

# Generation and Calibration of Transit Hyperpaths

Jan-Dirk Schmöcker<sup>a</sup>, Hiroshi Shimamoto<sup>a</sup>, Fumitaka Kurauchi<sup>b</sup>

<sup>a</sup>*Department of Urban Management, Kyoto University, Japan*

<sup>b</sup>*Department of Civil Engineering, Gifu University, Japan*

---

## Abstract

This paper introduces a new discrete choice model aimed at describing behaviour of public transport passengers at stops. We assume that passengers choose a set of buses from which they take the first arriving. This leads to a nested model formulation in which the upper level (choice set formation) is based on utility maximization. The lower level choice of a specific bus from the choice set is given by the frequency distribution of the bus arrivals. We further consider hyperpath characteristics in the choice set formulation which means that the utility of the choice set in general increases with the addition of further options due to a reduction in the reduced waiting time. We discuss model properties and apply our model to some selected OD pairs of the bus network of a local city in Japan where we could observe passenger behaviour due to the availability of smart card data. We find that choice sets vary fairly significantly between some passenger groups and discuss implications for transit assignment models.

*Keywords:* Hyperpath, choice set calibration; discrete choice modelling; smart card data; bus network

---

## 1. Introduction

Understanding and predicting travel patterns and travel demand is important for transport providers, in particular those of public transport services. Therefore there is by now a significant body of literature describing and modelling the expected demand as well as the random variation and adaptation of travel patterns over time. Assuming that choice is based on the utility concept, random variation in behaviour is usually dealt with by including stochastic factors into the model, leading to random utility models.

In particular in transit networks random variation and route selection are, however, often difficult to distinguish. A passenger might alter his/her route choice on subsequent days not because of any learning process but because of service inherent uncertainties. Whereas for drivers usually a change in route is explained with a (perceived) change in attractiveness of the road conditions (e.g. Jotisankasa and Polak, 2005), this explanation is not required for transit passengers. That is, a passenger might switch from (his/her preferred) bus A taken on day 1 to an alternate bus B on day 2 simply due to reverse order of arrival. On day 3 the passenger might, however, return to bus A if it arrives again earlier. This random variation can be usefully described with the concept of hyperpaths and strategy (Nguyen and Pallentino, 1988; Spiess and Florian, 1989).

A hyperpath of transit passengers is generally defined as a choice set consisting of a number of paths out of which any could be optimal depending on the arrival of buses that form the paths. Different strategies to choose a specific path from options among the hyperpath are possible. Spiess and Florian proposed the simple strategy “take whichever bus comes first”. This strategy can be shown to be useful for risk averse passengers fearing that any bus at the stop might arrive only after a time equal to its headway (Schmöcker et al, 2009). Alternative more complex strategies are, however, also possible, in particular if more information about the time until the next arrival of the attractive lines are available (Nöckel and Weckek, 2010).

Though there is ample literature discussing that different passenger groups attach different values to on-board travel time, waiting time, transfers or seat availability, this has not been much reflected in hyperpath based transit assignment models. Obviously, different values will result in different path sets. Generally, higher values attached to waiting times at stops compared to on-board travel times will

lead to more complex hyperpaths. That is, passengers will avoid being fixed to a specific line and rather increase their choice set (with some lines that are potentially longer) in order to reduce their expected waiting time. Kurauchi et al (2012) provide evidence for this with a stated preference survey. They show that there are some significant differences in the hyperpaths constructed by elderly. Fonzone et al (2012) extend this line of research by providing further evidence for differences in hyperpaths among passenger groups based on a survey asking respondents to describe their actual travel patterns as well as asking them to choose their strategy in hypothetical bus networks. It appears that only some passengers choose the hyperpaths predicted by the Spiess and Florian model; a significant number of passengers appear to prefer simple choice sets. Furthermore, the choice of strategy seems to be influenced by the actual experiences made by the passenger during their daily commute.

Both of these studies have used “artificially” collected data for the estimation of hyperpaths, i.e. transit passengers have been asked to recall their travel patterns and to answer hypothetical choice situations. Through the advance in smart card data nowadays there are improved possibilities though. A large number of cities have introduced such systems offering possibilities for better understanding of passenger behaviour, service planning and evaluation (Pelletier, 2007). For our purposes, of importance is that such data store the actual lines boarded by passengers over longer time periods. This allows us to observe their actual choices and derive some conclusions about their strategy.

This forms the motivation for this paper. Our objective is to develop a model that estimates the choice set of passengers in dense networks where it is often optimal to form complex choice sets. Since choice sets are latent we assume that we can construct these by observing passengers repeated choice when they travel from the same origin to the same destination. We obtain such time series route choice data from the bus service provider of a local city in Japan and provide some example results for different OD pairs where we could observe that different passengers have taken different routes to reach the same destination. The paper is structured as follows.

We firstly review previous research on choice set construction and point out two important differences that make our problem different to the existing literature. We then develop our methodology, emphasising assumptions we make in the process. In the final parts of our paper we develop our case study. We first describe the bus network data, demand patterns and then apply our methodology. We discuss our model estimations as well as implications for transit assignment and further research.

## **2. Choice Set Generation Methods**

### *2.1. Introduction*

Various authors have pointed out the importance of generating an appropriate choice set from the totally available options and potential errors that might arise otherwise (e.g. Swait and Ben-Akiva, 1986; Ortuzar and Willumsen, 2001). In particular Bovy (2009) reviews the generation of “consideration sets” from the “universal sets” and points out difficulties of choice set generation for route choice modelling. In comparison to for example mode choice, the universal set can be extremely large, choices can be “hidden”, i.e. need to be extracted from a network and one needs to consider overlap between options. Many of these “universal options” the traveller might a priori reject and just focus his choice on a few potentially attractive options. Dial’s (1971) algorithm is one common way to reduce the path set to “reasonable options” that could then be included into a choice model though the set might still be (too) large. Fonzone et al (in press) use the term “potentially optimal” routes and show that in road networks even assuming that links have only a congested and uncongested state, the number of options can be extremely large. In large cities transit passengers sometimes face similar problems to choose between several lines that could bring them to their destination, in particular if one considers that passengers can choose their starting station as well.

Furthermore the choice set (and its size) will depend on the underlying choice behaviour. If the traveller chooses the route from the origin to the destination at once he might consider a different set

than travellers choosing their path sequentially or strategically anticipating possible downstream delays. One might argue that the more uncertainty there is downstream, the more important it becomes to choose one's path sequentially and strategically in order to minimise overall travel time. This has been the main motivation for the development of optimal hyperpaths in transit assignment where waiting time at stops is the main source of uncertainty. Schmöcker et al (2009) for example show that the most commonly used transit assignment approach following Spiess and Florian (1989) is equivalent to an optimal strategy fearing a specific type of worst case scenarios at each node.

Consequently, this has led to the development of two sets of literature relevant to our problem and reviewed in the following, which have not been overlapping much to our knowledge. On the one side the literature emanating from authors' often based in discrete choice modelling who have developed various approaches to generate approximations to universal choice sets and then person specific consideration choice sets. We will argue though that this set of literature has omitted the consideration of some specific issues for generating choice sets relevant for route choice within high frequency transit networks. On the other side the transit assignment literature has focused on finding optimal strategies, though mostly ignoring the specification of person-specific hyperpaths and inclusion of stochastic effects.

## *2.2. Discrete Choice literature*

A large number of methods have been proposed to create choice sets and a range of issues for route choice applications have been reviewed in the aforementioned paper by Bovy (2009) as well as by Bekhor et al (2006). A general discussion remains whether the choice set generation should be separated or combined with the choice model itself.

Proponents of a two-step process argue that this helps to better reflect constraints such as maximum size of considered options as well as exclusion of unrealistic options. Horowitz and Louviere (1995) argue that one step approaches can be sufficient for reflecting consumer choices but that generating choice sets as a separate step can provide the analyst with additional information for choice predictions. Within transport applications though, sometimes modelling constraints will be of primary importance so that in several cases 2-step models are more appropriate. For example for complex mode choice decisions within trip chains excluding unlikely options such as choosing public transport followed by private car on the same journey will generally improve the model fit (Fiorenzo-Catalano, 2007; Hoogendorn-Lanser, 2007).

For public transport route choice applications also choice sets have been created as a separate step. Friedrich et al (2001) use branch and bound techniques to create sets of possible routes for railway passengers, also with the aim to reflect the constraints imposed by (long) waiting times and departure patterns. Kato et al (2010) develop route choice models for Tokyo's large urban rail network and discuss different options to develop choice sets. They use a combination of generating the k-shortest paths combined with expert knowledge limiting the choice set to those that appear realistic. They then test the applicability of different discrete choice methods to replicate observed flows in the network. In Nuzzolo (2003) further transit assignment models are presented where travellers are assigned stochastically to the shortest routes according to a fixed service schedule and previously fixed choice sets.

Bovy (2009) classifies the Friedrich et al (2001) approach as "constrained enumeration procedure" since the starting point for the consideration set generation is the full network. Bovy as well as Bekhor et al (2006) review further methods that have been proposed to generate choice sets for (driver) route choice which rather increase the choice set until a sufficient size is reached. Given a single shortest path several heuristics can be employed to generate additional paths. Bekhor et al distinguish heuristics depending on whether they use link penalties or eliminate links from the shortest path. Alternatively links might be "labelled" according to various attributes. A fourth method is to draw a number of network conditions and find according to these shortest paths.

Note that one might also distinguish route sets depending on whether they have been created “globally” for all users or considering user preferences. Manski (1977) and other authors have long highlighted the role of (individual) constraints as determinants of choice feasibility and attractiveness. Also Gaudry and Dagenais’s (1979) development of the dogit model is motivated by the fact that some users will be captive to some options. They introduce “dogits” into the choice model (not the choice set generation) to overcome the IIA assumption for all options. Specific for route choice and considering “utilities” within the choice set generation Freijinger et al (2009) hence propose an approach where route sets are determined “probabilistically” by using an importance sampling approach. Based on a random walk biased towards the shortest path, they determine a correction term which reflects that the more the link deviates from the (subjective) shortest path, the less likely it is considered by the traveller. These probabilities can then be directly utilised for assignment purposes.

Proponents of one-stage approaches hence mainly argue that the attributes used for consideration of the choice set and for the choice itself overlap. In line with this, Swait (2001) follows the 2-stage approach but generates the attractiveness of choice sets endogenously with the same attributes that determine the choice. His model, termed “Generation Logit”, hence develops the choice set attractiveness also based on utility or “tastes” whereas previous authors, such as Manski (1977), have highlighted the role of constraints as determinants of choice feasibility and attractiveness. In Swait (2001) the possible (exogenously determined) choice sets could include all  $2^i - 1$  options. Swait discusses that this, often very large, number could be significantly reduced by for example making assumptions on the maximum number of options per choice set. Also Morikawa (1991) discusses this problem and, similar to Swait, he determines choice sets attractiveness based on utilities. In contrast to Swait, however, choice set attractiveness is determined by the attractiveness of including the option as well as excluding it.

### *2.3. Transit hyperpath characteristics*

As noted above the starting point for hyperpath generation is a dynamic programming approach where the traveller sequentially chooses the shortest path. The advantage is that the choice set at each node becomes, in many cases, manageable. In the context of transit passengers, where choices are made at stops, usually there are not more than at most 5 or 6 lines that one might consider, unless at some major hubs in large metropolitan networks. This means that for our problem the universal choice set is easily obtainable but that generating the by the traveller considered options is the main issue.

In contrast to discrete choice approaches frequency-based transit assignment choice sets are generated truly endogenously. Following the seminal work of Spiess and Florian (1989) the above defined hyperpaths are created by solving a linear program at each node. The problem can be solved with a variant of Dijkstra’s shortest path algorithm by backward search from the destination. We emphasise two important aspects of choice set generation for high frequency transit networks. These are justified and reflected in the work of Spiess and Florian but are not captured within the above described choice set generation methods.

**Transit Choice Characteristic (TCC) 1:** Whereas the transit passenger has full control over the selection of the choice set, (s)he might leave the choice of a specific option from the choice set to some degree “up to nature”. Especially in the absence of countdown information passengers might take a strategy of taking whichever bus comes first. This means that the choice depends to some degree on the (unknown) arrival time of the bus and only partly on utility maximisation at this 2nd choice level.

**Transit Choice Characteristic (TCC) 2:** The utility of a choice within a set is depending on the choice set itself. This is the case for passengers at bus stops due to the expected waiting time effect. The more choices are included in the choice set, the shorter the total expected waiting time becomes. Therefore passengers will include fast but infrequent transit lines into their choice set if the risk of

potentially long waiting times can be compensated by including also other buses with potentially longer on-board travel times into their choice set. Focusing on the fast bus only would be too risky.

In summary, it can be concluded that there are two sets of literature which tackle different aspects of a similar problem. On the one side the literature reviewed in Section 2.2 has developed various approaches to describe person specific choice sets. This set of literature has though, mostly omitted the above two choice characteristics typical for choices in high frequency transit networks where passengers have not full information about waiting time. On the other side the transit assignment literature has focused on these two characteristics, though mostly ignored the specification of person-specific hyperpaths and inclusion of stochastic effects. One exception is though the approach presented in Nguyen et al (1998) where passengers choose a specific path from a hyperpath with a logit approach. The path set in their work is though not depending on person-specific attributes.

In the following it is therefore our aim to offer an approach that combines both sets of literature. We develop a model that captures the above two transit characteristics and captures (person group specific) evaluation of the importance of travel time and weighting time already in the choice set generation. As will be shown in the following we therefore take the “opposite approach” compared to Nguyen et al (1998). In the approach presented here the choice set generation is stochastic (based on utilities) and the choice of a line itself is deterministic (based on line frequencies only). In contrast in Nguyen et al, the choice set generation is deterministic (based on the efficiency principle) whereas the choice of a line from a choice set is stochastic (based on utilities).

### 3. Logit Hyperpath Choice Set Generation with Random Bus Choice

#### 3.1. General Framework of choice at a stop

Given our transit characteristic TCC1 and TCC2 we require a joined modelling approach of choice set and choice since the traveller has only control over the choice set. Once the choice set is determined the choice probabilities are given by the service arrivals. Therefore, partly following the decision framework and notation in Swait (2001), the choice probability  $P(i, n\tau)$  can be described as

$$P(i, n\tau) = \sum_{k \in K_i} P(i\tau | C_k) \cdot Q_{n\tau}(C_k) \quad (1)$$

where  $i$  denotes the chosen option from a bus stop,  $n$  the “person type” and  $\tau$  the time period in which the traveller departs. Consideration of time is required since we consider that the service level varies during different periods of the day due to changes in the service frequency as well as longer travel times during peak traffic hours. Further,  $C_k$  represents the choice set or “nest” and  $Q_{n\tau}$  describes the attractiveness of the choice set as explained below.  $K$  denotes the number of choice sets and  $K_i$  the choice sets that include choice  $i$ .

The conditional choice probability for our bus choice problem of an option (i.e. bus line) from the choice set is formulated as

$$P(i | C_k) = \Pr(a_i + E(u(i)) < a_j + E(u(j)) \mid \forall j \in C_k; j \neq i) \quad (2)$$

where  $a_i$  denotes the waiting time until bus  $i$  arrives at the stop and  $E(u(i))$  the expected disutility of choosing option  $i$  from the current stop once the bus arrives (i.e. ignoring the waiting time at this stop). This equation hence fulfils our above stated TCC1. The passenger can estimate  $E(u(i))$  according to his/her taste but  $a_i$  is out of the passenger’s control. Following Spiess and Florian (1989) we simplify (2) by assuming that a) buses arrive with exponentially distributed headways and b) that passengers

only know the service frequency (no countdown information). In that case, and considering that service headways differ during the day, the above formulation reduces to

$$P(i\tau|C_k) = \frac{f_{i\tau}}{\sum_{j \in C_k} f_{j\tau}} \quad (3)$$

where  $f_{i\tau}$  denotes the service frequency of option  $i$  during departure time interval  $\tau$ . Note that we assume in (1) that this equation is only applied if  $i$  is included in  $C_k$ . Equation (3) illustrates that the lower choice problem is dependent only on the service frequency and is independent of personal characteristics of the decision maker.

In line with nested choice models as well as Swait's "Generation Logit" we assume that the selection probability of a choice set is determined by a general cost or inclusive value associated with this nest / choice set. We assume a dispersion parameter  $\mu$  and logit choice structure leading to

$$Q_{n\tau}(C_k) = \frac{\exp(\mu I_{kn\tau})}{\sum_{r=1}^K \exp(\mu I_{rn\tau})} \quad (4)$$

Accordingly the inclusive value should reflect the perceived disutility of choosing this choice set  $C_k$ . Eq. (5) hence includes on-board travel time, waiting time and expected number of transfers. In line with TCC2, however, one needs to consider that these attributes depend on the choice itself. This is in contrast to other nested discrete choice models where the utility of a nest can be determined as the logsum of the utility of the options within a nest. We therefore obtain the nest specific expected values  $T_{k\tau}$ ,  $W_{k\tau}$ , and  $Y_{k\tau}$  for travel time, waiting time and expected number of transfers respectively. We further include specifically path set size as a value for the inclusive value. This is based on findings in Kurauchi et al (2012) that some passenger groups seem to prefer simple hyperpaths *per se*. In other words, even if including an additional line would reduce the overall expected travel time, some passengers might prefer to limit their choice set, possibly to avoid having to track and check the arrival of multiple lines at the stop.

$$I_{kn\tau} = \beta_{in} T_{k\tau} + \beta_{wn} W_{k\tau} + \beta_{yn} Y_{k\tau} + \beta_{zn} |C_k| \quad (5)$$

### 3.2. Expected values for service attributes

Following our assumption of a frequency based service we obtain the expected nest specific service attributes as weighted average over the likelihood of taking a service within the choice set as in Equations (6) to (8).

$$T_{k\tau} = \frac{\sum_{i \in C_k} t_{i\tau} f_{i\tau}}{\sum_{i \in C_k} f_{i\tau}} \quad (6)$$

$$W_{k\tau} = \frac{1}{\sum_{i \in C_k} f_{i\tau}} \quad (7)$$

$$Y_{k\tau} = \frac{\sum_{i \in C_k} y_{i\tau} f_{i\tau}}{\sum_{i \in C_k} f_{i\tau}} \quad (8)$$

$t_{i\tau}$  and  $y_{i\tau}$  denote the expected travel time and number of transfers respectively if the traveller is boarding line  $i$  at the current node. We emphasise that these values are for the whole path from the current boarding point to the destination and not just for the travel time until the next decision point. This implies that our choice problem has to be solved in a network context recursively backward from the destination. Only if the passenger's strategy (choice likelihood of nests) at the downstream nodes is determined the expected travel time and the expected number of transfers can be determined.

### 3.3. Model Properties

We note that cross-nested logit models and Swait's "generation logit" model include scale factors  $\mu_k$  for each nest. These are not included in our model as we assume that the sensitivity to choice on the lower level is not determined by utility but is fixed and given by the bus frequency. Therefore our model only includes the scale parameter  $\mu$  which can be interpreted as the sensitivity to utility for the hyperpath set choice. As in MNL choice models we can, however, fix one parameter among  $\{\mu, \beta\}$  and in the following choose this to be  $\mu$  (Lerman and Ben-Akiva, 1985).

Swait (2001) notes that the generation logit model collapses to the MNL model in case all  $\mu_k$  equal  $\mu$  and if all alternatives appear in the same number of choice sets. In contrast, our model does not collapse to the MNL since the lower choice probability is not determined by the utility. As boundary conditions we can only establish a special case.

**Proposition:** Assume 1) all line frequencies  $f_{i\tau}$  take the same value 2) all  $2^i - 1$  hyperpaths are possible and one of following conditions is fulfilled: 3a)  $\beta_{tn} = \beta_{yn} = 0$  or 3b)  $t_i$  and  $y_i$  are the same for all options. In this case the choice probability for each option is identical with  $1/(\text{number of choices})$ .

**Proof:**

Condition 1 ensures that the model structure collapses to

$$P(in \tau) = \sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{\exp(\mu I_{kn\tau})}{\sum_{r=1}^K \exp(\mu I_{rn\tau})} \right) = \sum_{k \in K_i} \left( \frac{1}{|C_k|} \cdot \frac{\exp(\mu I_{kn\tau})}{\sum_{r=1}^K \exp(\mu I_{rn\tau})} \right) \quad (9)$$

Condition 3a or 3b is required so that the "hyperpath effect" vanishes for expected travel time and number of transfers, i.e. the effect of the service attributes, is not depending on the nest composition. With either 3a) or 3b) applying we can reduce Eq. (5) to

$$I_{kn\tau} = \beta_{wn} W_{k\tau} + \beta_{zn} |C_k| = \beta_{wn} \frac{1}{\sum_{i \in C_k} f_{i\tau}} + \beta_{zn} |C_k| \quad (10)$$

Since, according to condition 2) all options  $i$  appear in the same number of nests with same cluster sizes this ensures that  $\sum_{k \in K_i} \left( \frac{1}{|C_k|} \cdot \exp(\mu I_{k\tau}) \right)$  is the same for options  $i$ . Hence  $P(is)$  is the same for all options. Qed.

Note that in this case the multinomial logit model of choice between each single line gives the same result. This is since the size of the choice set is not considered in such a MNL model and condition 1) assumes that all line frequencies are identical. To better understand the properties of our model in the following we compare the MNL model with our proposed model assuming that only travel time and line frequency influence choice.

Consider choice between three lines at a stop. Line 1 is infrequent but fast ( $f_1 = 5$  services per hour,  $t_1 = 20$  min), Line 3 is frequent but slow ( $f_3 = 15$ ,  $t_3 = 30$ ) and Line 2 is a compromise between both ( $f_2 = 10$ ,  $t_2 = 25$ ).

Figure 1 compares the predicted probability of a traveller choosing line 1 as a function of  $\beta_{wait}$  assuming  $\beta_{travel\_time} = 0$ . At  $\beta_{wait}$  the MNL predicts that the traveller is equally likely to choose each line, whereas in our proposed “hyperpath logit” model it is considered that other lines are more frequent so that the choice of the fast line is only 0.25. With increasing importance of waiting time in both models the likelihood of taking Line 1 reduces. Our proposed model is, however, less sensitive to  $\beta_{wait}$ . This is because the likelihood of choosing Line 1 only reduces on the “upper choice” level, i.e. the likelihood of choosing a hyperpath that includes Line 1. Figure 2 shows the similar effect for a decrease in  $\beta_{travel\_time}$  assuming that  $\beta_{wait} = 0$ . With increasing importance of travel time in the choice the attractiveness of Line 1 increases, though in our proposed model less fast. Finally Figure 3 shows the effect of increasing the service frequency of Line 1 assuming  $\beta_{wait} = \beta_{travel\_time} = 1$ . Our proposed model is more sensitive to an increase in service frequency, as frequency influences choice at the lower level as well as on the upper level as it increases the utility of hyperpaths that include Line 1.



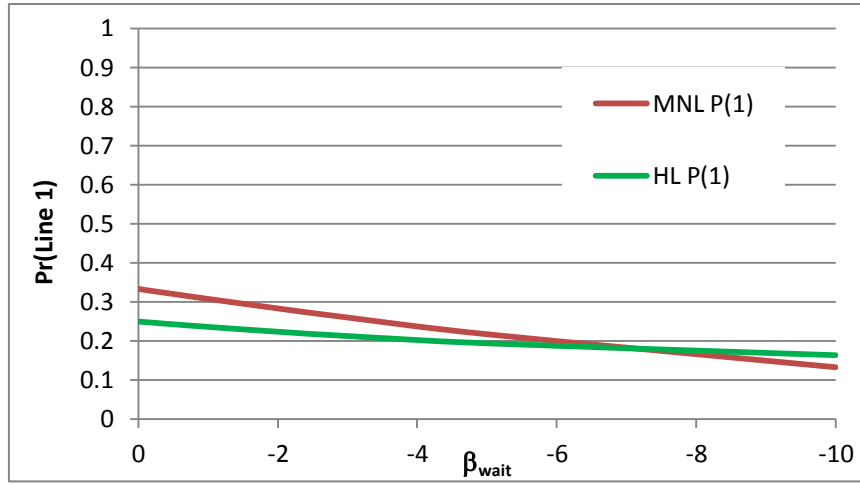


Fig. 1. Influence of waiting time sensitivity on the MNL and our proposed “HL model”

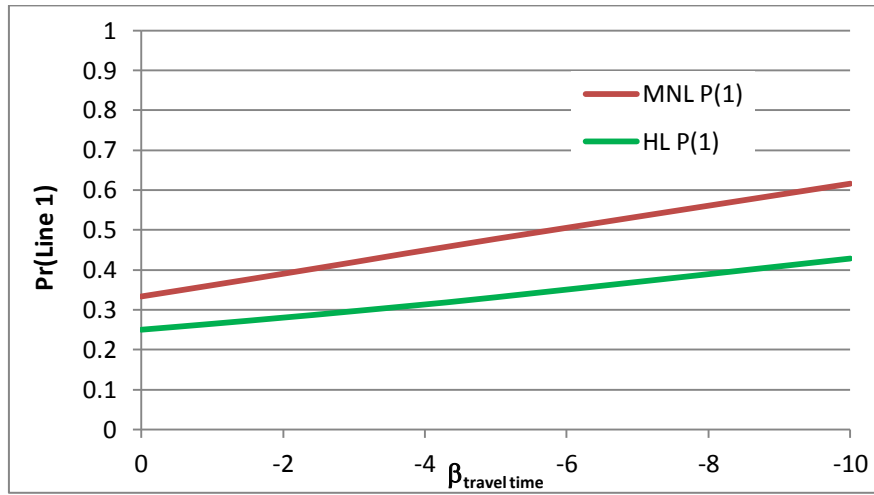


Fig. 2. Influence of travel time sensitivity on the MNL and our proposed “HL model”

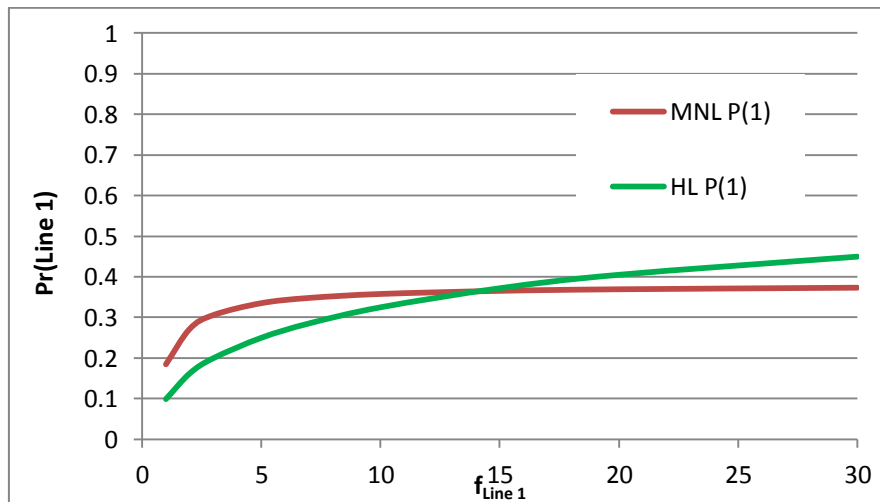


Fig. 3. Influence of service frequency on the MNL and our proposed “HL model”

#### 4. Maximising the Likelihood Function

For generality and notational simplicity we refer to the service characteristics that determine the nest attractiveness in (5) as  $X_{akn\tau}$  and  $Y_{ak\tau}$  where  $a$  denotes travel time, waiting time, number of transfers or nest size.  $X_{akn\tau}$  hence denote the attributes that are estimated passenger group specific, whereas the values of attributes  $Y_{ak\tau}$  are estimated for the whole sample. The estimated probability that line  $i$  is chosen by sample  $s$  is then (11) where  $I_{kn\tau}$  is determined with (12).

$$P(i, n\tau) = \sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{\exp(\mu I_{kn\tau})}{\sum_{r=1}^K \exp(\mu I_{rn\tau})} \right) \quad (11)$$

$$I_{kn\tau} = \sum_a \beta_{an} X_{akn\tau} + \sum_b \beta_b Y_{bk\tau} \quad (12)$$

Let us denote the observed choices as set  $S$  consisting of samples  $s$ . Each sample is associated with a person type  $n$  and a travel period  $\tau$ . With this formulation the likelihood function  $L$  in (13) and the log likelihood function  $L^*$  (14) can be formulated as follows where  $n(s)$  and  $\tau(s)$  denote the person types and travel period of sample  $s$  respectively.  $\delta_{is}$  is 1 if sample  $s$  chooses option  $i$  and 0 otherwise.

$$L = \prod_i \prod_s \left( \sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{\exp(\mu I_{kn(s)\tau(s)})}{\sum_{r=1}^K \exp(\mu I_{rn(s)\tau(s)})} \right) \right)^{\delta_{is}} \quad (13)$$

$$L^* = \sum_i \sum_s \delta_{is} \ln \left( \sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{\exp(\mu I_{kn(s)\tau(s)})}{\sum_{r=1}^K \exp(\mu I_{rn(s)\tau(s)})} \right) \right) \quad (14)$$

We aim to maximise (14) with respect to our parameters  $\beta_{an}$ . For this we establish the gradient and Hessian of this log likelihood function in Appendix A. Unfortunately, especially the Hessian takes a complex nonlinear form. Similar to other cross-nested logit models we cannot establish that our objective function is concave and hence test convergence with different initialisation for our parameters (Bierlaire, 2006). In the following tests we use Matlab<sup>R</sup> for estimation of (14) as well as the calculation of t-values of our parameters and model fit statistics at convergence.

## 5. Case Study with Smart Card Data

### 5.1. Data Overview

To check the validity of the proposed model, we use smart card data obtained by a bus operator of a local city in Japan. The dataset includes 2,100,285 records made by 82,320 cards over two months from the beginning of September 2011 until the end of October 2011. The smart card has been used mainly for the route bus services but also for some community-run bus services within this Prefecture. Therefore we picked up the records of route bus services within the city. Consequently, we have 2,005,421 trip records made by 44,310 cards.

The bus company applies a flat-fare system in the central part of the city with a time-independent fixed cost per ride. Once the bus leaves the central area, an additional distance-based fare is applied. For this reason, passengers have to tap the smart card twice, when boarding and alighting a bus. If passengers transfer to another bus route, the fare on the subsequent bus is discounted if the transfer time is less than 45 minutes. Because of this fare structure, we can accurately identify the boarding and alighting bus stop. The advantage of our dataset compared to smart card data from other cities can be summarised as follows: 1) The card ID has been kept and individual behaviour can be tracked; 2) the whole city is covered only by the bus services and there is no rail service; 3) more than 70% of travellers use the smartcard data; and 4) boarding and alighting bus stops can be identified since travellers have to tap at boarding as well as alighting. Since the smartcard data also contain date/time of boarding and alighting as well as route and bus ID we can therefore identify the data needed for our model.

A fifth characteristic is that the bus service is schedule based. This has the advantage that we can estimate the effect of the delay of service. It has the disadvantage though that our assumptions made in (3) and (5) are possibly too simplistic. If passengers know (and trust) the schedule the experienced traveller will not arrive random at the bus stop but instead time his/her arrival. This will have an effect on the expected waiting time and line choice probability (see Nökel and Wekeck (2009) for a detailed discussion on this.)

## 5.2. Data Extraction

From the journey dataset we pick up some OD pairs where there is choice between different routes and where we can observe repeated choices. This limits our data choice for this network fairly stringent as there are few OD pairs for which there are reasonable distinguishable alternatives and a significant number of observations. We pick up the 3 OD pairs and construct the hyperpaths. The destination of all three ODs is the railway station as many passengers arriving there transfer to rail lines. Figures 4 a)-c) illustrate the ODs together with the chosen routes and their passenger share. We note that there are six chosen routes from Origin B to the station, leading to  $63 (=2^6 - 1)$  choice sets.

Table 1 summarises the service characteristics of all lines. For some lines the service attributes differ significantly depending on time of day. In particular line 2 of OD *a* is only operated in the morning peak hour. Detailed service characteristics by time of the day are reported in Appendix B. Limiting our sample further to those time periods where passengers face a choice, we obtained 4,033 journeys made by 257 cardholders for OD *a*, 1,589 journeys made by 122 cardholders for OD *b* and 958 journeys made by 123 cardholders for OD *c*. Note that on none of these three routes we could find passengers who transfer and hence in the following we cannot estimate  $\beta_{ym}$ . For other OD pairs where one transfer is required there is no reasonable alternative route without transfer (or with more than one transfer) or sample sizes are very small, so that these data are not useful for our illustration. This shows a second disadvantage of our data set as this city is not as big and the bus network not as complex as that of other large metropolitan cities, so that transfer is often not required.

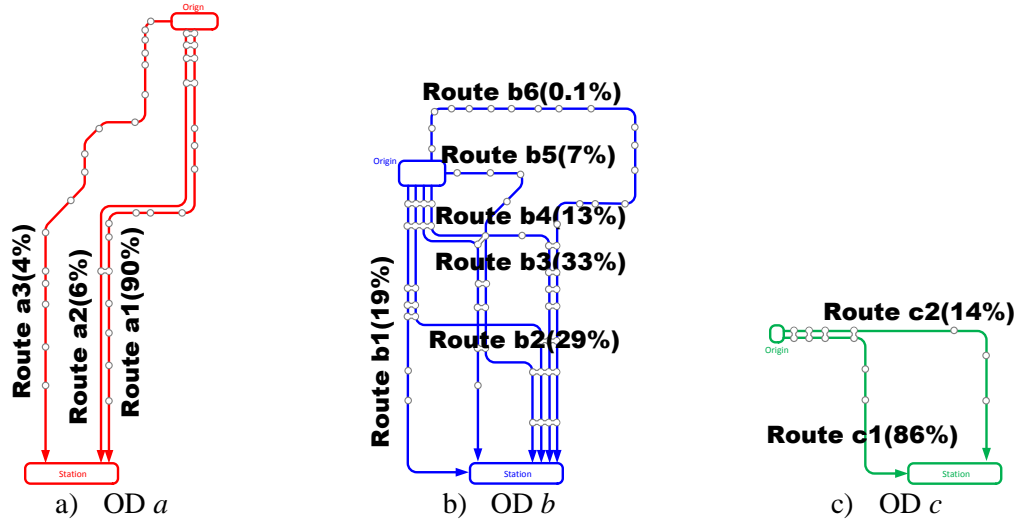


Fig. 4. Three OD pairs with share of passengers for each route

Table 1. Summary of service characteristics of the lines

	OD <i>a</i>			OD <i>b</i>						OD <i>c</i>	
	a1	a2	a3	b1	b2	b3	b4	b5	b6	c1	c2
Services per hour	9-17	4	1-3	1-3	7-9	7-8	3-5	2-3	3-4	4-7	1-4
Operating hours	5am-11pm	6-7am only	6am-8pm	6m-8pm	6am-11pm	6am-11pm	6am-10pm	6am-7pm	6am-11pm	5am-11pm	6am-9pm
Travel time (min)	18-26	18	26-31	10-14	14-16	14-23	12-17	17-23	28	10-12	10-15

### 5.3. User Groups

In a separate analysis we use all bus user data to establish whether we can distinguish some user groups. We use information such as whether the user holds a seasonal ticket, if yes what kind of seasonal ticket, as well as data on his/her “general aggregated” behaviour such as usual day of time travelled and number of trips per month. We further include characteristics on how often they make journeys that include transfers. We employ cluster analysis and find that we can distinguish four passenger groups described in Table 3. Details of this behavioural analysis are reported in Kurauchi et al (2012). With four user groups this means that in total we can estimate up to 12 parameters ( $\beta_{tn}$ ,  $\beta_{wn}$  and  $\beta_{zn}$  for all clusters).

Table 2. Characteristics of the four distinguished user groups.

User group	Characteristic of user group
Commuter	Hold commuter pass, travel often and mostly during weekday, include a large number of students.
Elderly	Hold elderly season pass, travel not often, mostly during day time, make almost no trips that include transfers.
Irregular	Passengers that fairly often make journeys that include transfers (23.6% of all journeys). Fairly few total journeys. Irregular OD patterns.
Other	Not passholders, fairly few total journeys, very few journeys that include transfers.

#### 5.4. Estimation Results

Tables 3 and 4 illustrate our model results for the three OD pairs shown in Figure 4. Travel time and waiting time parameters all have the expected sign. We further note that the model fit varies significantly depending on the OD pair. With larger choice sets the model fit reduces as one would expect. In particular for OD *b* there are six lines with often fairly similar travel times so that for a passenger arriving at the stop without prior knowledge of the exact departure time it is indeed reasonable to choose the line whichever comes first, which explains our lower  $\rho^2$ . (We remind that our model fit measure is an index of the model estimating the specific chosen line correctly, not the choice set, since this is obviously not measureable.)

For OD *a* we estimate two models; in the first one we include a group independent waiting time parameter as well as choice set size. We find that choice set size is not at all significant for this OD pair as well as for all other OD pairs so that we omit it for other estimations. A reason for this is likely the strong correlation with our waiting time parameter. We secondly find that group specific estimates of waiting in general lead to slightly better model fits. Further, we find some fairly consistent differences in the waiting time estimates across the OD pairs. Older persons appear to value waiting time more than commuters. As a result of this we estimate the probability for older persons to choose choice sets including more lines higher than for commuters as shown in Table 4. We believe there are two explanations for this, which we cannot distinguish with data available to us. Firstly, older persons might indeed prefer to spend time in the bus than at the bus stop. Secondly, commuters might have more accurate knowledge of the precise departure time of the services. Therefore they target their arrival time at the bus stop to the arrival of the faster bus services, meaning that service frequency is less of a criterion for their line choice.

We finally note that we find that the parameter estimates to some degree vary depending on the starting point in the maximisation of our log likelihood function due to the above discussed issue that our optimisation might be trapped in local optima.

Table 3. Model Estimation Results

	OD a				OD b		OD c		All OD pairs	
	beta	t-value	beta	t-value	beta	t-value	beta	t-value	beta	t-value
Travel time $\beta_t$	<b>-10.7</b>	-18.6	<b>-183.0</b>	-1062	<b>-43.9</b>	-57.1	<b>-65.8</b>	-441.3	<b>-87.6</b>	-218.6
Waiting time $\beta_{wn}$										
Commuter	<b>-10.8</b>	-6.83	<b>-62.7</b>	-351.1	<b>-3.58</b>	-5.98	<b>-1.48</b>	-3.74	<b>-33.8</b>	-219.4
Elderly			<b>-75.7</b>	-862.0	<b>-34.0</b>	-523.8	<b>-26.8</b>	-420.2	<b>-46.6</b>	-427.3
Irregular			-69.7	-173.0	-0.59	-0.07	-1.20	-1.59	<b>-34.1</b>	-35.1
Other			-79.6	-474.9	-2.99	-1.16	<b>-18.5</b>	-124.65	<b>-48.1</b>	-256.2
Choice set size $\beta_{zn}$	-1.61	-0.23								
sample size	4033		4033		1589		958		6580	
$\rho^2$	0.51		0.51		0.15		0.31		0.27	
LL(0)	2385.6		2385.6		2701.1		576.9		5663.5	
L*	1177.7		1174.1		2305.5		394.9		4137.0	

Table 4. Examples of Estimated Choice Set Probabilities

<b>OD a, 6-7am</b>	(a1)	(a2)	(a3)	(a1,a2)	(a1,a3)	(a2,a3)	(a1,a2,a3)
Commuter	0.01	0.18	0.00	0.64	0.00	0.01	0.17
Elderly	0.00	0.02	0.00	0.75	0.00	0.00	0.22
Irregular	0.01	0.06	0.00	0.73	0.00	0.01	0.20
Other	0.00	0.01	0.00	0.75	0.00	0.00	0.23

<b>OD c, 7-8am</b>	(c1)	(c2)	(c1,c2)
Commuter	0.71	0.04	0.25
Elderly	0.34	0.00	0.66
Irregular	0.71	0.04	0.25
Other	0.47	0.01	0.52

## 6. Conclusions

This paper presented a discrete choice model with explicit choice set generation aimed specifically at transit line choice at stops. A main feature of our model is that it is only on the upper level, the choice set formation, a RUM model. On the lower level the user is assumed to not control his/her choice but simply board which bus from the choice set arrives first. A second aspect of our model is that the inclusive value of the nest considers the “hyperpath effect”. For example in the “generation logit” model of Swait (2001) the inclusive value of a nest is estimated as the log sum of the utilities of the options within the nest. We discuss that this is not appropriate in the transit case, in particular due to the reduced expected waiting time when several lines are included in the nest.

Considering these transit characteristics hence leads to a model which cannot be reduced to other simpler discrete choice models easily. We illustrate that we expect passengers' line choice to be more sensitive to service frequency and less sensitive to other factors, given similar user preferences. To estimate our model we establish the log likelihood function and its first and second order derivatives.

Our model formulation was motivated by the common practice in frequency-based transit assignment models to assign passengers to the shortest hyperpath in line with the "take whichever attractive line comes first" assumptions. We believe that our results, despite some shortcomings mentioned below, illustrate that these assumptions are often too simplistic. We present an approach to estimate the relative value of waiting time compared to on-board time in order to find (person-group specific) attractive sets. With such calibrated hyperpaths the model accuracy of transit assignment models might improve.

Smart card data from a local city in Japan allowed us to illustrate choice behaviour. In initial results we find that choice behaviour between passenger groups vary, in particular we find that older persons dislike waiting times relatively more compared to on-board travel time and other person groups. We can observe that some passengers form smaller choice sets than the attractive set proposed if evaluating travel time and waiting time equally. We acknowledge that our model results should be considered with some care as some passengers might perceive the service as schedule-based rather than frequency-based and hence our data might not fully fit our model assumptions. In other cities with less reliable bus services we would expect that passengers form larger choice sets, highlighting the need to estimate hyperpaths supply specific.

This research can be continued in several directions. Firstly, a detailed analysis comparing our estimation results with that of other choice models has not been carried out yet. Secondly, sensitivity to model assumptions such as service regularity assumptions should be investigated in more detail. Thirdly, we do not consider panel effects in our model nor do we consider that users change preferences depending on time of day. Fourthly, convergence and solution uniqueness should be further investigated. Finally, we believe with nowadays more smart card data becoming available our estimations could be repeated in more complex networks. This would allow us to also estimate the impact of possible different numbers of transfers to reach the destinations on the choice set formation.

## **Acknowledgements**

This research is supported by Grant-in-Aid for Challenging Exploratory Research from Japan Society for the Promotion of Science (No. 23656312, Project Leader: Fumitaka Kurauchi, 2011-2012). We further would like to thank the anonymous bus operator for allowing us to use their data.

## References

- Bekhor, S., Ben-Akiva, M.E. and Ramming, M.S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operation Research*, 144, 235-247.
- Bierlaire, M. (2006). A theoretical analysis of the cross-nested logit model. *Annals of Operations Research*, 144(1), 287-300.
- Bovy, P.H.L. (2009). On Modelling Route Choice Sets in Transportation Networks: A Synthesis. *Transport Reviews*, 29(1), 43-68.
- Dial, R.B., 1971. A probabilistic multi-path traffic assignment mode which obviates path enumeration. *Transportation Research*, 5, 83-111.
- Fiorenzo-Catalno, M. S. (2007). Choice Set Generation in Multi-Modal Transportation Networks. TRAIL Research Thesis. Available online from < [http://repository.tudelft.nl/assets/uuid:ef3b9c22-b979-4f46-9b02-110c82d67535/ceg\\_fiorenzo\\_20070605.pdf](http://repository.tudelft.nl/assets/uuid:ef3b9c22-b979-4f46-9b02-110c82d67535/ceg_fiorenzo_20070605.pdf) >
- Fonzone, A., Schmöcker, J-D., Kurauchi, F., Hemdam, S.M.H. (2012). Determinants of Hyperpath Choice in Transit Networks. Presented at 12<sup>th</sup> Conference on Advanced Systems for Public Transport (CASPT), July 2012.
- Fonzone, A., Schmöcker, J.-D., Ma, J.S. and Fukuda, D. (2013). Link-based Route Choice Considering Risk Aversion and Regret. *Transportation Research Records* 2322, 119-128.
- Frejinger, E., Bierlaire, M. and Ben-Akiva, M. (2009). Sampling of alternatives for route choice modelling.
- Friedrich, M., Hofsass, I. and Wekeck, S. (2001). Timetable-based transit assignment using branch and bound techniques. *Transportation Research Record*, 1754, 100-107.
- Gaudry, M. and Dagenais, M. (1979). The dogit model. *Transportation Research* 13B, 105-112.
- Jotisanakasa, A. and Polak, J. W. (2005) "Modelling learning and adaptation in route and departure time choice behaviour: Achievements and prospects", in *Integrated Land-Use and Transportation Models*, M. Lee-Gosselin & S. Doherty, eds., Elsevier, Oxfords, 133-157
- Hoogendorn-Lanser, S. (2006). A Rule-Based Approach to Choice Set Generation. *Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 2006*. Available online from <<http://www.cvs-congres.nl/cvspdfdocs/cvs06.22.pdf>>.
- Horowitz, J.L. and Louviere, J.J. (1995). What is the role of consideration sets in choice modelling? *International Journal of Research in Marketing*, 12, 39-54.
- Kato, H., Kaneko, Y. and Inoue, M. (2010) Comparative analysis of transit assignment evidence from urban railway system in the Tokyo Metropolitan Area. *Transportation*, 37(5), 775-799.
- Kurauchi, F., Schmöcker, J.-D., Fonzone, A., Hemdam, S.M.H., Shimamoto, H., and Bell, M.G.H., (2012). Estimation of Weights of Times and Transfers for Hyperpath Travellers. *Transportation Research Records* 2284, 89-99.
- Kurauchi, F., Schmöcker, J.-D. and Shimamoto, H. (2012). Understanding demand/supply variations on transit network using smartcard data. Paper submitted for presentation at the 5<sup>th</sup> International Symposium on Transportation Network Reliability, Hong Kong, December 2012.
- Manski, C. (1977). The structure of random utility models. *Theory and Decision* 8, 229-254.
- Morikawa, T. (1991). Destination choice analysis of vacation trips considering attractiveness of regions and probabilistic choice sets. *Infrastructure Planning Review*, 9, 117-124 (In Japanese).
- Nguyen, S., Pallottino, S. (1988). Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research*, 37, 176-186.
- Nguyen, S., Pallottino, S. and Gendreau, M. (1998). Implicit Enumeration of Hyperpaths in a Logit Model for Transit Networks. *Transportation Science*, 32(1), pp. 54-64.
- Nökel, K. and S. Wekeck (2009). Boarding and alighting in frequency-based transit assignment. Paper presented at 88<sup>th</sup> Annual Transportation Research Board Meeting, Washington D.C., January 2009.
- Nuzzolo, A. (2003). Schedule-Based Transit Assignment Models. In: *Advanced Modeling for Transit Operations and Service*; edited by William H.K. Lam and Michael G.H. Bell. Pergamon.



- Pelletier, M., Trepanier, M., and Morency, C. (2011). Smart Card Data Use in Public Transit: A Literature Review. *Transportation Research*, 19C, 557-568.
- Schmöcker, J.-D., Bell, M.G.H., Kurauchi, F. and Shimamoto, H. (2009). A Game Theoretic Approach to the Determination of Hyperpaths in Transportation Networks. *Selected Proceedings of the 18th International Symposium on Transportation and Traffic Theory (ISTTT)*. Hong Kong, July 2009.
- Spiess, H., Florian, M. (1989). Optimal Strategies: A new assignment model for transit networks. *Transportation Research*, 23B, 83-102.
- Swait, J. and Ben-Akiva, M. (1986). An analysis of the effects of captivity on travel time and cost elasticities. *Annals of the 1985 Int. Conference on Travel Behavior*, 16-19 April, Noordwijk, Holland, 113-128.
- Swait, J. (2001). Choice set generation within the generalized extreme value family of discrete choice. *Transportation Research*, 25B, 643-666.

## Appendix A. Derivation of Gradient and Hessian of the Log Likelihood Function

For brevity we derive in this appendix the derivatives of a log likelihood function consisting only of group specific parameters  $\beta_{an}$ . The derivatives with respect to the non-group specific parameters  $\beta_a$  can be obtained in a similar fashion.

To simplify our notation we firstly denote  $X_{akn(s)\tau(s)}$  as  $X_{aks}$  and similarly  $I_{kn(s)\tau(s)}$  as  $I_{ks}$ . Further, we simplify our notation of the log likelihood function (14) by introducing

$$A_{ks} = \exp(\mu I_{ks}) = \exp\left(\mu \sum_{a=1}^A \sum_{n=1}^N \eta_{ns} \beta_{an} X_{ksa}\right)$$

$$B_s = \sum_{r=1}^K \exp(\mu I_{rs}) = \sum_{r=1}^K \exp\left(\mu \sum_{a=1}^A \sum_{n=1}^N \eta_{ns} \beta_{an} X_{rsa}\right)$$

Where  $\eta_{ns}$  equals 1 if sample  $s$  belongs to person group  $n$  and 0 otherwise. Then

$$L^* = \sum_i \sum_s \delta_{is} \ln \left( \sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{A_{ks}}{B_s} \right) \right)$$

With this we can establish the gradient of our log likelihood function as:

$$\begin{aligned} \frac{\partial L^*}{\partial \beta_{an}} &= \sum_i \sum_s \delta_{is} \frac{1}{\sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{A_{ks}}{B_s} \right)} \cdot \frac{\partial}{\partial \beta_{an}} \left( \sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{A_{ks}}{B_s} \right) \right) \\ &= \sum_i \sum_s \delta_{is} \frac{1}{\sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{A_{ks}}{B_s} \right)} \cdot \left( \sum_{k \in K_i} \left( \frac{f_i}{\sum_{j \in C_k} f_j} \cdot \frac{\partial}{\partial \beta_{an}} \left( \frac{A_{ks}}{B_s} \right) \right) \right) \end{aligned} \quad (15)$$

Note that the partial derivatives of  $A_{ks}$  and  $B_s$  are:

$$\frac{\partial A_{ks}}{\partial \beta_{an}} = \eta_{ns} X_{ksa} \exp \left( \mu \sum_{a=1}^A \sum_{n=1}^N \eta_{ns} \beta_{an} X_{ksa} \right) = \eta_{ns} X_{ksa} A_{ks}$$

$$\frac{\partial B_s}{\partial \beta_{an}} = \sum_{r=1}^K \eta_{ns} X_{rsa} \exp \left( \mu \sum_{a=1}^A \sum_{n=1}^N \eta_{ns} \beta_{an} X_{ksa} \right) = \eta_{ns} \sum_{r=1}^K X_{rsa} A_{rs}$$

For simplicity, we define further

$$E_{ksan} = \frac{\partial}{\partial \beta_{an}} \left( \frac{A_{ks}}{B_s} \right) = \frac{\eta_{ns} X_{ksa} A_{ks} B_s - \eta_{ns} A_{ks} \sum_{r=1}^K X_{rsa} A_{rs}}{B_s^2} \text{ and}$$

$$F_{ks} = \frac{A_{ks}}{B_s}$$

Then

$$\begin{aligned} \frac{\partial E_{ksan}}{\partial \beta_{a'n'}} &= \frac{\frac{\partial}{\partial \beta_{a'n'}} \left( \eta_{ns} X_{ksa} A_{ks} B_s - \eta_{ns} A_{ks} \sum_{r=1}^K X_{rsa} A_{rs} \right) B_s^2 - 2 B_s \frac{\partial B_s}{\partial \beta_{a'n'}} \left( \eta_{ns} X_{ksa} A_{ks} B_s - \eta_{ns} A_{ks} \sum_{r=1}^K X_{rsa} A_{rs} \right)}{B_s^4} \\ &= \frac{\left( \eta_{ns} \eta_{n's} X_{ksa} X_{ksa'} A_{ks} B_s + \eta_{ns} \eta_{n's} X_{ksa} A_{ks} \sum_{r=1}^K X_{rsa'} A_{rs} - \eta_{ns} \eta_{n's} X_{ksa'} A_{ks} \sum_{r=1}^K X_{rsa} A_{rs} - \eta_{ns} \eta_{n's} A_{ks} \sum_{r=1}^K X_{rsa} X_{rsa'} A_{rs} \right) B_s^2 - 2 B_s \left( \eta_{n's} \sum_{r=1}^K X_{rsa'} A_{rs} \right) \left( \eta_{ns} X_{ksa} A_{ks} B_s - \eta_{ns} A_{ks} \sum_{r=1}^K X_{rsa} A_{rs} \right)}{B_s^4} \end{aligned}$$

$$\begin{aligned} \frac{\partial F_{ks}}{\partial \beta_{a'n'}} &= \frac{\partial A_{ks} / \partial \beta_{a'n'} \cdot B_s - A_{ks} \partial B_s / \partial \beta_{a'n'}}{B_s^2} \\ &= \frac{\eta_{n's} X_{ksa'} A_{ks} B_s - \eta_{n's} A_{ks} \sum_{r=1}^K X_{rsa'} A_{rs}}{B_s^2} \\ &= E_{ksa'n'} \end{aligned}$$

Therefore the Hessian of our log likelihood function takes following form

$$\begin{aligned}
\frac{\partial L^*}{\partial \beta_{an} \beta_{a'n'}} &= \sum_i \sum_s \delta_{is} \frac{\partial}{\partial \beta_{a'n'}} \left( \frac{\sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} E_{ksan}}{\sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} F_{ks}} \right) \\
&= \sum_i \sum_s \delta_{is} \frac{\left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} \frac{\partial E_{ksan}}{\partial \beta_{a'n'}} \right) \left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} F_{ks} \right) - \left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} E_{ksan} \right) \left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} \frac{\partial F_{ks}}{\partial \beta_{a'n'}} \right)}{\left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} F_{ks} \right)^2} \\
&= \sum_i \sum_s \delta_{is} \frac{\left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} \frac{\partial E_{ksan}}{\partial \beta_{a'n'}} \right) \left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} F_{ks} \right) - \left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} E_{ksan} \right) \left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} \frac{\partial F_{ks}}{\partial \beta_{a'n'}} \right)}{\left( \sum_{k \in K_i} \frac{f_i}{\sum_{j \in C_k} f_j} F_{ks} \right)^2} \quad (16)
\end{aligned}$$

We note though that  $\frac{\partial L^*}{\partial \beta_{an} \beta_{a'n'}} = 0$  whenever  $n \neq n'$ . Nevertheless, Eq. (16) appears difficult to characterise, in particular convexity cannot be established. Accordingly it is also not surprising that our estimation results suggest that local maxima exist.

# Appendix B. Detailed Service Characteristics of Lines Used In Case Study

Table 5. Frequency of each line.

Time	OD <i>a</i>			OD <i>b</i>						OD <i>c</i>	
	L1	L2	L3	L1	L2	L3	L4	L5	L6	L1	L2
5	4									1	
6	12	4	3	1	3	7	4	3	3	4	1
7	17		3	1	12	8	5	3	4	6	4
8	14		1	3	9	7	3	3	4	6	2
9	10		2		7	8	3	3	4	7	2
10	15		1	1	7	8	3	3	4	6	2
11	10		2		7	8	3	3	4	6	1
12	13		1		7	7	3	3	4	6	1
13	12		1		8	7	3	3	4	5	2
14	12		2	1	7	7	4	3	4	5	2
15	12		1	1	7	7	4	3	4	6	1
16	15		1	1	7	7	4	3	4	7	2
17	13		2	1	9	7	4	3	4	6	2
18	13		2	1	8	7	3	2	4	7	2
19	10		1	1	8	7	3	1	4	6	1
20	9		1		6	7	2		4	4	1
21	9				5	6	2		4	4	
22	3				4	1			3	3	

Table 6. Average travel time of each line by time of day

Time	OD <i>a</i>			OD <i>b</i>						OD <i>c</i>	
	L1	L2	L3	L1	L2	L3	L4	L5	L6	L1	L2
5	0:21										
6	0:23	0:18	0:26	0:10	0:14	0:15	0:13	0:19	0:28	0:10	
7	0:26		0:30	0:12	0:16	0:15	0:16	0:23	0:28	0:11	0:13
8	0:26		0:31	0:14	0:15	0:16	0:17	0:22	0:28	0:11	0:13
9	0:24		0:29	0:13	0:15	0:15	0:16	0:19	0:28	0:11	0:13
10	0:24		0:27	0:12	0:15	0:14	0:16	0:19	0:28	0:12	0:15
11	0:24		0:26		0:14	0:14	0:15	0:21	0:28	0:12	0:14
12	0:24		0:27		0:14	0:15	0:13		0:28	0:11	0:13
13	0:24		0:29		0:14	0:17	0:15	0:17	0:28	0:12	
14	0:25		0:26	0:12	0:13	0:16	0:14		0:28	0:11	0:12
15	0:24			0:13	0:16	0:23	0:16	0:22	0:28	0:11	0:12
16	0:24				0:15	0:22	0:16	0:17	0:28	0:12	0:14
17	0:26			0:13	0:16	0:17	0:12	0:17	0:28	0:11	0:14
18	0:24				0:13		0:14	0:25	0:28	0:12	0:14
19	0:22					0:15			0:28	0:12	0:15
20	0:18				0:14		0:13		0:28	0:10	0:12
21	0:21								0:28	0:10	0:10
22	0:21								0:28	0:10	