11TH CONFERENCE ON COMPETITION AND OWNERSHIP IN LAND PASSENGER TRANSPORT

Delft University of Technology (The Netherlands)
20-25 September 2009

MODELS FOR COMPETITION BETWEEN PUBLIC TRANSPORT ROUTES AND MODES*

Kjell Jansson
Dept. of Transport and Economics, Royal Institute of Technology, Stockholm
kjell.jansson@infra.kth.se

Harald Lang
Dept. of Mathematics, Royal Institute of Technology, Stockholm
lang@kth.se

ABSTRACT

For assessment of infrastructure measures and find appropriate ways to reduce environment and climate damages etc., forecast models are of utmost importance. The aim of such models is, for assumed transport measures, to forecast demand for various modes and calculate consumer surplus and other components in a cost-benefit analysis.

In Sweden one model is comprised of a combination of one network model for routes within each mode and a structured logit model for the modes. The alternative applied in Sweden is a network model that handles all routes and modes simultaneously. This paper is based on an ongoing project, from which we present the basic principles of the two models, some tentative judgments and research issues that remain. Hopefully the conference will give us further ideas for the rest of the study, which is due to be finished in April 2010.

* This research is financed by the Swedish Governmental Agency for Innovation Systems (Vinnova), the Swedish National Rail Administration, the Swedish Institute for Transport and Communications Analysis (SIKA) and the Stockholm County Public Transport Authority (SL).

1 INTRODUCTION

During the last twelve years the so-called Sampers model has been developed and used in Sweden. This model uses 3 steps: i) the network model Emme/2 for assignment on routes within each mode and for estimation of travel time components for each mode, ii) a multinomial logit model for assignment on modes and for demand forecasts concerning modes and destinations, iii) Samkalk for calculation of consumer surplus, revenues, costs etc. for cost-benefit analysis.

In parallel a network model that we denote RDT (stemming from **R**andom **D**eparture **T**imes), has been used for simultaneous assignment on both routes within each mode and on modes and for calculation of revenues, costs and consumer surplus.

The aim of this paper is to analyse the two models in theory and by use of examples. The concerns in Sweden and the findings of this paper are probably relevant for researchers and planners in other countries even if other models are used.

A crucial matter is that the two types of models employ different assumptions concerning the stochastic element with respect to choice of alternatives. The logit model in Sampers assumes individual randomness depending on measurement errors, preference differences etc. while RDT instead assumes that the passengers' ideal departure (or arrival) times are randomly and uniformly distributed.

When dealing with competition between operators or modes a crucial issue is what factors affect the passengers' choice. Clearly travel time components and price matter, among other factors, and time and price are not valued the same by all individuals. These are important facts that should not be ignored. There are at least three methods to take care of variations of travel time and price, as well as other factors:

a) Apply separate analyses for passenger groups with different values of travel

time components. This segmentation would take care of some of the\taste" variation in

terms of varying willingness to pay for reduction of travel time components.

b) Apply randomness to reflect passengers' different ideal departure or arrival

times.

c) Apply randomness as a model of taste variations and other unknown factors.

Of course, one can consider various combinations of these methods. To find a reasonable combination may be seen as the main aim of the project.

Another matter in focus is how the two models handle calculation of consumer surplus.

Finally, there are features in public transport that differ from private transport. For instance in making a journey with public transport modes there is usually no one main mode but the journey is made up of a combination of modes, e.g. Train-Coach, Bus-Train-Flight etc. There is also the issue of specifying the fare systems, since in practice many different fare structures are employed. These special features of public transport make it complicated to model.

Note that in this paper we are only dealing with features that both approaches handle, i.e., assignment between routes and modes, generalised cost and consumer surplus. Sampers, but not the RDT approach that we are referring to, also forecasts destination choice and trip generation, i.e., changes of the structure of O-D matrices.

2 MODELLING UTILITY, DEMAND AND CONSUMER SURPLUS

In this section we analyse how assignment of passengers and consumer surplus can be derived when there is more than one service or mode to choose among.

Without loss of generality we confine to two alternative routes (or modes), one O-D pair and one passenger group. Each group should be as homogenous as possible with respect to ticket prices and valuation of time components in relation to price. We will not discuss this matter further but evidently a large number of segments will provide more reliable results than few segments.

For the purpose of this paper we do not need to discuss the various travel time components in detail. Each of these components is assumed to have a specific constant value of time.

In order to concentrate on the issues addressed in this paper it is sufficient to distinguish between on the one hand wait time, V, which depends on frequency of service, and on the other hand all other travel time components and price, R. The sum of V and R is called generalised cost, G. All elements are expressed in minutes by use of values of time (VoT). For convenience R is here often referred to as travel time only.

2.1 Basic micro-economic model

We assume that the substitution quotient between time and money is the same for all individuals, i.e. that all have the same valuation of time. We ignore the income effect, which is standard in transport analysis.

Each individual is assumed to choose the alternative with minimum G. This G is, however, not the same for each individual due to stochastic influence. In order to simplify notation and calculations, without affecting the general aspects, we assume that there are two alternatives, 1 and 2.

The generalised cost of alternative j (j=1,2) for each individual i is composed of the following elements. Travel time R^j (including all travel time components plus price, except wait time) plus a stochastic variable, t^j_i, that varies among individuals with taste, measurement errors etc. plus a stochastic variable, x^j_i, that varies among individuals with ideal departure or arrival time in relation to actual time. We define x^j_i as time to departure, i.e., the difference between actual and ideal departure time. The generalised cost of alternative j is then:

(1) $G^j = R^j + t_i^j + x_i^j$

When each individual chooses the alternative with the minimum generalised cost the realised “joint” (or combined) generalised cost of individual i is:

(2) $G_i = \min\left[R^1 + t_i^1 + x_i^1, R^2 + t_i^2 + x_i^2\right]$

The average joint generalised cost of both alternatives over all individuals in a segment is then:

(3) $G = E\left[\min\left[R^1 + t_i^1 + x_i^1, R^2 + t_i^2 + x_i^2\right]\right]$

where E denotes the expected value corresponding to the distribution of individuals.

We have thus defined one single G for a journey from door to door when there are several alternatives to choose among. The deviation ε_i from the joint G for an individual could be composed of t_i and/or x_i. The individual G_i is then defined by:

(4) $\mathrm{G_i} = \mathrm{G} + \varepsilon_\mathrm{i}$

Each individual is assumed to have a utility of travelling from origin to destination, i.e., the utility of the journey itself, which is denoted v_i.

The net utility for individual i, when taking G into account, is:

(5) $$v_i - G_i = v_i - \varepsilon_i - G \equiv u_i - G$$

Let f(u) be the density function over u_i among the individuals.

The individual chooses to travel if $u_i \geq G$, where u_i has a distribution f(u) over all individuals. The choice is illustrated in the figure below.

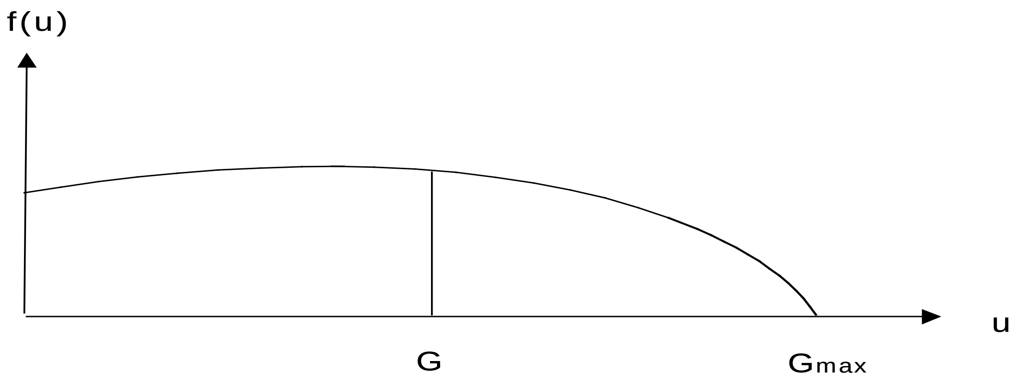

Figure 1: Distribution of utility

The aggregate demand, X, is the integral over f(u) between G and the reservation price G_{max}.

(6) $$X = \int_G^{G_{max}} f(u)\,du = X(G)$$

The consumer surplus, S, is thus:

(7) $$S = S(G) = \int_G^{G_{max}} (u - G)\,f(u)\,du$$

It then follows that:

(8) $$\frac{\partial S}{\partial G} = -(G - G)\,f(G) + \int_G^{G_{max}} -f(u)\,du = -X$$

Observe that consumer surplus is a function of the joint generalised cost, which in turn is a function of the generalised cost of both alternatives.

(9) $$S = S(G) = S(G(G_1, G_2)) = S(G_1, G_2)$$

3 THE RDT MODEL

3.1 Introduction

Here we refer to the models applied in for example Sweden, Vips and Visum[*], which has the RDT property.

Assuming that passengers know the timetable RDT estimates assignment on routes and modes and calculates all travel time components and price in one single step.

The passengers in each origin zone normally have a choice between various walk links to various services and modes.

RDT assumes no stochastic variation with respect to preferences. In practical applications instead the model allows a) substantial segmentations for passenger categories with respect to different values of time, b) that services and modes are given specific characteristics in terms of comfort, price etc., which may differ between passenger categories.

It is the model itself that generates all possible travel paths from origin and destination, using a number of combinations of services and modes. The number of travel paths (each with a combination of services and modes from origin to destination) can be very large for the Swedish national network, even up to around 50.

The model itself calculates the fare for each travel path by adding the (user specified) price of each service and mode from origin to destination. The price structure is defined separately for each service, composed of a base price and a variable price dependent on distance or zones. Each passenger segment can then have a specific price structure for each mode and service within mode.

3. 2 Principles

When departure times of all routes are known all routes and stops are considered simultaneously, but all cannot be acceptable. Assume that different routes i have travel times Ri and headway Hi. Expected wait time is then not Hi/2. Expected wait time when the timetable is known is the difference between ideal and actual departure time. The basis for choice of acceptable routes is walk time to the stop plus travel time after boarding, here denoted Ri, plus all of the headway, Hi. Assume that route 1 is best, has the lowest value

[*] Vips was originally developed in Sweden. The ptv AG in Germany purchased it and implemented the basic Vips algorithms in Visum.

R1+H1. Other routes m are acceptable if Rm < R1+H1. This means that it is not worthwhile to wait for a route that has travel time only that is longer than travel time plus the whole headway of the best route.

The RDT approach ignores the stochastic element t that varies with taste differences among individuals, measurement errors etc. We are thus left with the stochastic element x, difference between actual and ideal departure time, often called schedule delay. This delay is here based on expected delay based on average frequencies of services and not on exact departure times.

It is assumed that (x1, x2) has a uniform distribution on [0, H1]×[0, H2]. This assumption is in turn based on the assumption that we do not now anything about the true distribution about ideal departure times for the period of time (peak hours or non-peak hours for example) we are analysing. It is also assumed that departure times of alternative routes are uniformly distributed.

Notation

H^1 headway of route1.

H^2 headway of route2.

R^1travel time (including price expressed in minutes) of route1.

R^2 travel time(including price expressed in minutes) of route2.

t^1 time to departure of route 1.

t2 time to departure of route 2.

Expression (3) is then:

$$(10) \qquad G = E\left[\min\left[R^1 + x_i^1, R^2 + x_i^2\right]\right]$$

It has then been shown, see Jansson, Lang, Mattsson (2008) and Hasselström (1981), that the probability of choice of alternative 1, Pr(1), is:

$$(11)) \qquad \Pr(1) = \frac{1}{H^1 H^2} \int_0^{H^1} \int_0^{H^2} h\left[R^2 - R^1 + x^2 - x^1\right] dx^2 dx^1$$

where h(s) is the heaviside function defined by:

$$(12) \qquad h(s) = \begin{cases} 1 & \text{if } s > 0 \\ 0 & \text{if } s \le 0 \end{cases}$$

Note that the probability for choice of a specific route depends on travel times, prices and intervals of all acceptable routes.

Note that R^1 and R^2 may have a different weight in relation to the weight of the headway. Jansson, Lang, Mattsson (2008) and Hasselström (1981) also show that the expected wait time, V, is:

$$(13) \qquad V = \frac{1}{H^1 H^2} \int_0^{H^1} \int_0^{H^2} h\left[R^2 - R^1 + x^2 - x^1\right]\left(x^1 - x^2\right) + x^2 dx^2 dx^1$$

The average expected travel time when there are several acceptable routes is found by the weighted travel time for all routes where the weights are the calculated probabilities. If there are j acceptable routes and the travel time for route j is Rj and the probability of choice of route j is denoted Pr(j), the average expected travel time, R, is:

$$R = \sum_{j=1}^{k} \Pr(j) \mathrm{R}^{j} \tag{14}$$

The generalised cost is simply the sum of the joint expected wait time and the average expected travel time: G=V+R

The two figures below illustrate the choice probabilities.

Figure 2 illustrates points of time, headway and travel times of the two routes. The total bar lengths represent the maximal costs associated with the routes. We assume that the two routes arrive at the same time (0). The passengers' ideal departure times are along the x-axis; the wait time for each alternative is a uniformly distributed random variable over the light coloured parts of the bars. The passengers chose the alternative where this wait time variable is closest to the origin, having smallest total cost. For ideal departure times between 15 and 20 min only route 1 is chosen. For other times the passengers are split between the two alternatives according to the proportions Pr[1jt1] and 1 -Pr[1jt1].

The conditional probability of selecting route 1 is shown in figure 3.

For each alternative in figure 2 the wait time is a random variable uniformly distributed over the yellow (light coloured) parts of the bars.

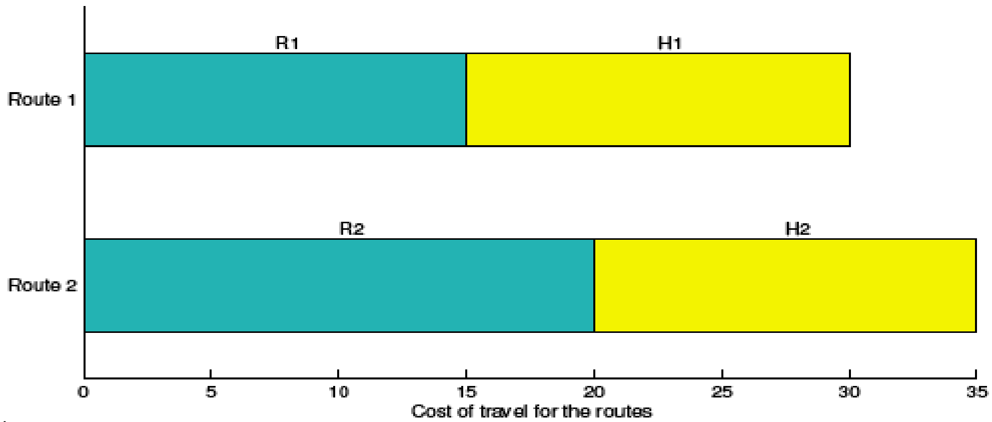

Figure 2: Travel time and headway for routes 1 and 2

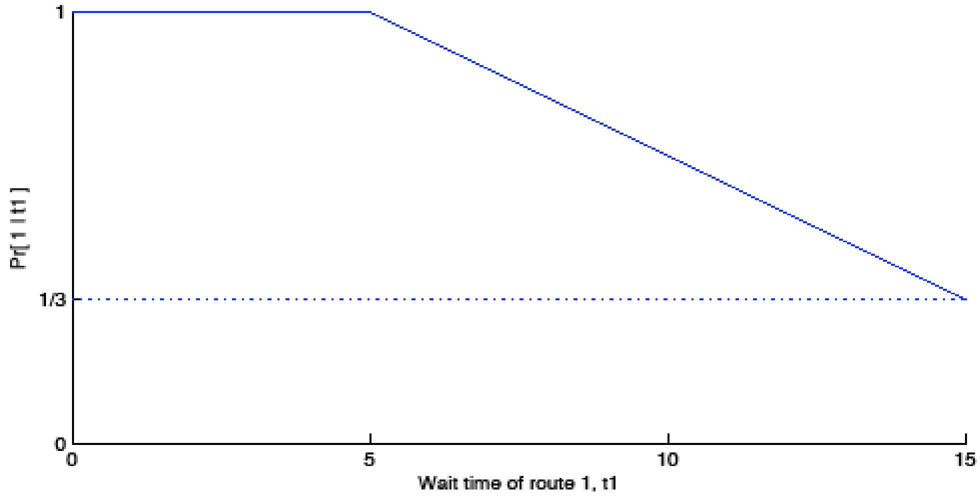

Figure 3: The conditional probability of selecting route 1 as a function of t[1]

4 THE SAMPERS MODEL

4.1 Introduction

While RDT works in one step the Sampers model works in three steps:

Emme/2 (version based on frequency-based assignment) is used in the Sampers model in order to give the travel time components of each alternative mode. The services belonging to each specific mode contribute to a “joint” generalised cost of this mode.

The joint generalised cost of each mode is put in to the multinomial logit model for assignment on modes.

For the estimation of consumer surplus changes, due to changes of network, prices etc. a special complementary program, Samkalk, is used.

The next three sections will describe each of these three steps.

4.2 Basic features of Emme/2 in Sampers

The Emme/2 model used in Sweden is based on average frequencies of services and not on real timetables. The passengers are assumed to know the travel time components and headway of all routes. But they do not know the timetable (the actual departure times) or behave as if they do not. They therefore are assumed to arrive uniformly distributed to the stop. Each passenger’s ideal departure time is when arriving at the stop. Also the departure times between services are assumed uniformly distributed.

The basic behavioural assumption is that passengers choose the best alternative, the alternative with minimum weighted travel time components, i.e., generalised cost except price.

The consequence of not knowing the timetable is that passengers walk to the stop with the shortest expected total travel time. Since only one stop is taken into account typically only one mode can be chosen (unless where for example a bus stop and a train station can be regarded being the same stop). Since typically only one stop and mode is chosen another complementary model, e.g., the logit model is needed for assignment on modes. A consequence is that Emme/2 cannot assign passengers to more than one airport or one railway station from an origin zone.

All alternatives are not acceptable. Assume that different routes i have travel time R_i and headway (interval) H^j, both expressed in minutes. Frequency of service, F^j, i.e., the number of departures per hour is then $60/H^j$. Expected wait time, V, if only route i is available is then

$V = H^j/2$. Assume that route 1 is best, i.e., has the lowest value $R^j+H^j/2$. The second best route m is accepted if $R^m \leq R^j+H^j/2$. This means that it is not worthwhile to wait for a route that has travel time only that is longer than travel time plus half the headway of the best route.

The probability of choice of route j, Pr(j), among k acceptable routes is calculated according to:

$$\text{(15)} \qquad \Pr(j) = \frac{F^j}{\sum_k F^k} \equiv \frac{60/H^j}{\sum_k 60/H^k}$$

Since the timetable is assumed not known the wait time is dependent on frequencies only. Other travel time components are ignored. The probabilities are actually proportional to frequencies.

The expected, joint, wait time, V, is calculated as half the average headway according to:

(16) $$V = \frac{60}{2\sum_k \mathrm{F}^k} \equiv \frac{60}{2\sum_k 60/\mathrm{H}^k}$$

Note that this way to calculate wait time is valid only if the departures are perfectly co-ordinated (evenly phased), i.e., that the gaps between the departures of different routes are the same. Such perfect co-ordination is theoretically possible only if all alternatives have the same headway. In practice perfect co-ordination is difficult to achieve even if all routes have the same headway, but for certain segments of parallel routes. The reason is that a route often runs parallel with different routes along different sections. The consequence is that Emme/2 underestimates wait time.

The expected travel time when there are several routes to choose among is calculated as the weighted travel time for all routes where the weights are the calculated probabilities for choice. If there are j acceptable routes and the travel time for route j is R^j and the probability of choice of route i is denoted Pr(i), the average expected travel time, R, is:

(17) $$R = \sum_k \mathrm{Pr(j)R}^{\mathrm{j}}$$

Generalised cost for a mode, G, is the sum of the calculated joint wait and travel time costs over all services belonging to this mode: G= V+R.

The generalised costs, G, from origin to destination for each alternative mode are then put in to the next step, he multinomial logit model.

It should also be observed that Emme/2 does not distinguish between weight for (first) wait time and weight for transfer time. This is a defect especially for long-distant infrequent transport since first wait time is normally taken at home etc. The weight for transfer time is typically 4 to 5 times the weight for first wait time since the first wait time is not spent at the stop.

4.3 Basic features of the logit model in Sampers

Sampers applies a standard multinomial logit model dealing with modes, using generalised cost (except price) of each mode from Emme/2. The price of each so-called “main mode” is estimated exogenously and added to the generalised cost.

With this structure of the Sampers model one thus has to define “main modes” between each O-D pair. The model thus ignores the fact that many journeys between origin and destination need several modes. One combination could be commuter bus plus train plus regional bus. Another combination could be regional bus plus train plus flight, etc.

While in RDT the wait time, x, was a stochastic variable, it is here a deterministic value calculated by Emme/2. Now instead the stochastic variation with respect to preferences etc., t, is taken into account, as a deviation from G. When each individual is assumed to choose the alternative with the minimum generalised cost the realised generalised cost of individual i is (also see Jansson, Lang, Mattsson, Mortazavi (2008), Ben-Akiva and Lerman (1985) and Louviere et al., (2000)):

(18) $$\mathrm{G} = \mathrm{E}\left[\min\left[G^1 + t_{\mathrm{i}}^1, G^2 + t_{\mathrm{i}}^2\right]\right]$$

The logit model assumes that the error term is so-called Gumbel distributed with a scale factor $\mu>0$, which has the inverse dimension of G, i.e., 1/minutes or 1/money. The share of the passengers that will choose alternative j, Pr(j), among k alternatives is calculated according to:

$$\text{(19)} \qquad \Pr(j) = \frac{e^{-\mu G^j}}{\sum_{i=1}^{k} e^{-\mu G^i}}$$

The joint G for the two alternatives is represented by the so-called logsum (see for example Small and Rosen (1981), Ben-Akiva and Lerman (1985)), expressed as:

$$\text{(20)} \qquad G = \frac{1}{\mu}\ln(e^{-\mu G^1} + e^{-\mu G^2})$$

The logit model thus produces not only measures for probabilities but claims also to calculate joint generalised cost. The difference in generalised cost between two alternative public transport scenarios should thus be represented by the difference between the logsums of these scenarios.

This is something that may be questioned with respect to public transport. In an E-mail dialogue in an earlier research work Andrew Daly (2004) wrote "*One problem is that a model of choice among routes may yield a logsum that is not a representation of the total quality of the combined service – this is a standard feature of hierarchical models*."

Let us look at the following simple example in order to illustrate the problem. Assume that originally there is only one alternative, 1, where the joint G equals G^1. In this original situation the joint G according to the logsum is simply:

$$\text{(21)} \qquad G = \frac{-1}{\mu}\ln(e^{-\mu G^1}) \equiv \frac{1}{\mu}\mu G^1 \equiv G^1$$

Assume now that we double the number of alternatives so that there are two alternatives with the same G. The new joint G* is then:

$$\text{(22)} \qquad G^* = \frac{-1}{\mu}\ln(2e^{-\mu G^1}) \equiv \frac{-1}{\mu}\ln 2 + \frac{-1}{\mu}\ln(e^{-\mu G^1}) \equiv G^1 - \frac{1}{\mu}\ln 2$$

The change of the joint G is thus $(1/\mu)\ln 2$. If we have k alternatives with the same G the joint G would be $(1/\mu)\ln k$.

The conclusion is that the logsum cannot be use for a representation of combined generalised cost of a number of modes.

Some other problematic features of the multinomial logit model are:

a) The cross elasticity with respect to generalised cost, G, or any component in G, is uniform, i.e., the cross elasticity of the probability of alternative i with respect to a change of G^j are equal for all alternatives $i \neq j$, i.e., $\varepsilon^{ij} = \Pr(j)\mu G^j$.

b) The direct elasticity with respect to generalised cost, G, or any component in G, is proportional to the level of G or any other component, and proportional to the scale i.e., $\varepsilon^i = -(1 - \Pr(i))\mu G^i$.

c) The probability of choice of each alternative depends only on the difference between the generalised cost levels irrespective of headway. Assume that in one situation there are two alternatives with generalised cost 10 minutes and 20 minutes respectively. Assume that in another situation there are two other alternatives with generalised cost 350 minutes and 360 minutes respectively. The logit model calculates the same probabilities for the two alternatives in both situations.

4.4 Basic features of the Samkalk step in Sampers

For estimation of consumer surplus the logsum from the logit model is in fact not used in Sampers. In the supplementary Samkalk model, which is a model for calculation of consumer surplus (and other components for cost-benefit analysis) instead ”rule-of-the-half” is employed.

Consumer surplus is then based on change of generalised cost only for the alternative mode that has been subject to change of price, travel time components or frequency. For existing/remaining passengers the change in consumer surplus is the change in generalised cost (G) multiplied by the number of passengers. For new or lost passengers on this mode the change in consumer surplus is the change in generalised cost (G) multiplied by the number of new/lost passengers, divided by 2.

The change of consumer surplus is thus assumed independent of the generalised cost components of other modes and the change of shares of these modes.

The set of values of time used is not consistent. One set is used in the Emme/2 step and another set in the logit model.

5 SUMMARY OF FEATURES OF THE TWO MODELS

5.1 Characteristics of Sampers

In summary the features of Emme/2 are:

- It does not handle fares,
- The model assumes that passengers do not know the time table, which implies that only one mode is chosen and passengers are assigned only in proportion to frequency, ignoring other travel time components. This means a problem in general but especially if fast modes such as skip-stop-services or high-speed trains are evaluated.
- It assumes perfect co-ordination between services, which means underestimation of wait time,
- Weight of first wait time and weight of transfer time cannot be separated,
- Travel time components and generalised cost may be severely wrong.
- Since prices and ride times are not considered (except for elimination of unacceptable routes) one cannot assess the effects of faster versus slower trains or expensive versus cheap flights.

In summary the features of the logit model in Sampers are:

- It deals with main modes only, not with combinations of modes for the whole journey, e.g., bus plus train or bus plus train plus air.
- Prices are put in exogenously as a matrix for all O-D pairs, without any concern for prices of combinations of modes.

- It cannot distinguish between various airports and prices of various routes or various operators.
- It provides a logsum over public transport modes that have dubious features since it may not reflect the combined generalized cost for a set of public transport modes. Nevertheless the logsum is used for estimation of demand including choice of destination, something that might be questioned.
- A basic feature is that all alternative modes are taken into account simultaneously. However, for calculation of consumer surplus, the logsum is not applied, for good reasons since it does not reflect the combined generalized cost. Instead this calculation is carried out in Samkalk.

In summary the features of the Samkalk model are:

- Samkalk takes into account only the effect of the mode that is subject to a change, in ride time, headway, price etc. This way of calculating consumer surplus is not correct if headway or headway and other travel time components or price are changed. This matter is further discussed in section 6 below. Note thus that the basic philosophy in the logit model that all modes are regarded together is violated in the Samkalk step.
- Samkalk cannot calculate the consumer effects if more than one mode is subject to some change. Neither is it possible to estimate the consumer effect of a change of one mode in a first step and of another mode in a second step.
- A severe problem is that so-called “main modes” have to be specified outside the model. In reality there is seldom a main mode or it is difficult to specify, since for many O-D pairs there may be many combinations of services. Some journeys need regional bus plus train plus regional buss. Another journey may need regional bus plus train plus flight plus airport bus. In addition there may be a variety of travel paths to reach the destination each with a specific combination of services and modes.
- A related problem is that the price for each “main mode” also is specified outside the model. The fact that the price depends on combinations of services is not taken into account.
- One cannot specify different prices for different types of trains, e.g., commuter train, Intercity train, high-speed train. One cannot specify different prices for different airlines. One cannot specify different prices for different bus or coach services.
- All passengers under and over 100 km respectively are assumed to have the same value of time for all modes. No concern is taken for the fact that passengers may perceive different modes as more ore less comfortable, in other ways than using a mode constant.

5.2 Characteristics of RDT

- The RDT model estimates assignment on routes and modes and calculates all travel time components, price and generalised cost in one single step.
- The passengers in each origin zone normally have a choice between various walk links to various stops and modes.
- It does not take into account individual randomness depending on for example dispersion of starting and destination points within areas, taste variation, various measurement errors.

- In practical applications the model takes into account variation between passenger categories by a) substantial segmentation with respect to different values of time per mode depending on perceived comfort, b) specific prices for various services.
- No "main modes" are defined. It is the model itself that generates all possible travel paths from origin to destination, using a number of combinations of services and modes.
- The model itself calculates the fare for each travel path by adding the price of each service and mode from origin to destination. The price structure is defined separately for each service, composed of a base price and a variable price dependent on distance or zones. Each passenger segment can then have a specific price structure.
- It does not contain ways to forecast destination choice and trip generation, i.e., changes of the structure of O-D matrices.

6 THEORY AND EXAMPLES ON CONSUMER SURPLUS CALCULATIONS

6.1 Theory

The issue addressed here is whether one can calculate change of consumer surplus by use of the change of generalised cost of each alternative respectively or whether one shall use the change of the joint generalised cost.

Assume the case where the total demand is constant X, so that we deal with demand variations between alternatives only. The gross consumer surplus (net of generalised cost) is denoted S*. The consumer surplus for the joint generalised cost G is then:

$$\text{(23)} \qquad \mathrm{S} = \mathrm{S}^* - X\,\mathrm{G}$$

By use of (11) and (8) we know that the derivative of S with respect to a change of travel time of alternative 1 is:

$$\text{(24)} \qquad \frac{\partial \mathrm{S}}{\partial \mathrm{R}^1} = \frac{\partial \mathrm{S}}{\partial \mathrm{G}}\frac{\partial \mathrm{G}}{\partial \mathrm{R}^1} = -\mathrm{X}\Pr(1)$$

The total differential of S is:

$$\text{(25)} \qquad \mathrm{dS} = -X(\Pr(1)\,\mathrm{dR}^1 + \Pr(2)\,\mathrm{dR}^2)$$

In order to compute the change ΔS in consumer surplus due to a change ΔR of R^j, one can integrate the demand function X^j for mode j:

$$\text{(26)} \qquad \Delta \mathrm{S} = -\int_{R^j}^{R^j+\Delta R} X^j\,\mathrm{dR}^j$$

and similarly for a change in the headway H^j by ΔH:

$$\text{(27)} \qquad \Delta \mathrm{S} = \frac{1}{H^j}\int_{H^j}^{H^j+\Delta H} X^j E\left[x^j \middle| j\right]\mathrm{d}H^j$$

This means that the change of consumer surplus can be calculated with respect to change of travel time of alternative 1 only, but it does not hold for change of the interval of alternative 1. The proof of this is seen in Jansson, Lang, Mattsson (2008).

A principle example and schematic examples of consumer surplus calculations are given in the next two sections.

6.2 Principle example

Let us consider two alternative traffic modes (between two given locations A and B). One of the modes has the headway h_1 and the other the headway h_2, which are assumed not to be synchronized. The travellers are assumed to consult the timetable and to choose the mode that has the shortest waiting time between preferred and actual departure time. These waiting times are stochastic, independent and uniformly distributed over the interval $(0, h_1)$ viz. $(0, h_2)$. We standardise the total number of travellers to 1 and keep it constant while changing the intervals h_1 and h_2. We further assume that $h_1>h_2$ (See illustration below; the x and y axes represent the stochastic waiting times for travellers choosing mode 1 (with headway h_1) viz. mode 2 (headway h_2.)

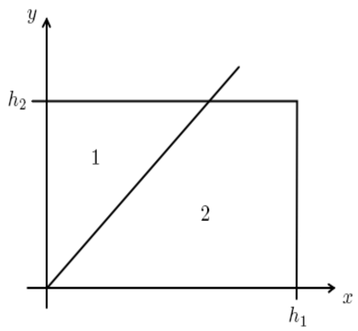

Figure 4: Choice of mode 1 and 2

We assume for simplicity that all other travel time components and prices are the same for both modes. We can now calculate the total number of travellers with mode 1 (see figure above.)

(28) $$X_1 = \frac{1}{2}\frac{h_2^2}{h_1 h_2} = \frac{1}{2}\frac{h_2}{h_1}$$

and all remaining travellers will choose mode 2 i.e.,

$$X_2 = 1 - \frac{1}{2}\frac{h_2}{h_1}$$

We now calculate the total waiting time V_i for travellers travelling with mode i, i=1, 2:

$$V_1 = \frac{1}{h_1 h_2}\int_0^{h_2}\int_0^{y} x\,dxdy = \frac{h_2^2}{6h_1}$$

$$V_2 = \frac{1}{h_1 h_2}\int_0^{h_2}\int_0^{y} y\,dxdy = \frac{h_2}{2} - \frac{h_2^2}{3h_1}$$

which gives the total waiting time (V_1+V_2):

(29) $$V_1 + V_2 = \frac{h_2}{2} - \frac{h_2^2}{6h_1}.$$

We may now calculate the average waiting time per traveller on both modes:

(30) $$\frac{V_1}{X_1} = \frac{1}{3}h_2$$

(31) $$\frac{V_2}{X_2} = \frac{h_2(3h_1 - 2h_2)}{3(2h_1 - h_2)}$$

from which we observe that (since $h_1 > h_2$ by assumption) the average waiting time per passenger is longer for mode 2 than for mode 1 (sic!)

Now it is also easy to see what happens with the consumer surplus (i.e. the negative waiting time) if we decrease the headway for the mode with the longest interval (mode 1 with headway h_1;) however we assume that $h_1 > h_2$ holds also after the change.

Such a decrease will lead to

1. Unchanged average waiting time for travellers with mode 1 (sic! See (30))

2. Shorter average waiting time on mode 2, see (31)

3. More travellers will choose mode 1 and hence fewer will choose mode 2, see (28)

4. The total waiting time for all travellers decreases (that is the consumer surplus increases,) see (29)

We observe that the increase of the consumer surplus stems from two sources: first, the number of travellers on mode 1 increases, which has the shorter waiting time per passenger, and second, the average waiting time decreases for passengers on mode 2.

This example illustrates that the method applied by Samkalk to calculate the consumer surplus is not applicable. Samkalk would calculate in the following way; on one hand those who already travel by mode 1 will get a reduction of total travel time by

(32) $$X_1(h_1 - h_1^*)/2$$

and on the other hand new passengers on mode 1 will get a total reduction of travel time by

(33) $$(X_1^* - X_1)(h_1 - h_1^*)/4$$

In formulas (32) and (33) h_1^* and X_1^* denote the headway viz. the number of travellers after the change.

The relation (32) is based on the idea that the average waiting time for those travelling with mode 1 is $\frac{1}{2}h_1$. As we have seen this is not correct; the average waiting time is $\frac{1}{3}h_2$, which is considerably shorter. Moreover, it is independent of h_1 as long as $h_1 > h_2$. Already (32) causes Samkalk to exaggerate the reduction of waiting time by at least 50 per cent, since the actual reduction of waiting time are, as shown in the calculations above,

$$\Delta V = \frac{h_2^2}{6h_1h_1^*}(h_1 - h_1^*)$$

while (32) gives

$$\Delta V_s = \frac{h_2}{4h_1}(h_1 - h_1^*)$$

which gives the ratio

$$\frac{\Delta V_s}{\Delta V} = 1.5\frac{h_1^*}{h_1} \geq 1.5$$

When (33) is further added to the Samkalk calculation, the error will become even greater.

6.3 Schematic examples

Before we provide a few schematic examples we will mention some results from a previous study that was financed by Swedish Institute for Transport and Communications Analysis (SIKA). In this study we had the whole Swedish network, with roads and all routes and modes coded in both Sampers and Vips, and we ran the models for various assumed changes of ride times, frequencies and prices for various modes. We found extraordinary differences between the results of the models in terms of consumer surplus changes. Sampers could give the double change compared to Vips and vice versa. One goal in this project is to find the reasons for these differences. The cases where Sampers gives a larger change of consumer surplus seems so far easier to understand, and the schematic examples here refer to this case.

Assume that there are two modes, M9 and B7, where M9 could be an airline and B7 a railway line for example.

In the original reference situation, denoted reference alternative (RA) B7 has ride time 60 minutes. M9 has headway 180 minutes and B7 headway 60 minutes.

For a new alternative solution (NA) the headway of M9 is reduced to 120 minutes. We evaluate the consequences in terms of generalized cost and consumer surplus (CS) of this change of headway for various ride times of M9, in the range 30 – 90 minutes.

The fixed demand between origin and destination is 1 000 passengers.

The figure below shows the relationships between consumer surplus (CS) according to Samkalk and Vips respectively, for each of the assumed ride times for M9.

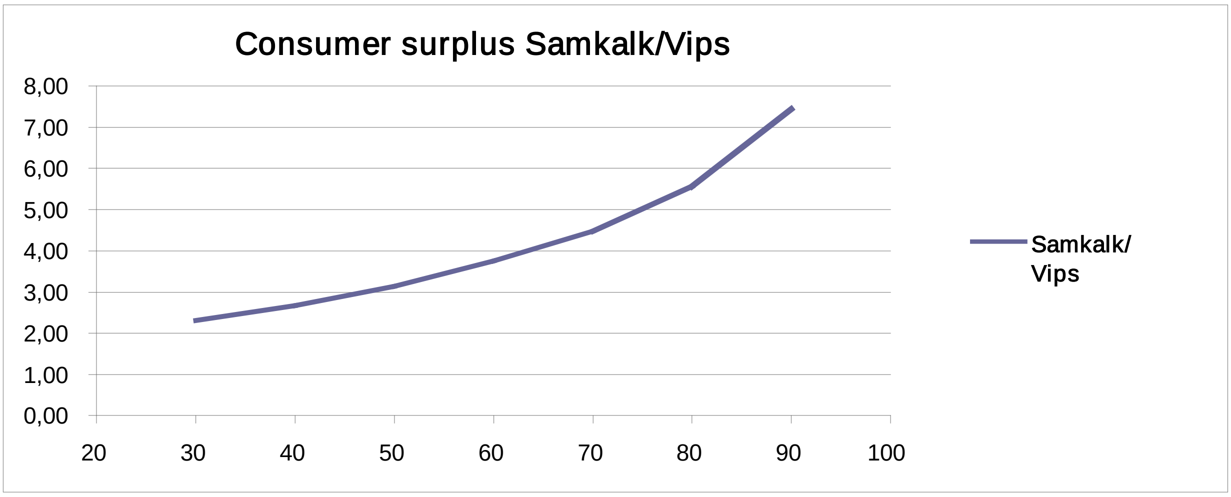

Figure 5: Consumer surplus Samkalk/Vips for various ride times of mode M9

CS is around 2-7 times larger according to Samkalk compared to Vips. The reasons are the following:

Samkalk takes into account only the change of M9 in terms of reduced wait time, assuming that the existing passengers on M9 gain half the reduction of the headway and that the additional, new, passengers gain half of this (rule of the half).

Vips calculates the composite change of both wait time and ride time for all 1 000 taking into account both modes together.

The table below shows the results for all scenarios.

Table 1: Results of all scenarios

Prerequisities	**RA**				**NA**				**Consumer surplus Samkalk/ Vips**
	Ride time M9	**Head-way M9**	**Ride time B7**	**Head-way B7**	**Ride time M9**	**Head-way M9**	**Ride time B7**	**Head-way B7**	
Ride time	30	180	60	60	30	120	60	60	2,31
Ride time	40	180	60	60	40	120	60	60	2,68
Ride time	50	180	60	60	50	120	60	60	3,14
Ride time	60	180	60	60	60	120	60	60	3,76
Ride time	70	180	60	60	70	120	60	60	4,47
Ride time	80	180	60	60	80	120	60	60	5,55
Ride time	90	180	60	60	90	120	60	60	7,44

The table below gives some more details for two scenarios, that M9 has ride time 30 minutes and 90 minutes respectively.

Table 2: Detailed results for two scenarios)

Mode M9 has ride time 30 minutes

Prerequisities	RA		NA	
	M9	B7	M9	B7
Ride time	**30**	60	30	60
Headway	180	60	120	60
Samkalk	Ride time	Wait time	G	CS
existing 333	0	-30	-30	10 000
new 167	0	-15	-15	2 500
Sum				12 500
Vips	Ride time	Wait time	G	CS
All 1 000	-5,00	-0,42	-5,41	5 410
CS Samkalk/ Vips				
2,31				

Mode M9 has ride time 90 minutes

Prerequisities	RA		NA	
	M9	B7	M9	B7
Ride time	**90**	60	90	60
Headway	180	60	120	60
Samkalk	Ride time	Wait time	G	CS
existing 42	0	-30	-30	1 249
new 21	0	-15	-15	313
Sum				1 562
Vips	Åktid	V-tid	G	KÖ
All 1 000	0,63	-0,83	-0,21	210
CS Samkalk/ Vips				
7,44				

The case where M9 has ride time 30 minutes

Samkalk means that the existing 333 passengers gain 30 minutes wait time and that the new 167 passengers gain 15 minutes. The change in generalised cost, G, is the same. Since only M9 is taken into account there is no change in ride time. The change in CS is thus in total 12 500.

Vips takes into account that both modes contribute to the level of service. Since more passengers choose M9 due to the shorter headway the composite ride time is reduced, since M9 is faster than B7. The reduction of the composite wait time is lower than that according to

Samkalk since B7 has a shorter headway than M9. The change of CS according to Vips is 5 410.

CS is 2.31 times higher according to Samkalk compared to Vips.

The case where M9 has ride time 90 minutes

Samkalk means that the existing 42 passengers gain 30 minutes wait time and that the new 21 passengers gain 15 minutes. The change in generalised cost, G, is the same. Since only M9 is taken into account there is no change in ride time. The change in CS is thus in total 1 562.

Vips takes into account that both modes contribute to the level of service. More passengers choose M9 due to the shorter headway. In this case, however, the passengers will lose in terms of ride time since M9 is slower than B7. The change in CS is 210.

CS is 7.44 times higher according to Samkalk compared to Vips.

7 CONCLUSIONS AND SUBSEQUENT RESEARCH

The conclusions drawn here are preliminary and we mainly outline subsequent issues in this research project, and maybe in future ones.

Preliminary we have found that the Sampers model has a number of problematic features and that the RDT approach does not fulfill all features that one would wish .

The RDT-approach seems superior at least in the following respect, which Odd Larsen in a comment within the project (2008) expresses as: "*The most important feature of a RDT-model is the ability to handle properly the combined waiting time.*"

We can also cite John Bates from an E-mail conversation in an earlier research work (2004) where he wrote*: "In line with what I have said above, suppose we have a particular traveller i with preferred arrival time PATi (scheduling preference could also be in terms of departure time), scheduling parameters bSi, generalised cost parameters (relating to money and various time components including interchange) bGi, and access times to appropriate points on the network ai. Note that the generalised cost parameters could vary with mode, and it would also be possible to include "modal constants" (eg a predisposition to prefer tram over bus etc). If we knew all this, then we could use the "route finder" to generate the preferred route, using entirely deterministic principles. This assumes that the timetable is known.*

Of course we will never know all this information for each individual, and we must therefore consider how to take account of the distribution of the relevant variables among individuals. It seems to me that key aspects are: the variation in PAT, the variation in scheduling parameters (eg using a Small/Vickrey type function), the variation in generalised cost parameters, the variation in access conditions, and the assumption that the timetable is known or knowable. Although you also raise the question of fare variation by individual categories, it should be possible to handle this by segmentation."

As far as we understand the RDT approach in Vips/Visum works in accordance with that citation.

Also in the following respects the RDT approach seems advantageous:

- It works in one single step with one consistent set of values of time.
- It takes into account that various modes may be more or less comfortable.

- It does not deal with "main modes", but generates all possible travel paths from origin to destination, taking into account prices of each route, using a number of combinations of services and modes.

However, a discrete choice model approach has one advantage, even if it may be difficult to make use of standard logit due to problems that we have described here, but there may be other more suitable logit or discrete choice models that work better. The advantage of discrete choice models is that they take into account individual randomness.

Even if the RDT approach has several advantages concerning representation of combinations of routes and modes, comfort of modes, prices of routes, calculation of consumer surplus etc., the following features are missing:

individual randomness depending on for example dispersion of starting and destination points within areas, taste, various measurement errors etc.,

forecasting destination choice and trip generation, i.e., changes of the structure of O-D matrices.

Ideally one would therefore want to combine the RDT approach with structuring of O-D matrices and with individual randomness, besides the randomness with respect to differences between ideal and actual departure or arrival times that RDT takes care of. Maybe this can be seen as the main goal of this project.

It is, however, not obvious that a logit model should take care of this individual randomness. Larsen and Sunde (2008) write:

"*Logit models have been proposed and discussed as an alternative assignment principle both for transit systems and more generally for choice between different public transport modes. We will not attempt a review of the different approaches in this paper, but a recent example is Nguyenet al. (1998). In our opinion a satisfactory scheme for use of logit models has so far not been demonstrated…… A major problem by using the logit model is caused by the fact that the main component in the random term of an alternative will be due to the random waiting time even if we allow for heterogeneous transit users. Headways that vary between routes then imply that we will have heteroscedastic error terms in the utility function*"

The aim of the paper by Larsen and Sunde (2008) is to try to combine the RDT approach with individual randomness and they propose one way, based on what they call a "heuristic" type of logit model. Their approach is certainly a good example of making an effort to get forward. In this project both this approach and other approaches will be investigated.

REFERENCES

Bates, J. (2004). Comments to our basic report. Unpublished working paper.

Ben-Akiva, M., and Lerman, S (1985). Discrete Choice Analysis, Theory and Application to Travel Demand. MIT Press, Cambridge, Massachusetts.

Daley, A. (2004). Comments to our basic report. Unpublished working paper.

Hasselström, D. (1981). Public transportation planning – a mathematical programming approach. Ph.D. thesis, Department of Business Administration, University of Gothenburg, Sweden.

Jansson, K., Lang, H. and Mattsson, D. (2008). Optimal Economic Interventions in Scheduled Public Transport. Research in Transportation Economics, Transit Economics. Elsevier.

Jansson, K., Lang, H., Mattsson, D, Mortazavi, R. (2008). Models For Public Transport Demand and Benefit Assessments. In Hensher, D.A. and Button, K (eds), Transport Modelling Handbooks in Transport, Second edition. Vol. 1, Elsevier,

Larsen, O. I. and Sunde, Ø. (2008). Waiting time and the role and value of information in scheduled transport. Research in Transportation Economics, Transit Economics. Elsevier

Larsen, O. I. (2009). Estimating market shares and consumer benefits with random departure time, version 2. Unpublished working paper.

Louviere, J.J., Hensher, D.A. and Swait, J. (2000) Stated Choice Methods and Analysis. Cambridge UniversityPress, Cambridge.