Analysis of currently used models in Sweden for long-distance public transport and possible ways to improve*

Kjell Jansson, Royal Institute of Technology. Stockholm, Sweden

Harald Lang, Royal Institute of Technology. Stockholm, Sweden

Odd I. Larsen, Molde University College, Norway

Reza Mortazavi. Dalarna University, Sweden

ABSTRACT

For assessment of infrastructure measures and for finding appropriate ways to reduce environmental and climate damages etc., forecast models are indispensable tools. The aim of such models is, for assumed transport supply and policy measures, to forecast demand for various modes and calculate consumer surplus and other data that are inputs to a cost-benefit analysis.

An important task for a model is in this context to take into account all lines and modes as correctly as possible.

In Sweden the passenger transport model that is predominantly used for assessment of transport policy measures is comprised of a combination of one network model for travellers' line choice within each "main" mode and a structured logit model for the choice between these modes. An alternative model that is also applied in Sweden is a network model that handles all lines and modes simultaneously. In this paper we present the present status of a project finished in June 2010 aiming to assess the pros and cons of these two modelling approaches, including some tentative judgments and ideas on how to improve passenger transport models. The result is recommendations on how to go forward in order to proceed in modelling of long distance travel.

Note that we exemplify with the concrete Swedish models in actual use, in order to illustrate the general ideas. Since we exemplify with the concrete Swedish case we use the commercial software names where these are part of the models, The principles behind the models and softwares illustrated by these examples should be of wide general interest.

One preliminary conclusion is that for mode and line choice in long distance public transport a network model that handles all lines and modes simultaneously has certain advantages On

---

* This project is financed by The Swedish Governmental Agency for Innovation Systems (Vinnova), The Swedish National Rail Administration (BV), the Swedish Institute for Transport and Communications Analysis (SIKA) and the Stockholm County Public Transport Authority (SL). We are grateful to Henrik Swahn for valuable viewpoints on this paper.

the other hand the pure network model disregards randomness with respect to taste, measurement errors etc. We therefore suggest a combination of a network model and discrete choice approach as one of several important ways forward.

Keywords: long distance public transport, forecast models, infrastructure, cost-benefit analysis, transport policy, network model, logit model.

1 INTRODUCTION AND BACKGROUND

During the last twelve years a model called Sampers has been developed and used as the standard national model system for passenger travel in Sweden. Sampers is a comprehensive planning tool that models total transport demand, demand distribution between destinations as well as distribution between modes and lines. For assignment on lines within mode Sampers uses the Emme/2 system. Consumer surplus is calculated in a separate module based on input data from Sampers. In parallel a network model Vips (or Visum) applying an assignment principle that we denote RDT (Random Departure Times), has been used for certain applications. This model applies simultaneous assignment on both lines and modes in one step, and for calculation of revenues, costs and consumer surplus. Note that we here are dealing with models for long distance travel, i.e. trips exceeding 100 kilometres. This means that travellers do not go to a stop and wait there. Already “at home” they are considering various railways stations, airports, coach stations and choose the stop, mode and line that suit them best.

The aim of this project has been to find out whether there are possible combinations of elements from the two models, and possibly also from other sources, that would give considerable improvements of predictions of travel distribution between modes and lines as well as more precise and valid consumer surplus calculations.

Note that even if the paper focuses on specific models applied in Sweden, it illustrates general features of two distinct approaches that are widely used and that should be of general interest for researches and planners in the passenger transport area.

One important basis for the work is the problems related to Sampers that have been observed over the last ten years; another is that Vips has been used in parallel in several studies, which meant that results from the two models have been subject to analyses and comparisons. The comparisons have often shown remarkable differences in social net benefit results of the same transport measure, which of course has been one strong argument for this project.

We are only investigating the two types of models where they aim to do the same thing, that is calculate demand for lines/modes and consumer surplus of changes in the transport system for a fixed origin-destination matrix containing the total demand over all modes. The issue on destination choice and how the total number of journeys are affected by changes in the transport system is discussed in section 9, dealing with possible development steps.

The terminology is as follows. With line we mean a public transport service from a defined terminal to another defined terminal, each with a specific stop pattern, speed and price. For example: Rail line 60a Stockholm-Gothenburg has specific stop points along the travel path, takes 4 h and costs SEK 450. Rail line 60b Stockholm-Gothenburg has other stops takes 3 h and costs SEK 700. With mode in Vips we mean for example intercity trains, high-speed train, aircraft, coach, local bus etc., where each of these modes has various lines. One reason to distinguish modes from /lines is that it is often useful to assume that travellers think that comfort as well as other properties, e.g., safety, differs between modes. Car is, of course, also a mode with a specific convenience parameter. Sampers uses only “main” modes, which are train, bus, air and car, with no concern for that different modes within each main mode can differ with respect to convenience. With routes (or travel paths) we mean the various combinations of lines and modes that the travellers need to go between origin and destination

An important aspect is that there are features in public transport that differ from private transport. For instance, in making a journey with public transport modes the journey often consists of a combination of modes, e.g. Train-Coach, Regional bus-Train-Flight etc. There is also the issue of specifying the fare systems, since in practice there are many different fare structures. These special features of public transport makes it more complicated to model than car transport.

As a starting point, section 2 includes a discussion on what should be the basic requirements for a long-distance public transport passenger model. In section 3 we give an overview of the features of the two model systems. Section 4 describes the theoretical background of the models. In sections 5 and 6 we present the principles of Sampers and Vips/Visum respectively. Section 7 describes practical experiences of the two models. Section 8 includes preliminary conclusions and section 9 preliminary recommendations on how to go forward.

2 REQUIREMENTS FOR A PUBLIC TRANSPORT MODEL

2.1 Public agencies

The basic requirements should of course come from users, mainly Governmental or Municipal agencies, who are involved in public sector planning processes, design of regulatory policies, design of economic support/subsidies as well as in planning of traffic supply and supply policies. From time to time alternative measures or combinations of measures are considered for implementation – both by the public and the private sector.

The role of the long distance travel demand models is to supply relevant and reliable information about various aspects of travel demand to the users' decision processes. Of course, it is conceivable if not likely that the need for information differs between various users, e.g. between users interested in infrastructure development, public regulation wiz private public transport companies.

Necessary key outputs from a model should be information about travel demand as well as how travel demand is affected by alternative lines of action and how travellers, operators, the public sector finances and the environment are affected.

The main purpose of the model is thus to:

1. Generate a reliable estimate of total travel demand (the flow matrix above) and how this demand will be distributed among modes, lines and travel paths, given the total traffic supply.
2. Estimate consumer surplus of individual or packages of public or private measures relating to infrastructure for long distance traffic supply
3. Compute the net benefit to society as a whole (could be negative) of individual or packages of public or private measures relating to infrastructure for or long distance traffic supply.

It should be possible to:

- separate demand between different journey purposes such as work, business and leisure.

- separate demand between different approximately homogenous groups of travellers, which have different valuations of travel time components for each mode and different prices for lines or modes.
- estimate demand and consumer surplus for each such purpose and group.

2.2 The travellers

The model should be able to take into account the following behavioural assumptions:

- Individual travellers minimize the generalised cost (G) of each journey.
- Since otherwise “identical” travellers may have different ideal departure or arrival times and fares they do not all choose the same path and departure.
- At the origin of a trip travellers have full access to and use time timetable information on all combinations of modes, lines and travel paths in the course of planning and deciding on their journey., but also with respect to how well actual departure or arrival times of lines/modes suit their ideal departure or arrival times.
- Travellers may have different preferences for different modes dependent on comfort etc., which also means that they do not all choose the same path.
- In a perfect model the travellers may have some prior knowledge about probability of delay and/or cancellation and/or fare adjustments.

2.3 Model approaches that satisfy the requirements

No model can satisfy all wanted requirements. Let us start with a sketch of an “ideal model” from the modeller’s point of view and then discuss what features that may have to be relaxed.

2.3.1 Features of an ideal model

- The modeller knows the exact timetables of all lines and modes.
- The modeller knows at what time each traveller wants to travel.
- The modeller knows the valuations of all passengers.

The ideal method for assessment of transport measures in order to calculate changes of demand of generalised cost (G) and of consumer surplus (CS) would then be to segment all travellers for any origin – destination pair into a finite (but possibly huge) number of groups, where each group is (nearly) homogeneous as to their preferences for modes, all sorts of value of time (ride time, waiting time, delay time, deviations from preferred departure (or arrival) times etc.. Within each segment the travellers would thus have nearly the same preferred departure (or arrival) times. All travellers in the same group will chose the same mode and line and have the same G. The basis for both demand and CS would then be to assess the number of travellers in each such group. This procedure is of course infeasible in practice, since the number such groups would be exceedingly large

2.3.2 Simplifications

Instead of assuming that one knows all the valuations of all passengers, the travellers are segmented in maybe 10-15 (nearly) homogenous groups according to value of ride time for each mode, mode specific constant (comfort dependent), value of wait time, value of transfer time, value of transfer penalty (constant cost of the transfer per se), price for each line.

When one does not know all passengers preferred departure or arrival time one can divide each day into time slices, such as 05.00-10.00, 10.00-15.00, 15.00-20.00, 20.00-05.00, each with separate demand levels and uniform distribution of preferred departure or arrival times.

Though the ideal model is based on the assumption that travellers always now the exact time tables – for any present of hypothetical future situation – it may be both cumbersome and difficult to define the exact time tables and fares, especially for future or alternative scenarios. When the modeller and/or the user of the model do not define exact timetables, simpler modelling alternatives have to be defined that are still compatible with the behavioural assumption that travellers have full information about the timetables., Defining headways per line is one possible simplification.

In section 9 we discuss what may be plausible development steps, even if one cannot get close to the “ideal model”.

3 OVERVIEW OF FEATURES OF THE SWEDISH MODELS

None of the Swedish models examined in this paper fulfil the requirements of the “ideal model”.

- The modeller and or user do not define the exact timetables of all lines and modes, only the headways and travel times between the stops of each line.
- The modeller only knows demand per day.
- The modeller does not know the valuations of all passengers.

The Sampers model uses three steps: i) the network model Emme/2 for assignment on lines within each mode and for estimation of travel time components for each mode, ii) a multinomial logit model for demand projections concerning modes, destinations, and travel frequency, iii) Samkalk for calculation of consumer surplus, revenues, costs etc. for cost-benefit analysis. The models Vips and Visum[1] uses one step for simultaneous assignment on all combinations on lines and modes and for consumer surplus computation. Below we summarize characteristics of the models.

Network representation

Sampers	**Vips**
Lines/lines for each mode	Lines for each mode
Only “main modes” between each O-D pair; combinations of modes not included	Combinations of all lines/lines and modes between all O-D pairs
The average ride time for all lines of a main mode in each O-D pair	Individual ride times for all lines of a mode in each O-D pair
The average price for all lines of a main mode in each O-D pair	

---

[1] Vips and Visum are partly similar models, at least with respect to the RDT principle. Vips is no longer maintained meaning that in practice only Visum is available.

Parameters

Sampers (Emme/2 for assignment within main modes)	**Vips**
Mode specific constants and parameters for different time components. Same weight for wait time and transfer time in Emme/2 assignments In Sampers headway has a stepwise weight. which affects mode and destination choice Separate relative values of different time components in Emme/2 assignments and in Sampers logit models => inconsistency Headway has a stepwise weight, which affects mode and destination choice	Mode specific constants. Separate weights for wait time and transfer time. Separate weights for wait "at home" (plus a margin) and at stop. These affect both assignment and level of service. Convenience parameters per mode (specific ride time weights) and per stop (specific wait time weights) Parameters for wait time and ride time delay at each stop on each line. Affects both assignment and level of service Consistent values of time since there is simultaneous calculation step for path and mode.

Fares

Sampers	**Vips**
One average fare for each main mode in each O-D pair, for private and business journeys respectively. These are coded exogenously as a price matrix for each main mode.	Fares for each line for all modes for 12 travellers segments. For each segment, many different fare structures for rail and bus and hundreds for airlines. Prices as a) concave or convex kilometre fares plus a base fare, b) zone fares plus a base fare, c) stop-stop fares for each line. Free transfers can be coded. The traveller's price for a journey between A and B is endogenously calculated as the weighted average for all combinations of lines and modes.

Algorithms

Emme/2	**Vips**
Assumes no knowledge of timetables, only of travel times and headways. Passengers choose the stop (for the mode given by the choice model) with minimum generalized cost. At this stop assignment on (acceptable) lines in proportion to frequency, with no concern for travel times and price. Wait time is calculated as half the combined headways, which typically means underestimation of wait time. **Sampers** Takes from Emme/2 average travel times for each mode plus the exogenous price matrix and assigns on modes by use of the logit model. Calculates the logsum for each scenario of network, which is used for travel matrix generation, but not for consumer surplus calculation. Sampers takes implicit into account variation w r t taste and measurement errors but not w r t RDT.	Assumes knowledge of timetables, travel times and headways. Passengers are assigned to all acceptable stops and modes. Concern for all headways, travel times and prices of all lines and modes. Assumes random ideal departure or arrival times and random gaps between departures of lines. These two features can be called RDT (for Random Departure times). One may, however, specify that certain lines are coordinated, i.e., that they have equal gaps between departures on certain sections of the lines. Shares for lines and modes and wait times are thus calculated by use of integration over all alternatives (lines and modes) with their respective travel times, prices and headways. Vips takes into account variation w r t RDT but not w r t taste and measurement errors.

Consumer surplus (CS)

Samkalk	**Vips**
The module Samkalk calculates CS by regarding the lines of the mode that is subject to change, Samkalk then distinguishes between existing/remaining travellers and new/lost travellers, where the latter gain/lose half of the existing/remaining. However, when headway is changed this principle is wrong. Wait time is an endogenous variable, dependent on headway of all lines or modes.	CS is calculated with concern for all lines and modes simultaneously. One can thus calculate CS for changes of one or several lines and modes at the same time. A limitation compared to Sampers/Samkalk is that the OD-matrix for the total number of trips is unaffected.

4 THEORETICAL BACKGROUND

4.1 Introduction

In fact, both the logit model in Sampers and Vips/Visum have to some extent the same theoretical background in the sense that each individual is assumed to choose the alternative with minimum generalized cost (G). This G is, however, not the same for each individual due to the variation of individual preferences – stochastic influence, but this stochastic influence is totally different in the two models.

In Sampers there is a variation among individuals due to taste, measurement errors, omitted variables etc. In Vips/Visum there is a uniform variation among individuals due to ideal departure or arrival time in relation to actual departure or arrival time. It is also assumed that departure times of alternative lines are uniformly distributed. Both the departure times of travellers and lines are thus assumed randomly (uniformly) distributed, which has given rise to the term RDT (Random Departure Times).

In order to deal with 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.

In order to simplify notation and calculations, without affecting the general aspects, we assume that there are two alternatives, 1 and 2.

4.2 Basic micro-economic model

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 from the ideal departure time, i.e., the difference between actual and ideal departure time, which can also be called schedule delay. The generalised cost of alternative j for individual i is then:

$$(1) \qquad G_i^j = R^j + t_i^j + x_i^j$$

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

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

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

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

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 generalised cost G_i of individual i is then defined by:

$$G_i = G + \varepsilon_i \tag{4}$$

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:

$$v_i - G_i = v_i - \varepsilon_i - G \equiv u_i - G \tag{5}$$

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.

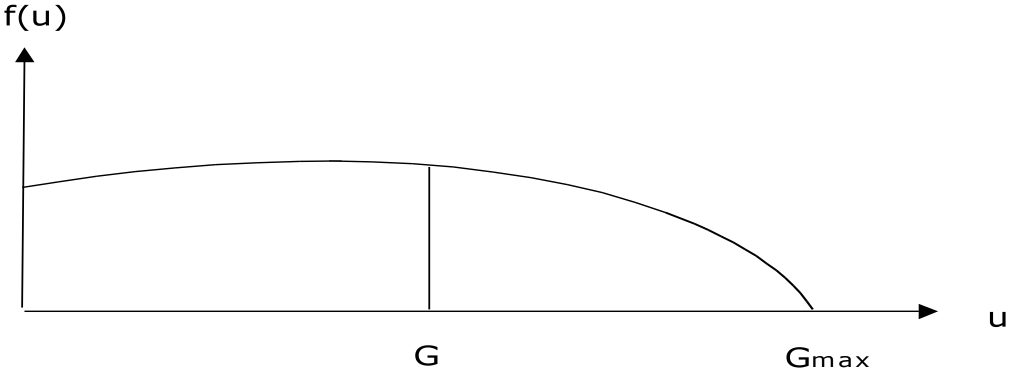

Figure 1: Distribution of utility

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

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

The consumer surplus, S, is thus:

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

It then follows that:

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

5 PRINCIPLES OF THE SAMPERS MODEL

5.1 Step 1: Emme/2 in Sampers

The Emme/2 model is based on average frequencies of services. The passengers are assumed to know the travel time components and headway of all lines, but not the timetable (the actual departure times) or behave as if they do not. The basic behavioural assumption is that passengers choose a strategy that minimizes the weighted travel time components, i.e., generalised cost except price, which is not specified. The price variable is specified exogenously as an O-D matrix in the multinomial logit model described later; this way of handling the price variable has the effect that price differentials between different services within one mode cannot be considered.

From the assumption that passenger do not know the time tables, only frequencies, follows: i) that travellers go to the stop with the shortest expected total travel time, which typically has only one specific mode, ii) that passengers are assigned to lines in proportion to frequency only, without concern for ride time and price. This decision rule is correct given the assumption that departure times are not known. When timetables are known, which is our basic prerequisite, the Emme/2 assumption means a problem in general but especially if fast modes such as skip-stop-services or high-speed trains (maybe also with higher prices) are to be assessed.

5.2 Step 2: The logit model in Sampers

The logit model in Sampers assumes stochastic variation with respect to preferences etc., t, 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 the joint generalised cost in (3) is (see for example Ben-Akiva and Lerman (1985) and Louviere et al., (2000)):

$$(9) \qquad G = E\left[\min\left[G^1 + t_i^1, G^2 + t_i^2\right]\right]$$

Without going into all the details in the Sampers system, the – for our purpose – important issues can be stated as follows:

For each defined main mode the Emme/2 assignment gives level of services (LOS) variables that are weighted averages for the alternatives that the network model picks for each OD-pair. Subsequently parameters are estimated for the different LOS-variables that enter the utility functions of each main mode, Each estimated utility function, less one, will then also get an alternative specific constant. The utility functions also include other variables that do not concern us here.

This approach will in general be acceptable if we deal with high frequency services as in most large urban areas. Then travellers may base their decision of mode choice on expected values and forego the use of time tables. Thus the LOS- variables in the utility functions is in this case also represents the magnitudes of what affects the decisions of travellers. When dealing with long distance travel and people who use timetables and in addition are concerned with preferred arrival or departure time they do not base their decisions on mean values for each mode, say for bus, train and airplane, but consider the arrival and departure time as well as cost and onboard time for the options available under each mode more or less simultaneously.

What are the consequences of such behaviour in relation to logit models where the mean values of different variables enter as arguments in the utility function for the different modes, directly or with some non-linear transformation?

In the estimation phase we certainly get a problem with the IID (independent and identical distributed) assumption for random terms.

For each observation we use in the estimation, the implicit random term will absorb the deviation between the mean values of the LOS-variables used for a mode and the actual values experienced by the traveller. Especially for the headway variable in scheduled transport modes we must also expect that the distribution of deviations from the mean depend on the headways and the number of alternatives available for travelling with a main mode between an origin and a destination.

But the situation is even worse, for this particular variable the deviations from the mean for one mode will in general also depend on the deviations from the mean for the other modes. Thus the implicit random terms of the model that we estimate will have a special component in the error term that is heteroskedastic and cross-correlated over modes. Or to state it in another way, we apply "standard assumptions" to estimate a model with random measurement errors where at least some measurement errors in addition are heteroskedastic and correlated across alternatives. This may have quite severe and unknown implications for estimated parameters (Bhatta and Larsen, 2008). In any estimation process, different specifications are usually tested. Some specification of the model might correct somewhat for the initial misspecification of the model and become the accepted specification, but out of sample predictions and demand predictions related to changes in the transport system may still be far off the mark.

These problems will be present whichever assignment model is used to produce LOS-variables for the estimation of a logit model as long as the travellers have preference for arrival and/or departure time that reaches across modes. However, any problems in estimation and later application of the model may be amplified by the use of an assignment model that is not suited for the task.

The seriousness of the problems pointed out here will depend on several factors. If a "standard" (IID) error term is the dominating component in the implicit composite error terms of the utility functions the impact of the other error components may be less severe.

In an E-mail dialogue in an earlier research work Andrew Daly (2004) wrote:

> "One problem is that a model of choice among lines 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."

Larsen and Sunde, Ø. (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 Nguyen et 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 lines then imply that we will have heteroscedastic error terms in the utility function."

However, in the article they propose a heuristic logit-formulation that seems to overcome some of the problems inherent in the use of logit models for choice among alternatives where one or more of the alternatives are scheduled services.

Ben-Akiva and Lerman (1987) write (pages 108-109) that the core of the problem with the IIA property (Independence of Irrelevant Alternatives) is:

> "the assumption that the disturbances are mutually independent.", and that "In some instances it can give rise to somewhat odd and erroneous predictions."

Greene (2008, pages 852-854) describes a study on a logit model for travel choice between Sydney and Melbourne based on 210 choice observations, with four modes available, air, train, bus and car. The IIA property was rejected.

Then a nested logit model was examined. The first choice level was FLY and GROUND. The next level was AIR under FLY, and TRAIN, BUS and CAR under GROUND. The null hypothesis of homoscedasticity was rejected below the level 0.5 per cent.

One may wonder whether mixed logit or HEV (Heteroskedastic extreme value) could solve some of the problems. Even if IIA is relaxed in terms of heteroskedastic random parameters, the model would probably still suffer from an independence problem. This problem occurs if, as we assume, that passengers have different ideal departure or arrival times in relation to actual departure or arrival times. A discrete choice model ignores that alternative lines or modes are dependent via their headways and the subsequent implication for choice. One can also see it from the operators´ point of view. The higher frequency they choose, the larger is the possibility that their departures will suit the passengers compared to the competitors´ departures.

Appendix 1 includes a somewhat more detailed discussion on logit models.

Note finally that the set of values of time used is not consistent in the Sampers model since the set used in the Emme/2 step is not the same as the set of values that are estimated by the logit model and subsequently used when the model is run.

5.3 Step 3: Samkalk

The Samkalk module uses “rule of one half” (ROH) for calculating changes in consumers’ surplus. ROH is an approximation of the logsum originating from the logit model.

This way of calculating consumer surplus (CS) is wrong (overestimates CS) if headways are changed. See a mathematical proof in Jansson et al. (2008) and a brief discussion in appendix 3. In fact, when headways are changed there is nothing like existing or new travellers, since changes of headways make people move around between departures. The wait time in total and the wait time per line/mode are endogenous and depend on the headway of all lines/modes. This problem is mitigated in practical applications of Sampers since a stepwise function is used for the weight of wait time.

That Samkalk takes into account “main modes” only means that measures cannot be evaluated where combinations of modes are needed to reach the destination. This is illustrated by the following example:

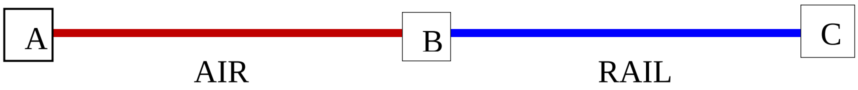

Figure 2: Travelling between A and C

Between A and B there are airlines. Between B and C there are rail lines. Assume that we want to evaluate the effects of introducing high-speed rail between B and C for passengers going from A to C via B. The change in consumer surplus according to Vips is quite large but zero according to Samkalk. The reason is that Sampers does not take into account the possibility to go by air between A and B and then by rail between B and C. Since air is not taken into account when rail is subject to change, Samkalk cannot calculate the effect.

There is also another important aspect related to the Sampers main mode concept, their relations with access/egress links. Vips/Visum regards commuter train lines, regional bus lines and coach lines that feed to/from train stations, airports and coach line stops. In addition Vips/Visum can have access/egress links representing travel by car or taxi to the same railway stations, airports or coach line stops. So Vips/Visum gives the travellers a choice between for example a regional bus and car/taxi for the ride to the station or the airport.

To/from the main modes Sampers instead only uses access/egress links, each with a specified time, i.e., no public transport lines. The model assumes very low speed for these egress/access links in order to prevent too extensive use of access/egress links. If for example one can go by train to a city but there are two possible alighting stations Sampers only accepts the egress link with the shortest time, meaning that only one alighting stop is selected..

Assume that we want to evaluate the benefit to travellers from Stockholm to Örebro of reducing the ride time of rail line R1 (blue) Stockholm-Örebro from 2h 15 minutes to 1h 30 minutes while the speed on a more or less parallel line R2 (black) Stockholm-Oslo via Degerfors is unchanged. See the figure below.

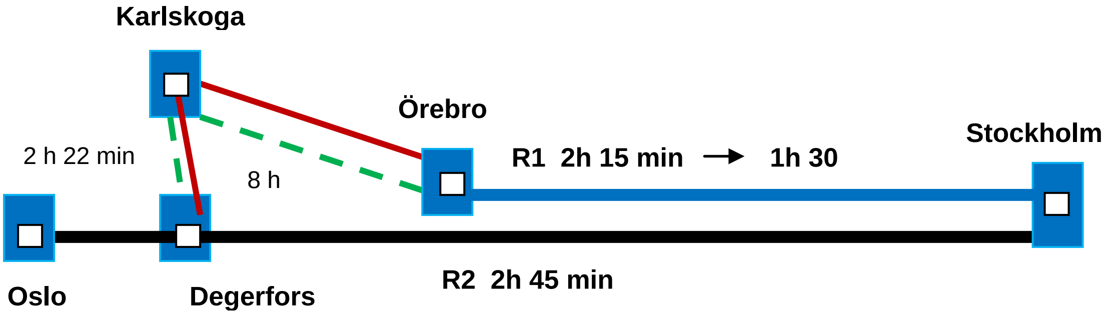

Figure 3: Travelling between Stockholm and Karlskoga

The red lines indicate regional buses according to Vips and the green dotted lines egress links according to Sampers. The Sampers egress link Örebro –Karlskoga takes 8 hours and the Sampers egress Degerfors-Karlskoga takes 2h 22 minutes according to the actual Sampers coding of the network. Since Sampers does not take into account the real-world regional buses, only the egress links with excessively long times, nobody will choose to alight in Örebro and no travel time gain is calculated. Vips/Visum on the other hand takes the regional buses into account and calculates time gains for the travellers Stockholm-Karlskoga.

6 PRINCIPLES OF VIPS/VISUM AND THE RDT APPROACH

6.1 Introduction

The RDT principle in Vips and Visum[2] assumes that travellers know the timetables. The models Vips and Visum work in one simultaneous step for lines and modes, where all feasible combinations of lines and modes are taken into account, and where each line has a specific ride time and price (kilometre, zone, or stop-stop based). Each passenger segment can then have a specific price structure for each mode and service within mode. The price in

---

[2] We refer to the model Visum for the RDT principle, since we know that this is implemented in the Visum software. There may, however, be other softwares that include this principle.

each O-D pair for a specific group is then calculated endogenously by the program as the weighted average of the prices of all combinations of lines and modes.

The passengers in each origin zone are assigned to various stops with various services and modes.

RDT assumes no stochastic variation with respect to preferences and measurement errors. 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.

Since the model takes into account 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, tens or hundreds.

6. 2 Principles

When departure times of all lines are known all lines and stops are considered simultaneously, but all cannot be acceptable. Assume that different lines ^{j} have travel times R^j and headway H^j. Expected wait time is then not $H^j/2$. Expected wait time when the timetable is known is the difference between ideal and actual departure time. The basis for choice of acceptable lines is the time to reach the boarding point plus travel time after boarding, here denoted R^j, plus all of the headway, H^j. Assume that line 1 is best, has the lowest value R^1+H^1. Other lines m are acceptable if $R^m < R^1+H^1$. This means that it is not worthwhile to choose a line that has travel time only that is longer than travel time plus the whole headway of the best line.

The RDT approach is based on two assumptions. Firstly it ignores the stochastic element t (see (3) in section 4.2) that varies with taste differences among individuals, measurement errors etc. It uses 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 thus assumed that (x^1, x^2) has a uniform distribution on $[0, H^1]\times[0, H^2]$, exemplified by two alternatives. This assumption is thus based on the mentioned assumption that the modeller does not now anything about the true distribution about ideal departure times for the period of time (a whole day, peak hours or non-peak hours for example) we are analysing. Secondly it is also assumed that departure times of alternative lines/modes are uniformly distributed.

Notation

H^1 headway of line 1.

H^2 headway of line 2.

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

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

x^1 time to departure of line 1.

x^2 time to departure of line 2.

Expression (3) in section 4.2, reflecting the joint generalised cost, is then:

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

It has been shown, see Jansson, Lang and 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:

$$(12) \qquad h\left[s\right] = 1 \quad if \quad s > 0; 0 \quad if \quad s \leq 0$$

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

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:

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

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

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

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

More details on the principles are found in Jansson et al. (2008). Some details of the implemented Vips model are found in Jansson and Ridderstolpe (1992).

7 EXPERIENCES OF THE TWO MODELS

7.1 Uses in practice

For assessment of public transport the Sampers model has been employed by the Swedish National Railway Administration (BV) and the Swedish Institute for Transport and Communications Analysis (SIKA) about a decade, especially for the Governmental requirements to produce “action plans” for the transport sector. The model has also been used in a number of studies on various specific railway projects.

The Vips model was used at the Swedish Institute for Transport and Communications Analysis (SIKA) for assessment of a Governmental inquiry on whether deregulation of coach services is socially beneficial, also taking into account whether the railway would “suffer” too much. The assessment led to a recommendation of deregulation and that also became the Governmental decision in 1998.

Vips and Sampers have also been used in parallel on behalf of BV, not least for various assessments of high-speed rail in Sweden.

In 2008 a hybrid model was used for a Governmental inquiry on high-speed rail in Sweden. The hybrid meant use of Sampers for creation of the travel matrix, Vips for estimation of demand on lines and modes and for calculation of all travel time components and price in all origin-destination (O-D) pairs. Finally Samkalk was used for calculation of consumer surplus. In a parallel study, however, Vips was also used for calculation of consumer surplus.

The interesting but somewhat “troublesome” outcome was the substantial differences in cost-benefit outcome from the hybrid method using Samkalk and the Vips method based on the same travel matrix. Using common project cost estimates, the net present value ratio ((Benefits-costs)/investment costs) was 0.15 according to the hybrid and 0.78 according to Vips. ‘

This remarkable difference partly depends on the fact that Samkalk cannot calculate gains where travellers have to use also other modes than train to go from origin to destination; combinations of modes are disregarded due to the “main mode” concept.

7.2 Sampers according to users

Since Sampers is the standard model tool in Swedish transport planning it has been used extensively by all transport agencies since it was first introduced around the year 2000 Over the years there has been an ongoing amendment process often triggered by the users' experience and suggestions. Thus there have been many amendments to the Sampers model since it was first introduced. These amendments have mostly concerned other things than the demand model. The demand model has been updated only once, and then constrained with respect to model specification (Transek 2004). In this context, also some specification tests were carried out.

In 2009 a project on Sampers experience and long and short term development possibilities was financed by the Swedish Road and Rail Administrations (Algers et al 2009) This work also included interviews with users. Among critical viewpoints were that the model cannot realistically enough describe effects of relevant transport measures, , that it shows unreasonable small shifts between modes, air to train, car to train when high-speed trains are assessed. It also appeared that some users said that they have little experience of running the model, which may be one reason for the critical viewpoints. Some users, according to the report, argued that it may have to be further developed or even replaced by a new system

One recurring criticism over the years has addressed the currently applied headway based version Emme/2 which operates with the assumption that travellers do not know the departure times for long-distance journeys, that they are assigned in proportion to frequency only (for the set of acceptable lines). Some criticism, although not from the users, has also addressed the problem that the model uses fares per OD rather than considering that, various connections may have different prices. However, the version of Emme/2 currently being used remains the same.

8 PRELIMINARY CONCLUSIONS

- The Emme/2 step in Sampers assumes no knowledge of timetable, which is not realistic for long-distance transport and which makes it impossible to take into account that lines within the same mode have different ride times and prices. The reason is that without knowledge of timetable it is from the travellers´ point of view optimal to choose the stop with the lowest expected generalised cost, with the consequence that the passenger chooses line in proportion to frequency only.

- The logsum concept in Sampers cannot be used as a representation of combined generalised cost of a number of modes if concern for schedule delays is an important factor.

Let us now recall our mentioned behavioural requirement that passengers use timetables, meaning that they “already at home” compare various combinations of lines and modes not only with respect to price, ride time, transfer time etc. but also with respect to how well departure or arrival times suit their ideal departure or arrival times.

The Sampers model in fact violates this requirement in each of the steps Emme/2, logit model, Samkalk.

The RDT principle in Vips/Visum accords with this requirement. Odd Larsen in a comment within this project (2009) writes: “The most important feature of a RDT-model is the ability to handle properly the combined waiting time.”

However, The Vips and Visum models do not take into account variation with respect to preferences for modes and for measurement errors by use of a random term. This will affect - at least mode choice and evaluation of changes and implies that also this approach may give more or less erroneous results.

The fact that different passengers have different valuations of modes can be handled in four ways:

1. Mode specific constants,
2. Different weights of ride time for different modes,
3. Segmentation of passengers w r t different valuations of travel time components and prices,
4. A random term for taste, measurement errors etc.

The ways 1, 2 and 3 are handled in Vips/Visum while 1 and 4 are handled by Sampers. In practical use of Vips for the Swedish national network, passengers are often segmented in 10-12 groups, each with assumed specific homogenous valuations of time for different modes, price for different lines, weights for travel time components car availability etc.

Vips/Visum can thus take into account certain distributions of valuations of modes by mode specific constants and segmentation. But can this be considered sufficient? Of course these parameters have to be estimated in some way, which is not in focus in this paper. This is a well developed field for discrete choice models and an RDT approach would need a sound scientific method that can do the same thing.

In addition to segmentation, one should add to the RDT approach also variation with respect to taste, measurement errors, omitted variables etc., given of course that RDT shall simultaneously handle both lines and modes. This would imply a combination of the RDT and the discrete choice paradigm. How this can be done is left for a future project.

9 PRELIMINARY RECOMMENDATIONS WITH RESPECT TO FURTHER RESEARCH

9.1 New user requirements call for development of existing model tools for long distance travel

The aim of the project is to give recommendations as to the way ahead for the further development of long distance travel demand models in Sweden. These recommendations should be based on general considerations of user requirements of such models, the findings in the present project as well as on recent research and development within the field.

Of course there is a need of continuous and gradual development of all model tools that are used in transport planning. However, as for demand modelling for long distance travel, there are particular reasons to consider more substantial reviews of existing model tools, i.e. the current standard Sampers implementation, Vips, and Visum. These reasons are to a certain extent related to some weaknesses of the long distance Sampers model that have been observed by the users and were briefly touched upon in section 7.2 above.

In addition to this and possibly also related to the criticism that has been voiced, rather fundamental changes to long distance transport supply are now envisaged. High speed train lines are seriously considered, regular train traffic continues to attract considerable political interest, the role of air transport may have to be reassessed, and car transport, which is still the dominating single mode also for long distance travel is put under pressure for various reasons. The national public transport agencies as well as regional and local authorities will have to be well informed to be able to respond in a rational way to the challenges that will come up in the course of this development. Against this background there is an urgent need to reassess and improve also available model tools for long distance demand modelling.

The planning system will require all model tools to be more flexible as well as versatile, since there will be less time and resources available for large scale up-dating of comprehensive set of input data. It will be even more important than before to put emphasis on journeys, movements, and interconnections rather than the transport modes themselves.

9.2 Development should be planned in different time perspectives

Typically, short run development will use existing algorithms and largely also rely on existing data and the basic effort will be to do the programming, testing etc. that is required to implement such improvements.

In the medium run it is realistic to envisage development of new algorithms, programs that could be based on new but limited sets of data. Combining elements, e.g. algorithms, from different model implementations and introducing new or amended model features within existing model implementations e.g. ,Sampers, Visum, Vips.

In the long run it is realistic to consider more sweeping development strategies such as developing totally new model systems with new algorithms, newly estimated behavioural models based on new, possibly large, data sets. It would also be possible in the long run to re-consider the overall approach to manage and coordinate models, data, and scenarios. We will here discuss short and medium term developments only.

9.3 Suggestions for short and medium term developments

We consider it wise in the short to medium term to pursue several parallel development lines.

Development of logit models

The project has cast some doubt about the suitability of the specification of the MNL that is implemented at the mode choice level in the standard Sampers model. There are many alternative logit model specifications and model structures that may lead to a better model specification than the current one as has been demonstrated in research projects related to the development of long distance logit models (see for example Beser Hugosson 2003). This development may have to draw on supporting studies of travellers' decision structure as well as a thorough assessment of the available evidence on how various combinations of modes and sub-modes have been and possibly will be used in the future

Development relating to the RDT-algorithm

Though the RDT algorithm operates with the assumption that travellers know the timetable it has some weaknesses, namely that it does not take passenger heterogeneity into account and that it assumes uniform distributions of travellers desired departure time viz. departure time of services. One way to improve and expand the properties of the RDT-algorithm in the context of modelling long distance travel that we have identified and taken first steps to develop in this project is to amend it with a mechanism that allows various aspects of traveller heterogeneity directly to influence the algorithm. The proposal is therefore to promote development of theory, algorithms, and practice for a random term relating to "mode" that could be combined with the RDT-algorithm. This random term methodology could be used to improve the present RDT-approach with some influence from various random factors. One preliminary idea is presented in appendix 2.

Another potential improvement of the RDT-algorithm is related to the basic assumptions about uniform distributions of traveller desired departure time and the uniform distribution of supply of public transport departures. The question that has to be studied is whether it would be possible to use alternative assumptions about these distributions based on empirical evidence and if and how such distributions could be implemented in the algorithm. A special case is to do this only for the desired departure time for travellers. Development efforts along these lines should be supported by empirical studies giving further evidence about travellers´ actual departure times, maybe in time slices over the day.

As of today there is no systematic and scientifically sound way to calibrate the parameters of the Vips model with the RDT-algorithm. We have pointed out that there may be difficult issues to deal with here. We suggest that it is investigated whether one can find a sound scientific method to estimate parameters with the RDT approach, as is done for discrete choice models. However, both for discrete choice models and models using RDT there is a problem if the "true" model is something in between. Whether we estimate or calibrate we will have observations that are generated by a different mechanism than assumed by either model and the estimation/calibration will produce biased parameters

Timetable based options

In the medium term we think that timetable-based options should be investigated carefully. However, there are two options at hand here that could be addressed in a pre-study already in the short term namely the Sampers timetable model and the Visum timetable based assignment. With a timetable-based model one can abolish one of the two assumptions behind the RDT principle, i.e., uniformly distributed departure times of lines.

The development of the Sampers departure time model is based on supporting studies of the asymmetry between early and late arrivals. As is expected, late arrivals are valued more negatively than early arrivals. These results are available if one would decide to go ahead with the timetable based approaches. Visums timetable-based version also takes into consideration this asymmetry between early and late arrivals.

In Germany already the timetable-based version is dominating. In the near future it may be that also for the whole of Sweden all timetables for all lines and modes can be exported to Visum or some other timetable-based software.

Calculation of consumer surplus change

Though an important role of long distance travel demand models is to provide projections/estimates about various aspects of demand such as the distribution of long distance travel over modes and lines, the models have an important additional role which is to provide the basis for reliable estimates of change of consumer surplus. The latter is a primary input to the social cost benefit analyses that have to be prepared for all major policy measures relating to infrastructure, traffic supply, and other transport policy measures.

In the project we have argued, based on theoretical considerations as well as practical model runs that it is necessary that calculation of consumer surplus change must be fully synchronised with the demand model. Failure to do this may result in erroneous estimates of consumer surplus change. We have identified this type of synchronisation problems for the two model systems we have studied in detail in this project, the standard Sampers implementation and the Vips implementation. Based on these findings it is clear that the short/medium term development needs to address also the issue of calculation of consumer surplus change. See a discussion on consumer surplus calculations in appendix 3.

REFERENCES

Algers, S., Mattsson, L-G., Rydergren, C. and Östlund, B. (2009) Sampers – erfarenheter och utvecklingsmöjligheter på kort och lång sikt. (Sampers – experiences and development possibilities in short and long run).

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

Beser Hugosson, M. (2003) Issues in estimation and Application of Long distance Travel Demand Models. Doctoral dissertation, ISBN 91-7323-0044-8, Royal Inst. Of Technology, Stockholm Sweden 2003.

Bhat, C. (1995), A Heteroskedastic Extreme Value Model of Intercity Travel Mode Choice, Transportation Research, 29.

Bhatta B.P and O I Larsen (2007). Measurement errors in discrete choice modeling: a simulation study based on a multinomial logit model for travel to work. Paper presented at WCTR, San Fransisco.

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

De Jong, G., Pieters, M, Daly, A. Graafland, I., Kroes, E. and Koopmans, C. (2005) Using the Logsum as an Evaluation Measure: Literature and Case Study. RAND Working Paper.

De la Barra, T. (1989). Integrated land use and transport modeling. Decision chains and hierarchies. Cambridge Urban and Architectural Studies.

Greene, W (2008). Econometric Analysis, Prentice Hall, 6th Edition.

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

Jansson, K, Ridderstolpe, B. (1982). A Method for the Route-Choice Problem in Public Transport Systems" Transportation Science, vol. 26, no. 3.

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.

Nguyen, S., Pallottino, S., & Gendreau, M. (1998). Implicit enumeration of hyperpaths in a logit model for transit networks. Transportation Science, 32(1), 54–64.

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

Train, K. (2003), Discrete Choice Methods with Simulation, Cambridge University Press, Cambridge, MA.

Small, K.A. and Rosen, H.S. (1981) Applied welfare economics with discrete choice models. Econometrica, 49.

Small, K.A. (1992) Urban Transportation Economics. Harwood Academic Publishers.

Transek (2004) Utveckling av Sampers 2.1. (Development of Sampers 2.1).

Widlert, S. (2003) Ger Sampers rimliga resultat? (Does Sampers provide reasonable results?)

Zeng, L. (2000), A Heteroskedastic Generalized Extreme Value Discrete Choice Model. Sociological Methods & Research.

APPENDIX 1 THE LOGIT MODEL AND LOGSUM AS A MEASURE OF CONSUMER SURPLUS WHEN THE STANDARD ASSUMPTIONS ARE NOT VALID:

Discrete choice situations are often modelled based on the random utility framework in which it is basically assumed that the consumers compare different alternatives based on the characteristics of the alternatives and their own preferences and choose the alternative that gives maximum amount of utility. Since all the relevant information is not observed by the analyst and also due to the existence of randomness the choice models used to predict consumer behaviour are probabilistic in nature. In a travel to work context, for instance, the following specification based on the random utility framework is often used: $U_{in} = V_{in} + \varepsilon_{in}$. Here *i* is the index for the alternative and *n* is the index for the individual traveller. V_{in} is the systematic part of the utility function and ε_{in} is the random unobserved part. The individual traveller is then assumed to choose among different alternatives so as to maximize his utility. Assuming different distributions for ε will lead to different models. For instance, when ε is assumed to be IID (Independently and Identically Distributed) Gumbel we get the multinomial logit model (MNL). It is well known, however, that the IID assumption regarding the distribution of the error terms of the multinomial logit model is restrictive. In the utility function above, the systematic part, V_{in}, may be written as a linear function of the attributes of the alternative, $V_{in} = x'\beta$, where, x_{in} is a vector of observable characteristics of the alternatives and β is a set of parameters, which, in reality may vary over individuals. But, if these vary among individuals then it would mean that we cannot maintain the constant error variance assumption. The variance would vary over alternatives and individuals. This type of pattern may be even more apparent in more complex situations when the alternatives themselves are not independent.

A1.1 Non-constant error variance and the logit model

Suppose we have a situation where the schedule delay (difference between the desired and actual departure time) is unobserved. In the above random utility framework then this would make a part of the random error, ε_{in}, or the non-systematic part of the utility function. One can then imagine that in this case the variance of this error term may be different from individual to individual. For instance those with a flexible starting working time may exhibit larger variance than those with a fixed starting working time. To simplify, assume that we can divide people into two distinct categories (1 and 2) in our example. Then we have: $U_{in} = x'_{in}\beta + \varepsilon_{in}^{k}$ where $k = 1, 2$ and $Variance(\varepsilon_{in}^{1}) \neq Variance(\varepsilon_{in}^{2})$. Assume for instance that $Variance(\varepsilon_{in}^{1}) > Variance(\varepsilon_{in}^{2})$. This would mean that the observed factors explain the choice better for group 2 than group 1 relative to the unobserved factors. Not accounting for possible heteroskedasticity may have severe implications. Zeng (2000, p.120) referring to Horowitz (1981) writes: "the damage of the specification error is often serious enough to destroy the practical value of the model". In general there are two basic approaches to deal with heteroskedastic errors in logit based discrete choice models. One is to apply the basic philosophy of the GLS (Generalized Least Squares) in ordinary regression, i.e., to reweigh the observations so that one gets a homoskedastic error term. See Bhat (1995) for dealing with heteroskedasticity across alternatives and Zeng (2000) for a more general approach that also includes heteroskedasticity across individuals. A problem is to find the correct specification in explicitly modelling the heterskedasticity by using the observed characteristics. Another approach is to use more sophisticated models, like the Mixed Logit Model, which permits different parameter values for different individuals by assuming a distribution of these over the individuals (Train, 2003). However, even here one needs to make distributional assumptions that may or may not represent reality well. Another type of problem that remains is the fact that in many practical situations the set of alternatives may be very large, for instance in a public transportation choice context.

Note also that the hetroskedasticity issues discussed above mainly concerned non-constant error variance due to individual differences. In cases when observed factors are correlated with the error term the problem is even more severe. Larsen and Sunde (2008), for instance, discuss the problem when headway or estimated waiting time are correlated with unobserved factors implying that the error variance varies with them. They suggest a heuristic logit model that to some extent gets around the problem.

A1.2 Logsum as a measure of consumer surplus

A by-product of the standard logit model, often used as a measure of welfare or consumer plus, is the so called logsum. Suppose that there are two alternatives, 1 and 2 and let us ignore the individual index *n*. Utility functions can be written as: $U_i = V_i + \varepsilon_i$ where $i = 1, 2$. Again V_i is the representative utility (function of observables) associated to alternative *i* and ε_i = random term (unobservable). Consumer surplus (CS) is defined to be the maximum of U_1 and U_2. Because of the random error term we have to take the expectation over all possible values for the random error term. So $E(CS) = E\big(Max(U_1 = V_1 + \varepsilon_1, U_2 = V_2 + \varepsilon_2)\big)$. This measure is in the units of utile so to have it in money terms it must be divided by marginal utility of income, say, λ,which is often assumed to be a constant. For the case of the logit model the probability associated to the alternatives are:$P_1 = \frac{e^{V_1}}{e^{V_1}+e^{V_2}}$ and $P_2 = \frac{e^{V_2}}{e^{V_1}+e^{V_2}}$ respectively. Then, the expected consumer surplus can be written as: $E(CS) = \frac{1}{\lambda}\ln(e^{V_1} + e^{V_2}) + C$. Where, the term $\ln(e^{V_1} + e^{V_2})$ is the log of the denominator in the probability expressions. Also since preference ordering is not changed by addition and/or multiplication of constants to the utility function the unidentifiable term C is added. This then would give a measure of welfare for the representative consumer. For policy analysis we are often interested to calculate the change in consumer surplus brought about by a change in the attributes of an alternative. Suppose for instance that we have had a change in the representative utility for alternative 1. So we have a before (b) and after (a) situation. Then the change in expected consumer surplus is given by: $\Delta E(CS) = \frac{1}{\lambda}\ln[\frac{(e^{V_{1a}}+e^{V_2})}{(e^{V_{1b}}+e^{V_2})}]$. Note that the additive constant, C, vanishes when calculating the change in consumer surplus. (De Jong et al, 2005, Small 1995).

A question is whether this measure, i.e. logsum, is a suitable measure of consumer welfare. The literature seems to point to the fact that whenever a logit model is used to estimate the choice probabilities one also should use logsum as the welfare measure (De la Barra, 1989). This actually is seen as an advantage since then the estimation and evaluation parts are done in one step. However, this also implies that whenever the logit model is not representing the real situation under consideration well then the calculated logsum is also wrong. For instance if we have individual variation that is not captured by the logit model because of simplifying assumptions (for instance ignoring possible heteroskedasticity as discussed above) then the logsum will also give unreasonable results.

APPENDIX 2 ALTERNATIVE WAYS TO DEAL WITH HETEROGENEITY IN THE CONTEXT OF MODELLING LONG DISTANCE MODE AND ROUTE CHOICE – COMBINING RDT AND DISCRETE CHOICE (DC) MODELS.

An RDT – model can handle the schedule delay aspect related to scheduled traffic in an approximately correct way. The concern with deviations from preferred departure and/or arrival time is also relevant for the choice of mode. Travellers might for various reasons also have mode-specific preference that follows some statistical distribution across individuals and modes. This is the concern of discrete choice (DC) - models, which - at least in traditional applications – are unable to handle the schedule delay issue in a correct way.

A more correct model might be an RDT-model with the discrete choice aspect included for modes. Another option is a DC-model with the schedule delay aspect of route choice included in a correct way. While the two approaches should give approximately the same results, at present it seems that the first option is the most flexible and is easiest to implement.

It can be proved rigorously that most continuous statistical density functions can be approximated very accurately by a discrete five-point distribution with appropriate weights for each point. We write the (observable) generalized cost of an alternative by:

$$G_{m,i} = G^{*}_{m,i} + H_{m,i} v_{m,i}$$

$$\sum_{i,j,k} W_{1,j,k} = 1$$

Where m denote mode and I an alternative within a mode, i.e. a route. $G^{*}_{m,i}$ includes costs in terms of monetary cost, in-vehicle time etc, all normalized to the mean value of minute deviation from preferred arrival and/or departure time. $H_{m,i}$ is the headway and $v_{m,i}$ is a random variable on the unit interval. The product is a fraction of the headway. An RDT-model will assign trips on modes and routes for the integral over all values of $v_{m,i}$ across all modes and routes.

A discrete choice assumes preferences for modes. Thus for each alternative within a mode we will have an additional “cost” ε_m. This cost which can be both positive and negative can be considered as a random variable with a continuous distribution over modes and travellers. If this distribution is approximated by a 5-point distribution we have the number of combinations of points and modes equal to 5^modes. We can make an assignment for each combination, which in the case of 3 modes will come to 125 assignments.

Let the 5 data points be $d_1,\ldots,d_5$ and the corresponding weights be $w_1,\ldots,w_5$.

For a particular assignment with points i,j,k respectively for the three modes respectively, we add the term $\sigma \cdot d_i$ to all fixed costs for routes belonging to mode 1, $\sigma \cdot d_j$ to routes belonging to mode 2 and $\sigma \cdot d_k$ to routes belonging to mode 3. σ is the standard deviation assumed for the distribution.

We then make an RDT-assignment with these adjusted costs and weigh the results by

$W_{ijk} = w_i \cdot w_j \cdot w_k$. $\sum_{i,j,k} W_{ijk} = 1$ and we can thus add up the sum of weighted assignments to get the results for a combined RDT/DC – model. In practice, the necessary number of assignments is (5^modes – 5) and for symmetric distributions even less.

Conceptually this is straightforward if we know the distribution and the standard deviation to apply. However, for a full implementation also involving multimodal trips, it becomes somewhat more involved and tedious to implement if the same mechanism shall be applied at every possible transfer point.

APPENDIX 3 COMPUTING CONSUMER SURPLUS IN LONG DISTANCE PUBLIC TRANSPORT

Long distance public transport

One important property with long distance public transport, as opposed to short distance (or urban) transport, is that departures take place with a low frequency, which implies that it is reasonable to assume that travellers find out the actual frequency delay (“generalised” cost for the fact that the departure doesn’t take place at the ideal time) rather than the expected frequency delay. The implication of this is that the generalised cost (G) for a particular choice of travel alternative (modes, lines) is endogenous in the sense that it depends not only on the ticket price, time schedule etc. for that particular choice, but on the time schedules for all substitute alternatives.

An example

To get an illustration, consider the following example:

- Travellers want to go from destination P to destination Q.
- There are two buses, Black and Red, which both go from P to Q,
- they both have the same ride time and fare.
- However, Black has a headway of 60 minutes, Red 90 minutes. Black leaves at 12pm, 1pm, 2pm etc. Red leaves at 12:15pm, 1:45pm, and 3:15pm etc.
- There are 300 travellers, each with a personal ideal departure time. These departure times are uniformly distributed between 12pm and 3pm (note that the schedules will repeat in the same manner with a period of 3 hours.)
- Travellers have to be at their destination in time, so they take the bus that leaves at the latest time before (or exactly at) their ideal departure time.
- Travellers consult a time table to decide which line to chose.

The schedules are illustrated below:

Hence, travellers with ideal departure time between 12pm and 12:15pm will chose Black at 12pm, travellers with ideal departure time between 12:15pm and 1pm will chose Red at 12:15pm, etc.

Let us denote the time from actual departure to ideal departure "waste time", which is a kind of generalised cost, G. It is now easy to see, that

- 200 travellers will chose Black, with a mean waste time, i.e. "mean generalised cost" MGC=24.375 minutes.
- 100 travellers will chose Red, with a MGC 18.75 minutes.

Now let us consider a change of headway for the Red buses. Assume that they too leave with an interval of one hour: 12:15pm. 1:15pm, 2:15p,. etc.

Similar calculations show:

- 75 travellers will chose Black, with a mean generalised cost MGC=7.5 minutes.
- 225 travellers will chose Red, with a MGC 22.5 minutes.

Hence,

- the demand for Black goes down from 200 to 75, MGC goes down from 24.375 min. to 7.5 min.
- the demand for Red goes up from 100 to 225, MGC goes up from 18.75 min. to 22.5 min.

Note three important features

1. The MGC for travellers *with Black* goes down by 69%, although the change in schedule is for done for *Red*.

2. MGC and demand goes in the *same direction* i both cases: when price (MGP) goes up, demand also goes up; when demand goes down, demand goes down.

3. An apparent *improvement* in the Red schedule—a reduction of headway with 33%—leads to *higher* MGC for that mode.

Consequences for the Computation of Consumer Surplus

The "traditional" way to compute (changes in) consumer surplus (CS) is to employ the rule that CS is the "area under the demand curve". But in the current context, there is no such thing as “a demand curve”. The very notion of “demand curve” is that price is exogenous to the consumer (he is price taker) whereas he chooses the quantity to buy according to a utility maximising scheme. In the current context this is still true, but the actual “price” differs among travellers (because of their heterogeneity as to ideal departure time,) and since they have full information of all alternatives (substitutes,) they will allocate themselves in such a way that on the aggregate, the mean generalised cost is not exogenous, but endogenous, to the aggregate of travellers.

In the example above, what would the “demand curve” for Red be? What would the “demand curve” for Black be? Wouldn’t any candidate for such a curve be *upward sloping* according to the previous paragraph? How would we calculate the total change in CS due to the change in schedule for Red by calculating areas under some “demand curve”?

Two conclusions

1. Consumer surplus, as well as demand, must be computed from “first principles” where all alternatives (substitutes) are taken into account simultaneously, even if a change in schedule is done only for one of them. (In particular, any attempt to employ the “rule of one half” would be erroneous.)

2. Since the MGC is typically not directly observable (because of its endogenous nature) a theoretical model that sufficiently well describes the real situation is imperative, such that the MGC and hence consumer surplus can be accurately assessed.