models.

Liu, Bunker, and Ferreira (2010) review the studies on route choice within public transport. Several other works not cited in this review were inspiring for our research. We mention them here briefly.

The main factors affecting travel demand are discussed by de Jong and van de Riet (2008). They analyse the impact of household incomes, population density, labour participation, travel cost, etc. on the various choices that travellers make (choice of activity type, destination, mode, time-of-day and route) and the resulting impact on the level of mobility.

Route choice has also been the subject of a series of studies by researchers from the Institute for Transport Planning and Systems (IVT), part of the Swiss Federal Institute of Technology (ETHZ) in Zurich. Vrtic et al. (2007) deal with the impact of different mobility pricing schemes on a route, mode and departure time choice in Switzerland. Two of their stated preference experiments include route choice. In the former experiment, respondents were supposed to travel by car and they had to choose a route with or without pricing. In the latter experiment, the respondents decided between routes using private or public transport. The authors employed the binary logit model structure.

Axhausen et al. (2006) present a study examining the valuation of travel time savings. Route choice is a part of the multinomial logit model with continuous interactions between respondents’ tastes and socio-demographic attributes, such as trip distance and income. The model is estimated using the stated preference survey data.

Route choice in public transport differs from route choice in private car transport in several aspects: first, the universal choice set is much smaller than the choice set for car drivers, since the transportation network is determined by line routes; second, under the assumption that the transportation service is reliable and has sufficient capacity, the alternatives can be regarded as independent, even if their paths overlap. If travel times are consistent, a passenger sitting on a bus is not interested in the traffic conditions, road quality and other unobserved system and traffic attributes that the overlapping routes share. Another source of similarities among the routes may arise from overcrowding. If this occurs, people tend to avoid routes on which they suffer from discomfort in the vehicle or failure to board. But if the vehicle capacity is sufficient, overlapping routes can be regarded as distinct and independent. Having consulted the published experience with logit models for route choice in the context of urban public transport (cf. Hunt 1990; Pursula and Weurlander 1999; Eboli and Mazzulla 2008), we find the simple multinomial logit model to be the most appropriate.

As far as the authors are aware, the route choice model estimations so far have been based on survey data. The widespread implementation of smart card-based fare payment systems around the world offers the new potential of using these data for other purposes different from the original intention. Several types of analyses are possible with the smart card data, including ridership monitoring, revenue estimation and service performance measurement. Recently, the applications to analysing travel behaviour have appeared, for example, for OD matrix estimation (Wang, Attanucci, and Wilson 2011). The benefits of public transport smart card data in comparison with traditional transportation data sources (such as surveys) are summarised by Bagchi and White (2005) as follows: larger volumes of personal travel data; the possibility of linking trips to individual cards; covering longer
periods of time than it is possible to obtain by surveys; allowing panel data analysis
techniques; and classification of different customer market segments.

The goal of the research on which this paper is based is to propose a methodology for
the choice set generation and the estimation of the route choice model, using transaction
data from an electronic payment system. A multinomial logit model is described in the
following section. Then the data-set is introduced, and the methodology of choice set
generation is presented, followed by estimation results and their discussion. In the
conclusion, the utilisation of the model and the contribution are summarised.

Model and choice set description

In the multinomial logit model, the probability that the \( i \)th individual chooses alternative \( j \)
is defined as:

\[
P(i,j) = \frac{\exp(\beta_1 x_{ij1} + \ldots + \beta_p x_{ijp})}{\sum_{r \in A_i} \exp(\beta_1 x_{ir1} + \ldots + \beta_p x_{irp})}
\]

where \( p \) is the number of attributes, \( A_i \) is the set of the alternatives feasible for individual
\( i \), \( x_{ij1}, \ldots, x_{ijp} \) are the values of the attributes of the alternative \( j \) feasible for the individual
\( i \), and \( \beta_1, \ldots, \beta_p \) are the parameters which define the direction and importance of the effect
of the attributes on the utility.

The probabilities \( P(i,j) \) contain unknown parameters \( \beta_1, \ldots, \beta_p \). They can be estimated
by the maximum likelihood method using the attribute values \( x_{ijk} \) for \( k = 1, 2, \ldots, p \) and
from the known number of those individuals who chose alternative \( j \).

In route choice modelling, the following attributes of a route were taken into account:

- in-vehicle travel time,
- walking time needed for transfers and to pass from the alighting stop to the trip
destination,
- number of transfers (0 or 1),
- time between consecutive vehicles (headway) of the line (the first line in the case
of a linked trip with the transfer).

We will assume that the symbol \( i \) does not denote just a single individual but a group of
individuals with identical characteristics and the same set of feasible alternatives. Let us
explain what is meant by ‘identical characteristics’. Since we did not survey passengers,
their individual characteristics cannot be represented by exogenous variables in the
model. Instead of using variables, we generalise the characteristics of travellers (such as
travel purposes, demographic characteristics) and suppose that they change in the course
of a day: in the morning and afternoon peak times, passengers are mostly regular public
transport users, like commuters, schoolchildren and students, while in the off-peak time
most passengers are irregular passengers who do not have a precise schedule, e.g., people travelling to a doctor or for shopping. In addition to the passengers’
characteristics, the line routes and frequencies are different in peak and off-peak time too,
so it is necessary to estimate the model separately for each time period that is usually
observed in the traffic. The same set of feasible alternatives means that \( i \) denotes a group
of passengers who travel from the same origin to the same destination stop. Then \( A_i \) is the
set of feasible routes between these stops.
The choice set, as well as the number of passengers having chosen each alternative, can be inferred from archived smart card transactions data. The data are collected by the electronic ticketing machines which are placed in the vehicles and used to validate the passengers’ travel (paper or electronic) tickets. When a passenger validates their electronic ticket at boarding, the stamp machine records the identification number of the smart card, together with the line number, the identification number and name of the boarding stop and the date and time of the ticket validation. Alighting is not validated, so we do not know the trip destination. However, using the transactions data we can follow all trips the passenger made during a particular day, and estimate their destinations. The boarding stop of the first trip is regarded as the origin of the passenger’s journey. If only one transaction for the card is recorded on that day, the destination cannot be inferred. Otherwise, the trip-chaining method similar to the one used by Trépanier, Tranchant, and Chapleau (2007) or Wang, Attanucci, and Wilson (2011) can be applied with the following assumptions:

- passengers do not use a private transportation mode (car, motorcycle, bicycle) between consecutive transit trip segments in a daily trip sequence,
- passengers do not walk a long distance to board at a stop different from the one where they previously alighted,
- passengers end their last trip of the day at the stop where they began their first trip of that day.

If the same card is used at least twice a day and the interval between the first and the second boarding is longer than an hour, then the boarding stop of the second trip is the destination of the first trip subject to the constraint on the distance limit. It means that the boarding stop must be close to the previous alighting stop (within 15-minute walking distance at a speed of 5 km/h). On the other hand, if the time interval is less than an hour, then we suppose that the traveller changed line. If the boarding stop of the next line is not one of the stops of the previous line, then the alighting stop is supposed to be the nearest stop with regard to the boarding stop, and the traveller had to walk between them. In this research, the distances and the walking times were computed on the basis of the city street network. Trépanier, Tranchant, and Chapleau (2007) offer another way of destination estimation, based on the transit use regularity and the Euclidean distances between the stops.

Once the destinations have been derived, it is easy to obtain all quantities needed for the choice model estimation, namely the attribute values for all alternative routes, the total number of the passengers between a couple of stops, and how many of them used each feasible alternative of the route.

**Description of data-set and estimation method**

As a case study for route choice modelling, we chose the urban public transport system in the city of Žilina. Žilina is a middle-sized city situated in the north-western part of the Slovak Republic. It had 85,302 inhabitants in 2009, and covers an area of 80 km². The transportation service in the city is provided by the transportation operator *Dopravný podnik mesta Žiliny* (DPMŽ). During the day, 8 trolleybus lines and 10 bus lines operate. At night, the city area is served by one bus line.
The number of passengers needed for the model estimation was derived from the smart card transaction data provided to us by DPMŽ. The contactless smart cards used to make fare payments are owned by individuals and are non-transferable.

We had the data collected during one week (from 12 to 18 October 2009). The whole database contains 115,007 transactions on 13,361 smart cards. Using the trip-chaining methods, the destinations for 80.72% trips could be identified. From the recognised OD pairs, only those were selected for model estimation where the travellers took at least two different routes between the given origin and destination stops. As it has already been mentioned, the model should be estimated separately for peak and off-peak time periods. We concentrated on the morning peak (from 6:00 to 8:00) and the off-peak period during the day (8:00–14:00). Table 1 gives the statistics of the relevant attributes for all alternatives included in the model estimation.

The impact of the attributes on the passengers’ decision-making process is expressed by the parameters $\beta_1, \beta_2, \beta_3$ and $\beta_4$ that correspond to particular attributes. So that the parameters can identify the relevant importance of the attributes, the attributes need to be modified to have the values of the same order. Therefore, the in-vehicle times, walking times and headways were normalised to the value in the range (0; 1) by dividing them by the maximal value of the given attribute from all alternatives.

### Estimation results

Tables 2 and 3 report the estimates, standard deviations and $t$-ratios of the parameters $\beta_1, \beta_2, \beta_3$ and $\beta_4$ for the morning peak period and the following off-peak daytime. For both of the investigated periods, the statistical test evaluated the parameters $\beta_1, \beta_2, \beta_3$ and $\beta_4$ as significant at the 0.05 significance level. It means that all considered attributes affect passengers’ decisions. As expected, all four parameters are negative. The in-vehicle travel time plays the most significant role in the route choice for both periods. The order of the other attributes differs for particular periods, and it reflects different passengers’ characteristics in the peak and off-peak periods. In the morning, most passengers commute to work and schools, and time is the most significant factor for them. That is why the line frequency is the second most important attribute in the morning peak time. Later, in the off-peak time, most passengers are irregular passengers who do not have a precise schedule. They may travel for shopping, to see a doctor or to visit an institution, and many of them may be elderly people. They appreciate comfortable travel, and perceive walking and transfers negatively.

Since commuters travel every day, the weekly data-set for the morning period contains multiple observations of the same persons, so the data-set is an unbalanced panel. The
choices made by the same traveller are not independent due to the habit. The correlation in
the unobserved utility across repeated choices can be corrected by resampling. From several
available resampling methods, the bootstrap method was chosen. Table 4 compares the
resampling results with the original estimates. One can see that the estimates on the pooled
sample undervalue those attributes that can be determined precisely from the smart card
data, such as in-vehicle travel time, number of transfers and headway. Walking times were

Table 2. Estimation results for the pooled sample with relative attribute values, morning peak
period.

<table>
<thead>
<tr>
<th>Parameter number</th>
<th>Description</th>
<th>Coefficient estimate</th>
<th>Standard error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>In-vehicle travel time</td>
<td>−4.55</td>
<td>9.27E-2</td>
<td>−57,90</td>
</tr>
<tr>
<td>2</td>
<td>Walking time</td>
<td>−2.64</td>
<td>4.97E-2</td>
<td>−53.14</td>
</tr>
<tr>
<td>3</td>
<td>Number of transfers</td>
<td>−3.04</td>
<td>3.70E-2</td>
<td>−82.23</td>
</tr>
<tr>
<td>4</td>
<td>Headway</td>
<td>−3.92</td>
<td>6.61E-2</td>
<td>−59.30</td>
</tr>
</tbody>
</table>

Summary statistics
Number of OD pairs = 513
Number of observations = 24,722
$L(0) = −18,862.51$
$L(\hat{\theta}) = −5453.33$
$\rho^2 = 0.71$

Table 3. Estimation results with relative attribute values, off-peak period.

<table>
<thead>
<tr>
<th>Parameter number</th>
<th>Description</th>
<th>Coefficient estimate</th>
<th>Standard error</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>In-vehicle travel time</td>
<td>−5.37</td>
<td>9.27E-2</td>
<td>−57,90</td>
</tr>
<tr>
<td>2</td>
<td>Walking time</td>
<td>−2.64</td>
<td>4.97E-2</td>
<td>−53.14</td>
</tr>
<tr>
<td>3</td>
<td>Number of transfers</td>
<td>−3.04</td>
<td>3.70E-2</td>
<td>−82.23</td>
</tr>
<tr>
<td>4</td>
<td>Headway</td>
<td>−3.92</td>
<td>6.61E-2</td>
<td>−59.30</td>
</tr>
</tbody>
</table>

Summary statistics
Number of OD pairs = 598
Number of observations = 23,808
$L(0) = −18,692.74$
$L(\hat{\theta}) = −7040.46$
$\rho^2 = 0.62$

Table 4. Original and bootstrap estimation, morning peak period.

<table>
<thead>
<tr>
<th></th>
<th>Original estimation</th>
<th>Bootstrap estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-vehicle travel time</td>
<td>−5.37</td>
<td>−6.33</td>
</tr>
<tr>
<td>Walking time</td>
<td>−2.64</td>
<td>−2.15</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>−3.04</td>
<td>−3.61</td>
</tr>
<tr>
<td>Headway</td>
<td>−3.92</td>
<td>−4.00</td>
</tr>
</tbody>
</table>
just estimated, and they seem to be even less important than the original parameter suggested.

Using the bootstrap estimates for the morning peak period and pooled sample estimates for the off-peak period, the direct aggregate elasticities were calculated (see Table 5). The elasticities for both periods indicate that in-vehicle travel time and headway have the greatest negative impact on the public transport demand. Lower off-peak elasticities can be explained by less flexibility in route choice, compared with the peak time.

The trade-offs among the estimated parameters can help identify those service attributes that have the greatest impact on customers’ satisfaction. Table 6 and Table 7 contain the relative ratios of parameters for the absolute attribute values. These figures reflect passengers’ aversion to transfers, since they are willing to accept the greatest increase in the travelling time, walking time and headway in return for transfer elimination. From Table 6 one can derive that a change in the line network that eliminates the need to transfer but does not increase travelling times by more than 33 minutes, will increase the attractiveness of the system. Regarding walking time, one transfer is an equivalent to a 21-minute walk. In the off-peak period, the corresponding trade-offs are approximately 30 and 18 minutes, respectively.

We can compare our results with previous studies regarding public transport route choice. We focus on two of them that use the multinomial logit model and revealed preference data, namely Hunt (1990) and Pursula and Weurlander (1999). Hunt estimates the route choice behaviour of employees travelling to work in the morning (the data were collected in the central business district of Edmonton in June 1983). The trade-offs computed from the parameter estimates reported by Hunt are given in Table 8. Pursula and Weurlander analyse the importance of different level-of-service factors in the public

### Table 5. Aggregate elasticities.

<table>
<thead>
<tr>
<th></th>
<th>Morning peak</th>
<th>Off-peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-vehicle travel time</td>
<td>-0.76</td>
<td>-0.58</td>
</tr>
<tr>
<td>Walking time</td>
<td>-0.09</td>
<td>-0.08</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>-0.07</td>
<td>-0.11</td>
</tr>
<tr>
<td>Headway</td>
<td>-0.48</td>
<td>-0.29</td>
</tr>
</tbody>
</table>

### Table 6. Relative ratios of parameters with panel model and absolute attribute values, morning peak period.

<table>
<thead>
<tr>
<th></th>
<th>In-vehicle travel time (min)</th>
<th>Walking time (min)</th>
<th>Number of transfers (–)</th>
<th>Headway (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-vehicle travel time (min)</td>
<td>N/A</td>
<td>0.64</td>
<td>0.03</td>
<td>1.61</td>
</tr>
<tr>
<td>Walking time (min)</td>
<td>1.57</td>
<td>N/A</td>
<td>0.05</td>
<td>2.52</td>
</tr>
<tr>
<td>Number of transfers (–)</td>
<td>33.23</td>
<td>21.19</td>
<td>N/A</td>
<td>53.36</td>
</tr>
<tr>
<td>Headway (min)</td>
<td>0.62</td>
<td>0.40</td>
<td>0.02</td>
<td>N/A</td>
</tr>
</tbody>
</table>
transport in the Helsinki metropolitan area (the survey was performed in 1992). The trade-offs computed on the basis of their estimates are given in Table 9.

Our mutual trade-offs between travelling time, walking time and transfers generally agree with the results reported by Hunt. Hunt’s travelling and walking time equivalents of a transfer are almost 50% lower than our results for the morning peak. The reason may be that bus/trolleybus arrivals and departures from transfer stops are not coordinated in Žilina, and so the transfer waiting times are higher and transfers are more bothersome here than in Edmonton. The transfer waiting times were not modelled either in Edmonton or in Žilina, as their measurement is imprecise. The relatively large error associated with the estimates of transfer waiting times might worsen the fit of the model. The headway trade-offs computed from Hunt’s parameter estimates are by an order different from our results because Hunt’s estimate of the headway parameter is 10 times higher than ours (−0.6 vs. −0.06). It may be caused by the different definition of the headway for a linked trip with a transfer, as well as by the lower reliability of the service almost 30 years ago, which prolongs waiting times.

The headway trade-offs in Žilina are consistent with those in Helsinki; however, the other ratios are quite different – for example, one transfer is equivalent to only 8.85 minutes of the in-vehicle time in Helsinki (see Table 9). We believe that our results are more precise due to a more extensive data-set (compare approximately 24,000 observations in Žilina with just 250 in Edmonton and 562 in Helsinki) and higher likelihood ratio index (compare 0.71 for the morning peak in Žilina with 0.38 in Edmonton and 0.336 in Helsinki, respectively). However, we fully agree with the conclusions by Pursula and Weurlander that transfers form a substantial disutility to

Table 7. Relative ratios of parameters with absolute attribute values, off-peak period.

<table>
<thead>
<tr>
<th>In-vehicle travel time (min)</th>
<th>Walking time (min)</th>
<th>Number of transfers (–)</th>
<th>Headway (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-vehicle travel time (min)</td>
<td>N/A</td>
<td>0.59</td>
<td>0.03</td>
</tr>
<tr>
<td>Walking time (min)</td>
<td>1.70</td>
<td>N/A</td>
<td>0.06</td>
</tr>
<tr>
<td>Number of transfers (–)</td>
<td>29.73</td>
<td>17.52</td>
<td>N/A</td>
</tr>
<tr>
<td>Headway (min)</td>
<td>0.17</td>
<td>0.10</td>
<td>0.01</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>In-vehicle travel time (min)</th>
<th>Walking time (min)</th>
<th>Number of transfers (–)</th>
<th>Headway (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-vehicle travel time (min)</td>
<td>N/A</td>
<td>0.66</td>
<td>0.09</td>
</tr>
<tr>
<td>Walking time (min)</td>
<td>1.53</td>
<td>N/A</td>
<td>0.09</td>
</tr>
<tr>
<td>Number of transfers (–)</td>
<td>17.91</td>
<td>11.73</td>
<td>N/A</td>
</tr>
<tr>
<td>Headway (min)</td>
<td>3.80</td>
<td>2.49</td>
<td>0.21</td>
</tr>
</tbody>
</table>

travellers, and direct connections should be offered where possible. On the other hand, it is not economically feasible to base a transportation network in an urban area on direct connections only. By providing high-quality transfer service, the overall coverage and quality of service can be increased.

**Conclusions**

This paper has focused on travel behaviour with respect to route choices in the context of urban public transport. It presented a methodology for estimating the model parameters from transaction data archived in the electronic fare payment system. The multinomial discrete choice model was used for the representation of route choice. The attributes of route alternatives were in-vehicle travel time, walking time needed for transfers and to pass from the alighting stop to the trip destination, number of transfers and headway of the first line. We have shown that the choice set, the attribute values, as well as the number of passengers having chosen each alternative can be easily inferred from the smart card data collected by the electronic ticketing machines, which is undoubtedly less expensive and a more accurate way than the conventional passenger survey. The statistical analysis revealed that all chosen attributes are significant, and the travel time plays the most important role in route choice.

The model is an important tool for transit planners and transport operators. It allows to predict the passengers’ behaviour in the case when the values of the route attributes (e.g., a line frequency) change. The change can cause a shift of passengers to another alternative of their trip. The model also allows predicting passenger link flows resulting from the modification of the whole line network. Using the model will improve the input data for public transport planning and optimisation at both strategic (line routes) and tactical (line frequencies) levels. Moreover, the estimated parameters can help identify those service attributes that have the greatest impact on customers’ satisfaction. The improvement of important attributes can help make public transport more attractive, resulting in a market share increase.

**Funding**

This research was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences under project VEGA 1/0339/13 ‘Advanced microscopic modelling and complex data sources for designing spatially large public service systems’ and by the Slovak Research and Development Agency under project APVV-0760-11 ‘Designing of Fair Service Systems on Transportation Networks’.
References


