https://doi.org/10.3311/PPtr.15864

Periodica Polytechnica Transportation Engineering, 51(1), pp. 22–30, 2023

A New Route Choice Model for Urban Public Transit with Headway-based Service

Peter Horbachov[1*], Oleksandr Makarichev[1], Olha Svichynska[1]

[1] Department of Transport Systems and Logistics, Faculty of Transportation Systems, Kharkiv National Automobile and Highway University, str. Yaroslava Mudroho, 25, 61002 Kharkiv, Ukraine
[*] Corresponding author, e-mail: gpf@khadi.kharkov.ua

Received: 16 March 2020, Accepted: 26 September 2022, Published online: 07 November 2022

Abstract
A new approach to create the model of individual behavior in urban public transit systems with headway-based service that is built on the same principles of rational decision making as discrete choice models is presented. To describe a passenger's decision-making, the attractiveness function which reflects the difference between trip results and costs was used. The attractiveness of the transit route was determined by solving the system of Fredholm's integral equations based on observed frequencies of choosing the alternative routes by each individual. The frequencies were determined based on multi-day survey results and range between 0 and 1 but equal neither 0 nor 1. To estimate the significance level of the attractiveness function, the regression analysis was used and the accuracy of the choice probability forecast was evaluated. Computation of model coefficients was carried out based on the parameters of home-based work trips obtained via the travel survey that was conducted in the city of Kharkiv, Ukraine.

Keywords
choice modeling, urban public transit, observed frequency, rational decision making, regression analysis

1 Introduction

Public transit systems in Ukrainian cities are characterized by highly developed route networks that serve more than 60 percent of all trips taken by residents and visitors. Effective planning of a public transit system operation is a very important task under conditions of a low level of service in Ukraine that is quite different from the typical level of passenger service in developed countries. In most cases, the timetables are not available for passengers, and most of them are not aware of the time of vehicles' departures from the stops. Municipal transportation agencies do not effectively control the performance of transit operators. It often allows drivers to neglect the schedule and promotes excessive freedom in the choice of a type of service in an urban transit system. As a consequence, there exist self-regulated relationships between drivers at many urban routes that make a headway-based service the main way of the public transit operation. In this case, passenger waiting time becomes a random part of travel costs (or "impedance") and deserves special attention during modeling of the passengers' assignment to the routes. The absence of regular surveys of the transportation process decreases the level of public transport services for passengers in Ukrainian cities and, as a consequence, there is a lack of input data to make scientifically grounded decisions to improve servicing.

In recent years, public transport planning in Ukraine has been advanced due to the use of state-of-the-art transportation planning technologies. They are based on mathematical models and transport modeling software like VISUM. The feature of software of this type is that it provides transport engineers with an advanced toolkit for transport modeling, but there is no support for the methods to estimate model parameters. This issue is one of the main reasons to develop the methods of the passenger assignment to a public transport network based on factual information about the individual's behavior in a transit system and reasonably describe the complexity and relativity of the route choice in the urban transit systems.

2 Literature review

VISUM uses a variety of models for the assignment of the passenger's trips to the transit routes – Kirchhoff, Box-Cox, EVA, etc. – and their parameters are always the subject of calibration (de Dios Ortúzar and Willumsen, 2011). The major

Cite this article as: Horbachov, P., Makarichev, O., Svichynska, O. (2023) "A New Route Choice Model for Urban Public Transit with Headway-based Service", Periodica Polytechnica Transportation Engineering, 51(1), pp. 22–30. https://doi.org/10.3311/PPtr.15864

shortcoming of any choice model is a lack of the theoretical explanation of a random choice of transport mode by individuals. The most reliable and theoretically sound base for the development of the models of individual behavior in a route network is the microeconomic theory. According to this theory, individuals behave in such a way as to take maximum advantage from the defined set of alternatives (Ben-Akiva and Bierlaire, 2003). To describe the attractiveness of the alternative, the term "utility" is used. The term is a convenient theoretical structure defined as something that the individual wants to maximize. Since the researcher often does not have comprehensive knowledge of all elements affecting an individual choice, the utility is divided into two parts (de Dios Ortúzar and Willumsen, 2011). The first part is presented by the observed variable V_{ij} where $i = 1, \ldots, N$, $j = 1, 2, \ldots, z$ (N is the number of individuals, z is the number of alternatives available for a passenger). The second part of the utility is a random variable ε_{ij} which reflects the impact of all factors having a negligible influence on the individual's choice and also covers possible errors of the observation during mathematical modeling (Kjaer, 2005). The random variable makes it possible to explain the reasons for different choices because the probability that the individual i will choose the alternative j is the probability that the utility of this alternative will be the maximum among other alternatives:

$$P_{ij} = \Pr\left\{U_{ij} > U_{ir}, \forall r = 1, \ldots, z\right\}, \tag{1}$$

where U_{ir} is the utility of the alternative r from a set of alternatives available and convenient for the individual i.

Discrete choice models (DCM) became the main tool to solve the above-formulated problem. Such models can be divided into two classes, depending on the number of alternatives – binary and multinomial models. DCM have been applied in different fields, and many models have been developed specifically for transportation modeling (Ben-Akiva and Lerman, 1985; Ben-Akiva et al., 1986; Koppelman and Pas, 1980; McFadden, 1978). Koppelman and Wen (1998) identified a specific variable that influences the choice among transport modes. De Palma and Rochat (2000) considered another type of DCM – Nested Logit Model – for the analysis of passengers' choice of transport mode in Geneva. Chen and Li (2017) supplemented the DCM with latent variables, and the history of the development of these variables in the transportation field can be found in the sources (Ben-Akiva and Boccara, 1989; Train et al., 1987). Chen and Li (2017) proposed to include five latent variables into the utility function to explain the public transport mode choice behavior and identify which latent variables influence the passenger's choice.

Based on the presented review, we can conclude that the use of the abovementioned approach to model the individual route choice behavior continually increases with the development of the mathematical modeling apparatus. Eq. (1) is a basis for all discrete choice models, and the main difference between them is defined by distinct assumptions about the distribution of random variable ε. Logit models are based on the assumption that errors ε are independent and identically distributed according to the I-type extremum values distribution (the Gumbel distribution). Probit models are based on the assumption about the normal distribution of the random part of utility. There also exist Mixed Logit models that are based on the assumption that the random component of the utility consists of two parts – the first part corresponds to any distribution given by the researcher, and the second one consists of ε errors that are independently and identically distributed according to the Gumbel distribution. The DCM with latent variables are based on the assumption that error distribution for each utility function follows Gumbel distribution with a mean of 0 independently, while error distribution for the rest of the stochastic factor functions follows a normal distribution (Chen and Li, 2017).

Maximum likelihood method estimates DCM parameters, and its inputs are the binary choice probabilities and parameters of alternatives. This method is easy to understand as it maximizes the product of probabilities of chosen alternatives. That is why the DCM are the universal modeling tool. However, the DCMs are based on a strict assumption that random components of all parameters of alternatives are identically distributed, although they are of a different nature.

Among the indices for estimation of DCM parameters, the most widespread one is the likelihood ratio ρ^2 developed by McFadden (1974). It is used in many works, but it characterizes the result of coefficients estimation rather than the predictive capability of a choice model. The same is true for other standard measures of DCM quality that make us think about the development of an alternative measure that will explicitly show the discrepancy between the observed data and the data obtained during modeling.

Another property of DCM, in our opinion, is that these models are not convenient for the analysis of the results of the route choice modeling in urban transit systems, especially taking into account that in this case, the sets of alternatives are different for each pair of transport zones. They differ in both the number of alternatives and competitive transport modes (or routes). It becomes clear from the results presented by Senk (2010). Here to evaluate the respondents' route choice two alternative routes between given zones were proposed to consider. Such a situation does not allow creating an ordered array of parameters of alternatives and requires evaluating differences between travel options in every set of alternatives. Also, it is necessary to consider the relative nature of alternative parameters in every finite set.

Taking into account a relatively simple issue of how individuals select routes in urban transit systems, there is a possibility to simplify a problem formulation and develop a special modeling method. This method should be based on the features of a headway-based public transit service when the individual makes a decision directly at the stop according to the experienced transport situation.

3 Methodology to create a new route choice model

The method being developed is based on the same principle of the individual's rational behavior as in DCMs. A basis of modeling is the assumption that an individual maximizes efficiency which represents the quality of the trip. The quality is defined as a set of trip parameters that determines the ability of route alternatives to meet the individual's need to change location. Then, a choice efficiency must fully reflect the travel purpose and can be determined as a difference between the result and costs of travel. Taking into account the maximization principle, the efficiency of trip between a pair of transport zones using the alternative *j* can be determined as

$$E_j = \max_r \left\{ E_r = R_r - C_r \right\}, \; j, r \in [1, z], \tag{2}$$

where *j* is the index of the chosen alternative route between the origin and destination zone, *r* is the index of one of the alternative routes between zones; *z* is the number of alternative routes between zones; E_j, E_r are the travel efficiency of the alternatives *j* and *r* respectively; C_r are all travel costs by the alternative *r* between zones; R_r is the result of using the alternative *r*.

A travel result is determined by its necessity and urgency which depends on the passenger's attitude to the travel purpose. These factors are not directly related to the transportation process and may vary widely.

For the correct use of the principle of travel efficiency maximization, it is necessary to make an additional assumption that the individual has a clear understanding of the route choice results. This assumption seems to be quite valid if modeling of individual behavior in the public transportation system is limited to home-based work trips. Such restriction also allows a substantial reduction of the possible range of R_r values due to the similarity of passenger trips since the travel purpose is the same for everyone – to be at work in time. Therefore, for the considered trips we can assume that

$$R_r = \text{const} = R, \; \text{for } \forall r \in [1, z]. \tag{3}$$

Regardless of Eq. (3), incorporation of trip results into Eq. (2) is an important step, not only from the methodological point of view but also from a perspective of the proposed model since it provides the actual content for the *y*-intercept of the regression model.

Research of the individual's behavior in the context of public transportation significantly differs from the situation which is standard for the DCMs and implies that traveling by car and traveling by public transport are considered competing alternatives. This choice is often made before leaving home. If an individual always uses public transit for home-based work trips, then taking into account the headway-based service in Ukrainian cities, the decision on the route choice is usually made directly at the stop. This leads to the special set of travel costs that differ from the typical quantitative parameters of the DCMs. Firstly, not all the quantitative transit route parameters are deterministic that especially relates to the waiting time. Secondly, the individual does not know the exact result of their trip when deciding on choosing one from all available routes. That is why it is necessary to introduce the measurable factors of the transit trip into the model. It is assumed that an individual is aware of the values of the measurable factors at the moment of the route choice, and they are constant for every route alternative.

It is quite clear that fare, number of transfers and walking time can be considered measurable factors of the route choice. At the same time, other types of travel impedance require more detailed analysis. At the moment of decision making an individual does not know the values of such parameters as in-vehicle time, transfer time, and in-vehicle crowding during the trip. But the individual has a clear notion about the expected values of these parameters due to the permanent use of public transport for home-based work trips. These expected values are determined by the experience of the use of one route versus another. The mentioned parameters can be taken as measurable factors of the route choice because of considering home-based work trips only.

Besides the measurable factors mentioned above, there are two random factors in the headway-based transit service. They are the waiting time of the vehicle at the stop and in-vehicle crowding at the moment of boarding which can be represented by the load factor. The latter factor differs from the average load factor during the trip, which can be considered as deterministic one. At the moment of boarding the transit vehicle, it may vary from trip to trip. But it turned out that this is not a significant factor for home-based work trips, and it is not considered at this modeling stage.

Also, there are other factors that can be both quantitative and qualitative and cannot be defined via observations. So, there are three categories of travel impedances that the individual considers during the route choice:

- measurable (fare, number of transfers, travel time, and load factor);
- random (waiting time);
- unspecified (non-identifiable and perceptional non-measurable factors).

Measurable and random costs are the base for the model being developed because the impact of unspecified factors can be estimated only after the statistical processing of the experimental data. So, the question of this research is how to determine the probability of the route choice when the decision significantly depends on passenger waiting time at the stop, which under conditions of headway-based public transport service is a random variable.

The result of the travel costs analysis can be written as

$$C_r = \sum_{i=1}^{m} c_i \times g_{ir} + c_t \times t_r \ , \tag{4}$$

where m is a number of measurable attributes considered by individuals when choosing among the best alternative routes between a pair of transport zones; c_i is the weighting factor of the variable i (estimated parameter); g_{ir} is the value of the variable i for the alternative r; t_r is the waiting time (random variable) till the first boarding when choosing the alternative r; c_t is the weighting factor of the waiting time.

Taking into account the constancy of the travel result in Eq. (3), the modeling goal is to obtain the objective values of weighting coefficients of the measurable attributes c_i, $i \in [1,m]$, that will allow adequate description of the individual's attitude toward the corresponding travel parameters. The first step in this direction is the transformation of Eq. (4). It is convenient to introduce a variable y_r which is determined by dividing the difference between travel result and measurable costs from Eq. (4) by a coefficient c_t of the random waiting time t_r:

$$y_r = \frac{R}{c_t} - \sum_{i=1}^{m} \frac{c_i}{c_t} \times g_{ir} \ . \tag{5}$$

Taking into account Eq. (2) it can be written that

$$\frac{E_r}{c_t} = y_r - t_r \ . \tag{6}$$

Let a new variable X_r be introduced as follows:

$$X_r = \frac{E_r}{c_t} \ . \tag{7}$$

In this case, it is possible to rewrite Eq. (6) as follows:

$$X_r = y_r - t_r \ . \tag{8}$$

New variables X_r and y_r have a theoretical sense. X_r is the travel efficiency in units of time including a random part of the costs (waiting time). In the considered problem, X_r is a positive random variable due to the influence of the waiting time. Variable y_r is the travel efficiency in units of time that does not refer to the waiting time. This positive variable hereinafter will be called "attractiveness of the route". According to the formula of the joint probability for a continuous random variable (Gnedenko, 1962), the probability of choosing the alternative j is determined by the integral

$$p_j = \int_D P\left\{ X_j = \max_r \left\{ X_r \right\} \right\} \times f\left(x_j \right) dx_j \ , \tag{9}$$

and according to Eq. (8)

$$p_j = \int_D P\left\{ y_j - t_j = \max_r \left\{ y_r - t_r \right\} \right\} \times f\left(t_j \right) dt_j \ , \tag{10}$$

where D is the domain of integration of random variables X_j and X_r; $f(x_j)$ is the probability density function of the waiting time if choosing the alternative j.

Equation (10) has the same form as a base equation of DCM, e.g. Eq. (4.14) in the (Kjaer, 2005). The differences lie in distinct measurable attributes and random variables.

In general, it can be assumed that under headway-based service the arrival time of the vehicles operating on a certain route does not depend on the arrival time of the vehicles that operate on another route. Then, random events $A_r = \left\{ y_{r \neq j} - t_{r \neq j} < y_j - t_j \right\}$ and $A_s = \left\{ y_{s \neq j} - t_{s \neq j} < y_j - t_j \right\}$ for $r \neq j \neq s \in [1, z]$ are independent and

$$P\left\{ A_r \middle| A_s \right\} = P\left\{ A_r \right\} , \tag{11}$$

$$p_j = \int_D \prod_{r:r\neq j}^{z} P\{A_r\} \times f\left(y_j - t_j\right) dt_j \, . \tag{12}$$

Using Eq. (12), it is possible to describe the probability of choosing all real alternative routes. The route variants which have a non-zero statistical assessment of choice probability are the real alternatives. This is a difference from Discrete-choice models that deal with individual choice probabilities that equal either 0 or 1. But the equation system for one pair of alternatives can be directly solved only when observed relative frequencies of the choice equal neither 0 nor 1. So, we need to observe the behavior of every individual in more than one choice situation and involve only equations with the non-zero statistical assessment of choice probability in the corresponding system. When there are z real alternatives between a pair of transportation zones, it is possible to create a system of z integral equations:

$$\begin{cases} p_1 = \int_D \prod_{r=2}^{z} P\{y_r - t_r < y_1 - t_1\} \times f\left(t_1\right) dt_1 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ p_j = \int_D \prod_{r:r\neq j}^{z} P\{y_r - t_r < y_j - t_j\} \times f\left(t_j\right) dt_j \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ p_z = \int_D \prod_{r=1}^{z-1} P\{y_r - t_r < y_z - t_z\} \times f\left(t_z\right) dt_z \end{cases} . \tag{13}$$

In this system, it is required to compute the values of y_r, $r \in [1,z]$ using the statistical probabilities p_j and the distribution function $f(t_r)$ of the actual waiting time till the first boarding. The system of Eq. (13) is a system of integral equations with a multiplicative integrand, where the number of unknown variables y_r equals the number of equations z. Nevertheless, it is necessary to prove that Eq. (13) has a unique solution. The need is caused by the special property of the attractiveness of the routes y_r.

The comparison condition of random variables in each equation in the system of Eq. (13) has a form

$$y_r - t_r < y_j - t_j; \;\; j \neq r; \;\; j,r \in [1,z] \tag{14}$$

that can be rewritten as follows:

$$t_j < t_r + y_j - y_r; \;\; j \neq r; \;\; j,r \in [1,z] . \tag{15}$$

Equation (15) contains two unknowns – y_r and y_j which are constant values allowing to rewrite the equation as follows:

$$t_j < t_r + \xi; \;\; \xi = \text{const} = y_j - y_r \, . \tag{16}$$

It necessary to introduce a new designation for the constant ξ. Taking into account that indices r and j vary within the range $[1,z]$ depending on the equation in the system of Eq. (13), it is reasonable to define a comparison base, and the first alternative in the set of all alternatives under investigation is suitable for this purpose

$$\Delta y_r = y_1 - y_r, \;\; r \in [2,z] , \tag{17}$$

where Δy_r is the relative attractiveness of the first alternative in the set compared to the alternative r.

A computational meaning of Δy_r is a shift of the distribution $f(t_r)$ of the random waiting time that leads to the change of the probability $P\{t_j + \Delta y_j < t_r + \Delta y_r\}$ (Horbachov and Svichynskyi, 2014; 2018).

When the statistical probabilities of choosing the alternatives p_j and the waiting time distribution $f(t_r)$ are known, $z - 1$ values of Δy_r can be obtained. These values mean a relative attractiveness of the first alternative in the set compared to alternative r, i.e. how much worse is alternative r compared to the first one. So, the value of the real attractiveness of the first alternative in the set remains unknown. It completely corresponds to the relativity of the choice among alternative routes between a pair of transport zones but does not allow to produce a final decision at this stage. This decision is a vector of the relative attractiveness of the alternatives that can be compared with the measurable travel costs through the forming of the system of Eq. (18).

This system is the basis for the multiple regression analysis of the set of alternatives between a pair of transport zones and a_0 in this system is the constant term of the regression model, a_i – the slope term of i^{th} type of the measured travel costs.

An undefined variable y_1 appears on the right side of the system of Eq. (18), and it does not affect the probability of the choice of a certain route:

$$\begin{cases} a_0 + \sum_{i=1}^{m} a_i \times g_{i1} = y_1 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_0 + \sum_{i=1}^{m} a_i \times g_{ir} = y_1 - \Delta y_r \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_0 + \sum_{i=1}^{m} a_i \times g_{iz} = y_1 - \Delta y_z \end{cases} . \tag{18}$$

Thus, the attractiveness of the first alternative can be determined freely. The existence of this independent variable is logical and reasonable enough because an individual compares travel options and the result here can be relative only. Moreover, the presence of an unspecified "base" variable confirms the correctness of this approach since it is fully consistent with the relative nature of the choice of the route between origin and destination.

When using a standard variant of the regression analysis to solve the system of Eq. (18), the constant term a_0 linearly depends on the y_1 value:

$$a_0 = y_1 + \varpi \ , \tag{19}$$

where ϖ is a constant defined via regression analysis.

Slope terms for factors g_{ir} do not depend on the y_1:

$$a_i = f\left(g_{ir}, \Delta y_r\right); \ i \in [1, m]; \ r \in [2, z]. \tag{20}$$

So, in this case, the value of a_0 cannot be considered as the desired solution. It can freely vary jointly for y_1 and a_0. There exists one logical restriction only – the calculated attractiveness for all alternative routes has to be positive that follow from $E_r = R_r - C_r > 0$.

This situation is understandable considering the value of R/c_t represented by a_0 in the regression equation. According to Eq. (3), R/c_t is a constant. As was noted above, this constant could not be determined based only on transportation factors because it is not enough to have information about choice probability and route parameters to uniquely define a level of satisfaction with the home-based work trip.

Thus, considering one trip it is possible to determine the regression coefficient a_0 through setting a value of the attractiveness of the first alternative in the set of all alternatives under consideration. Other regression coefficients and statistic parameters of the model are independent of a_0. This conclusion is very important for expanding the data array to many trips that are usually observed during the surveys, and it is a prerequisite to get the route attractiveness model.

System of Eq. (18) does not provide enough opportunities to determine the regression coefficients. To obtain the desired result, i.e. the route attractiveness function, it is necessary to enlarge the data set from one respondent to a full array of observed trips:

$$\begin{cases} a_0 + a_1 \times g_{111} + \ldots + a_m \times g_{m111} = y_{11} \\ \cdot \\ a_0 + a_1 \times g_{1z_1 1} + \ldots + a_m \times g_{mz_1 1} = y_{11} - \Delta y_{1z_1} \\ \cdot \\ \cdot \\ a_0 + a_1 \times g_{11N} + \ldots + a_m \times g_{m1N} = y_{1N} \\ \cdot \\ a_0 + a_1 \times g_{1 \ldots N} + \ldots + a_m \times g_{mz_N N} = y_{1N} - \Delta y_{Nz_N} \end{cases} . \tag{21}$$

The last index of travel parameters and attractiveness that appears in this equation denotes an ordinal number of respondents. System of Eq. (21) consists of m unknown values of the attractiveness of the first route y_{1k}, $k \in [1,N]$, that should vary for every pair of origin and destination zones. Thus, it is incorrect to specify the same attractiveness of the first route for different pairs of transport zones. The difference between specified attractiveness is more explicit when Eq. (3) is valid and parameters of the route can vary due to different locations of origins and destinations. A priori, an increase in measured costs leads to the reduction of travel attractiveness and vice versa. Owing to this, the unknown value of the attractiveness of the first route is a relative factor, which helps to determine a mutual placement of the attractiveness of the different travels. In other words, all trip attractivenesses are additional unknown values, which affect modeling results. At the same time, it should be understood that during the modeling of individual behavior in the transit system the relation of individuals to the transport parameters should be identical. So,

$$a_i = \text{const}, \ i \in [0, m] \tag{22}$$

for all individuals for whom statistical probabilities of the work travel choice were determined.

Taking into account the expression in Eq. (20), the constancy of the coefficients for all respondents is provided automatically. An intercept of the regression line, based on Eq. (19) and Eq. (22), can be expressed as follows:

$$y_{1k} + \varpi_k = y_{1q} + \varpi_q; \ k \neq q; \ k, q \in [1, N], \tag{23}$$

where ϖ_k, ϖ_q are the constant part of the intercept of the regression line in Eq. (19) for respondents k and q respectively; y_{1k}, y_{1q} are the specified attractiveness for the first alternative route for respondents k and q respectively.

Expression (23) can be rewritten as follows:

$$y_{1k} = y_{1q} + \varpi_q - \varpi_k; \ k \neq q; \ k, q \in [1, N]. \tag{24}$$

So, if Eq. (21) is used to determine coefficients in the model of the route attractiveness, it is enough to specify only one basic value of the attractiveness for the first alternative route, e.g. for the first respondent, $q = 1$. For all other equations of Eq. (21) values of y_{1k} are determined by Eq. (24).

Thus, the proposed model to predict the route choice probabilities provides an assessment of individual attitude toward travel costs related to home-based work trips in a city using the mixed analytical-statistical approach.

4 Experimental results

The experiment was carried out in the second biggest city in Ukraine, Kharkiv, with a population of almost 1.5 million people. In the beginning, the simulation experiment allowed us to discover that when the public transport service is headway-based the distribution of the passenger waiting time at the stop corresponds to the gamma distribution with scale and form parameters that linearly depends on the average headway on the route. At the same time, the major characteristics of the transit system, such as measurable parameters and scheduled headways of the routes, were determined.

Then 866 commuters were surveyed concerning their home-based work trips made during five days. 307 of the obtained questionnaires appeared to be acceptable for further research because they contained information about more than one alternative route and were correctly filled in. The survey results include questionnaires that were filled in by the passengers who used 2 or 3 routes. All set of observed routes counts 681 alternatives.

This survey largely coincides with the investigation of the transportation process described by Raveau et al. (2010). However, the distinction of the survey is that the number of binary probabilities estimated for each respondent was less than five, depending on the number of route alternatives chosen by an individual and the frequency of their choice.

In the next stage, values of the attractiveness of route alternatives were calculated numerically by the Gauss method. It was defined, that waiting time was distributed by the gamma distribution with parameters based on the scheduled route headway. Calculation results are illustrated for one respondent who used three routes with choice probabilities $P_1 = 0.2$; $P_2 = P_3 = 0.4$ and headways of the corresponding routes $J_1 = 16$ min; $J_2 = 6$ min; $J_3 = 5$ min. The initial distribution of waiting time with the equal route attractiveness is shown in Fig. 1.

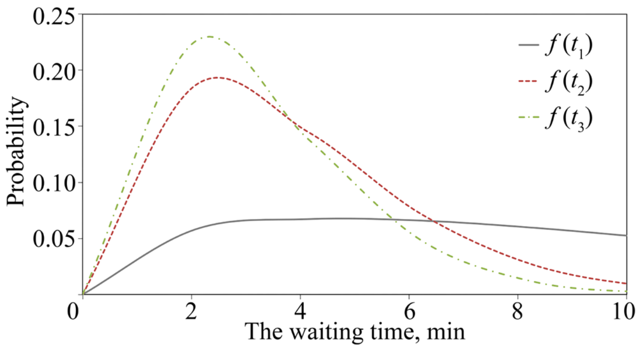

Fig. 1 Distribution of the waiting time when using alternative routes

The probabilities of choosing the routes, that correspond to this base mutual placement of the waiting time distribution, equal $P_1^0 = 0.13$; $P_2^0 = 0.36$; $P_3^0 = 0.51$ according to the numerical integration results. To obtain the probabilities that were observed in the survey, the corresponding waiting time distribution was corrected by adding the route attractiveness Δy_2 and Δy_3.

When solving the system of Eq. (13), the objective function is the minimum magnitude of the vector of disparities between empirical and calculated values. The calculations resulted in the following values of the relative route attractiveness: $\Delta y_2 = 1.186$ min for the second route and $\Delta y_3 = 1.668$ min for the third route. Taking into account relative attractiveness, the corresponding distributions of the waiting time are shown in Fig. 2.

Under these distributions, calculated values of route choice probabilities are $P_1^0 = 0.204$; $P_2^0 = 0.399$; $P_3^0 = 0.397$ and they can be considered as final ones. Based on these calculations, a data array for regression analysis of the attractiveness function for home-based work trips was formed. Processing of the data allowed obtaining the route attractiveness model which is as follows:

$$y = 103.25 - 1.029 \times t - 9.227 \times \gamma - 83.544 \times C \\ -11.430 \times T , \tag{25}$$

where t is in-vehicle time travel from home to work, min; γ is the load factor at the moment of boarding; C is the travel cost, \$; T is the number of transfers on the way to work.

Characteristics of Eq. (25) are given in Tables 1 and 2. Model parameters and attractiveness of the routes show a high correlation, and regression coefficients in the model turned out to be significant, Table 2. However, high statistical results cannot guarantee the accuracy of computed route choice probabilities. To reach that, the actual values of probabilities should be compared with the estimated ones, and then it will be possible to make conclusions

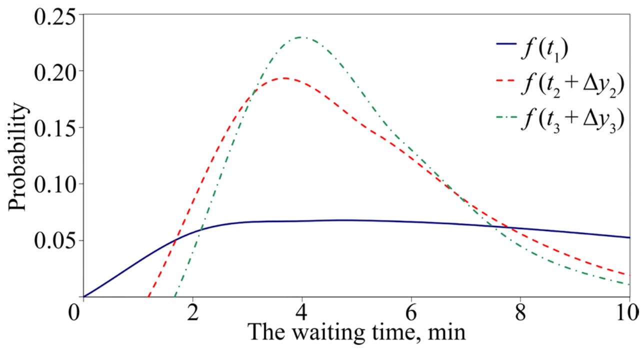

Fig. 2 Distribution of the waiting time when using alternatives routes taking into account the relative attractiveness Δy

Table 1 Regression analysis of the route attractiveness model

Factor	Value
Multiple Correlation Coefficient (*R*)	0.903
R squared	0.816
Standard error	9.159
Number of observations	681

Table 2 Characteristic of regression coefficients

Factors	Coefficients	Standard error	*t*-value	*P*-value
Y-intercept	103.25	4.47	9.68	7.44E-21
Travel time	−1.03	0.09	−12.71	2.16E-33
Load factor	−9.23	3.07	−3.01	0.002706
Travel cost	−83.44	12.20	−4.75	2.47E-06
Number of transfers	−11.43	2.33	−4.91	1.14E-06

regarding their accuracy. To do this, a chi-squared test is acceptable:

$$s_N^2 = \sum_{k=1}^{N}\sum_{j=1}^{z}\left[\frac{\left(\upsilon_{jk}-P_{jk}\right)^2}{\left(\frac{P_{jk}}{n}\right)}\right], \quad (26)$$

where s_N^2 is the measure of deviation between the observed frequencies and calculated choice probabilities – the criterion of model accuracy which is asymptotically χ^2-distributed; υ_{jk} are observed frequencies of the choice of alternative routes j by individual k; P_{jk} is the theoretical probabilities of the choice of alternative routes j by individual k; n is the number of choice situations which were described in the questionnaire of each individual k, $n = 5$.

The criterion of Eq. (26) was applied to make two assessments of choice probabilities. Besides the proposed model presented by Eq. (25), the equiprobable forecast was made via the simplest model when $P_j = 1/z$.

The results were not encouraging. For the proposed model $s_N^2 = 673.3$ that is more than the analogous value for the simplest model $s_N^2 = 579.6$. The significance level for both values of the criterion is practically equal to 0. It means that the direct transition from the observed frequency to regression between the route attractiveness and the route parameters gave acceptable results, but the reverse transition from the route parameters to the choice probability appeared to be unsatisfactory. Taking into account that the proposed method is based only on the confirmed assumptions, a weak predictive ability of the developed model leads to the following conclusions.

5 Conclusions

The proposed model describes an individual behavior in a specific public transit system with a definite type of service. This narrowing during the research allows getting a straight assessment of an individual's attitude to travel parameters taking into account the relative character of alternatives and the mechanism to join trips of different individuals into single array for the regression analysis.

The obtained distribution of the waiting time for the case when public transport service is headway-based and the passenger departure time is random gives a specific sense to this random variable in the expression for the travel efficiency and solves the problem of the relative attractiveness of route alternatives based on integral equations.

In-vehicle travel time, load factor at the moment of boarding the vehicle, fare and the number of transfers appeared to be significant factors which an individual considers when choosing between the routes to get from home to work.

Unsatisfactory results in predicting the probability of the choice of alternative route may be caused by high requirements for accuracy of the waiting time description. An additional reason is in the used approach when the most significant and stable part of the route attractiveness is determined by comparing the relatively small random elements in travel costs. Subject to the presence of unaccounted factors in the model, this approach can lead to serious forecasting errors.

It is necessary to develop an approach to model individual's behavior in a transport system that would be based on the classic principles of data alignment when the modeling purpose is to minimize the difference between predicted and factual data. It should be realized that in contrast to the classical approach, the dependent variable – the quantitative estimate of alternative attractiveness – cannot be surveyed, but the frequency of the use of alternatives can be fixed instead.

References

Ben-Akiva, M., Bierlaire, M. (2003) "Discrete choice models with applications to departure time and route choice", In: Hall, R. W. (ed.) Handbook of Transportation Science, [e-book] Springer, pp. 7–37. ISBN 978-0-306-48058-4
http://doi.org/10.1007/0-306-48058-1_2

Ben-Akiva, M., Boccara, B. (1989) "Integrated framework for travel behavior analysis", In: International Association for Travel Behaviour (ed.) Travel Behaviour Research: Fifth International Conference on Travel Behaviour, Aix-en-Provence, France, 17. ISBN 9780566070624

Ben-Akiva, M., de Palma, A., Kanaroglou, P. (1986) "Dynamic model of peak period traffic congestion with elastic arrival rates", Journal of Transportation Science, 20(3), pp. 164–181. https://doi.org/10.1287/trsc.20.3.164

Ben-Akiva, M., Lerman, S. R. (1985) "Discrete Choice Analysis: Theory and Application to Travel Demand", The MIT Press. ISBN 0-262-02217-0 Available at: https://books.google.com.ua/books?id=oLC6ZYPs9UoC&printsec=frontcover&hl=de&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false [Accessed: 03 September 2022]

Chen, J., Li, S. (2017) "Mode Choice Model for Public Transport with Categorized Latent Variables", Mathematical Problems in Engineering, 2017, 7861945. https://doi.org/10.1155/2017/7861945

de Dios Ortúzar, J., Willumsen, L. G. (2011) "Modelling Transport", 4th ed., Wiley. ISBN 9780470760390 http://doi.org/10.1002/9781119993308

de Palma, A., Rochat, D. (2000) "Mode choices for trips to work in Geneva: an empirical analysis", Journal of Transport Geography, 8(1), pp. 43–51. https://doi.org/10.1016/S0966-6923(99)00026-5

Gnedenko, B. V. (1962) "The Theory of Probability", Chelsea Publishing Company, New York, NY, USA.

Horbachov, P., Svichynskyi, S. (2014) "Dependence of trip length frequency distribution on characteristics of route network elements", Archives of Transport, 30(2), pp. 21–30. https://doi.org/10.5604/08669546.1146974

Horbachov, P., Svichynskyi, S. (2018) "Theoretical substantiation of trip length distribution for home-based work trips in urban transit systems", Journal of Transport and Land Use, 11(1), pp. 593–632. https://doi.org/10.5198/jtlu.2018.916

Kjaer, T. (2005) "A review of the discrete choice experiment – with emphasis on its application in health care", [pdf] University of Southern Denmark, Denmark. Available at: https://scholar.google.com/scholar?oi=bibs&cluster=11494660393827772782&btnI=1&hl=en [Accessed: 03 September 2022]

Koppelman, F. S., Pas, E. I. (1980) "Travel-choice behavior: models of perceptions, preference and choice", [pdf] Transportation Research Record, 765, pp. 26–33. Available at: http://onlinepubs.trb.org/Onlinepubs/trr/1980/765/765-006.pdf [Accessed: 03 September 2022]

Koppelman, F. S., Wen, C.-H. (1998) "Alternative nested logit models: structure, properties and estimation", Transportation Research, Part B: Methodological, 32(5), pp. 289–298. https://doi.org/10.1016/S0191-2615(98)00003-4

McFadden, D. (1974) "Conditional logit analysis of qualitative choice behavior", [pdf] In: Zarembka, P. (ed.) Frontiers in Econometrics, Academic Press, pp. 105–142. ISBN 0-12-776150-0 Available at: https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf [Accessed: 03 September 2022]

McFadden, D. (1978) "Modeling the choice of residential location", [pdf] In: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (eds.) Spatial Interaction Theory and Planning Models, North Holland, pp. 75–96. ISBN 9780444851826 Available at: https://eml.berkeley.edu/reprints/mcfadden/location.pdf [Accessed: 03 September 2022]

Raveau, S., Álvares-Daziano, R., Yáñez, M. F., Bolduc, D., de Dios Ortúzar, J. (2010) "Sequential and Simultaneous Estimation of Hybrid Discrete Choice Models: Some New Findings", Transportation Research Record: Journal of the Transportation Research Board, 2156(1), pp. 131–139. https://doi.org/10.3141/2156-15

Senk, P. (2010) "Route choice Under the Microscope: Process-Oriented Analysis of Decision Making", Transportation Research Record: Journal of the Transportation Research Board, 2156(1), pp. 56–63. https://doi.org/10.3141/2156-07

Train, K. E., McFadden, D. L., Goett, A. A. (1987) "Consumer Attitudes and Voluntary Rate Schedules for Public Utilities", The Review of Economics and Statistics, 69(3), pp. 383–391. https://doi.org/10.2307/1925525