BACKGROUND FACTORS EXPLAINING TRAIN CHOICE IN EUROPEAN LONG-DISTANCE TRAVELLING

Goeverden, C.D. van, Delft University of Technology, Transport&Planning Department, c.d.vangoeverden@tudelft.nl

Arem, B. van, Delft University of Technology, Transport&Planning Department, b.vanarem@tudelft.nl

ABSTRACT

Long distance travel is not well developed in transport research. Though its volume is small in terms of journey numbers (1-2%), it is substantial in terms of person kilometres. A rough estimation suggests that the share of long distance travel could be about 50%. Moreover, long distance travel is growing relatively fast, in particular travelling by the energy-inefficient modes airplane and car. Both modes have by far the highest shares for either the medium distances (car) or the very long distances (airplane). The shares for train and bus are modest with 13% and 6% respectively. The European Commission observes an increasing imbalance in modal use and aims to increase the market shares of sustainable modes, in particular the train. Efficient policy for achieving this goal needs understanding about modal choice in long distance travelling. Most studies on modal choice examine the influence of modal attributes. However, other factors play a role as well. These can be indicated as background factors. The paper analyses the influence of background factors on modal choice. The focus is on train choice.

Based on data from the Dateline-project, a survey on long distance travelling by residents of the European Union, the impact of 17 background variables on train choice is examined by binary logistic regression. All variables proved to have statistically significant impacts, though the explanatory power varies largely between different variables. Most powerful variables are, in decreasing order: number of participants in journey, car ownership, size of the destination city, home country, the need to cross a national border, employment status, gender, size of the home city, and distance. Looking at the choice for other modes, number of participants in journey is the most important variable for explaining bus choice as well, while distance is most important for explaining both car and plane choice.

The conclusion is that background variables contribute substantially to modal choice in long distance travelling. Though such variables generally cannot be influenced by policy makers or train operators, they still can be useful for developing marketing strategies.

Keywords: modal choice, long distance travel, train, binary logistics

INTRODUCTION

Long distance travel is a small travel segment in terms of trip numbers but significant in terms of person kilometres. Nearly all kilometres travelled are by motorized modes, and within these modes to a large extent by the energy-inefficient car and airplane. Moreover, long distance travelling is increasing rapidly, in particular travelling by airplane. Table 1 gives an indication of the growth rates in the 1990s and modal shares in 2001. More recent data on holidays abroad of the Dutch (that are not comparable with the older data due to a change in the survey set-up) suggest that the increase in car and plane use as well as the decrease in coach use continue, while the decrease in train use has been stopped.

Table 1 – Developments in long distance travel

mode	annual growth rates in the 1990s		modal shares in 2001	
	international tourist arrivals[1]	holidays spent abroad by the Dutch[2]	journey numbers[3]	person kilometres[3]
car	+2.9%	+1.2%	66%	32%
bus, coach		-0.9%	6%	4%
train	+1.9%	-2.1%	13%	8%
airplane	+5.4%	+9.1%	12%	54%
other	+4.1%	-0.0%	2.0%	1.4%
total	+3.6%	+2.4%	100%	100%

1: source: Cabrini (2002)
2: source: Statline database of CBS
3: source: Dateline databases; the figures relate to journeys >100 km crow-fly

The European Commission has observed an increasing imbalance in the modal use in long distance travelling and aims to increase the market shares of sustainable modes, in particular the train (European Commission, 2001). Efficient policy for achieving this goal needs understanding about modal choice in long distance travelling. However, long distance travel is not well developed in transport research. Most research on travel behaviour focuses on regional and local problems, like road congestion and local air pollution. Moreover, the major principals commissioning travel research are regional and national authorities, and they are mainly interested in travel and related problems with regard to their own territories.

Probably most existing studies on long distance travelling have been carried out in the context of the evaluation of new high speed railway lines. Some focus on monitoring the impacts on travel demand (Bonnafous, 1987, Ettema *et al*, 1998, Steer Davies Gleave, 2006); they report growths in train use that vary from some tenths of percents to some hundreds of percents. Other studies include an analysis of factors inducing the observed changes in demand (Gonzáles-Savignat, 2004, Román *et al*, 2007). Assessments of modal choice in long distance travelling are also done in a more general context and then not restricted to the distances served by high-speed lines (300-700 km). Examples are Mandel *et al* (1997) and Zumkeller (2005) that are based on long distance travel surveys, and van Goeverden (2006) that is based on ticket sales of train operators. In analyzing the impacts of factors on modal choice, the focus is on the influence of modal attributes, characteristics of the relevant modes that define the level of service and costs for the traveller. The studies produce sometimes different results, but give the general impression that travel time, travel costs, and number of interchanges are important variables, while frequency of train services is unimportant in long distance travelling.

Only few studies examine the influence of background factors, like characteristics of the traveller and general features of the journey. Zumkeller (2005) assessed the influence of background variables on the generation of choice sets, i.e. sets of modal alternatives that are considered by the individual travellers. He found a large number of significant variables. Most significant were car ownership and number of destinations in a journey (all purposes), luggage volume (only business and leisure), and gender and distance (only business). Catalani (2006) observed substantial impacts of income, profession and travel purpose on the choice between three train types with different speeds and fares.

This paper intends to contribute to knowledge about the influence of background variables on modal choice in long distance travelling, with focus on the choice for the train. This knowledge gives an indication about the potential of the train and can be used by train operators when developing marketing strategies.

The organization of the paper is as follows. First, a short overview of volume and characteristics of long distance travelling is given. Then the analysis of the influences of background variables on train choice is described and, briefly, the choices for other long distance modes. Finally, the applicability of the results is briefly discussed.

VOLUME AND CHARACTERISTICS OF LONG-DISTANCE TRAVELLING

Dealing with the volume of long distance travelling is a difficult job. In Europe a large number of surveys on long distance travelling have been performed, and they give quite varying results (Kuhnimhof *et al*, 2009). Differences can partly be explained by different definitions of long distance travelling or by different survey periods (as we found, long distance travelling is increasing rapidly), but a different survey set-up is another important part of the explanation. Long-distance journeys are rather rare events, and an efficient way to capture them in a survey is to ask people to report the long distance journeys they made in a longer period, at least a few weeks. However this creates the problem of recall effects: respondents forget journeys that they made some time ago and do not report them. To which extent this effect leads to underregistration of long distance journeys is highly dependent on the survey set-up. The number of long distance journeys per person per day registered by different surveys ranges from 2 to 9.

The survey set-up also influences the type of unreported long distance trips. Kuhnimhof *et al* found that conventional mobility diary surveys perform well at shorter distances (< 400 km), while typical long-distance surveys give better results for the larger distances. Recently a new survey design is tested in the European KITE-project that seems to produce a rather complete registration of long distance journeys over the whole range of distances (Frei *et al*, 2010). Defining the minimum distance at 100 km crow-fly (which is the definition used by Eurostat), the estimated number of journeys per person per year is between 8 and 9 in 2008/2009 in three European countries. Assuming that people make on average about 1.5 round-trips per day, long distance journeys would account for about 1.5% of all round-trips.

The share in total distance travelled is considerably higher. Frei *et al* give no figures about that. We estimated the distance travelled by long distance journeys by analyzing data of another European project, Dateline. The Dateline-project was a survey on long distance travelling by Europeans in 2001/2002. By now it is the only EU-wide complete long distance travel survey. In this project the minimum distance is defined as 100 km crow-fly as well. The survey periods are long: one year for holiday travel (journeys for leisure with at least 4 overnight stays), three months for business and other private travel (including short holidays), and four weeks for commuting. Due to the long periods the recall effects are large, especially in holiday travel (Hautzinger *et al*). We estimated expansion factors for forgotten holidays in the first nine months of the survey-periods and found that forgotten journeys are highly concentrated on the shorter distances. Using these factors, the calculated number of long distance journeys per person per year is 3.3 for all countries involved, while similar figures apply for the two countries that are examined in both the Dateline and the KITE surveys. The large difference between the Dateline and KITE results is partly due to the earlier survey period. Assuming an annual growth rate of 2.5-3%, the number of long distance journeys could have increased by about 20% between the two survey periods. The KITE-design would then have produced 6.5% to 7% in the Dateline-period, still about double of the Dateline figures. The ratios between the figures of both surveys are relatively large for the larger distances (> 500 km) suggesting a relatively large underregistration by Dateline for these distances. However, we hypothesize that this observation can be ascribed to a relatively large growth of long distance travelling on the longer distances between the two survey periods, and assume that the underregistration in Dateline is not dependent on distance.

The number of kilometres per European resident per year made by long distance travelling is, if directly calculated with the Dateline data, 3,400 (crow-fly). Assuming that the underregistration is 50% for the whole range of distances and that the average detour factor is 1.25, the total distance travelled per person per year in 2001 on long-distance trips would have been 8,500 km. Total distance travelled by Dutch persons on short-distance trips (< 100 km crow-fly) in 2001 was about 10,000 km per person per year (Dutch NTS). If this was also the European average, over 40% of person kilometres in 2001 should have been made by long distance travellers. Today long distance travelling may account for 50% or more of all travelled kilometres. However, due to the uncertainties in constructing these figures, the margins are large.

According to the Dateline-data, 31% of long distance trips are made for holidays, 15% for business travel, 39% for other private purposes, and 15% for commuting, either for work or education. Holiday travel is relatively large on the long distances, while other private and commuting are rather concentrated on the shorter distances. This is shown by Figure 1 that presents the shares of the purposes for the middles of the distance classes with a width of 100 km in the range from 100 to 1500 km. In this range the train is a competitive mode.

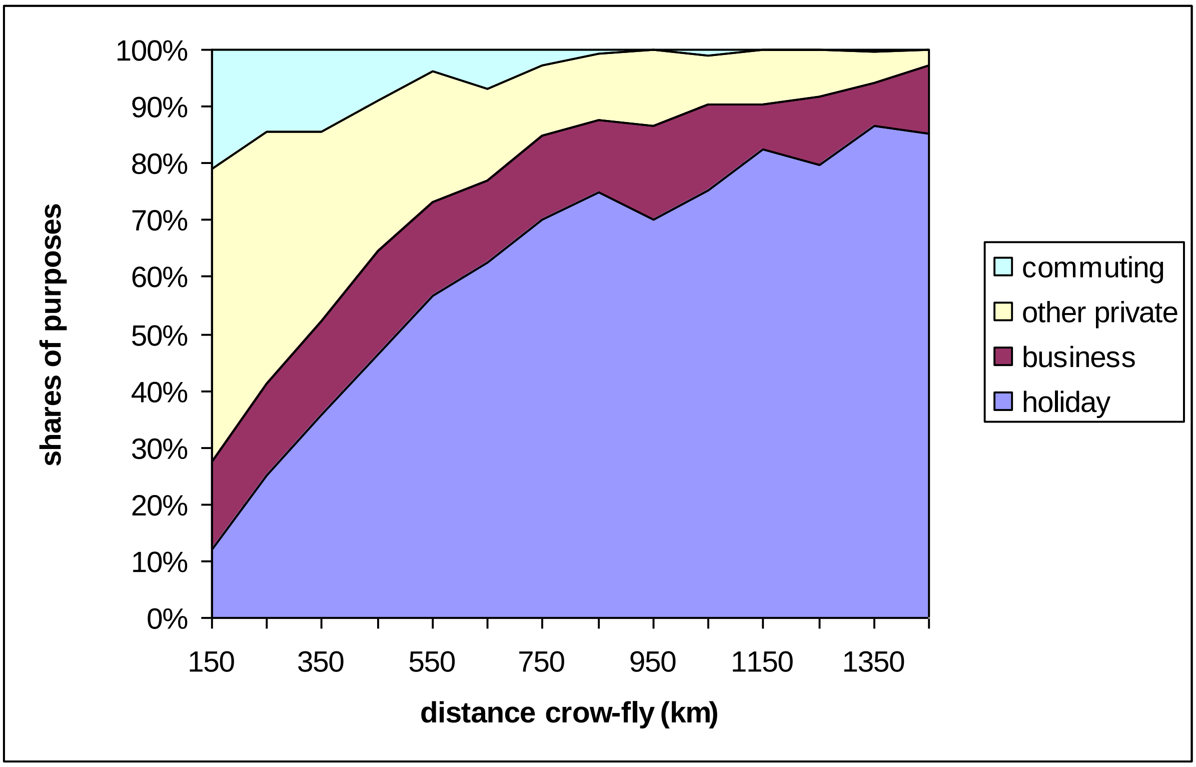

Figure 1 – Shares of travel purposes by distance

Table 1 showed that the car is the most frequently used mode for long distance journeys (66%), followed by train and airplane (13% and 12%). The bus has a smaller but still substantial share (6%). The shares are strongly related with distance, in particular for car and airplane (Figure 2). Both modes seem to be communicating vessels. The train and bus shares are more stable. Interestingly, over 1000 km the bus share exceeds the train share. One explanation is that the bus performs well in international journeys while the train has a strong position in domestic journeys. Selecting domestic journeys, the train has a substantial higher share than the bus for all distances. Selecting international journeys, the share of the bus is slightly higher than the share of the train for all distances up to 1000 km, and start to become substantially higher at longer distances.

The train's share is relatively high in commuting and business travel (22% and 18% respectively) and low in holiday travel and journeys for other private purposes (9% and 11% respectively). The car has high shares, over 50% for all modes. The airplane is mainly used for business and holiday travel, while the share of the bus is rather constant at 6-7% for all purposes except business, where its share is only 3%.

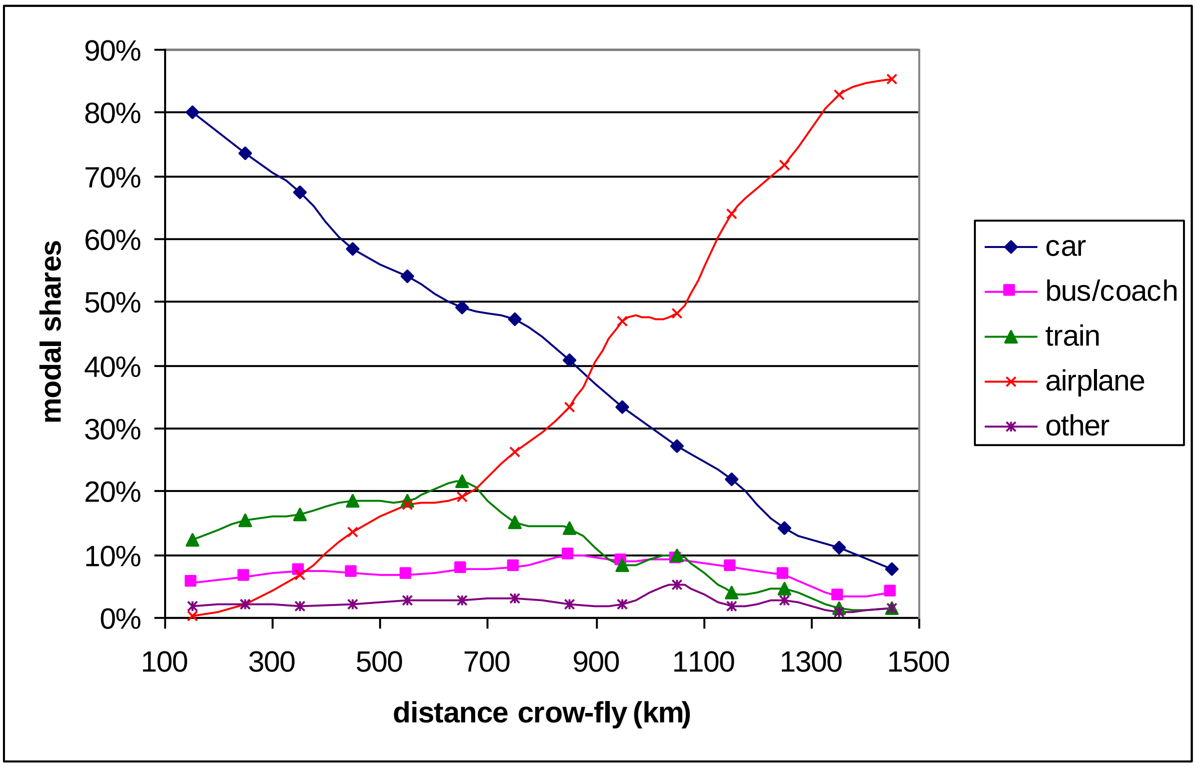

Figure 2 – Modal shares by distance

EXPLAINING FACTORS FOR MODAL CHOICE

This section discusses the influence of background factors on train choice and pays in addition briefly attention to the influence on the choice for other long-distance modes, i.e. bus/coach, airplane and car. The analysis is based on data from the Dateline-survey. These data include information on household level, person level, journey level, and trip level. We investigated the influence of variables on the first three levels. Next variables are included in the analysis:

- household characteristics:
  - number of persons in household;
  - children aged 5-15 or less than 5 in household;
  - car ownership;
  - home country;
  - size of home city, including distinction between living in the periphery and core of urban agglomerations; we added information about this variable to the Dateline data.
- person characteristics:
  - age class;
  - gender;
  - employment status;
  - driver's licence.
- journey characteristics:
  - journey purpose;
  - distance class;
  - destination country;

- size of destination city, including distinction between periphery and core of urban agglomerations;
- indication whether journey is domestic or international;
- number of participants in journey;
- number of nights in journey;
- season.

Income is not included because information about this variable is not provided by Dateline. Two other variables that have significant impacts according to Zumkeller (2005), luggage and number of destinations visited during a journey, are not included as well. Information about luggage is missing in Dateline. The number of destinations is provided by Dateline. Still we decided not to include this variable in the analysis because we found no significant effect on train choice and the variable is missing for 25% of the observations. In the estimation procedure all observations with at least one missing value for the variables included are omitted; inclusion of number of destinations would reduce the number of analyzed observations considerably.

Not all of the listed variables are pure background variables. Some of the variables have strong correlations with the modal qualities, in particular the sizes of the home and destination cities, and the distance class. Therefore, the results of the analysis will give some indication regarding the influence of modal qualities.

Only journeys where surface modes are feasible alternatives are selected. Journeys that have to overcome an important sea barrier are excluded from the analysis. Examples are journeys to Ireland and those from southern Finland in south-western direction. Additionally, we excluded all journeys with a crow-fly distance of more than 1500 km. For longer distances the plane has a near monopoly and the shares of train and other surface modes are small.

The method

The influence of the variables is examined simultaneously using binary logistic regression. Parameters of the following model are estimated:

$$p(m) = \frac{\exp(b_0 + \sum_k (b_k * x_k))}{1 + \exp(b_0 + \sum_k (b_k * x_k))} \qquad (1)$$

where:

$p(m)$: probability that mode m will be chosen (values are between 0 and 1)

x_k: k[th] explanatory variable

b_k: parameter describing the influence of x_k

b_0: constant

This model describes the influence on the choice for just one mode. Choices for different modes must be analyzed separately. The analyses produce parameter values for the explanatory variables, results about the significance of the variable, and a ranking of variables with respect to their influence. The explanatory variables can be both continuous and categorical. A categorical variable with n classes is split up into n-1 dichotomous variables where the n^{th} class is the reference class. Then the model estimation produces n-1 parameter values, each describing the difference between the influence of a certain class and the reference class.

Data considerations

The analyses are based on data from the Dateline-databases. The Dateline-survey is performed in 16 European countries: the 15 EU-countries at the time (2001) and Switzerland. Two kinds of data collection have been used: collection on household level and collection on person level. In the household survey a sample of households is selected and data about all members and their journeys are collected. In the person survey a sample of persons is selected and data about these persons and their journeys are collected. The household survey is performed in Austria, Flanders, Germany, Ireland, Italy, Luxembourg, .the Netherlands, Sweden and the United Kingdom. The person survey is performed in the Belgian regions Wallonia and Brussels, Denmark, Finland, France, Greece, Portugal, Spain and Switzerland.

Unlike the household survey, the person survey gives no information about accompanying household members at the registered journeys. Consequently, data of the person survey can not be used for analyzing the influence of the number of participants in a journey. This is a serious problem because the number of participants proves to be one of the most influential variables, even the most influential for train and bus choice. Surprisingly, the person survey does include information about the number of accompanying persons that are no member of the household. Moreover, the number of participating household members is asked for on the enquiry forms, but the answers seem not to have been coded. Because we disliked to throw away data about half of the countries, and also wanted to analyse the influence of the number of participants, we decided to do two analyses: one using all data and excluding the variable ‘number of participants’, and one using only the household survey and including the number of participants.

The observation that people sometimes travel together (that is valid for most of the long distance journeys) creates a second, methodological problem. If persons with different characteristics travel together, the characteristics of which person should be selected for the analysis of personal variables? Preferably, it is the person who decides about which mode will be used. However, who is it? We did three estimations with different selections of the assumed representative person. First, it was assumed that the oldest fellow-traveller in the age of 20-65 makes the modal choice. If there is no traveller in that age class, the oldest traveller was selected. If the oldest traveller was younger than 18, the journey was excluded from the analysis. This estimation could only be done for journeys in the household survey because in the person survey information about other participants is missing. Secondly, the persons in the person survey and their journeys were selected. In fact, this means a random selection of representative persons. Thirdly, single journeys were selected, i.e. journeys where persons travel alone. Here the problem is not existent, assuming that single travellers make the modal choice by themselves. A possible exception is when children travel alone; in that case the choice might be made by the parents. Therefore, single journeys were selected made by persons aged 18 years or older. Because single journeys cannot be identified in the person survey, only observations from the household survey could be used. The three estimations produced roughly the same results regarding the influence of the personal variables. However, the results of the third estimation were most significant and its impacts were largest, despite a strongly reduced sample size. The first estimation was second best. In this paper we present only the results of the third estimation of the influence of personal variables, based on the choice of single travellers.

Summarizing, due to missing information in the person survey and problems regarding the assessment of the influence of personal variables in group travel, three kinds of estimations are performed:

- estimation of the influence of all variables except for number of participants and the personal variables, using all data,
- estimation of the influence of all variables except for the personal variables, using all data of the household survey,
- estimation of the influence of all variables except for number of participants using only data of single journeys in the household survey; inclusion of number of participants in this analysis is not useful because this variable has always the value ‘1’.

Results regarding train choice

The results of the three estimations of the influencing factors for train choice are presented in the large table in the appendix. We decided not to present the estimated parameter values, but instead the impacts in terms of percentage differences between the probabilities for train use in the reference situation and the situation that the value of a variable changed, either by an increase by 1 unit (continuous variable) or by a change to another class (categorical variable). The reason is that such a percentage gives a more intuitively understandable representation of the impacts. The presented impacts are equal to $(\frac{p(m)_1^v}{p(m)_0} - 1) * 100$ where $p(m)_1^v$ and $p(m)_0$ are the probabilities for train use after and before the change in variable v. The ratio $\frac{p(m)_1^v}{p(m)_0}$ is calculated as $\frac{\exp(b_v)}{1 + (\exp(b_v) - 1) * p(m)_0}$; b_v is the value of the parameter representing the influence of variable v. In order to give the reader the opportunity to calculate the original parameter values, the values of $p(m)_0$ are included in the table. These probabilities relate to the unweighed sample.

An interesting outcome is that all examined explanatory variables have statistically significant impacts. Only in the estimation for single travellers that is based on a relatively small sample, two variables proved to be not significant (children in household and season). Next the variables are discussed, ranked to decreasing importance in explaining train use.

- The most important variable is number of participants in the journey. If more persons travel together, the probability of train use is about 60% lower than when one person travels alone.
- Second in importance is car ownership of the household. Car ownership reduces the probability of train use by about 70% when several persons travel together. For single travellers the reduction is smaller, approximately 50%. It might be that price plays here a role: travelling alone by car is relatively expensive. Another explanation is a lower availability of household cars for single travellers than for group travellers. The probability of a car being used by another member of the household is relatively large for single travellers.
- Third in importance is the size of the destination city and the location in either the core or the periphery of an urban agglomeration. The larger the city, the higher is the probability that the train will be chosen. The difference between a very large city (more than 5 million inhabitants) and a small city or rural area is a factor 3 to 4. People travelling to the core of an agglomeration are more inclined to choose the train than those travelling to the periphery.
- There are substantial differences between residents of different countries. The inclination for train use is lowest for the Irish and Greek, followed by the Portuguese and Spanish. Residents of Finland and Switzerland, and to a lesser extent, those of Belgium, exhibit a high preference for the train.
- Crossing national borders proves to reduce train use substantially. The probability that the train is chosen for an international trip is only about 30% of that for choosing the train for a similar domestic trip.
- The most influencing personal variable is employment status. Train use is low by persons who work full-time and high by those who are at school or university, looking for work, or belong to the "other" not specified category.
- Gender also has a substantial influence. Women are significantly more inclined to travel by train than men.
- Just like the size of the destination city, the size of the home city influences train use. Again, train use increases as the city size increases. However, the relation is weaker than for destination cities. The difference between living in the core of a very large agglomeration and a small city/countryside is roughly a factor 2. Living in the core results in higher train use than living in the periphery of an agglomeration.
- Not surprisingly, distance is important. The probability for train use is highest between 600 and 900 km and lowest in the highest of the analysed classes (up to 1500 km). It is also rather low between 100 and 200 km.
- Train choice is influenced by the country of destination. This variable correlates with home country, because in the case of domestic journeys home country and destination country are equal. Still, there are substantial differences between both variables for a number of countries. Most striking is that the two extremes in the home country list (Ireland with the lowest probability and Finland with the highest) are the opposite extremes in the destination country list. These should be due to strongly deviating behaviour of international travellers to these countries. In the case of both countries many incoming international journeys are excluded from the analysis because they have to overcome an important sea barrier. Train use must be very

high for the remaining international journeys to Ireland (these include only journeys made by British residents of Northern Ireland) and very low for the remaining journeys to Finland (journeys from the north of Sweden, or from other Europe to the north of Finland). A smaller discrepancy can be observed for the United Kingdom that appears not so attractive for international train users. Low train use can also be observed in travelling to non-Dateline countries, Norway and the Eastern European countries.

- A driver's license decreases the probability for train use.
- If there are children in a household, train use is somewhat reduced, in particular when the children are young.
- Persons travelling for education are much more inclined to choose the train than those who travel for other purposes. Interestingly, the 'all'-analysis (using all data and excluding the variable 'number of participants') and the 'HH'-analysis (using only data from the household survey and including number of participants), give opposite results regarding the probability for train use by business travellers compared to holiday travellers. The explanation is that number of participants and purpose are correlated. Unlike holiday travellers, business travellers frequently travel alone. If the number of participants is left out from the analysis, its influence will be reported by other correlated variables including purpose, and raise the probability for train choice by business travellers. The table in the appendix gives no results regarding commuting. Commuting journeys were skipped in the estimation procedure because some variables are always missing for commuting journeys (number of nights in journey, season, number of participants). We did an additional analysis where these three variables were left out and found a high probability for train use by commuters (ca 100% higher than for holiday travellers). One should note here that inclusion of number of participants would lower this figure because commuters frequently travel alone, just like business travellers.
- Regarding age, there is no age class that differs significantly from the reference class (18-25). Significant differences within other pairs of classes can neither be found. Still the variable as a whole is significant. Train use is relatively high for the classes 18-25 and 55-65 and low for persons between 25 and 45 and older than 75.
- There are small differences in train use between the seasons. The winter is favourable for train use, the summer unfavourable.
- Enlarging the duration of the journey increases the probability for train use somewhat.
- Finally, the number of persons in the household affects train choice. Just like for the travel purpose business, the 'all'-analysis and the 'HH'-analysis give opposite results. According to the first, the probability for train choice decreases, according to the latter, the probability increases. Again, the opposite results can be explained by correlation with the number of participants. Persons from small households will on average travel with less accompanying persons. If the number of participants is left out from the analysis, small households, in particular one-person households, will attract the large positive influence on train choice by single travellers and convert the 'true' positive relationship between household size and train use into a negative one.

The choice of other modes

Analyzing the separate choices for bus/coach, airplane and car shows that each mode has its own mix of important explanatory variables and direction of the influences. Most important variable for the bus is number of participants in the journey, just like for the train. Single travellers and larger groups are most inclined to choose the bus. Next most important variables for the bus are, in decreasing importance: purpose (high probability for holiday and education, very low probability for business), employment status (high probability by students, retired persons and persons that principally do home duties), car ownership (high probability for non-owners), country of destination (high probability for Spain, Norway and Eastern Europe), gender (high probability for women), and (inter)national character of the journey (high probability for international journeys).

Distance is the most important variable for explaining both plane use and car use. Increasing distance increases the probability for plane choice and decreases the probability for car choice (see also Figure 2). Second in importance for the plane is journey purpose; the probability is extremely high for business and commuting. Next in importance for the plane are: destination city size (increasing probability for increasing city size, in particular high probabilities for the cores of agglomerations), home country (high probability for residents of Luxemburg, the United Kingdom and Switzerland), and destination country (high probability for those travelling to one of the Scandinavian countries, Greece, and the United Kingdom). Unlike for bus choice, household and personal characteristics are of minor importance for plane choice. Car ownership, driver's license, and gender have even no significant effect.

Most important explanatory variables for car choice are, besides distance: number of participants (high probability for groups for 2-6 persons), car ownership and driver's license (high probability for car and license owners), destination city size (decreasing probability for increasing city size, in particular low probabilities for the cores of agglomerations), and home country (car use is relatively high for residents in Germany, the Benelux countries, Denmark, Spain and Greece).

DISCUSSION

Background variables contribute substantially to modal choice in long distance travelling. In contrast to variables that define the service level of train and alternative modes, background factors can generally not or just marginally be influenced by train operators or policy makers. The usefulness of knowledge about their influences is then restricted to predicting future train patronage. Nevertheless, looking at the list of significant variables for train choice, a few variables suggest opportunities for train operators to increase the number of their long distance travellers substantially. The first is the most influencing variable, number of participants in a journey. Train operators cannot change these numbers, but they can try to raise the train's share in group travelling, for instance by offering substantial fare reductions for the second, third, etc. traveller. Another variable that seems to give good opportunities for increasing the train's share substantially is the domestic/international character of a journey. It is unclear why crossing a national border has so a large impact on train choice. Possibly, part of the explanation is out of reach of the train operators. However, the relatively poor conditions for travelling internationally by train will certainly contribute to it. Compared to domestic travelling, travel information is less well provided, buying tickets can be much more complicated and time consuming, fares are generally higher, and train supply has a lower quality due to a low density of the international train network (forcing passengers to make larger detours and more interchanges) and low frequencies. A policy to reduce or level out the conditional differences between international and domestic travelling might increase international train use and the train's share in long distance travel substantially.

Sometimes, the strategy of public transport operators is to focus on the market where there position proved to be strong and neglecting the other markets of non-frequent users. Following this strategy, the results of the study are also useful. Then train operators should not try to attract persons travelling in a group, but just direct their strategy to attracting more single travellers. And instead of reducing the gap between international and domestic train services, they could concentrate on further improvement of the domestic supply, thus increasing the gap with international services. Such a strategy may sometimes be more efficient in enlarging the total market share, though it leaves a large potential market unexplored.

The results of the modal analyses of this paper give cause to recommendations on data collection in the field of long distance travelling. The lack of information about number of participants in a journey in the survey on person level is a serious deficiency in the Dateline data. This variable as well as the other variables that proved to influence the choice for one of the long-distance modes substantially should be included in each long distance survey. The most important variables that we detected in our analysis are: number of participants in journey, travel distance, features of the locations of residence and destination, travel purpose, car ownership, driver's license, employment status, and gender. The literature reports in addition income, luggage volume and number of destinations visited as highly influencing variables.

REFERENCES

Bonnafous, A. (1987). The regional impact of the TGV. Transportation. 14, 127-137.

Central Bureau of Statistics (CBS), Statline database, provided on internet at URL http://www.cbs.nl/nl-NL/menu/cijfers/statline/toegang/default.htm.

Cabrini, L. (2002). Trends and Challenges for the Tourism Industry in Europe. Speech at the CERTS Annual Conference on Hospitality and Tourism Industry in Europe, Dublin, November 27.

Catalani, M. (2006). The impact of the high speed system on the Naples-Rome railway link. Proceedings of the European Transport Conference, PTRC, London

Dateline project, reports and databases, provided on internet at URL http://cgiserver.uni-mb.si/elmis/.

Ettema, D., N. Cohn and F. Savelsberg (1998). Monitoring the effects of the Thalys high speed train. Proceedings Seminar G of the European Transport Conference, PTRC, London.

European Commission (2001). White paper, European transport policy for 2010: time to decide, Office for official publications of the European Communities, Luxemburg

Frei, A., T. Kuhnimhof and K.W. Axhausen (2010). Long distance travel in Europe today: Experiences with a new survey. Proceedings of the 89th Annual meeting of TRB, Washington D.C.

Goeverden, C.D. van (2006). Modelling capricious behaviour of international train users. Proceedings of the 9th TRAIL Congress, TRAIL Research School, Delft.

Gonzáles-Savignat, M. (2004), Competition in Air Transport, The case of the High Speed Train. Journal of Transport Economics and Policy. 38, part 1, 77-108.

Hautzinger, H., W. Stock and J. Schmidt (2005). Erstellung von Microdatenfiles zu Ein- und Mehrtagesreisen auf Basis der Erhebungen MiD und DATELINE, Schlussbericht. Institut für angewandte Verkehrs- und Tourismusforschung e.V., Heilbronn/Mannheim

Mandel, B., M. Gaudry and W. Rothengatter (1997). A disaggregate Box-Cox Logit mode choice model of intercity passenger travel in Germany and its implications for high-speed rail demand forecasts. The Annals of Regional Science. 31, 99-120.

Kuhnimhof, T., R. Collet, J. Armoogum and J.L..Madre (2009). Generating internationally comparable figures on long-distance travel for Europe. Transportation Research Record. 2105, 18-27.

Román, C., R. Espino and J.C. Martin (2007). Analyzing competition of the high speed train with alternative modes. The case of the corridor Madrid-Zaragoza-Barcelona, paper submitted to the 11th WCTR, Berkeley.

Steer Davies Gleave (2006). Air and rail competition and complementarity, case study report, prepared for the European Commission DG Energy and Transport, London.

Zumkeller, D. (2005). Die intermodale Vernetzung van Personenverkehrsmitteln unter Berücksichtigung der Nutzerbedürfnisse (INVERMO), Schlussbericht, Institut für Verkehrswesen, Universität Karlsruhe.

APPENDIX

The results of the analysis by binary logistic regression on train choice are summarized in Table A1. The table includes results of three estimations: an estimation using the whole Dateline database excluding the variable 'number of participants in journey' as well as personal characteristics (indicated by "All"), an estimation using only the household survey including the number of participants and excluding personal characteristics (indicated by "HH"), and an estimation using only single travellers in the household survey including the personal characteristics (indicated by "ST"). For each variable or class of (categorical) variables the impact on train choice, significance indication and observation numbers are presented. If a variable is numeric, its impact is the percentage increase in the probability for train use if the variable increases with one unit. So, according to the 'HH'-analysis, adding one person to a household would increase the probability by 5% (last variable in the table). If a variable is categorical, the impact refers to the difference in probabilities between a certain class of the variable and a predefined reference class. Each categorical variable has one reference class, indicated by *ref*. For example, if travel distance is between 500 and 600 km, according to the 'All'-analysis the probability of choosing the train is 73% higher than for distances between 100 and 200 km, the reference class. If the impact of a class of a categorical variable is considered as not significant (at a 5% level; the significance indicator exceeds 0.05), only a dash is shown in the impact-column. If a variable has no significant influence at all, "n.s." is indicated in the column. This is only valid for a few variables in the 'ST' analysis. Because of lack of space, only the observations used for the 'All'- and 'ST'-analyses are shown as far as variables are included, except for the number of participants which influence is only analyzed with the 'HH'-data.

The variables are ordered in decreasing importance regarding the contribution to the explanation of train choice, based on the sequence of inclusion of variables in the model in the stepwise estimation procedure. Though the three analyses produced roughly the same ranking, there were minor differences. In that case, the table displays the ranking of the 'All'-estimation.

Table A1 – Estimated impacts on train choice

Variables	Classes	Impact			Significance			Observations	
		All	HH	ST	All	HH	ST	All	ST
Number of participants in journey	*ref.*: 1 participant							6585*	
	2 participants		-59%			0.00		8698*	
	3-6 participants		-65%			0.00		7241*	
	7-29 participants		-53%			0.00		1186*	
	>= 30 participants		-74%			0.00		283*	
Car ownership househ.	*ref.*: no car							7012	1140
	car	-74%	-68%	-49%	0.00	0.00	0.00	49724	5296

Variables	Classes	Impact			Significance			Observations	
		All	HH	ST	All	HH	ST	All	ST
Destination city size	*ref.*: < 100,000 inh.							35058	2822
	1-500,000 core	72%	75%	60%	0.00	0.00	0.00	8476	1124
	periphery	54%	-	-	0.00	0.29	0.65	293	54
	.5-1,000,000 core	95%	88%	60%	0.00	0.00	0.00	2659	574
	periphery	74%	69%	-	0.00	0.01	0.72	206	35
	1-2,000,000 core	134%	113%	64%	0.00	0.00	0.00	3535	784
	periphery	58%	-	-	0.00	0.26	0.78	276	56
	2-5,000,000 core	153%	131%	75%	0.00	0.00	0.00	3009	593
	periphery	66%	-	-	0.01	0.29	0.14	215	53
	>= 5,000,000 core	299%	274%	176%	0.00	0.00	0.00	2775	320
	periphery	178%	-	-	0.00	0.12	0.64	234	21
Home country	*ref.*: Austria							1244	324
	Belgium/Flanders	54%	76%	-	0.00	0.01	0.78	1706	55
	Denmark	-			0.31			1504	0
	Finland	121%			0.03			1597	0
	France	-			0.92			9451	0
	Germany	-	-	-	0.77	0.15	0.24	9560	2715
	Greece	-76%			0.00			1238	0
	Ireland	-87%	-86%	-92%	0.00	0.00	0.00	134	36
	Italy	-	-	-	0.81	0.58	0.25	3192	706
	Luxemburg	-	-	-	0.24	0.28	0.13	174	40
	The Netherlands	-	-	-48%	0.56	0.62	0.01	3249	673
	Portugal	-57%			0.00			3269	0
	Spain	-42%			0.00			13479	0
	Sweden	-	-	-	0.23	0.75	0.41	1514	441
	United Kingdom	-	35%	-	0.34	0.04	0.11	4680	1446
	Switzerland	116%			0.00			745	0
(Inter)national journey	*ref.*: domestic							44000	4873
	international	-69%	-71%	-75%	0.00	0.00	0.00	12736	1563
Employment status	*ref.*: work full time								4184
	work part time			51%			0.00		449
	at school, univ.			62%			0.00		318
	retired			24%			0.03		1130
	home duties			-			0.36		169
	looking for work			68%			0.00		109
	other			88%			0.00		77
Gender	*ref.*: male								4008
	female			43%			0.00		2428

Variables	Classes	Impact			Significance			Observations	
		All	HH	ST	All	HH	ST	All	ST
Home city size	*ref.*: < 100,000 inh.							29897	3358
	1-500,000 core	31%	26%	16%	0.00	0.00	0.04	7987	789
	periphery	-	-	-	0.08	0.07	0.18	1415	169
	.5-1,000,000 core	24%	29%	-	0.00	0.00	0.34	2851	413
	periphery	-	-	-	0.48	0.29	0.49	1182	185
	1-2,000,000 core	55%	44%	27%	0.00	0.00	0.01	2605	394
	periphery	25%	41%	31%	0.00	0.00	0.03	1825	194
	2-5,000,000 core	40%	49%	47%	0.00	0.00	0.00	2853	504
	periphery	39%	-	-31%	0.00	0.36	0.04	1372	163
	>= 5,000,000 core	135%	109%	49%	0.00	0.00	0.00	2538	182
	periphery	76%	67%	-	0.00	0.00	0.42	2211	85
Distance (crow-fly)	*ref.*: 100-200 km							18601	2337
	200-300 km	33%	24%	36%	0.00	0.00	0.00	10634	1291
	300-400 km	50%	34%	35%	0.00	0.00	0.00	8245	809
	400-500 km	75%	30%	24%	0.00	0.00	0.01	5675	666
	500-600 km	73%	30%	-	0.00	0.00	0.30	3963	400
	600-700 km	121%	70%	-	0.00	0.00	0.09	2884	248
	700-800 km	127%	52%	-	0.00	0.00	0.61	2077	170
	800-900 km	118%	90%	-	0.00	0.00	0.43	1271	106
	900-1000 km	53%	-	-	0.00	0.22	0.50	887	116
	1000-1100 km	48%	43%	-68%	0.00	0.03	0.02	705	73
	1100-1200 km	-	-	-	0.08	0.72	0.98	524	64
	1200-1300 km	-	-	-	0.57	0.55	0.20	483	65
	1300-1400 km	-61%	-	-	0.01	0.05	0.15	423	44
	1400-1500 km	-	-	-	0.17	0.20	0.09	364	47
Destination country	*ref.*: Austria							1871	376
	Belgium/Flanders	-	-	-	0.47	0.77	0.12	833	93
	Denmark	-	-51%	-	0.92	0.04	0.87	1067	29
	Finland	-76%	-		0.00	1.00		1551	0
	France	-22%	-	-	0.02	0.83	0.28	11391	244
	Germany	-23%	-33%	-	0.01	0.00	0.65	7230	2247
	Greece	-	-	-	0.63	1.00	1.00	1230	12
	Ireland	131%	87%	127%	0.00	0.03	0.04	206	55
	Italy	-	-	-	0.41	0.16	0.71	4288	724
	Luxemburg	-	-	-	0.42	0.96	0.86	94	8
	The Netherlands	-	-	55%	0.96	0.15	0.03	1765	462
	Portugal	-50%	-	-	0.00	1.00	1.00	3256	6
	Spain	-57%	-74%	-	0.00	0.00	0.49	13861	98
	Sweden	-	-43%	-	0.23	0.02	0.87	1459	413
	United Kingdom	-53%	-62%	-	0.00	0.00	0.34	4249	1328
	Switzerland	30%	51%	94%	0.04	0.01	0.00	832	94
	Norway	-49%	-	-	0.03	0.18	0.46	245	23
	other Europe	-48%	-42%	-	0.00	0.00	0.59	1308	224

Variables	Classes	Impact			Significance			Observations	
		All	HH	ST	All	HH	ST	All	ST
Driver's license	*ref.*: no license								617
	license			-39%			0.00		5819
Children in household	*ref.*: no childr. <15							39737	5176
	children 5-<15	-21%	-14%	n.s.	0.00	0.02		10407	774
	children <5	-34%	-27%		0.00	0.00		6592	486
Purpose	*ref.*: holiday							34386	1665
	other leisure	-	-	-	0.55	0.54	0.11	13207	1520
	business	24%	-21%	-	0.00	0.00	0.39	6602	2894
	education	130%	99%	69%	0.00	0.00	0.04	190	37
	other	18%	-	-	0.01	0.13	0.30	2351	320
Age	*ref.*: 18-<25								356
	25-<35			-			0.24		1060
	35-<45			-			0.13		1410
	45-<55			-			0.77		1408
	55-<65			-			0.94		1272
	65-<75			-			0.81		627
	>= 75			-			0.15		303
Season	winter	19%	12%		0.00	0.05		9341	1111
	spring	-	-	n.s.	0.83	0.15		12897	1768
	summer	-9%	-		0.00	0.18		23343	1979
	ref.: autumn							11155	1578
Number of nights in journey		0.4%	0.7%	0.7%	0.00	0.00	0.05	56736	6436
Number of persons in household		-3%	5%	6%	0.04	0.01	0.01	56736	6436
constant		-59%	-35%	-58%	0.00	0.00	0.00	56736	6436
initial probability ($p(m)_0$)		0.121	0.135	0.254					
R^2 (Nagelkerke)		0.21	0.21	0.23					

* Observation numbers from the household enquiry