Traffic flow: Scenario, traffic forecast and analysis of traffic on the TEN-T, taking into consideration the external dimension of the Union.

TRANS-TOOLS version 2; Model and Data Improvements

Co-ordinator

Tetraplan A/S
Copenhagen, Denmark

Partners

BMT Transport Solutions
Hamburg, Germany

DTU-Transport, Department of Transport, Technical University of Denmark
Lyngby, Denmark

Institute for Regional Research, Christian Albrechts University
Kiel, Germany

Institute for Transport Studies, University of Leeds
Leeds, UK

ISIS, Institute of Studies for the Integration of Systems
Rome, Italy

Mcrit Srl
Barcelona, Spain

Rapidis
Charlottenlund, Denmark

Systema, Systems Planning and Management Consultants s.a.
Athens, Greece

TENCONNECT

Project funded by the
European Commission – DG TREN

TENCONNECT

Traffic flow: Scenario, traffic forecast and analysis of traffic on the TEN-T, taking into consideration the external dimension of the Union.

Report information

Report no: TenC408_001

Title: Report on Scenario, Traffic Forecast and Analysis of Traffic on the TEN-T, taking into Consideration the External Dimension of the Union – TRANS-TOOLS version 2; Model and Data Improvements

Authors: Jeppe Rich, Christian Overgaard Hansen, Goran Vuk, Otto Anker Nielsen (DTU Transport), Artem Korzenewych, Johannes Bröcker (CAU)

Version: 2.0

Date of Publication: June 23rd, 2009 (with editing corrections July 21st, 2009)

This document should be referenced as:
Rich J., Bröcker J., Hansen C.O., Korchenewych A., Nielsen O.A., Vuk G. (2009): *Report on Scenario, Traffic Forecast and Analysis of Traffic on the TEN-T, taking into Consideration the External Dimension of the Union – TRANS-TOOLS version 2; Model and Data Improvements,* Funded by DG TREN, Copenhagen, Denmark.

Project Information

Project Acronym: TENconnect

Project Name: Scenario, Traffic Forecast and Analysis of Traffic on the TEN-T, taking into Consideration the External Dimension of the Union

Contract Number: TREN B1/159-2007

Duration: 1.01.2008 – 28.02.2009

Commissioned by: European Commission DG TREN

Lead Partner: Tetraplan A/S, Copenhagen, Denmark

Partners: BMT Transport Solutions, Hamburg, Germany
DTU-Transport, Department of Transport, Technical University of Denmark, Lyngby, Denmark
Institute for Regional Research, Christian Albrechts University, Kiel, Germany
Institute for Transport Studies, University of Leeds, Leeds, UK
ISIS, Institute of Studies for the Integration of Systems, Rome, Italy
Mcrit Srl, Barcelona, Spain
Rapidis, Charlottenlund, Denmark
Systema, Systems Planning and Management Consultants s.a., Athens, Greece

Subcontractors:
Faculty of Transport Planning, University of Belgrade, Serbia
InformiGIS, Charlottenlund, Denmark

Ramböll, Espoo, Finland
SUDOP, Praha, Czeck Republic
Technical University of Vilnius, Lithuania
University of Szczecin, Poland

TENCONNECT

Contents

1 Introduction

This report gives an overview of the model structure and data sources of the TRANS-TOOLS version 2. The modelling work and data compilation can be seen as improvements to the TRANS-TOOLS version 1 and the present report will focus on explaining differences and improvements. A more detailed description, however, of the various sub-models can be found in the documentation of the respective sub-models. These are illustrated in table 1 below;

Report / task no.	Description
4.02-002	Long distance passenger demand model
4.02-001	Short distance passenger demand model
4.01-003	Demand matrices
4.01-001	Assignment
4.03-001	Trade prediction model
4.03-002	Economic assessment model

Table 1: Sub-tasks in the TRANS-TOOLS version 2.

As already indicated in the proposal, the departure point for the model building has been the TRANS-TOOLS version 1 model. This model represents the most recent and comprehensive transport model for the EU Commission. TRANS-TOOLS version 2 has focused on improving the model along several dimensions. The improvements can be decomposed into two parts; data improvements and structural model improvements.

Data improvements

Improve the geographical coverage of the TRANS-TOOLS model, by

- *Disaggregating of the zone system in some new Member States and neighbouring countries*

Updating and improve trip matrices, by

- *Compiling more traffic counts in order to improve the car matrix estimation by mean of MPME*
- *Adding more traffic counts for air traffic by using the leg-database in EUROSTAT for EU27 and compiling additional counts for the remaining countries, thus enabling an MPME matrix fitting.*
- *Re-estimate rail matrices based on national statistics*
- *Transforming from OD to a GA representation*

Update and improve the networks within the core area, in order to

- *Reflect networks in year 2005, rather than 2000*
- *Upgrade networks in new member states as well as include a more detailed network structure in the core modelling area*
- *Selected extensions needed to enlarge the coverage area*

Model improvements

Update a number of the sub-models of the TRANS-TOOLS model, thus

- *Improving and extending CGEurope in the version used in TRANS-TOOLS*
- *Replacing the existing trade model with the above mentioned improved version of CGEurope*
- *Replacing the existing passenger demand model*
- *Improving the existing assignment model, especially for air traffic.*

The present report will focus on these two sets of improvements.

1.1 Background of the TRANS-TOOLS model

Very briefly the development of the TRANS-TOOLS model was initiated in 2004 with the objective to become the key decision support model concerning transport impact analyses conducted by the EU Commission.

The TRANS-TOOLS model forms a consistent set of linked sub-models which can be operated without manual interference. It is built into an ArcGIS framework which facilitates data editing, model execution, and production of GIS based illustrations and maps.

The evaluation of the TRANS-TOOLS model for use in the TENconnect showed that it was necessary to improve and update the existing model as indicated above to achieve better forecasting accuracy and comply with the terms of reference.

The TRANS-TOOLS version 1 model is a merger of modified existing European level models. It includes:

- *An economic model based on a modified and simplified version of* CGEurope*,*
- *A* freight trade *and* freight mode choice model *developed from the NEAC model system,*
- *A* freight logistic model *based on SLAM, a module that can be appended to the SCENES model,*
- *A* passenger demand model *based on ASTRA and VACLAV,*
- Assignment models for road, rail, inland waterways and air transport *designed to the specific modelling based on a general framework developed by CTT.*

In the proposal it was decided to improve and extent the model following the recommendations of the TEN-CONNECT study.

1.2 Improving the data foundation

1.2.1 New zone system

As indicated a disaggregation of the TRANS-TOOLS zones was necessary. It includes a decomposition of Albania, Belarus, Bosnia-Hercegovina, Bulgaria, Croatia, Romania, Russia, Serbia-Montenegro, Turkey and Ukraine. The new zone system consists of 1441 zone, compared to 1269 in the previous model.

NUTS regions are available for the new Member States (Bulgaria and Romania), Croatia, and Turkey. NUTS 3 are used as the base zone level. The new zone system is illustrated in Figure 1 below.

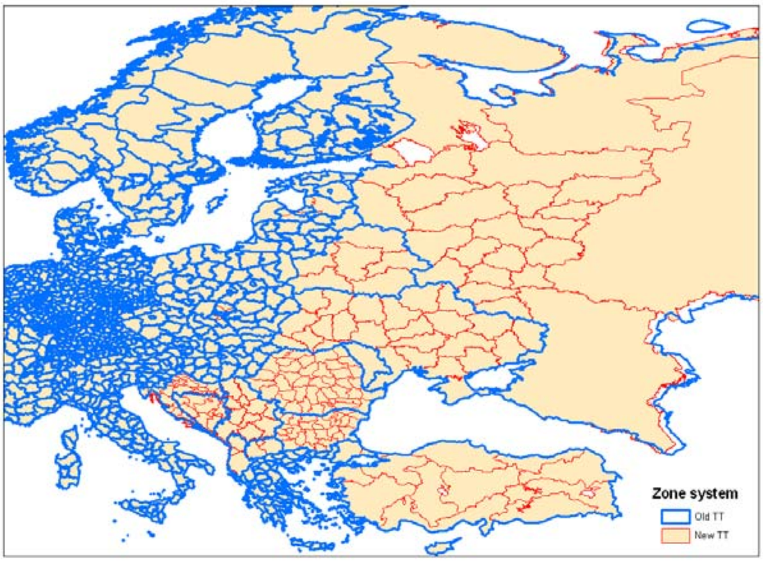

Figure 1: Zone system in the TRANS-TOOLS model version 2.

There have been made small corrections also in Latvia, Germany, and Poland.

1.2.2 Improved and updated base year car trip matrices

The car matrices of the first TRANS-TOOLS version demonstrated too much deviate from national trip rates and total volumes in some cases. In order to improve the car matrices, an extensive matrix fitting process has been carried out.

The approach has first been to fit the part of matrices within each country to the national trip rates. The matrices are then fitted to measured traffic counts using the so-called Multiple Matrix Estimation Method (MPME).

The idea is to use the average deviation of all counts along each route for the estimation, and to use all routes between each zone-pair according to the likelihood of choosing the specific route. The result minimises the weighted square deviation between counts and model, and conditional to this minimise the square deviation between the adjusted matrices and the original (2005) matrices. This fitting approach for the matrix update is in line with the final recommendations from the MOTOS EU-project given by Prof. Marc Gaudry at the final conference.

Although counts are missing in some new member states and at some parts of the road network, this adjustment of the matrices will improve matrix cells in areas where data is available, whereby the matrices will become consistent with available counts.

In addition, the following tasks have been conducted to finalize car demand matrices:

- Amendment of intra zonal passenger car trips

- Split of trip purpose
- Construction of symmetrical matrices
- Reformulation to GA-matrix
- Calculation of car passengers
- Fitting to statistics

A discussion of the GA versus OD representation is included in section 1.2.5.

1.2.3 Improved and updated air matrices

The air matrices for the TRANS-TOOLS version 1 model contain far too many short air trips, and too few long distance trips. The main problem was that air matrices were not calibrated due to an MPME matrix fitting algorithm.

In the new version of the model, a comprehensive matrix fitting process has been conducted, which will be briefly described in the following.

The initial air passenger matrix is estimated on basis of data from EUROSTAT supplemented by data collected from various internet sites (especially regarding East Europe) e.g.

- http://www.flightstats.com/
- http://www.aena.es/csee/ccurl/aeropuertos.pdf

The most important data source, the EUROSTAT database covers leg-loading from major airports in EU27 (excl. Czech Republic and Slovak Republic) Norway, Iceland, and Switzerland to all major airports in the World. If we assume a symmetrical flow pattern, the EUROSTAT database covers all intra European passenger flows as required in the study.

EUROSTAT database includes transfer passengers which introduces a double counting relative to an OD representation. It has been verified by comparisons with airport statistics. While a general down scaling of passenger flows has been conducted, passenger flows at few airports in East Europe was factored up.

The airport to airport flows have been converted to the TEN-CONNECT zonal structure. In principle, the mapping between airports and zones is many to many e.g., the hinterland to an airport may consists of many zones and several airports may share the same hinterland. In the initial phase of the development of an air matrix, airports were, however, assigned uniquely to zones unique.

Cells in the matrix were empty. First, 2005-data has been used to construct the matrix. If cells were not covered by the data due to lack in the statistics, an average of 2004- and 2006-data was applied. Second, if cells remained empty a small number of trips were added before MPME estimation to allow adjustments for those travel relations.

1.2.4 Updating of rail matrices

The matrix is constructed by first determining a country-wise OD matrix, which is then broken down by NUTS-3 level according to existing matrices and demographical data.

The country-wise matrix has been established based on different data sources. The “code” variable describes the process;

Code	Description
1	Determined from the national information in data set 2. However, for 8 eastern European countries, world bank data for 2003 has been used.
2-5	Determined from data set 3 and 4 for the years 2004 and 2005.
6	Transtool year 2000.
7	Are set to zero since unknown in EUROSTAT but part of EU25.

Table 2: Data usage for the estimation of the country-wise OD matrix.

The base matrix uses the level of the country-wise OD matrix and distributes traffic at the NUTS-3 level according to the existing Transtool matrix.

The different data sources applied to calibrate the rail matrices (refer to Table 2 above) is outline below;

1) Railway transport - Quarterly passengers transported
    a. This gives the quarterly national totals for EU25 (Cyprus and Malta has no rail, however, Norway and Turkey are added)
2) Railway passenger transport by type of transport (national/international) (in 1000 passengers)
    a. This gives total traffic national/international total for EU25 (remember that Cyprus and Malta has no rail). This is the number of passengers transported so transit traffic will be included.
3) International railway passenger transport from the reporting country to the country of disembarkation (in 1000 passengers)
    a. This provides off-diagonal elements reported by origin country.
4) International railway passenger transport from the country of embarkation to the reporting country (in 1000 passengers)
    a. This provides off- diagonal elements reported by destination country.
5) Accompanied passenger car railway transport, by type of transport (passenger cars)
6) Accompanied passenger car railway transport, by type of transport (number of passengers)

Due to a low number of counts, the matrix adjustments for the rail passenger matrixes has been carried out as a manually process where the matrix elements have been adjusted in regions to reproduce statistics on passenger km.

The few counts available were also taken into consideration. The post processing of the train passenger matrix has also included:

- Split of matrix by trip purpose
- Construction of symmetrical matrices
- Reformulation to GA-matrix

Initially, the passenger matrix was split by trip purpose in accordance with the existing TRANS-TOOLS model. The procedure is similar to the purpose split of air passenger matrix described in section 3.3.2.

After splitting into three purposes the purpose specific matrices were made symmetrical. Finally, the matrices were rearranged into GA-matrices to be used in the passenger demand models similar to the procedure explained in section 1.2.5.

1.2.5 Transformation from Origin-Destination (OD) to Generation Attraction (GA) matrices

It has been resolved that the matrices needs to be transformed from origin-destination only (the specific trip) to generation-attraction (the person that conducts the trip).

The core issue here is to add information to the matrices concerning the passenger travelling. Presently this is assumed equal, i.e. that 50% of tourists flying from the mining town of Kiruna in Northern Sweden to Malaga are people from Southern Spain flying back from holiday in Northern Sweden and 50% are people from Sweden flying out on holiday in Andalusia.

If change in GDP, occupation, migration etc. correctly should reflect demand for transport it is important that socio-economic drivers can be related to the person that is conducting the trip – regardless whether the trip is outward or backward. Furthermore this is important in order to correctly feed-back the impacts from pricing back into the CGEurope model.

Generally, the approach for transforming matrices from OD to GA is different for different trip purposes;

- Business trips are based on GDP and work place in the zone of origin and destination respectively

- Commuting is part of private trips in the TRANS-TOOLS model. Based on assumptions on trip lengths, commuting is separated from private trips, assuming commuting to be mainly short distance (yet several metropolitan regions in TRANS-TOOLS have 3-6 NUTSIII zones, and therefore between NUTSIII commuting). The separated OD is transformed to GA based on population, workplaces and GDP.

- The remaining private trips are transferred from OD to GS based on GDP and population in the zone of origin and destination respectively

- Tourism is based on population, GDP and a tourist attractiveness measure in the zone of origin and destination

1.2.6 Network update

The TRANS-TOOLS model includes separate networks for road transport, rail passenger and freight transport, inland waterways transport, and passenger transport by air. Since the base year of the existing networks is 2000, they are updated to 2005.

The air network has changed radically between 2000 and 2005 due to introduction of new low budget lines and September 11, 2001. Therefore is has been thoroughly revised and new attributes added to be used in the development of a new version of the TRANS-TOOLS model (Task 4A). EUROSTAT was deployed to gather information about flights between European airports. Then local airport web-sites were visited to extract information to fill caps concerning number of departures, identify connections operated by low budget lines, and add charter flights to tourist areas. Practically all links in the 2000-air network have been revised or deleted and new added.

The update procedure for road and rail networks was initially extensions of the networks into e.g. Turkey and Russia. Then they were distributed to local partners of the consortium for validation and updates from 2000 to 2005. The inland waterway network was distributed without extensions. Networks edits returned in various ways were then combined and implemented into GIS 2005-networks.

Table 3 summarizes number of edits conducted during the process of updating from 2000 to 2005.

Since the 2000 road network included almost 36,000 links, about 4% new links have been added and almost 10% edited. The edits mainly includes revisions to speed, number of lanes, and toll charges.

The rail passenger and freight networks include about 5,500 links. Thus, about 3% new links have been added and 30% edited. The edits concerns primarily speed revisions.

The inland waterways network includes about 800 links, and new few have been added and edited. The new added links are in East Europe and Balkans.

Network	**New added links**	**Deleted links**	**Edited links**
Road network	1374	12	3458
Rail passenger network	168	39	1660
Rail freight network	171	18	729
Inland waterways	6	10	4

Table 3: Number of network edits in updating from 2000 to 2005

2 Model improvements

2.1 Improving assignment of air passengers

2.1.1 Validating air assignment procedures

The air assignment model in the TRANS-TOOLS version 1 calculated the choice of departing airport, choice of air route, and choice of arrival airport. The choice parameters were related to time and cost, where the different trip purposes had different values of times.

The access- and egress to airports was simplified, as travel times and costs (as well as a border crossing dummy) have been pre-calculated for each link between a NUTS3 region and the possible airports. This meant that changes of the land network (road and rail) did not influence the choice of airport and air route.

Furthermore, the access and egress traffic was just allocated to the zone-connectors rather than onto the road and rail networks. This is likely to underestimate volumes on the main rail networks leading to the airports. Examples are the TGV network to Charles de Gaulle Airport in Paris, where a significant volume transfer passengers to the airport, or the main railway between Copenhagen and Sweden with high volumes going to and from the Kastrup airport in Copenhagen (20% of the passengers at the airport come by rail from southern Sweden). The de-coupling from the assignment models may also result in underestimation of congestion on the road networks near airports.

The proposed improvements of the assignment models for air transport include two main steps

1) Simulate access and egress mode choice more accurately

2) Assign the connector loads to the road and rail network.

2.1.2 Improving access and egress modelling

The first building block in the air access/egress mode choice is specified as a simple binary logit model that calculates choice probabilities for rail and car for each possible link between a NUTS3 region and relevant airports.

The utility function includes travel time, travel cost and border crossing. The improvements of the passenger model allow for differentiation between home-zone (where one can use ones own car) and destination zone (where one need to rent a car, or being collected by another person).

The second building block is calculation of access/egress travel times and costs for car and rail, respectively. The optimal paths by car and rail are calculated separately by a stochastic assignment model and the LoS is averaged over the paths and weighted by the binary choice model.

The weighed travel times and costs are transferred to the air network and the existing TRANS-TOOLS air assignment model is run. The procedure is repeated for each trip purpose, as they have different utility functions.

2.1.3 Assigning access egress mode to air onto road and rail network

The access and egress flow on each zone connector is split into rail and car transport based on the already calculated choice probabilities. Then the flows are simply preloaded onto the rail and road network, respectively, according to the pre-calculated paths.

The procedure is repeated for each trip purpose, as they have different utility functions.

2.2 Improving the passenger demand modelling

There is a need to improve the passenger demand modelling consisting of mode choice, destination choice and trip frequency. The modelling strategy is start by recognising that long and short trips tend to be very different in many respects (choice of mode, and destination choice). Moreover, different trip purposes (as for the old model) should be treated differently due to parameter-instability and different value-of-time estimates. As a result, the models have been segmented along the following dimensions;

- Trip-purpose: Business, Commuting, Private, and Holiday

- Distance: Under 100 Km, Over 100 Km.

A main difference compared to the TRANS-TOOLS version 1 is that the new model operates on a GA representation of the data.

In other words, the zone of the travellers' origin is explaining both the outward and return trip. This is especially important when the framework is extended to new member states and neighbouring countries, e.g. the trip frequency for a tourist going back from vacation in Bulgaria to Denmark is explained by the income (GDP) of that person (i.e. the income level in Denmark).

2.2.1 Short-distance model

The TEN-CONNECT short-distance model, i.e. up to 100 km, relates to an average weekday in 2005 (the model base year). The model consists of two sub-models; a frequency model at the top and a joined mode/destination choice model below. The sub-models are linked by an accessibility measure, i.e. logsums. Four modes are included in the modal split; car driver, car passenger, bus and train. Finally, the model includes four travel purpose segments; commuting, business, private and vacation/holiday trips.

An important issue in the model has been to calculate values of time (VOT) for short-distance trips across all zones in the TEN-CONNECT project. These values are differentiated across travel modes, travel purposes and income.

This has made it possible to do a parameter transfer from the Danish OTM model, which have enabled a complete specification of utility functions in the destination and mode choice model.

The model has been tested and validated in terms of elasticity estimates.

2.2.2 Long-distance model

The modelling strategy has followed a state-of-the-art strategy involving a large-scale discrete choice model (nested logit) for the choice of mode, destination, and trip frequency.

The “other” main difference compared to the previous approach has been extensive testing of non-linearities in the corresponding utility functions (as also recommended by MOTOS and Prof. Marc Gaudry at the final conference) and allowance for non-trivial substitution pattern (nested logit) among choices. This reduces the problems with underestimation of long air trips and overestimation of short air trips in the present Trans-tools models.

The new models are estimated using a full-information maximum likelihood estimation in the statistical package, SAS. As input are used matrices as described in 1.2.2, repeated preference data (DATELINE, TUdata), socio-economic data for the zones of origin and destination (EUROSTAT), and Level-of-Services variables (travel times and costs etc.) for the trips.

More over, for the long-distance model additional distance parameter-instability has been identified. As a result, separate functional forms for trips below 600 Km and above 600 Km has been used. The applied models have turned out to be stable across all trip purposes; business, private, commuting and tourism.

A so-called pivot point method is used to fit the demand model to the matrices in the base-year. Due to inaccuracies in the matrices for air and rail, the validation phase will determine whether a pivot-point procedure will be used for these segments, or only synthetically matrices from the demand model.

2.3 Improving the trade and economic model

2.3.1 Background

The inclusion of a multiregional regional economic model in the project is based on the fact that freight transport flows as well as passenger flows are determined by the regional economic activity. Demand for freight transport arises from the production and consumption of final goods, raw materials and intermediate goods. Demand for travel is derived from utility maximization, taking out-of-pocket costs for travel as well as a disutility due to time spent for travel into account.

Transport policies affect the prices for transport and travel times and therefore also the choice of location of industries. Changes in transport costs caused by policies have economic effects through their influence on patterns of interregional trade, on investments, on the location decisions of firms and employment of workers. It is expected that effects of transport policies are different in each region. This is especially true for transport infrastructure investments, which have immediate local, but also long distance impacts, and also for such measures as road pricing that can have different effects in different regions within a country. Therefore, it is important to forecast the economic effects not only at country level, but also at the regional level, taking into account interactions between the regions.

To account for these important drivers of demand, the CGEurope model captures the economic effects of transport policies, namely the effects on regional GDP, sectoral structure, and the resulting changes in the demand for transport generated in each region. CGEurope is a multi-regional computable general equilibrium model, in which transport costs explicitly appear as firms’ expenditure on freight traffic and households’ costs of travel (e.g. Venables and Gasiorek, 1998, Bröcker, 1998). It incorporates features from recent developments in the literature like product diversity and monopolistic competition, explicit modelling of out-of-pocket freight costs as well as costs of passenger transport. CGEurope is designed as a comparative static model: it takes changes in transport cost that arise from policies as instantaneously given, and calculates the effects of changes of transport cost and travel times on sectoral activity, transport flows, prices and households' utility.

CGEurope is used to compute regional welfare effects of policies in terms of GDP and real income changes. In addition, the response of regional unemployment rate will also be calculated. It has also replaced the unconstrained gravity trade model, which was part of the TRANS-TOOLS version 1 model.

2.3.2 Enlarging and refining the coverage

The primary inputs to CGEurope in the version of TRANS-TOOLS are predicted GPD growth rates and LoS information from the logistics and passenger model. In order to meet the study objectives, the geographical coverage of CGEurope should be enlarged and at the same time match the zonal system of the existing logistics model.

The existing TRANS-TOOLS version of CGEurope applies two world regions to account for flows outside the core area. The spatial resolution of the rest of world region is refined to correspond with the existing mode choice model. CGEurope is calibrated on the extended dataset including countries of Far-East, sub-Saharan, Africa and America. The main challenge for the inclusion of these countries is the absence of corresponding transportation data which are not included in the logistics and passenger models and bilateral trade data.

Because mainly ships are used in overseas goods transport, the consortium suggests to use the sea transport costs and times as approximates. Therefore, the existing maritime LoS matrices need to be expanded to include relations outside Europe and reflect the zonal structure of the enlarged and recalibrated CGEurope. Concerning passenger travel related to regions outside the core area of TRANS-TOOLS, future travel times and costs are assumed unchanged. It is reasoned that changes to overseas passenger travel times and cost only will have marginal economic impacts.

Bilateral trade information between Algeria, Egypt, Lybia, Lebanon, and Isreal, is missing. Therefore, we propose to include the relations to cover the whole world but in a less accurate way, since the bilateral trades are of less relevance to the study.

2.3.3 Replacement of trade model

A new freight demand model has been developed to replace the unconstrained gravity trade model, which is currently the part of the TRANS-TOOLS model. The motivation for this is the poor predictive power of the naive unconstrained gravity approach. A two-stage method, predicting origin and destination totals first and applying the doubly constrained model next, is proposed.

In calculation of origin and destination totals, money flows are converted into tonnes used in freight modelling. The consortium expects to use one production sector with multi-commodity outputs. To reduce the uncertainty a pivot-point procedure relative to the totals of ETIS 2000 base year matrix is used for all relevant regions within the core area of the TRANS-TOOLS model. The procedure should, however, in case of extreme pivot-point corrections or missing base year data use tonnes converted from monetary units directly.

If the value-to-weight ratios by commodity change in the future it will affect the pivot-point corrections. Therefore the developments of the value-to-weight ratios in forecasting future flows needs to be considered. Time-series of trade data will be analysed with the aim to develop likely future trends which can be used as basis for predicting future assumptions. An interface is developed to edit predicted changes to value-to-weight in forecasting.

A doubly constrained form will be used to distribute forecasted totals over zones where a deterrence function based on maritime LoS matrices, and outputs from the logistics and passenger models could be applied. The consortium proposes again to apply a pivot-point procedure relative to the individual cells of ETIS base year matrices to calculate predicted future flows. A set of rules will be applied to correct for extreme pivot-point values. Because the pivot-point correction of the individual cells is likely to change the totals, a normalization of rows and columns is added to the procedure.

This secures a better integration between CGEurope and the traffic model, as shown in the figure below which contains our proposed approach on the right side, and the present on the left. The figure is taken from the final recommendations from the recent EU-project MOTOS, which focussed on recommendations for traffic models with special focus on the New Member States;

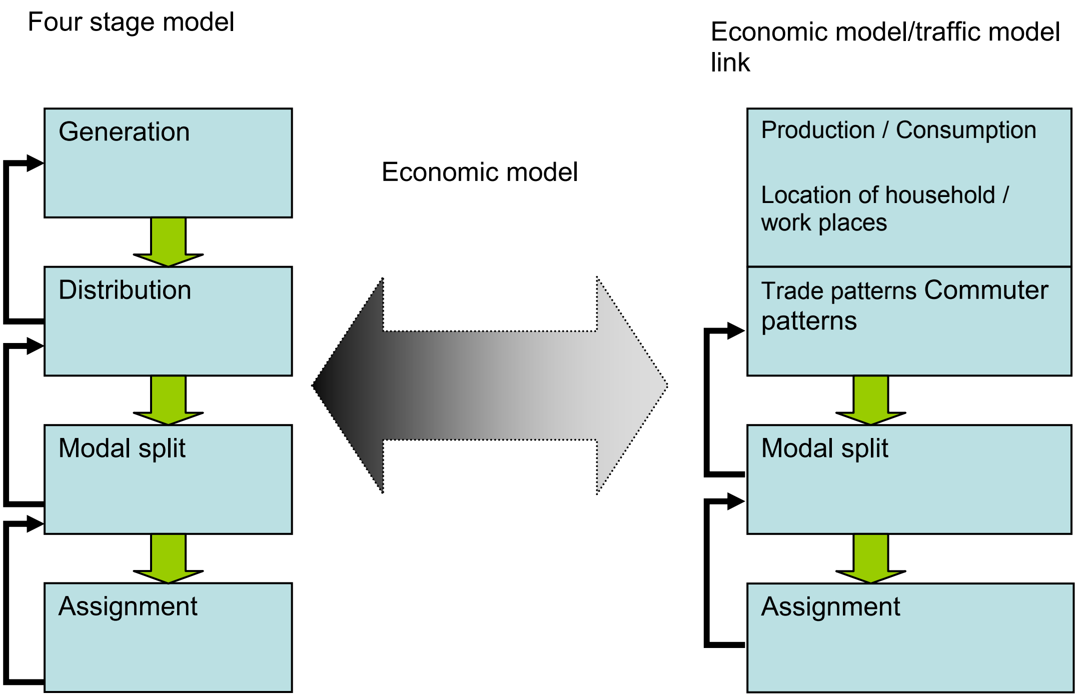

Figure 2: Present approach in the TRANS-TOOLS model (left) and the suggested improved method (right).

The figure is taken from the final conference from the MOTOS project, which developed good practice guidelines for the new Member States.

2.3.4 Upgrading the impact model

The trade impact model is essentially the *CGEurope model* (Bröcker, 1998/2000, 2002) frequently applied to transport policy evaluation before, with an important modification.

In the model version used so far, interregional trade was estimated as a by-product of the model calibration, assuming CES composition functions to be symmetrical and identical for all users. The strength of the approach was its applicability even without explicit interregional trade information. In the current project we have to keep consistency with the trade estimates of the TPM as far as possible. We therefore prefer to introduce so called Armington preferences in trade allowing calibrating the model such that it exactly reproduces the trade values of the TPM described before. This makes the model theoretically less elegant, but brings it closer to the data.

We briefly repeat the description of the basic structure of CGEurope. It is a static general equilibrium model for a closed system of regions covering the whole world. The regions are the 1441 regions on the NUTS3 level of the TEN-Connect project and 19 external zones. In each region reside identical immobile households owning the regional stock of production factors that are immobile as well. Their incomes stem from regional factor returns as well as from an interregional income transfer that can have a positive or negative sign. Income transfers are exogenous (in real terms) and add up to zero for the entire world. A further income source is introduced if we simulate charges to be paid by users of infrastructure. Revenues from these charges are redistributed to households. The redistribution shares of regions are imported from the transport model.

Households spend their income for buying goods and services partly produced in their own regions and partly produced in other regions. Households' demand represents total final demand, which means, private as well as public consumption, and investment. There is no separate public sector in the model; that is households have to be regarded as an aggregate of private and public households, their budget constraint is the consolidated budget constraint of private and public households in the region.

3 Passenger demand matrices

3.1 Introduction

This memo describes the development of passenger matrices which are used for estimation and in a pivot-point procedure for a new passenger demand model.

Matrices are established for an average day in 2005 describing five main modes:

- Car driver
- Car passenger
- Air
- Train
- Bus

And four trip purposes:

- Home-business (HB)
- Home-private (HP)
- Home-vacation (HH)
- Home-work (HW)

In the model, it is assumed that all trips are home-based. While this is almost correct for long distance trips, it is not entirely true for short distance trips. Since modeling of long distance trips are the main objective of TRANS-TOOLS model, it seems an acceptable solution to reduce computational requirements.

In total, 5 x 4 = 20 trip matrices are estimated. Different approaches have been applied in development of the matrices which are explained through the Sections 2-5. Whereas matrices for air and passenger car have been adjusted to fit count data using a Multi Part Matrix Estimation (MPME) method, matrices for car passengers, bus and train have been developed only on the basis of statistics and surveys.

Section 6 provides a summary of the matrices by mode and trip purposes.

3.2 Estimation of car demand matrices

3.2.1 Data and initial matrices

Initially, the 2005 passenger car matrix calculated by the existing TRANS-TOOLS model as part of the TEN-CONNECT Task 1 deliverable was used. It describes the flows between the 1269 zones in the existing version of TRANS-TOOLS. To apply the matrix to the TEN-CONNECT zonal structure with 1441 zones it is spilt by using a single-constraint gravity model (refer to: TRANS-TOOLS Deliverable 3, Chapter 2.7, 2006). The value for the α-parameter in the distance function formula has been estimated to -2. To describe the generations and attractions in the TEN-CONNECT zonal structure the population has been used.

The existing TRANS-TOOLS passenger model does not account for zone internal trips. In other words, a rigid use of the old model would result in no trips between new detailed zones located within a large original zone e.g., new zones within Russia and Turkey.

Before running the gravity model a preliminary estimate of the number of trips internally in those zones has, therefore, been carried out. Since most of the zone splits are carried out for countries in East Europe it was decided to base these estimates on the ratios between traffic going in/out of, and internally in other east European countries.

The traffic count data originates from various data sources. In most cases counts data are provided by national road authorities or found on various online applications/maps. First, volumes have been coded to the network. Second, extensive preparation of the counts has been conducted in order to use count data as input to matrix estimations. For instance,

- Counts not always represent the year 2005 and therefore have been corrected by annually average traffic growth factors.
- Some count data have only included the annual average daily traffic (AADT) in terms of vehicles (not divided into passenger cars and trucks).

The first issue has been handled by estimating different annual growth rates for each country and separated on different road class. The growth rates are shown in Table 4.

The second issue was handled by using the counts that actually were divided in different vehicle classes to estimate percentages of passenger cars. This was carried out on regional level and separately on each road class. The estimated percentages are shown in Table 5.

	Annual growth rates in passenger car km 2000-2005			
Country	Motorway	Rural	Urban	Total
Albania	1.45	1.22	1.17	1.27
Austria	1.08	1.02	1.00	1.03
Belarus	1.35	1.20	1.10	1.20
Belgium	1.05	1.01	1.02	1.03
Bosnia	1.45	1.22	1.17	1.27
Bulgaria	1.12	1.06	1.05	1.08
Croatia	1.45	1.22	1.17	1.27
Cyprus	1.12	1.05	1.02	1.06
Czech Republic	1.18	1.26	1.19	1.22
Denmark	1.17	1.03	0.99	1.06
Estonia	1.38	1.28	1.24	1.28
Finland	1.13	1.13	1.06	1.11
France	1.15	1.02	1.00	1.04
Germany	1.07	1.02	1.01	1.04
Greece	1.12	1.05	1.02	1.06
Hungary	1.20	1.28	1.23	1.25
Iceland	1.26	1.25	1.23	1.25
Ireland	1.17	1.13	1.12	1.15
Italy	1.12	1.05	1.02	1.06
Latvia	1.66	1.54	1.49	1.56
Liechtenstein	1.19	1.09	1.06	1.11
Lithuania	1.66	1.54	1.49	1.56
Luxembourg	1.14	1.04	1.01	1.06
Malta	1.20	1.10	1.07	1.12
Marcedonia	3.51	3.25	3.14	3.30
Moldova	1.59	1.48	1.43	1.50
Netherlands	1.14	1.04	1.01	1.06

Norway	1.11	1.10	1.08	1.10
Poland	2.65	2.40	2.35	2.49
Portugal	1.38	1.20	1.11	1.20
Romania	1.35	1.20	1.10	1.20
Russia	1.35	1.20	1.10	1.20
Serbia and Montenegro	1.45	1.22	1.17	1.27
Slovak Republic	1.12	1.18	1.15	1.16
Slovenia	2.06	1.73	1.66	1.88
Spain	1.39	1.20	1.11	1.20
Sweden	1.10	1.10	1.08	1.10
Switzerland	1.08	1.03	1.02	1.05
Turkey	1.24	1.14	1.30	1.23
Ukraine	1.35	1.20	1.10	1.20
United Kingdom	1.12	1.08	1.07	1.10

Table 4: Estimated annually growth rates for personal car km

Region	Road Class	Passenger car percentage of total volume
Albania, Bosnia, Bulgaria, Croatia, Greece, Hungary, Macedonia, Romania, Slovenia, Turkey and Serbia, Bosnia	Motorway	71
	Rural roads separated lanes	68
	Rural roads	69
	Urban road	70
Ireland and United Kingdom	Motorway	79
	Rural roads separated lanes	81
	Rural roads	81
	Urban road	85
Austria, Belgium, France, Germany, Italy, Luxembourg, Netherlands, Portugal, Spain and Switzerland	Motorway	84
	Rural roads separated lanes	93
	Rural roads	89
	Urban road	90
Belarus, Czech Republic, Estonia, Latvia, Lithuania, Moldavia, Poland, Poland, Slovak Republic and Ukraine	Motorway	69
	Rural roads separated lanes	73
	Rural roads	71
	Urban road	77
Denmark, Finland, Norway and Sweden	Motorway	87
	Rural roads separated lanes	87
	Rural roads	86
	Urban road	90

Table 5: Estimated personal car percentages of annual average daily traffic

3.2.2 Matrix adjustment procedure

The Multi Part Matrix Estimation (MPME) is a heuristic method where an existing trip-matrix is adjusted based on counts[1] . This is done so that each cell in the matrix is adjusted according to counts found on the chosen routes between each zone pair in the network. The method is integrated with the assignment model which makes it possible to use more complex route choice models. This ensures consistency between the adjust-

[1] Nielsen, Otto Anker (1998). Two new methods for estimating Trip Matrices from Traffic Counts. Chapter in Travel Behaviour Research: Updating the state of play. Edited by Ortúzar, H. D., Hensher, D & Jara-Díaz, D. Elsevier Science Ltd. Oxford, UK. 1998. pp. 221-250.

ments carried out and the route choice model in which the matrix is going to be used later on. As the assignment used in the road network is a stochastically user equilibrium and congestion is taken into account the results depend on the input OD-matrix. This means that MPME adjustments have to be carried out in iterative steps until the resulting matrix has stabilized.

Before the actual MPME adjustments where carried out an initial adjustment, in order to level-out big differences between counts and modeled flows in regions, was preformed. This initial adjustment was done at the country level by adjusting all matrix elements with the ratio between the sum of counted and the sum of modeled traffic on each link with a count.

The main focus of the initial level-adjustment has been the East-European countries as the levels where most uncertain in these regions due to zone splits. The initial adjustment factors are shown in Table 6.

Country	Adjustment Ratio
Albania	0,61
Belarus	1,27
Bulgaria	2,08
Latvia	0,71
Macedonia	1,23
Montenegro	0,71
Romania	1,12
Russia	0,37
Serbia	2,22
Turkey	4,77
Ukraine	0,88

Table 6: Initial level-adjustment ratios

The assignment model does not assign zone internal traffic. On the other hand, the provided counts do include local traffic. In order to secure consistency between the counts and the modeled traffic only counts on links crossing zone borders are included in the MPME adjustments. This has resulted in counts being available on 3110 of more than 35,800 links in the network (approximately 8%). In Figure a map over road links crossing zone boarders is shown. It is estimated that the amount of counts and the fact that these are well spread over the whole geographically area is sufficient to get valid results from the MPME adjustments.

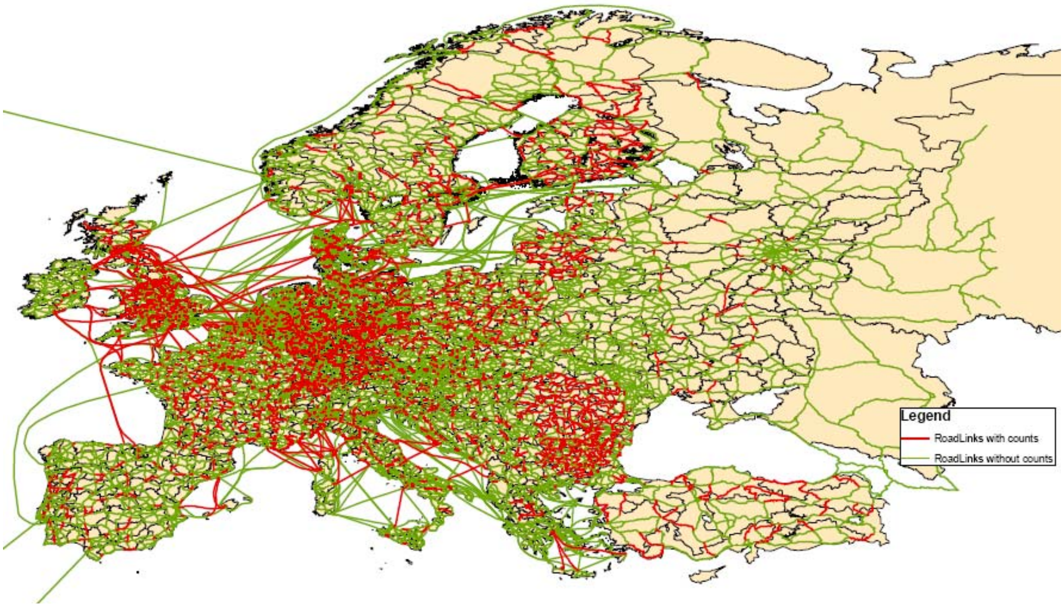

Figure 3: Road network with counts available for MPME adjustments.

The actual MPME adjustments have been carried out in 12 iterative steps where the traffic has been assigned using stochastically user equilibrium. Trucks have been included in the assignments to model congestion levels correct. Figure and Figure show differences before and after MPME adjustments when comparing counted and modeled flows.

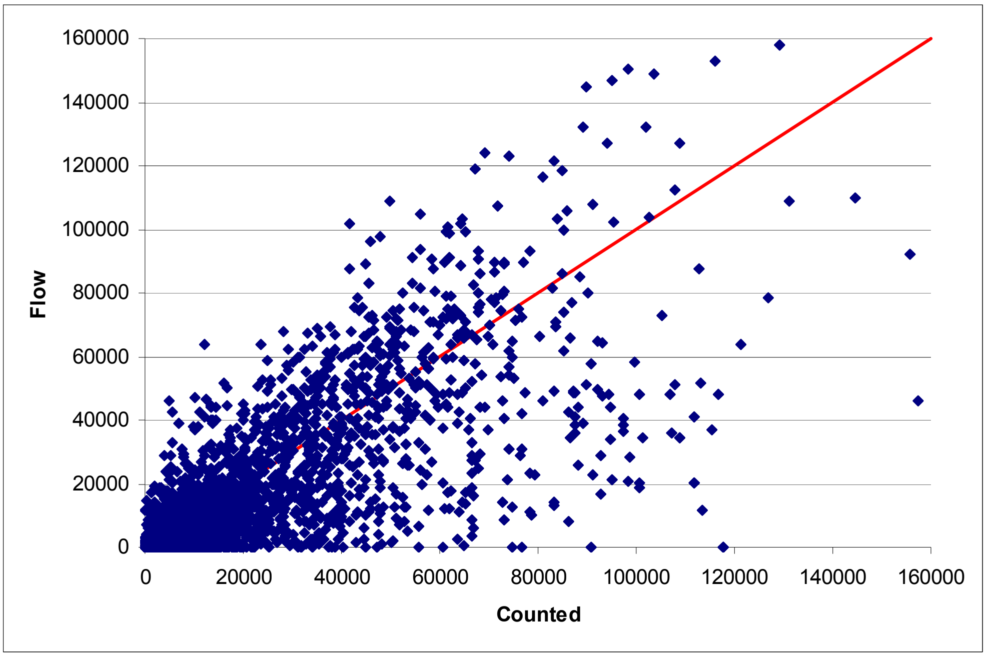

Figure 4: Modeled versus counted AADT before MPME estimation

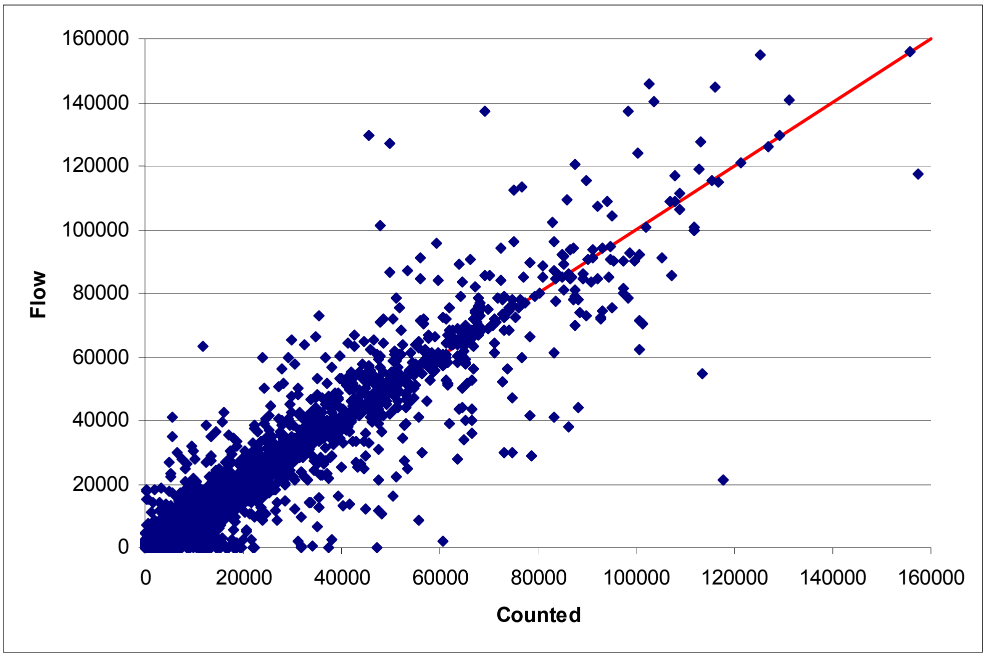

Figure 5: Modeled versus counted AADT after MPME estimation

3.2.3 Post processing car demand matrices

The following tasks have been conducted to finalize car demand matrices:

- Amendment of intra zonal passenger car trips
- Split of trip purpose
- Construction of symmetrical matrices
- Reformulation to GA-matrix
- Calculation of car passengers
- Fitting to statistics

Initially, intra zonal trips have been added to matrices to cover all trips by car. First, number of intra zonal passenger car trips T per purpose p was estimated from a simple formula,

(1) $$T_{ii,p} = \alpha_{1,p} x_{1i} + \alpha_{2,p} x_{2i} + \alpha_{1,p} x_{1i}$$

x_{1i} = passenger car trips to and from zone i

x_{2i} = passenger cars per 1000 inhabitants in zone i

x_{3i} = 1 if zone is an island else 0

Information about the external trips to and from a given zone was gathered from the MPME estimated matrix, while car ownership and geographical location was collected from the zonal databases. The parameter values (α's) were estimated based on the Danish Travel Survey giving very large Coefficients of Determination.

In the existing TRANS-TOOLS passenger model, trips are segmented by three purposes: business/commute, private, and holiday. The new passenger uses four trip purposes, therefore, the old business segment is split to business and commute trips. Typified shares by car per distance class have been computed on basis of the Danish Travel Survey and shown in Table 7.

Travel distance has been defined as the distance as the Euclidian distance multiplied with a detour factor of 1.2.

Distance class	Distance	Commute	Business
1	0- 10 km	90%	10%
2	10-20 km	87%	13%
3	20-50 km	85%	15%
4	50-100 km	75%	25%
5	100 km-	0%	100%

Table 7: Purpose shares to split trips between commute and business. Source: Danish Travel Survey

The table shows as expected that number of commuter trips are much larger than number of business trips for shorter travel distances. Above a travel distance of 100 km we assume that number of commuter trips can be neglected. The share factors have been applied to all zonal relations in the model to achieve an approximate split of the two purposes.

After splitting into four purposes the purpose specific matrices are made symmetrical by securing the number of trips form zone i to j equals the numbers trips form j to i on an average year day. Given the adjusted values $T_{ij(a)}$ and $T_{ji(a)}$ the symmetrically values T_{ij} and T_{ji} can be calculated as:

(2) $$T_{ij,p} = T_{ji,p} = \frac{T_{ij(a),p} + T_{ji(a),p}}{2}$$

The matrices are rearranged into GA-matrices to be used in the passenger demand models. It is done synthetically by a formula applying zonal generation and attraction:

(3) $$T_{ij,p}^{GA} = \frac{G_i A_j}{G_j A_i} T_{ji,p}^{GA}$$

The zonal population is used as generation variable for all trip purposes, while jobs are used as attraction variable for business, commute, and private trips. Hotel capacity available from the zonal database is used as attraction variable for leisure trips.

When passenger car matrices are processed, car passenger GA based matrices are estimated from average occupancy rates. Country and purpose specific occupancy rates used in estimation of passenger matrices are shown in Table 8: Occupancy rates used to estimate car passengers. Same occupancy rates are used for business and commute trips.

In the final step, matrices are fit to statistics. In an iterative process matrices have been assigned, passenger km computed and compared with statistical evidence. Then intra zonal trips and occupancy rates have been corrected and adjusted to match country specific statistics and preloaded traffic volume. It has been necessary to reduce occupancy rates compared with those used in the preliminary forecasts (Task 1 of TENconnect), in particular, East European countries to match statistics published in the Pocketbook[2] for year 2005.

Traffic is preloaded on road segments considered by the TRANS-TOOLS model. The number of intra zonal car passenger trips must exceed the preloaded traffic, because the model does not consider all roads within a zone. Based on average travel distances, intra zonal trips are (in the adjustment) restricted to meet this requirement.

Country	Occupancy rates per trip purpose		
	Home-Business Commute	Home-Private	Home-Holiday
Albania	1.45	2.29	2.48
Austria	1.22	1.55	1.69
Bosnia	1.45	2.29	2.48
Belgium	1.22	1.75	1.9
Bulgaria	1.31	2.06	2.23
Belarus	1.45	2.29	2.48
Switzerland	1.22	1.85	2.03
Czech Republic	1.45	2.29	2.48
Germany	1.22	1.55	1.69
Denmark	1.22	1.61	1.76
Estonia	1.45	2.29	2.48
Spain	1.29	2.03	2.48
Finland	1.22	1.8	1.97
France	1.22	1.82	2.23
Greece	1.22	1.74	1.89
Croatia	1.31	2.06	2.23
Hungary	1.31	2.06	2.23
Ireland	1.22	1.69	1.84
Italy	1.22	1.76	1.91
Liechtenstein	1.22	1.67	1.83
Lithuania	1.45	2.29	2.48
Luxembourg	1.22	1.55	1.69
Latvia	1.45	2.29	2.48
Moldavia	1.45	2.29	2.48
Macedonia	1.45	2.29	2.48
Netherlands	1.22	1.36	1.49
Norway	1.22	1.67	1.83
Poland	1.37	2.16	2.48
Portugal	1.22	1.63	1.77
Romania	1.45	2.29	2.48
Russia	1.45	2.29	2.48
Sweden	1.22	1.72	1.87
Slovenia	1.45	2.29	2.48
Slovak Republic	1.45	2.29	2.48
Turkey	1.22	1.74	1.89
Ukraine	1.45	2.29	2.48
United Kingdom	1.22	1.72	1.87
Serbia	1.45	2.29	2.48
Cyprus	1.45	2.29	2.48

[2] Directorate-General for Energy and Transport. *Engery and Transport in Figures 2007. Part 3: Transport*. 2008

Malta	1.45	2.29	2.48
Iceland	1.22	1.67	1.83
Montenegro	1.45	2.29	2.48

Table 8: Occupancy rates used to estimate car passengers

3.3 Estimation of air passenger matrix

3.3.1 Data

The initial air passenger matrix is estimated on basis of data from Eurostat supplemented by data collected from various internet sites (especially regarding East Europe) e.g.,

- http://www.flightstats.com/
- http://www.aena.es/csee/ccurl/aeropuertos.pdf

The most important data source, the Eurostat database covers leg-loading from major airports in EU27 (excl. Czech Republic and Slovak Republic) Norway, Iceland, and Switzerland to all major airports in the World. If we assume a symmetrical flow pattern, the Eurostat database covers all intra European passenger flows as required in the study.

Eurostat database includes transfer passengers which introduces a double counting relative to an OD representation. It has been verified by comparisons with airport statistics. While a general down scaling of passenger flows has been conducted, passenger flows at few airports in East Europe was factored up.

The airport to airport flows have been converted to the TEN-CONNECT zonal structure. In principle, the mapping between airports and zones is many to many e.g., the hinterland to an airport may consists of many zones and several airports may share the same hinterland. In the initial phase of the development of an air matrix, airports were, however, assigned uniquely to zones unique.

Cells in the matrix were empty. First, 2005-data has been used to construct the matrix. If cells were not covered by the data due to lack in the statistics, an average of 2004- and 2006-data was applied. Second, if cells remained empty a small number of trips were added before MPME estimation to allow adjustments for those travel relations.

Counted flows for the year 2005 have been coded in the flight network using data from Eurostat. The Eurostat data is leg based and gives the passenger per year between airports. The result of the coding process was reliable counts on 2,370 of 3,200 flight connections in the network (approximately 75%).

3.3.1.1 Air traffic counts

The most important data source is the EUROSTAT database. The database covers leg-loading between major airports in 27 European countries (EU25 + Bulgaria and Romania). In total, 4964 valid leg-counts are available from the database between EU27 countries (Matrix A in table 1 below). This database includes internal country traffic passenger flows, e.g. flows between Lyon and Nice are present in the database.

Moreover, the EUROSTAT database provides traffic flows to additional European countries. In fact, all of the countries in the Ten-Connect Model are represented. In the following, the countries within the Ten-Connect project, but outside EU27 will be named EU*. The flows to these countries (the B matrix in Table 9) are represented by 207 counts.

	EU27	EU*
EU27	A (4964)	B (207)
EU*	B' (207)	C (284)

Table 9: Data coverage with number of counts in parentheses.

The flow from EU* to EU27 (B') is in principle unknown. However, by the assumption of symmetry, the flows can be approximated by the reverse flow pattern in matrix B. As will be seen later, there is a strong degree of symmetry in the data.

What remains a challenge is to estimate internal traffic within in the EU* country matrix, e.g. matrix C. Since EUROSTAT does not provide any information in that respect, other sources of information have been applied. By means of the "flightstat" website a total of 284 counts have been established.

The extension of the zone structure causes the airport coverage to be weaker. This is especially true for Eastern Europe, Turkey, and Balkan.

In order to control and check the airport coverage, a gross table of European airports was constructed. The list consisted of more than 1000 airports for Europe. For EU* there were 244 additional airports in addition to those 417 already in the Ten-Connect database. In an analysis based on size, strategic importance, and geo location a total of 26 additional airports has been included in the model. These airports are shown in Table 10 below.

Abbrev	Country	IATA	Location	Airport name
RU	Russia	VOZ	Voronezh	Chertovitskoye Airport
RU	Russia	ROV	Rostov on Don	Rostov-on-Don Airport
RU	Russia	KRR	Krasnodar	Pashkovsky Airport
RU	Russia	MCX	Mineralnye Vody	Mineralnye Vody Airport
RU	Russia	ASF	Astrakhan	Astrakhan Airport
RU	Russia	KZN	Kazan	Kazan Airport
RU	Russia	KUF	Samara	Samara Kurumoch Airport
RU	Russia	GOJ	Nizhny Nov-gorod	Strigino Airport
RU	Russia	REN	Orenburg	Orenburg Tsentralny Airport
RU	Russia	UFA	Ufa	Ufa Airport
TR	Turkey	DLM	Mugla	Dalaman Airport
TR	Turkey	ADA	Adana	Adana Sakirpasa Airport
TR	Turkey	ASR	Kayseri	Kayseri Erkilet Airport
MK	Macedonia	SKP	Skopje	Skopje
MK	Macedonia	OHD	Ohrid	Ohrid
RS	Serbia and Monte-negro	INI	Nis	Nis
BA	Bosnien-Hertogovina	BNX	Banja Luka	Banja Luka International Airport
RO	Romania	CND	Constanta	Constanta "Mihail Kogalniceanu" International Airport [7]
UA	Ukraine	RWN	Rivne	Rivne International Airport
UA	Ukraine	CWC	Chernivtsi	Chernivtsi International Airport
UA	Ukraine	HRK	Kharkiv	Kharkiv International Airport
BY	Belarus	BQT	Brest	Brest Airport
BY	Belarus	GME	Gomel	Gomel Airport
BY	Belarus	VTB	Vitebsk	Vitebsk Vostochny Airport

Table 10: Added airports to the network.

The ten airports in Russia are chosen for their size and they are all to the west of the Ural Mountains. The other airports are added according to their size and geo location.

As noted in the introduction, the flow pattern within EU* is generally unknown. In order to compile an approximate leg-database another approach has been taken;

1. All major connections within EU* has been identified.
2. Then the number of flights for each pair of airports has been recorded by type of plane. The number of flights corresponds to a weekday in January 2007.
3. The number of available passengers on the different plane types has then been identified and joined with the flight database.
4. The recorded potential passenger load has then been adjusted according to an average cabin-factor of 65% and adjusted according to seasonal variation.
5. Subsequently, the data has been transformed to an annual basis by multiplying with 365 days.

In total, data for 138 legs has been compiled, which by the assumption of symmetry amounts to 276.

Clearly, the absolute level of the EU* matrix is expected not to be precise and is likely to under-estimate the true level. For instance, charter travel and private flights are not included. However, the structure between the major cities will most likely be reasonable. In other words, we apply an additional weighting of the EU* matrix before any further estimation in order to account for under/over-representation.

In addition to this, several bounds and restrictions have been applied to the solution procedure. Including logical restrictions are used to update and quality control the matrices;

1. Air trips below 100 KM have been set to 0.

2. Legs are assigned logical minimum number of passengers per flight. This secures unrealistic low volumes on long flights (based on minimum frequency and airplane size)

3. legs are assigned logical maximum number of passengers per flight (based on assumptions on frequency and maximum airplane size per leg).

4. legs, within these bounds, are assumed to have no counts.

A constrained MPME matrix estimation method is then applied given these inputs.

3.3.2 Matrix adjustments of air passenger matrix

First, a thoroughly manual calibration of the air network was conducted with the objective to level-out major differences between passenger flows and airport counts. This was done altering connections from zones to airports and adjusting fares and transfer penalties within the airports.

Figure 6 and Figure 7 illustrate estimated flows versus counted flows before and after MPME estimation.

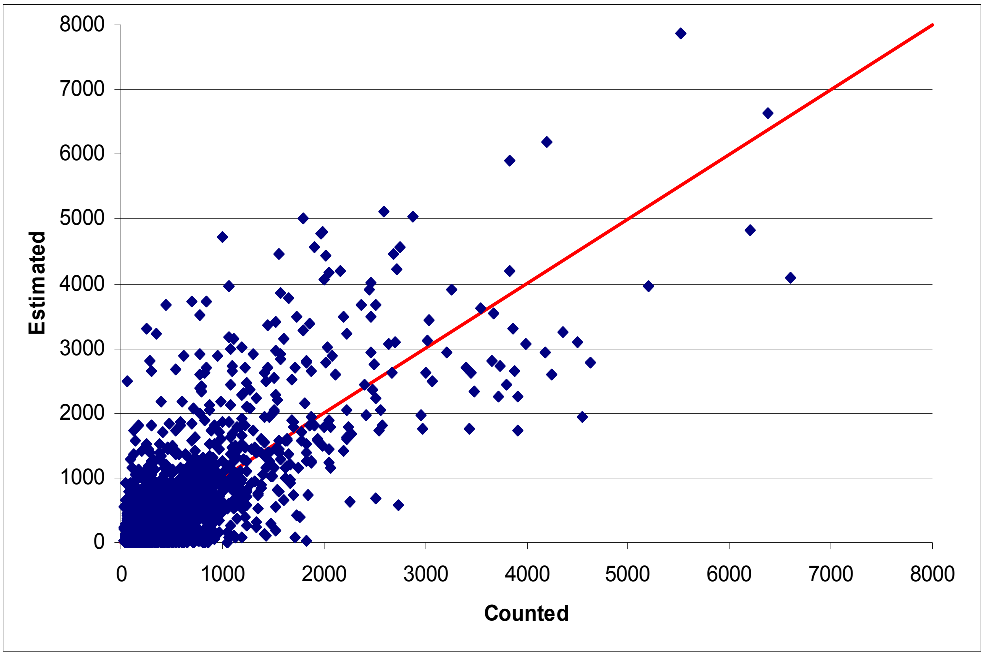

Figure 6: Modeled versus counted air passenger flows before MPME estimation

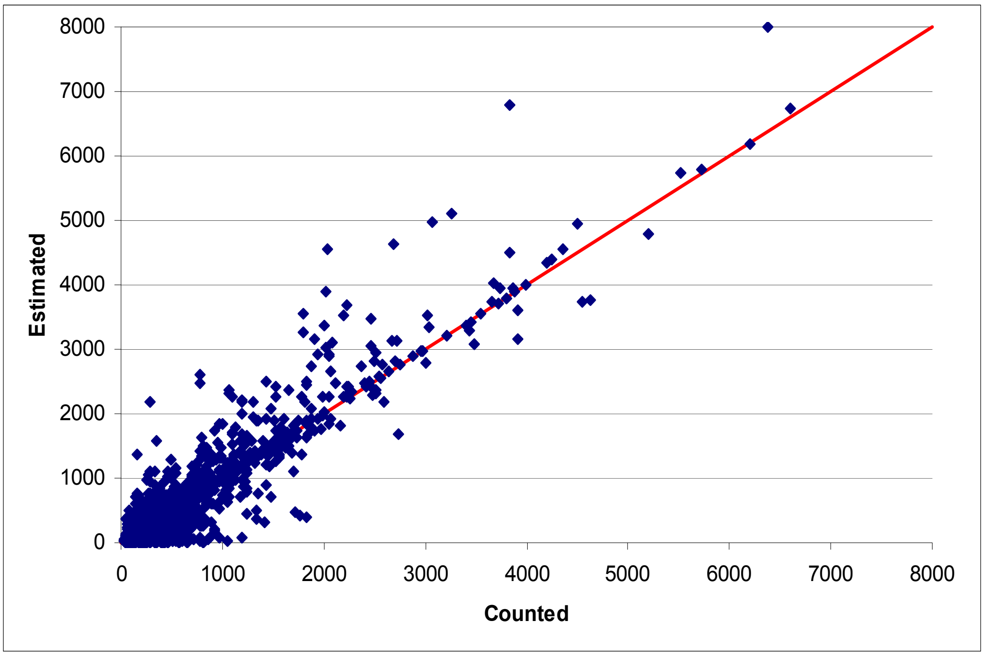

Figure 7: Modeled versus counted air passenger flows before MPME estimation

The air network contains few links compared with the road network. It has two consequences. First, the number of iterations can be reduced to fit the counted flows. Second, the estimated adjustments factors may for few zonal relations be very large or very small if the model is inaccurate.

The MPME estimation was completed in 5 iterative steps. After each iteration adjustments factors were manually checks and their values limited. In the first iteration, adjustment factors on cells with more than 10 trips per year were limited to values between 0.1

and 100. In the successive iterations adjustment factors were restricted to values within the range of 0.67 to 1.5 for cells with more than 10 trips per year and to values within the range of 0.2 to 5 for other cells in the matrix.

3.3.3 Post processing of air passenger matrix

The post processing of the MPME adjusted air passenger matrix has included:

- Removal of short distance trips
- Split of matrix by trip purpose
- Construction of symmetrical matrices
- Reformulation to GA-matrix

It is assumed that air trips below a travel distance of 100 km can be neglected. This includes intra zonal trips. We define travel distance as explained in Section 3.2.3.

The MPME adjustment was done without trip purpose segmentation to reduce computation time. Therefore, the post processing has to include a purpose split of the matrix. Three trip purposes are considered for air transport: home-business, home-private, and home-vacation. Due to a very limited number of commute trips by air they are added to the business segment in line with the existing passenger model of TRANS-TOOLS. First, a relative distribution among the three purposes were calculated per zonal relation based on the 2005-matrix produced by the existing TRANS-TOOLS model and converted to the new zonal structure. In addition, a default distribution per zone was calculated to take account of empty cells. Second, a comparison with the DATELINE Study indicated an underestimation of holiday trips and overestimation of business trips. Therefore, the purpose distribution was adjusted to be in accordance with the DATELINE Study.

After splitting into three purposes the purpose specific matrices were made symmetrical by applying the formula (2).

Finally, the matrices were rearranged into GA-matrices to be used in the passenger demand models similar to the procedure explained in Section 3.2.3.

3.4 Estimation of rail passenger matrix

The 1441-by-1441 base matrix is constructed in three steps. First, a country-level base matrix is constructed. Second, intra-country passenger flows are distributed into the 1441-zonal structure of TEN-CONNECT. Third, intra national (country-to-country) flows are distributed between zones of the new model.

Based on data from Eurostat supplemented by data from the World Bank, a country-level 42-by-42 base matrix is established. All entries for the 34 countries reporting to Eurostat where based on Eurostat. The internal traffic for the 8 remaining countries where based on World Bank data. Passenger flows between the 34 countries and the 8 countries are estimated from the 2005-forecast done by the existing TRANS-TOOLS model.

In the second step, the between country-level matrix is distributed to 1269 zones according to 2005-travel pattern forecasted by the in the existing TRANS-TOOLS model. Then, the matrix is further detailed to the new 1441 zonal structure using a single-constraint gravity model as explained in Section 3.2.1.

The third step considers intra national trips. Number of national trips by rail (excluded tram and metro) for 2005 is collected from Eurostat and World Bank. They are distributed among zones within the country based on a gravity model with population as attraction variable and distance[2] as travel impedance. While the inter zonal distance is calculated as the Euclidian distance, the intra zonal distance is estimated based on zonal area (A) and distance limits:

$$(4) \qquad 5\,\text{km} \le \frac{2}{3}\sqrt{\frac{A}{\pi}} \le 50\,\text{km}$$

It has been impossible to gather a solid dataset for rail passenger counts. In total only 678 counts were available. Furthermore they are clustered in geographically areas, mainly in central Europe (Germany, Czech Republic, Slovak Republic, and Hungary), Scandinavia and Turkey. A map with the counts crossing zone borders is shown in Figure 8. Because only 404 counts are available, and that these are clustered in specific geographically areas, it was concluded that a MPME adjustment could no be carried out properly.

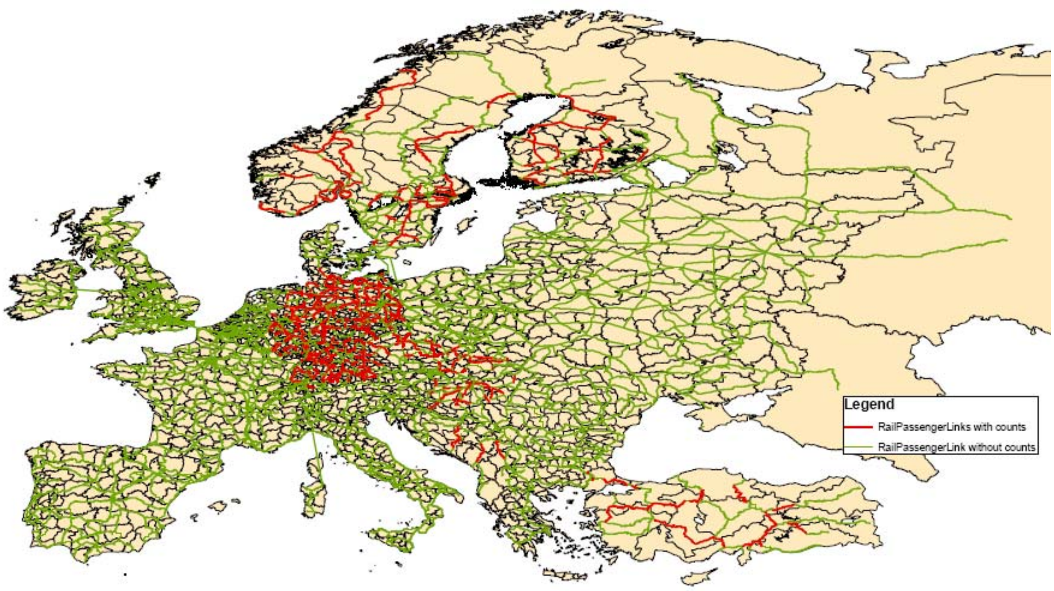

Figure 8: Rail passenger network with counts available for MPME adjustments.

The matrix adjustments for the rail passenger matrixes is therefore carried out as manually process where the matrix elements have been adjusted in regions to reproduce statistics on passenger km. The few counts available were also taken into consideration. The post processing of the train passenger matrix has also included:

- Split of matrix by trip purpose
- Construction of symmetrical matrices
- Reformulation to GA-matrix

Initially, the passenger matrix was split by trip purpose in accordance with the existing TRANS-TOOLS model. The procedure is similar to the purpose split of air passenger matrix described in Section 3.3.3.

After splitting into three purposes the purpose specific matrices were made symmetrical by applying the formula (2). Finally, the matrices were rearranged into GA-matrices to be used in the passenger demand models similar to the procedure explained in Section 3.2.3.

3.5 Estimation of bus passenger demand matrix

Travel by bus is included to capture the substitution effects between bus and train modes. Since there is no detailed information available about bus travel pattern, a matrix has initially been estimated on basis of the train passenger matrix (rail), car driver matrix (CD), and car passenger (CP) matrix,

$$T^{GA}_{ij,bus,p} = a_{car,p}\left(T^{GA}_{ij,cd,p} + T^{GA}_{ij,cp,p}\right) + a_{rail,p} T^{GA}_{ij,rail,p} \quad (5)$$

Shares have been estimated from the Danish Travel Survey and shown in Table 11.

Mode	Car share	Rail share
Home-business	0.007158	0.513768
Home-private	0.023447	0.865385
Home-vacation	0.022656	0.775191
Home-work	0.073548	0.252651

Table 11: Estimated shares of car and train matrices to calculate bus trips

Adjustments and correction have been applied to the roughly estimated bus passenger matrix to fit statistics on bus passenger km at country level.

3.6 Summary

Table 12 shows for an average day in 2005 the aggregated number of trips by mode and trip purpose covered by the TRANS-TOOLS model. In total, we have about 1,156 million trips per day of which 90% is carried out by car. Since the TRANS-TOOLS model covers a population of about 800 million, the average trip rate is 1.5 trips per day. While the trip rates in EU countries average about 2 to 3 trips per day, trip rates are much lower in some East European countries. A reason to low trip rates is due to the fact that not all modes are included in the TRANS-TOOLS model, e.g. walk, bike, tram and metro trips are not considered by the model. On the other hand, however, these modes contribute relative less to the number of kilometers produced.

Mode	Home-business	Home-private	Home-vacation	Home-work	Total
Car driver	30.602	225.032	157.892	245.062	658.589
Car passenger	8.481	161.043	139.142	68.212	376.878
Train	1.029	8.675	1.632	6.093	17.430
Bus	1.449	33.185	16.253	50.782	101.669
Air	0.177	0.056	1.091	0.000	1.324
Total	41.738	427.991	316.011	370.149	1155.890

Table 12: Million trips on an average day in 2005

4 Assignment Model

4.1 Introduction

This chapter describes the methodological specifications for the assignment model in TENCONNECT.

The main consideration of the methodological specification is to fulfil the following requirement to TENCONNECT;

- Fulfilment of contractual requirement
- Memory requirement
- Calculation speed

This has resulted in the following main challenges concerning the assignment modelling;

- The number of zones increases to 1441.
- The business demand matrices are split into commuting and business travel.
- The passenger demand models are transferred to GA models, i.e. trip generation for commuting is explained by the household and attraction by the workplace rather than by Origin and Destination (which is an erroneous causal relationship for the return trip). This also means that the assignment models should be transferred to GA, and trips are assigned out and back. This doubles the number of matrices/cells.
- A further issue is that the assignment models should work with generation-based utility functions, whereby preferences depends on the traveller (e.g. that the value of time and willingness to pay is higher for car drivers in Luxemburg than in Moldova).
- The assignment procedure must keep track of vehicle type with regard to fuel use (diesel versus petrol driven)
- Intra-zonal traffic must be included in the demand model, and pre-loaded traffic should depend on changes in intra-zonal traffic.
- The air assignment model should explicitly model access and egress mode.
- Revenue should be calculated in order to estimate the economic impacts.
- GDP influences VoT for each country, expressed in a VoTFactor
- GDP influences the VoT both in the base-year and future year
- Fuel price and fuel taxes should be modelled explicitly in the assignment model, and also in the feedback to the economic model.

These requirements increase the size of the matrices considerably beyond the capacity of 32 bit Microsoft Windows PC operating systems (3 GB per process). To solve this, the assignment software/algorithm has in addition been optimised in different ways.

The overall model flow consists of two main parts; a first run that creates Level of Services (LoS), and a second run where the demand model is run and the new demand is assigned again. In principle the second run could be repeated to secure convergence if second order effects are expected (e.g. relieved congestion in a road pricing scenario).

4.2 General modifications of the assignment models

All passenger assignment models are modified as described in the following. These changes reflect the requirements described in the introduction. It should be noted that this reflects the improvements by switching from an OD-based passenger demand model to a GA-based model.

As the project is not changing the freight mode choice and freight logistic model, the outputs of these models are still OD-based. This means that a number of the methodological improvements of the models are used in the passenger assignment models, but not in the freight assignment models. As an example, passenger cars use national VoT values, while trucks use an average European VoT value.

4.2.1 GA-based assignment

TENCONNECT uses generation/attraction (GA) assignment. This is in order to make sure that route choice for both the out-bound and homebound legs of longer trips are affected by national VoT values.

In practice a corresponding GA matrix is organised such that each matrix element designates both legs of a trip. Like:

FromZoneID	ToZoneID	Trips
2	5	23
5	2	17

Table 13: Matrix element.

These two matrix elements indicate that:

- Persons residing in zone 2 make 23 trips to zone 5 and back to zone 2. These trips are assigned by using VoT values appropriate for a person residing in zone 2.

- Persons residing in zone 5 make 17 trips to zone 2 and back to zone 5. These trips are assigned by using VoT values appropriate for a person residing in zone 5.

The assignment model accomplishes this without significant performance degradation through the use of backwards Dijkstra searches in the network.

The GA assignment also collects LoS information in a GA manner.

FromZoneID	ToZoneID	Length	Time	Cost
2	5	734.1	7.23	60
5	2	711.2	9.05	20

Table 14: Level-of-service matrix.

These two matrix elements indicate that:

- An average person residing in zone 2 that makes a trip to zone 5 and back to zone 2 have used 734.1 km, 7.23 hours, and € 60 in total for both legs of the trips.

- An average person residing in zone 5 that makes a trip to zone 2 and back to zone 5 have used 711.2 km, 9.05 hours, and € 20 in total for both legs of the trips.

This difference would indicate that persons residing in zone 2 to have a much higher VoT than persons residing in zone 5, since they elect to take a very quick route even if it is longer and incurs a toll. This might be a pay-motorway.

4.2.2 Origin-specific utility

The value-of-time estimates used in the model has been derived on the basis of three national studies (Denmark, Netherland, and UK) and has been transferred to other countries by mean of Purchase Power Parity (PPP).

4.2.3 Change in parameters in scenarios

In the future scenarios, GDP may change due to the scenario assumptions (defined by the user) or due to the results of CGEurope of the policy package (relatively to the future base-year scenario). These changes are then used to change the PPP's that are used in the assignment models.

A relationship between PPP (Purpose Power Parity) and GDP is made, which is used to forecast future PPP's based on the GDP growth.

4.2.4 Level-of-Service calculations for the demand models

The assignment models produce LoS matrices (e.g. time and cost) for trip purposes and modes. The passenger matrices are GA-based, and the freight matrices are OD-based. For each cell in the matrices, each type of LoS is averaged over the mode choice probabilities. The same is done within the assignment for each mode (given that several routes are used between the origin and destination).

The passenger demand models use each LoS component directly, while the trade model uses both the cost matrices and an overall weighted utility measure.

4.2.5 Interaction with CGEurope and the trade model

CGEurope receives LoS from the freight logistic model in TRANS-TOOLS. These LoS values seem to be quite unstable, meaning that small changes in a scenario may lead to huge (unrealistic) changes in the LoS. Naturally, this would also lead to unrealistic changes in CGEurope. To avoid this problem, TENCONNECT has changed the model flow significantly, as CGEurope now receives LoS from the assignment models.

For the passenger trips, this is averaged GA-based LoS (averaged over mode choices). For the freight transport this is averaged OD-based LoS (averaged over mode choices).

4.2.6 New trip purpose

TRANS-TOOLS has only one work-trip purpose. This has been split into business trips and commuting trips. The main efforts have been to establish new matrices and demand models. The assignment models need however also the extra trip purpose. These have been coded with utility functions that use parameters scaled relatively to business and private.

4.2.7 Calculation speed issues and convergence

Calculation Speed has significantly improved in the TENCONNECT assignment models due to

- Parallel processing.

- Calculation stops when destination is reached.
- Stochastic simulation per from zone can reduce the number of assignment iterations.

However, the model has much more zones and one more trip purpose. This adds calculation time. But the more zones also smoothes results in corridors, since it is now an origin based assignment procedure whereby prior large zones in Eastern Europe are now split and assigned separately.

The preference heterogeneity (random coefficients in the Mixed Probit assignment models) has been reduced, since some of this is now explained by the national-specific VoT correction. As some variation in preferences is now explained by this, the model become less “stochastic”. In order to improve convergence it has also been decided to reduce the variance of the error term in model. This seems to be necessary – especially for the long trips.

4.3 Intra-zonal traffic

TRANS-TOOLS does not include intra-zonal traffic. In TENCONNECT an effort has been made to include this in the modelling. This has been done by adding a short distance demand model for trips below 100 km. The long distance demand model describes trips over 100 km. The assignment model then uses a set of time-of-day factors to describe congestion in the rush-hours for trips below 250km.

4.3.1 Intra-zonal LoS (input to demand model)

To model intra-zonal trips, it is necessary to calculate intra-zonal LoS. This is not trivial, since the local network is not part of the model. The calculation of intra-zonal LoS must also consider that the majority of the trips are short, but there is a certain amount of longer trips. The intra zonal trip distance is estimated as a function of the zonal area (A) limited by minimum and maximum distances. The following expression is applied to estimate the average distance for passenger transport:

(3) $$d_{ii} = \min(3 + 0.2\sqrt{\frac{A_{ii}}{\pi}}, 12)$$

Given the average length is known, it is also needed to find average travel time. The average time (free flow as well as extra due to congestion) is calculated as a weighted average over the *preloaded* traffic at the individual links within a zone (weighted by length and traffic). One may also assume that a certain part of the intra-zonal trips are carried out on the local network, where the travel speed is lower. An assumption can then be that the average travel speed is 2/3 of the speed in the main model network. Travel times are then calculated as length / (2/3 x weighted averaged speed on model network).

If the links include tolls, etc., these are also calculated by the weighted average. This is not modified (divided by 2/3) since such tolls are often on a part of the network with much local transport (e.g. an urban tolling scheme).

4.3.2 Pre-loaded traffic

TENCONNECT uses an approach, where intra-zonal traffic has been pre-loaded (calibrated) in the base year on each link in the network. In the future year (scenario) the growth of the intra-zonal traffic (cell in the matrix) is used proportionally to forecast the pre-loaded traffic in the future year. This approach is used since the TENCONNECT network only includes the major roads, and a large part of the local (intra-zonal) traffic uses the local road networks that is not part of the model. Other European transport models assigns (loads) intra-zonal transport onto the model network within each zone by making a sub-model, and assigning between nodes here. However, if for example 75% of the intra-zonal traffic use local roads and 25% use the main roads in the model network, and all is assigned to this by such an approach, then this would overestimate the local transport on this network with 400%, whereby congestion would also be overestimated. It is assumed that using the average growth of intra-zonal transport is a better estimate of the growth of the share of transport that uses the main network.

If new roads are built in the scenario, local (pre-loaded) transport is estimated for this.

4.3.2.1 Base year notation

The new version of the TRANS-TOOLS model includes intra zonal passenger trips. Trips by passenger car are segmented into car drivers (CD) and car passengers (CP) The base year average occupancy rate for intra zonal trips is, since we have four trip purposes (p) in new model:

(1) $$B_{ii}^{0} = \frac{\sum_{i=p}^{4} CD_{iip}^{0} + CP_{iip}^{0}}{\sum_{i=p}^{4} CD_{iip}^{0}}$$

If we assume an average intra zonal trip distance (d), intra zonal passenger car km (CM) and passenger km (PM) can be calculated:

(2) $$CM_{ii}^{0} = d_{ii}CD_{ii}^{0}$$

$$PM_{ii}^{0} = CM_{ii}^{0}B_{ii}^{0} = d_{ii}CD_{ii}^{0}B_{ii}^{0}$$

The intra zonal trip distance is estimated as a function of the zonal area (A) limited by minimum and maximum distances. The following expression is applied to estimate the average distance for passenger transport:

(3) $$d_{ii} = \min(3 + 0.2\sqrt{\frac{A_{ii}}{\pi}}, 12)$$

The existing freight model produces truck trips at NUTS 2 level including intra zonal trips. Because they are converted into NUTS 3 zones before assignment, we know the number of intra zonal truck trips (G). If we assume that the average intra zonal trip distance is 20% longer for freight transport than for passenger transport, then we estimate intra zonal truck km (GM) as:

(4) $$GM_{ii}^{0} = 1.2 * d_{ii}G_{ii}^{0}$$

The preload base year table contains 2005 AADT figures for road links within a zone. Road links passing a zone border carry, in principle, exclusively inter zonal trips and therefore traffic is not preloaded. Since the preload figures are not split by vehicle type, we assume generally a split of 80% passenger cars and 20% goods vehicles (heavy and light goods vehicles). Busses are not considered, because their share is marginally on roads considered in the model. Intra zonal vehicle km (T) preloaded to the model network is then the product of AADT (x) and length (l) for links within the zone:

(5) $$T^0_{ii} = CT^0_{ii} + GT^0_{ii} = \sum_{j \in i} \left(0.8 * x^0_j l_j + 0.2 * x^0_j l_j\right)$$

Hence, passenger car traffic carried on local roads not included in the model is given by:

(6) $$CZ^0_{ii} = CM^0_{ii} - CT^0_{ii} = d_{ii} CD^0_{ii} - 0.8 * T^0_{ii}$$

Passenger km within a zone on local roads not included in the model is:

(7) $$PZ^0_{ii} = CZ^0_{ii} B^0_{ii}$$

For freight transport we follow a similar procedure to calculated local truck km on roads not considered by the mode:

(8) $$GZ^0_{ii} = GM^0_{ii} - GT^0_{ii} = 1.2 * d_{ii} G^0_{ii} - 0.2 * T^0_{ii}$$

If expression (8) gives a negative value due to inconsistency between demand matrices and preload assumptions, then it is set to zero. In the calibration of the passenger model, it is verified that expression (6) is non negative. Therefore,

(9) $$\begin{aligned} GZ^0_{ii} + GT^0_{ii} &\geq GM^0_{ii} \\ CZ^0_{ii} + CT^0_{ii} &= CM^0_{ii} \end{aligned}$$

4.3.2.2 Forecasting of preloaded traffic on existing links

The passenger demand model forecasts trips by mode and purpose which includes inter as well as intra zonal trips. Therefore, the ratio of predicted over base year intra zonal car driver trips may be computed (index P = prediction year):

(10) $$F_{CD,ii} = \frac{CD^P_{ii}}{CD^0_{ii}} \qquad CD^0_{ii} > 0$$

We also calculate a prediction factor for freight transport on basis of 2005 and predicted truck vehicle matrices:

(11) $$F_{G,ii} = \frac{G^P_{ii}}{G^0_{ii}} \qquad G^0_{ii} > 0$$

If base year number of trips is zero, then the prediction factor is set to 1. The average prediction factor for passenger and freight transport is:

(12) $$F_{ii} = 0.8 * F_{CD,ii} + 0.2 * F_{G,ii}$$

This factor is applied to estimate a future year table with preloaded traffic for all links existing in base year as well as in scenario year. The same factor is applied to all time, day, and directional figures in the preload table.

4.3.2.3 Forecasting of preloaded traffic on new links

If new roads are added or link connections changed compared with the base year network, preloaded volumes have to be estimated synthetically for those new or edited links crossing zonal borders. In the development of the TRANS-TOOLS model, preloaded traffic were initially derived from a formula using information about geographical location, road type etc.

It is suggested that AADT on new links is estimated in a more simple approach utilizing the level of traffic on existing road segments within the zone. After updating the preloads for existing road links to the future year, average AADT values stratified by link type and number of lanes is created for each zone. Let us assume we have a set n of existing links j within zone i and stratum h, the AADT on the new link j' within the zone then equals the average of the corresponding set of links:

(13) $$x^P_{j',h} \approx \frac{1}{n_{j,h}} \sum_{j \in i} x^P_{j,h}$$

The AADT estimates are subdivided into time and day segments by standard split factors. For days except weekdays outside summer an evenly directional split is computed. This is, however, not correct for am and pm peak flows. Therefore, an artificial directional split on the new road segments are computed based on assignment of available am and pm peak vehicle matrices.

4.3.2.4 Future year travel mileages

Future year intra zonal vehicle and passenger km based on demand matrices are

(14) $$CM^P_{ii} = F_{CP,ii} CM^0_{ii} = F_{CP,ii} d_{ii} CD^0_{ii}$$

$$PM^P_{ii} = CM^P_{ii} B^P_{ii}$$

$$GM^P_{ii} = F_{G,ii} GM^0_{ii} = F_{G,ii} 1.2 * d_{ii} G^0_{ii}$$

Preloaded vehicle km within a zone is:

(15) $$T^P_{ii} = CT^P_{ii} + GT^P_{ii} = \sum_{j,j' \in i} \left(x^P_j l_j + x^P_{j'} l_{j'} \right)$$

Passenger car traffic carried on local roads not included in the model is given by:

(16) $$CZ^P_{ii} = CM^P_{ii} - CT^P_{ii} = d_{ii} CD^P_{ii} - 0.8 * T^P_{ii}$$

Passenger km within a zone on local roads not included in the model is:

(17) $$PZ^P_{ii} = CZ^P_{ii} B^P_{ii}$$

Local truck km within a zone on roads not considered by the model is:

(18) $$GZ^P_{ii} = GM^P_{ii} - GT^P_{ii} = 1.2 * d_{ii}G^P_{ii} - 0.2 * T^P_{ii}$$

If expression (16) and (18) gives a negative value due e.g. preloads on new roads, then local transport values are set to zero. Therefore, we generally have:

(19) $$GZ^P_{ii} + GT^P_{ii} \geq GM^P_{ii}$$
$$CZ^P_{ii} + CT^P_{ii} \geq CM^P_{ii}$$

4.3.2.5 Implementation of preload for new roads

In the base year calibration of the model, a 2005 preload table is created. This table must be available and copied to all future scenarios because prediction factors are computed relative to the base year preload table.

After updating preloads on existing road links to future year, average AADT values are computed for links entirely within a given zone as shown in Table 15. The following link type attributes are used in the network:

1: Motorway
5/6: Rural roads
9: Urban roads
90: Ferry routes

We do not assign preloads to ferry routes. If the user has added a new road link with a link type and number of lanes that do not match the distribution of existing link types and lanes within the given zone, then the closest neighbour is chosen. The first choice if no exact match is possible would be within the same link type. The second choice is to apply the default average over all links within the zone. If preloaded traffic does not exist in base year for a given zone, default AADT is simply set to zero.

Link type	No. of lanes	Average AADT
1	4	
	>4	
	All	
5/6	2 or 3	
	≥4	
	All	
9	2 or 3	
	≥4	
	All	
Default	All	

Table 15: Average AADT figures applied to new links within a given zone

A new attribute is added to the road network to indicate zonal border crossing, because links crossing zonal borders are not preloaded. When the AADT estimates have been done for new intra zonal road links, the daily volumes are derived by applying the following standard factors to the estimated AADT value:

Weekdays outside summer:	1.12
Weekdays within summer:	1.00
Busy holidays:	0.90
Weekends:	0.80

Traffic on weekdays outside summer is divided into three day periods: am peak (7-9 am), pm peak (3-6 pm), and off-peak period. We assume a relative split of 30% in am peak, 35% in pm peak, and 35% in off-peak.

In am and pm peak periods the directional split of traffic on new links can be estimated from a provisional assignment. Since results only are used on few links for directional split any available preload table and am respectively pm vehicle matrices, may basically be used. To save time a limited number of iterations should be conducted to calculate the split factors for the two time periods.

4.4 The passenger air transport assignment model

The passenger air transport assignment model is an extended version of the model from TRANS-TOOLS.

At an upper level, it models access and egress links as zone connectors. The network is modelled in the exact same way as in TRANS-TOOLS.

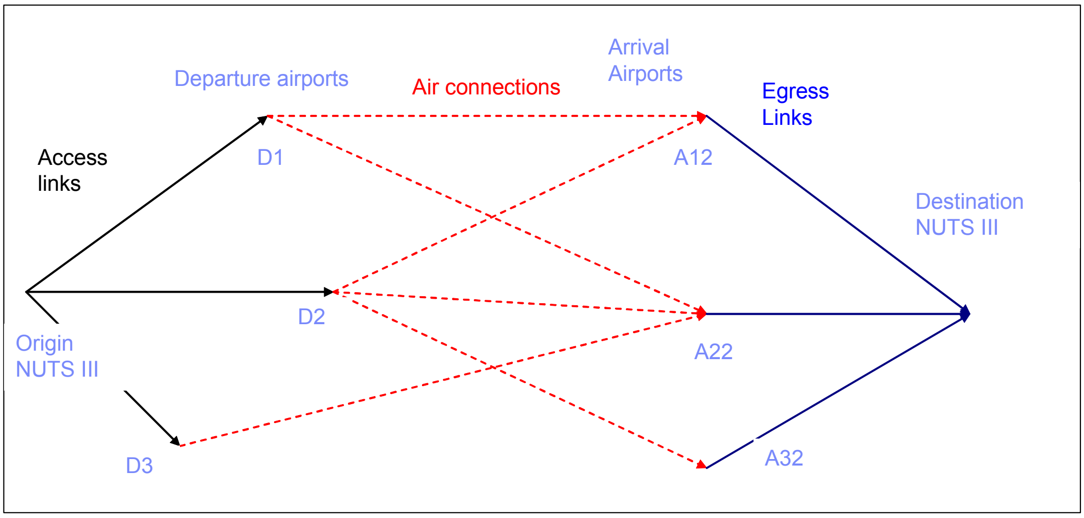

Figure 9: Air assignment overview.

At the lower level, however, the flows from a given connector is split into road and rail flows (binary logit model) and then assigned along the path from the origin zone to a pseudo destination zone that represents the airport. At this level, short connectors are attached from the origin zone to the network (the same as are used for road and rail traffic assignment) and from the network to the airport (new connectors – typically auto generated for connecting the airport to the nearest node in the road and rail network).

The LoS along the paths for road (connector, sequence of road links, connector) and rail are averages according to the likelihood from the binary logit model. This is used as LoS for the connector in the main air route choice model connecting the zone directly to the airport.

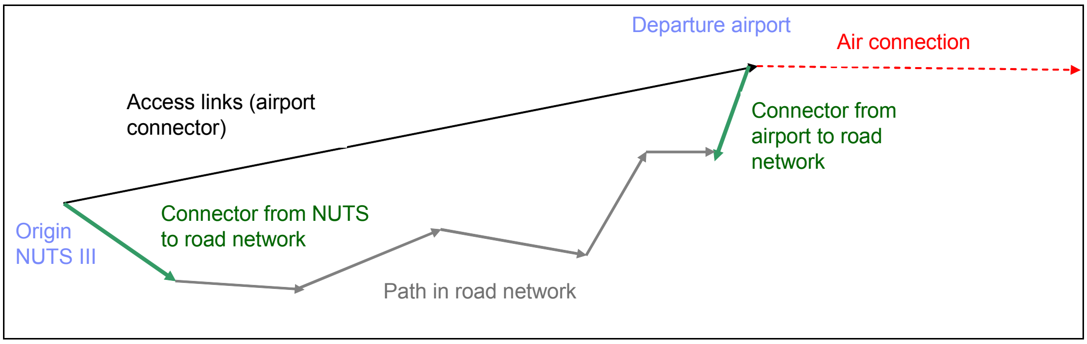

Figure 10: Airport access as connecter and through the road network.

Preconditions

1. All airports are attached with pseudo centroids, which are connected to the road and rail network by zone connectors. The pseudo centroids are placed in the same points (coordinates) as the nodes representing the airports in the air model network. However, the centroids are stored in the road and rail networks.
2. There exists a table with the following columns AirportID, zoneID.
3. Each airport is connected to a set of zones by zone connectors (the same way as in the TRANS-TOOLS model)

These data are normally fixed. However, if a new airport is created, then this airport needs to be connected to (NUTS3) zones by a set of zone connectors, a pseudo centroid needs to be created, and it needs to be connected to the road and rail network with a zone connector (only one for each).

The access and egress link matrices will be very thin, since they only contain the connectors. As an example, a full road matrix contains approx. $1{,}540^2$=2,371,600 cells, whilst the present TRANS-TOOLS model contains 7,437 connectors which equals the number of access mode cells (and similarly 7,437 connectors corresponding to egress mode cells). The Dijkstra search in the surface road and rail networks is also very fast, since all connector destinations are very close to the airport.

The air model runs as follow in the first iteration of the model (prior to the demand model);

1. **Create Pseudo Surface Mode Matrix, SM_{ij}.** An OD matrix with all cell values=1 is created based on the list of zone connectors, i.e. cells where there is a direct zone connector between the zone and the airport are assigned the value of 1. This matrix is very thin, since each airport is only connected to its hinterland. The matrix is asymmetric, since only from zones to airports are assigned.
2. **Rail LoS.** The matrix is assigned onto the rail network using an all-or-nothing assignment. A sparse rail LoS matrix is hereby created. The calculations are extremely fast, since all destination zones from a given airport is very close to this, and the Dijkstra search stops when a destination airport is reached.
3. **Road LoS.** The matrix is assigned all-or-nothing onto the road network using travel times from a prior assignment (equal to free flow times, if the model is run for the first time). A sparse road LoS matrix is hereby created. The calculations should be extremely fast, since all destination zones from a given airport is very close to this, and the Dijkstra search stops when a destination airport is reached.
4. **Access Mode Choice Model**. An access mode choice model (a simple binary logit) is run using the Rail and Road LoS matrices. This creates an average access mode LoS for each connection from a zone to an airport. This LoS is transferred to the zone connector.

5. **Egress Mode Choice Model**. An egress mode choice model is run using the Rail and Road LoS matrices. As the LoS is assumed symmetric, the LoS for a given egress zone connector is assumed to be the same as for access, however, the alternative specific constants may be different in the mode choice model. This creates an average egress mode LoS for each connection from an airport to a zone. This LoS is transferred to the zone connector.
6. **Air Assignment model**. The same air assignment model as in TRANS-TOOLS is run. The model is run separately for each trip purpose (since each trip purpose has different utility functions). The GA matrices are assigned.
7. **Creation of surface mode matrices.** For each trip purpose, the access mode and egress mode traffic on each connector is split into road and rail traffic using a binary logit model using the LoS from steps 2 and 3. The access and egress flows from the airport connectors are transferred into the centroid road and rail matrices. An access leg – e.g. to Copenhagen Airport – represent e.g. a person from Copenhagen travelling out or home. The same egress leg to Copenhagen Airport represents, e.g. an Italian person going out or home from a visit to Copenhagen.

The airport surface mode matrices are subsequently assigned onto the network together with the other road and rail matrices. This creates a new set of surface mode LoS matrices which is used as input to the demand models. At this stage separate LoS matrices have been produced for the airport surface trip purposes, which only contain the cells corresponding to the airport connectors.

Following the run of the demand models, new air transport matrices are produced. These are used as input to the final air assignment model.

However, the above step 1-3 can now be skipped, since LoS's have already been calculated for each trip purpose, while steps 4 and 5 may use purpose-specific access and egress mode choice models.

4. **Access Mode Choice Model**. An access mode choice model (simple binary logit) is run using the Rail and Road LoS matrices produced for that trip purpose in the prior assignment. This creates an average access mode LoS for each connection from a zone to an airport. This LoS is transferred to the zone connector.
5. **Egress Mode Choice Model**. An egress mode choice model is run using the Rail and Road LoS matrices, produced for that trip purpose in the prior assignment. This creates an average egress mode LoS for each connection from an airport to a zone. This LoS is transferred to the zone connector.
6. **Air Assignment model**. The same air assignment model as in TRANS-TOOLS is run. The model is run separately for each trip purpose (since each trip purpose has a different utility function) and with the specific LoS for the access and egress connectors.
7. **Creation of surface mode matrices.** For each trip purpose, the access mode (generation leg of a GA matrix) and egress mode (attraction leg of a GA matrix) traffic on each connector is split into road and rail traffic using a binary logit model using the LoS from for that trip purpose (from the prior assignment). The access and egress flows from the airport connectors are transferred into the centroid road and rail matrices.

4.4.1 Choice functions, air assignment

The route choice utility functions for air assignment in Euro are as follows (transferred from TRANS-TOOLS);

Business:
GenCost = 1 * LinkCostBusiness + 0.81 * VoTFactor(FromZone) + ConTime(min) + 1 * TotalConCost + 0.81 * VoTFactor(FromZone) * LinkTime(min) + 1.215 * VoTFactor(FromZone) * TransferTime(min) + 1.215 * VoTFactor(FromZone) * HeadwayTime(min)

Private:
GenCost = 1 * LinkCostPrivate + 0.23 * VoTFactor(FromZone) + ConTime(min) + 1 * TotalConCost + 0.23 * VoTFactor(FromZone) * LinkTime(min) + 0.345 * VoTFactor(FromZone) * TransferTime(min) + 0.345 * VoTFactor(FromZone) * HeadwayTime(min)

Holiday:
GenCost = 1 * LinkCostVacation + 0.23 * VoTFactor(FromZone) + ConTime(min) + 1 * TotalConCost + 0.23 * VoTFactor(FromZone) * LinkTime(min) + 0.345 * VoTFactor(FromZone) * TransferTime(min) + 0.345 * VoTFactor(FromZone) * HeadwayTime(min)

Commuters:
In the TenConnect model, commuters do not travel by air.

Time and cost are added on cross border crossing connectors reflecting that passengers may prefer local airports rather than airports in neighbouring countries.

Variance on preferences is 10%. Error term is 5%. Stochastic parameters are simulated for each fromzone.

Stochastic variance is only used on ConTime, LinkTime, TransferTime and HeadwayTime.

The first assignment uses 100 iterations. The second assignment uses 200 iterations.

4.4.2 Mode choice models for access and egress

A split is defined (calibrated) between rail- and “other” where “other” is described by a logit model per connector per category per access/egress, e.g. 40% use rail and 60% depends on a logit model for access for a specific category for a specific airport connector. The lower split factor for egress in the table below reflects that travellers often do not have access to car at the destination zone, but only at their home zone.

The split factors are defined in a table that is calibrated to fit counts. This is defined as follows;

ConnectorID	Category	Access (boolean value that shows whether it is access or egress) 1= access, 0 = egress	Split factor (defines the share that uses the logit model)	β
32	1	0	0.52	0.72
32	1	1	0.6	0.72

Table 16: Split factors for connectors.

The logit model then splits the OD cell in travellers by rail (public transport) and road. The total number of rail access/egress is then summed over the predetermined rail passengers and the logit model defined rail passengers. The logit model produces road passengers as direct result.

As input to the logit model is used a table with β values per combination of category and connectors:

$$P_{ijk} = \frac{e^{-\beta C_{ijk}}}{\sum_m e^{-\beta C_{ijm}}}$$

$$T_{ijk} = T_{ij} P_{ijk}$$

Assume that the formula gives P_{rail} = 30%.
In the example from before, the logit model split the traffic on 30% (of the 60%) use rail and 70% (of the 60%) used road, so 40% + 0.3*60% = 58% use rail in total and 0.7*60% = 42% use road.

The model flow is illustrated in the figure below;

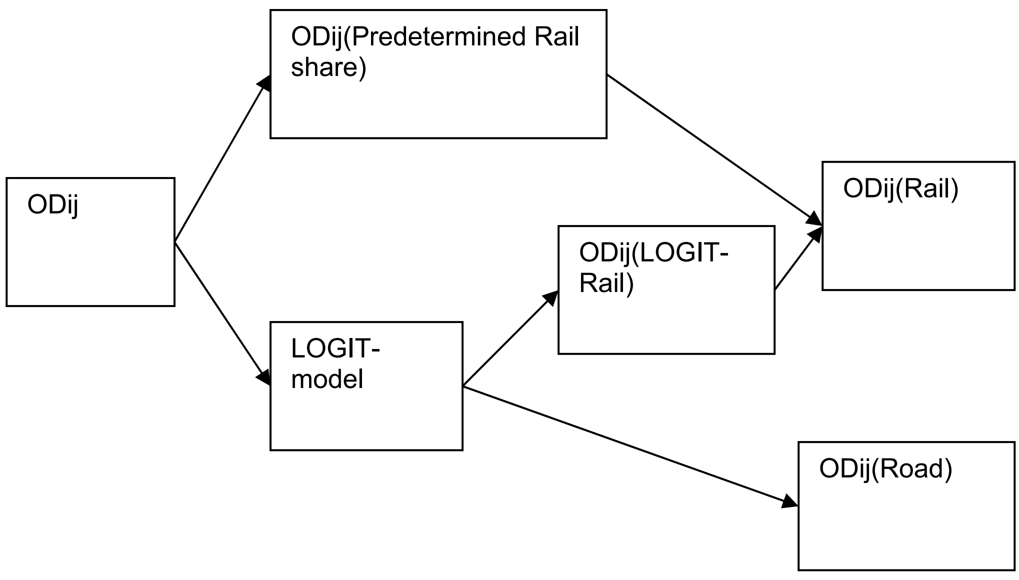

Figure 11: Model flow.

4.4.3 Connector LoS

One overall weighted LoS per connector per category per access/egress is calculated based on the Rail LoS and Road LoS from the prior assignment (the first run based on pseudo assignments) and their respective share of the overall traffic. From the example above, LoS for access for the specific connector and category becomes 0.58*RailLoS + 0.42*RoadLoS.

4.5 Road Traffic Assignment

The road traffic assignment models have been changed in a number of ways to reflect the required improvements in TENCONNECT.

4.5.1 Fuel and vehicle types

In the load on the roads for impact modelling, it is possible to distinguish between the numbers of diesel and petrol cars. These are based on national factors for the share of the two car types, and collected by so-called filter criteria, i.e. that Spanish cars visiting France is using the Spanish split-factors also when driving in France (rather than in TRANS-TOOLS, where all cars in France were assumed to be French).

4.5.2 Revenue generation

The model now explicitly calculates revenue of different taxation and toll systems. These are split into;

1. Private tolls that are used for construction and maintenance costs (e.g. French private motorways, Danish fixed links, ferries).
2. Urban toll rings (e.g. London) and tolling systems (e.g. German Maut).
3. Extra fuel tax beyond normal average value-added tax.

The factors are defined for both passenger cars and trucks.

The revenues are collected at NUTS3 level.

The purpose for this split is that the traffic model now explicitly can forecast the traffic impacts of different policies. A further benefit is that CGEurope can calculate the impact of the share of revenue that is recycled into a national or regional economy where it can be used to lower income taxes. This is the case for bullet 2 and 3 above.

4.5.3 Split into time periods

The demand model calculates an average daily traffic. This is split into time periods in order to model congestion. This split can be done in a more refined manner in TENCONNECT, since the passenger demand is GA-based (whereby one knows that a commuting trip mostly is going out in the morning and returning in the afternoon).

The freight models still use OD assignment.

After being split, the passenger model uses OD assignment for the weekday peak hour models. As these trips are mainly short trips (e.g. commuting), it is reasonable to assume that the VoT is approximately the same in the zone of origin and destination, whereby the OD and DO trips can use the origin and destination VoT rather than the VoT from the GA matrix for both.

To split the demand into main types of days, the following split factors are used;

Type of day	Number of days	Hours per period	Trip purpose					Short Distance (preload)
			Commuting	Business	Tourism	Private	Trucks	
Weekdays outside summer	200	See below	1,5	1,55	0,3	1	1,2	1,12
Weekdays within summer	35	12	0,9	0,8	1,8	1	1	1
Busy holydays	20	12	0,2	0,1	3	1	0,8	0,9
Weekends	110	12	0,27	0,23	1,65	1,00	0,70	0,80
Total	365							

Table 17: Demand split over the week.

Then for the short trips weekdays outside summer, the trips are further split into time of day using the following factors for short trips below 250 km;

Split of weekdays outside summer, short trips (<100 km)		Commuting	Business	Tourism	Private	Trucks	0
Outbound							
7-9 am peak	2	65	40	20	10	15	
3-6 am peak	3	0	10	15	20	15	
Off-peak	10	35	50	65	70	70	
Homebound							
7-9 am peak	2	0	0	0	0	5	
3-6 am peak	3	44	45	35	35	30	
Off-peak	10	56	55	65	65	65	
Averaged (OD-elements)	Note that GA passenger model can directy be split into						
7-9 am peak	outbound and homebound					10	20
3-6 am peak	The freight model is just OD, and we here assume symmetrícy					22,5	30
Off-peak	The short distance preload is just link-based split-factors					67,5	50

Table 18: Time-of-day split for short trips.

Finally the air access and egress trips by road have to be split into time periods, since these trips are usually quite short. The following factors are used for this.

Air access/egress by car		Business	Tourism	Private	
Outbound, access link					
7-9 am peak	2	45	15	15	
3-6 am peak	3	10	15	15	
Off-peak	10	45	70	70	
Outbound, egress link					
7-9 am peak	2	25	15	15	
3-6 am peak	3	10	15	15	
Off-peak	10	65	70	70	
Homebound, access link					
7-9 am peak	2	5	5	5	
3-6 am peak	3	50	25	25	
Off-peak	10	45	70	70	
Homebound, egress link					
7-9 am peak	2	0	0	0	
3-6 am peak	3	30	35	35	
Off-peak	10	70	65	65	

Table 19: Time-of-day split for access and egress trips.

4.5.4 Model specification in time-periods

Unlike in TRANS-TOOLS, all time periods run with full Stochastic User Equilibrium. This reflects that e.g. the holiday season may result in congestion on certain parts of the network in certain regions of Europe.

4.5.5 Choice function

The following choice function is used within the model;
Business:
GenCost = 3.25 * FuelCostPC + 0.598 * VoTFactor(FromZone) * FreeFlowTime(min) + 0.9802 * VoTFactor(FromZone) * CongestedTime(min) + 1 * LinkCostPC + 0.0598 * VoTFactor(FromZone) * FerrySailingTime + 0.9802 * VoTFactor(FromZone) * FerryWaitingTime

Private:
GenCost = 1 * FuelCostPC + 0.1392 * VoTFactor(FromZone) * FreeFlowTime(min) + 0.2168 * VoTFactor(FromZone) * CongestedTime(min) + 1 * LinkCostPC + 0.0136 * VoTFactor(FromZone) * FerrySailingTime + 0.2168 * VoTFactor(FromZone) * FerryWaitingTime

Holiday:
GenCost = 1 * FuelCostPC + 0.0928 * VoTFactor(FromZone) * FreeFlowTime(min) + 0.1448 * VoTFactor(FromZone) * CongestedTime(min) + 1 * LinkCostPC + 0.0096 * VoTFactor(FromZone) * FerrySailingTime + 0.1448 * VoTFactor(FromZone) * FerryWaitingTime

Commute:
GenCost = 1 * FuelCostPC + 0.1192 * VoTFactor(FromZone) * FreeFlowTime(min) + 0.1568 * VoTFactor(FromZone) * CongestedTime(min) + 1 * LinkCostPC + 0.01192 *

VoTFactor(FromZone) * FerrySailingTime + 0.1568 * VoTFactor(FromZone) * FerryWaitingTime

Trucks:
GenCost = 5 * FuelCostPC + 0.528 * VoTFactor(FromZone) * FreeFlowTime(min) + 0.732 * VoTFactor(FromZone) * CongestedTime(min) + 1 * LinkCostTR + 0.052 * VoTFactor(FromZone) * FerrySailingTime + 0.732 * VoTFactor(FromZone) * FerryWaitingTime

The periods AM, PM and OP use 25 iterations, DA use 15, the other periods use 10 iterations.

In the second assignment AM, PM and OP use 50 iterations, DA uses 30, the other periods use 20 iterations.

AM is 2 hours long, PM 3 hours, OP 15 hours and DA, WD, WE, HO are 20 hours long.

Variance on preferences is 2.5%. Error term is 2.5%.

Stochastic simulation is only used on FreeTime, CongestedTime, FerrySailingTime and FerryWaitingTime. The new parameters are simulated for each FromZone

4.5.5.1 LinkCost and FuelCost

LinkCostPC and LinkCostTR are calculated immediately before assignment.
LinkCostPC = Length(km) * TollCostPC + GenericCostPC
LinkCostTR = Length(km) * TollCostTR + GenericCostTR

TollCost is ferry prices and toll on roads, which a user can specify.

GenericCost is a general price for using the motorways in Austria, Switzerland and other countries.

FuelCostPC originates from fuel prices in 2005 for each country. Users can change this.

The road network includes a field to specify how much of the LinkCost generates revenue to the public budget. The field is called PublicBudgetShare. If the value is 0, the entire price (from TollCost and GenericCost fields) is used by a private entity to operate the link. If the value is 1, the entire price is recycled as revenue to the public budget. Ferries should have a value of 0. Toll roads are set to have a value of 1. The user can set any value between 0 and 1.

Fuel tax also generates revenue to public budget. This is calculated using a fixed tax share based on the fuel prices and fuel taxes in 2005.

The total public budget revenue for a road is calculated in a field called PublicRevenue which is collected in LoS.

4.6 Rail Traffic Assignment

The rail traffic assignment now uses zone-specific VoT factors (as the other assignment models), it also runs as a GA assignment, and it has been recalibrated to fit counts better. The utility functions are stated as follow;

Business:
GenCost = 0.2 * Length(km) + 0.597 * VoTFactor(FromZone) * FreeFlowTime(min)

Private:
GenCost = 0.15 * Length(km) + 0.148 * VoTFactor(FromZone) * FreeFlowTime(min)

Holiday:
GenCost = 0.15 * Length(km) + 0.109 * VoTFactor(FromZone) * FreeFlowTime(min)

Commute:
GenCost = 0.1 * Length(km) + 0.37 * VoTFactor(FromZone) * FreeFlowTime(min)

Variance on preferences is 5%. Error term is 3%.

The model runs with 100 iterations (due to origin-based assignment, this is reduced compared to TRANS-TOOLS) and in a stochastic version. There are no capacity restrictions. The rail cost is fixed and not related to the route choice. In the second assignment, 200 iterations are used.

4.7 Rail Freight Assignment

Rail Freight is assigned as OD like in TransTools.

The utility function:
GenCost = 1 * Length(km) + 0.86 * FreeFlowTime (min)

Variance on preferences for FreeFlowTime is 10%. Error term is 5%

First assignment uses 25 iterations, second assignment uses 50 iterations.

4.8 Inland Water Ways

IWW freight is assigned as 5 OD assignments, one for each CEMT.

The utility function:

CEMT2:
GenCost = 1 * Length(km) + 0.46 * FreeFlowTime (min)

CEMT3:
GenCost = 1 * Length(km) + 0.55 * FreeFlowTime (min)

CEMT4:
GenCost = 1 * Length(km) + 0.62 * FreeFlowTime (min)

CEMT5:
GenCost = 1 * Length(km) + 0.65 * FreeFlowTime (min)

CEMT6:
GenCost = 1 * Length(km) + 0.67 * FreeFlowTime (min)

The assignment is All-or-Nothing; there is no variance on preferences or error term. First assignment uses 1 iteration, second assignment also uses 1 iteration.

4.9 Freight considerations

4.9.1 Fixed ports

Unlike in TRANS-TOOLS, a NUTS2 zone can now have more than one fixed port, which is used in the disaggregation from NUTS2 to NUTS3 for assignment.

4.10 Level of Service (LoS)

For each assignment a set of LoS data is calculated. The specific LoS are outlined below.

4.10.1 Passenger Road

LoS Field	Explanation
GenCountryID	The country in which the tours are generated. A tour always consists of an outbound trip and a return trip. Also: This is the country in which the persons making the tours reside
TripPurpose	ID for passenger trip purpose (1-4)
vehicleKm	Total vehicle-km for all tours made by persons residing in the country
personFreeTime	Total driving time excluded congested time for all tours made by persons residing in the country (person-hours)
personCongTime	Total congested driving time for all tours made by persons residing in the country (person-hours)
personFerrySailingTime	Total sailing time for all tours made by persons residing in the country (person-hours)
personFerryWaitingTime	Total waiting time for all tours made by persons residing in the country (person-hours)
vehicleTollCost	Total toll costs including ferry costs for all tours made by persons residing in the country (Euro)
vehicleFuelCost	Total fuel costs for all tours made by persons residing in the country (Euro)

4.10.2 Passenger Rail

LoS Field	Explanation
GenCountryID	The country in which the tours are generated. A tour consists of an outbound trip and a return trip. Also: This is the country in which the persons making the tours reside
TripPurpose	ID for passenger trip purpose (1-4)
personAccessEgressLength	Total Access-Egress length for all tours made by persons residing in the country (person-km)
personAccessEgressTime	Total Access-Egress time for all tours made by persons residing in the country (person-hours)

personOnboardLength	Total Onboard length for all tours made by persons residing in the country (person-km)
personOnboardTime	Total Onboard time for all tours made by persons re-siding in the country (person-hours)
personFerrySailingTime	Total ferry sailing time for all tours made by persons residing in the country (person-hours)

4.10.3 Passenger Air

LoS Field	Explanation
GenCountryID	The country in which the tours are generated. A tour consists of an outbound trip and a return trip. Also: This is the country in which the persons making the tours reside
TripPurpose	ID for passenger trip purpose (1-4)
personAccessEgressLength	Total Access-Egress length for all tours made by per-sons residing in the country (person-km)
personAccessEgressTime	Total Access-Egress time for all tours made by per-sons residing in the country (person-hours)
personOnboardLength	Total Onboard length for all tours made by persons residing in the country (person-km)
personOnboardTime	Total Onboard time for all tours made by persons re-siding in the country (person-hours)
Price	Total fares for all tours made by persons residing in the country (Euro)
personHeadwayTime	Total waiting time due to service frequency for all tours made by persons residing in the country (person-hours)
personTransferTime	Total waiting time for other reasons than service fre-quency for all tours made by persons residing in the country (person-hours)

4.10.4 Generalised cost for passengers

The generalised costs for passengers for each mode are calculated as using the utility functions for that mode. The generalized cost is then a weighted average over the modes using the number of persons using the modes as weights.

4.10.5 Freight Road

LoS Field	Explanation
OriginCountryID	The origin country for the trips (one leg only)
NSTR	NST/R 1 digit commodity group, with crude oil separate
vehicleKm	Total vehicle-km for all trips originating in the country
vehicleFreeTime	Total driving time excluded congested time for all trips originating in the country (vehicle-hours)
vehicleCongTime	Total congested driving time for all trips originating in the country (vehicle-hours)
vehicleFerrySailingTime	Total sailing time for all trips originating in the country (vehicle-hours)
vehicleFerryWaitingTime	Total waiting time for all trips originating in the country (vehicle-hours)

TollCost	Total toll costs including ferry costs for all trips originating in the country (Euro)
tonKm	Total ton-km for all trips originating in the country
tonFreeTime	Total driving time excluded congested time for all trips originating in the country (ton-hours)
tonCongTime	Total congested driving time for all trips originating in the country (ton-hours)
tonFerrySailingTime	Total sailing time for all trips originating in the country (ton-hours)
tonFerryWaitingTime	Total waiting time for all trips originating in the country (ton-hours)

4.10.6 Freight Rail

LoS Field	Explanation
OriginCountryID	The origin country for the trips (one leg only)
NSTR	NST/R 1 digit commodity group, with crude oil separate
tonAccessEgressLength	Total Access-Egress length for all trips originating in the country (ton-km)
tonAccessEgressTime	Total Access-Egress time for all trips originating in the country (ton-hours)
tonOnboardLength	Total Onboard length for all trips originating in the country (ton-km)
tonOnboardTime	Total Onboard time for all trips originating in the country (ton-hours)

4.10.7 Freight Waterways

LoS Field	Explanation
OriginCountryID	The origin country for the trips (one leg only)
NSTR	NST/R 1 digit commodity group, with crude oil separate
tonAccessEgressLength	Total Access-Egress length for all trips originating in the country (ton-km)
tonAccessEgressTime	Total Access-Egress time for all trips originating in the country (ton-hours)
tonOnboardLength	Total Onboard length for all trips originating in the country (ton-km)
tonOnboardTime	Total Onboard time for all trips originating in the country (ton-hours)

4.10.8 Generalised cost for Freight

The generalized costs for freight for each mode are calculated as follows:

Rail freight:
For an average load of 325 tonnes, we get 5 € pr shipment.
VOT = 5 * 0.86 * 60 = 258 Euro pr. shipment if T is hours.
GenCost = (C*Load + VoT*T)/Load = C + (VOT/Load)*T

Road freight:
GenCost uses the utility functions for that mode.

IWW:
Average load for CEMT4 is 850 tonnes.
VOT = 4.34 * 0.62 * 60 = 161 Euro pr. shipment if T is hours
GenCost = C + (161/850)*T
C is the cost for a shipment taken from the LOS. The load-factor for the commodity group is used to get the cost pr ton.

The generalised cost is then a weighted average over the modes using the number of tonnes using the modes as weights.

4.10.9 Other

LoS data for freight sea transport also exists. This is not produced by assignment.

In the passenger model a bus mode exists. LoS for bus is inferred from the road LoS data by the passenger models.

4.11 Summary

To summarise, the new TENCONNECT assignment model is an improved version of the TRANS-TOOLS assignment model. The note also serves to clarify exactly how the two assignment models differ, i.e. GA-based versus OD-based assignment, national VoT values, and certain changes in sub-model interactions. An especially interesting improvement is the inclusion of intra-zonal traffic. The assignment model therefore has to deal with all traffic in Europe and not only inter-zonal. In the model network, it is modelled by adjusting the pre-loaded transport. As the model does not include the local networks, the assigned flows do not include that share of the intra-zonal transport that does not use the main roads. However, this is still part of the demand model, and LoS matrices. The total transport work (driven km) is therefore calculated at the matrix level by multiplying the demand matrix with the distance matrix.

<table>
<tr><th></th><th>LoS calculation</th><th>Demand model</th><th>Assignment model</th></tr>
<tr><td>Inter-zonal network</td><td>LoS matrices from assignment</td><td rowspan="3">Explicitly demand model (trip matrices) Explicit LoS matrices</td><td>Assigned onto the network</td></tr>
<tr><td>Intra-zonal model network</td><td>Weighted average over assigned link flows</td><td>Pre-loaded traffic adjusted proportionally to the demand model</td></tr>
<tr><td>Intra-zonal local network</td><td>Estimated by zone-based correction factor</td><td>Not modelled</td></tr>
</table>

Figure 12: Assignment overview.

Finally, the note describes the specifications for each of the modes, air, road, and rail.

5 Description of the Passenger Demand Model for Short Trips

5.1 Introduction

The **TEN**CONNECT short-distance model, i.e. up to 100 km, relates to an average week-day in 2005 (the model base year). The model consists of two sub-models; a frequency model at the top and a joined mode/destination choice model below. The sub-models are linked by an accessibility measure, i.e. logsums. Four modes are included in the modal split; car driver, car passenger, bus and train. Finally, the model includes four travel purpose segments; commuting, business, private and vacation/holiday trips.

This note begins with explaining a methodology of calculating values of time (VOT) for short-distance trips across all zones in the **TEN**CONNECT project. These values are differentiated across travel modes, travel purposes and income. In section 3, a description of the destination/mode choice model is given. Finally, section 4 describes the frequency model for short trips.

5.2 VOT for short trips

5.2.1 Approach

The mode/destination choice model includes VOT that vary across regions (countries and zones), travel purposes and modes. Due to the lack of choice data (needed for estimation) and the local VOT across Europe, we chose the following methodological approach in order to build the **TEN**CONNECT project's VOT:

1. The OTM 5.0 VOT are the starting values. The OTM (Ørestad Traffic Model) is an operational traffic model for the Greater Copenhagen Area (GCA), an area which is about 60x100 km large. The model operates with 5 travel modes, 6 travel purposes and 5 income groups.

2. Apply the Danish Transport Panel Survey data, i.e. TU-data, in order to aggregate OTM´s VOT into the values that can be compared with the Danish long-distance VOT. What is expected is that the OTM-aggregated VOT are lower than the long-distance VOT; i.e. VOT increases with travel distance. Step 2 is a quality insurance of how valid the chosen approach is.

3. Compare the long-distance VOT across 42 countries where the Danish VOT are the base values.

4. The obtained ratio from step 3 applies to the OTM values in order to calculated the country specific VOT for short trips, split across mode, purpose and income.

5.3 OTM 5.0 VOT

Table 20 shows the OTM values of travel time (VOT) across travel models and travel purposes* for the base income group, i.e. 0-200,000 DKK person gross income. Below the table are listed factors for higher income groups; they are all greater than 1.0, which is to say that VOT increases with income. For instance, VOT for the highest income group, (over 500,000 DKK) approximately twice the VOT presented in table 1.

Time components	BA	BU	BI	BF	nUU	EE
Invehicle time (rail)	21.3	23.3	16.5	16.9	14.0	85.7
Invehicle time (bus)	32.0	23.3	16.5	28.7	23.8	128.6
Invehicle time (metro)	14.9	23.3	16.5	23.6	19.6	60.0
PT access/egress	*22.1*	*23.3*	*23.3*	*46.6*	*25.3*	*119.9*
PT waiting time	*57.2*	*153.2*	*110.7*	*85.0*	*77.1*	*200*
PT interchange time	*57.2*	*153.2*	*110.7*	*48.9*	*77.1*	*200*
Car free flow time	33.5	24.4	14.9	26.4	24.5	46.8
Car congestion time	69.4	24.4	14.9	71.6	54.4	130.8

Table 20: VOT in OTM 5.0 for income groups 0-200,000 kr. (personal gross income), 2005 DKK/hr. BA=home-work, BU=home-education, BI=home-shopping, BF=home-private, nUU=non home-based private trips, EE=business

Income group 2-300,000 DKK: factor +1.27

Income group 3-400,000 DKK: factor +1.60

Income group 4-500,000 DKK: factor +1.93

Income group over 500,000 DKK: factor +1.98

5.3.1 TU data

In order to perform the aggregation of the OTM VOT we need some information from the Danish Transport Panel Survey, so called TU-data. The latest year we have in the TU-data is 2006. The following tables show distribution of the GCA intern sample across income groups and travel modes.

Table 21 shows the absolute and relative values of the GCA TU sample across the 5 income groups.

Income group	Absolute share	%share
0-200,000 DKK	1200	41.8
2-300,000 DKK	855	30.8
3-400,000 DKK	408	14.2
4-500,000 DKK	158	5.5
> 500,000 DKK	220	7.7

Table 21: Income distribution, gross annual value in 2006

Table 22 presents the absolute and relative values of the GCA TU sample across travel modes.

Mode choice	Absolute share	%share
Car driver	4301	64.9
Car passenger	1145	17.3
Bus	524	7.9
Train	657	9.9

Table 22: Use of travel modes across the GCA TU sample in 2006

5.3.2 Aggregation

When combining the VOT across 5 income groups with the income distribution across the greater Copenhagen based on the 2006 TU sample, we get the weighted mean of VOT over income categories. These values are shown in *Table 23*.

Time components	BA	BU	BI	BF	nUU	EE
Invehicle time (rail)	27.6	30.2	21.4	21.9	18.1	111.0
Invehicle time (bus)	41.4	30.2	21.4	37.2	30.8	166.5
PT access/egress	22.1	23.3	23.3	46.6	25.3	119.9
PT waiting time	57.2	153.3	110.7	85.0	77.1	200.0
PT interchange time	57.2	153.3	110.7	48.9	77.1	200.0
Car free flow time	43.4	30.3	19.3	34.2	31.7	60.6
Car congestion time	89.9	30.3	19.3	92.7	70.4	169.4

Table 23: Weighted mean of VOT over income categories, 2005 DKK/hr

When combining the figures from the above table with the TU distribution on transport modes we get the weighted average VOT over transport modes, shown in *Table 24*.

All 4 modes	BA	BU	BI	BF	nUU	EE
VOT	49.3	30.3	19.7	42.8	36.7	91.8

Table 24: VOT, aggregated across modes, 2005 DKK/hr.

The **TEN**CONNECT travel purposes are Commuting, Private, Holiday and Business. While the Commuting and Business purposes are straightforward we need to make some assumptions for the other two purposes by combining the OTM VOT for BI, BF and nUU purposes and based on TU data. The reason why BF VOT is higher than the other two VOT measures is that holiday trips (visiting summerhouse) are included. Let us assume that the non-holiday private trips VOT is 36.7 DKK/hr, instead of 42.8 DKK/hr. The 2006 TU split between shopping and private is 35% shopping and 65% private (cinema, sport, visiting friends/family). Combining the above information, we get that VOT for private short-distance trips is equal to 31 DKK/hr.

As mentioned above, part of the BF VOT is related to holiday, while the other part is related to private trips, i.e. we should therefore split the VOT for BF (i.e. 42.8 DKK/kr) into the part for private trips (31 DKK/hr) and holiday trips. The 2006 TU split between private trips and holiday trips is 90% BF and 10% holiday. Combining the above information, we get that VOT for holiday trips is 149 DKK/hr. As this value seems to be too high, we chose to assume that VOT for holiday short-trips is the middle value between commuter and private VOT, i.e. 40.1 DKK/hr.

The following table (*Table 25*) shows a comparison between the OTM-based Danish VOT for trips up to 100 km on one side and Danish VOT for long trips on the other side. The values are aggregated across travel modes and income, because these are the values we could get for long-distance trips.

Distance	Commuter	Private	Holiday	Business
< 100 km	49.3	31	*40.1**	91.8
> 100 km	110	71	71	146
Ratio	0.45	0.44	0.56	0.63

Table 25: VOT, comparison between the short and long trips in DK, 2005 DKK/hr. * assumed value

As expected, the short-distance VOT are smaller than the long-distance VOT. The conclusion of the exercise is that short-distance VOT, originating from the OTM, correspond well to the Danish long-distance VOT.

5.3.3 Application guideline; Example of Spain

In this chapter we make a guideline for producing countries VOT for short trips.

Step 1:

The following table shows the ratio between the Danish and Spanish VOT for long-distance trips. The ratio of 0.67 is constant across travel purposes and we will call this factor for **country factor**.

Country	Commuting	Private	Holiday	Business
Danish	14.64	9.52	9.52	19.48
Spanish	9.84	6.40	6.40	13.09
Ratio	0.67	0.67	0.67	0.67

Table 26: Comparison between the Danish and Spanish long-distance VOT, Euro

When applying the Spanish county factor of 0.67 we get the following Spanish short-distance VOT (*Table 27*). These values are obtained by: *values in Table 24 x 0.67.*

All 4 modes	BA	BU	BI	BF	nUU	EE
VOT	33.0	20.3	13.2	28.7	24.6	61.5

Table 27: Spain: VOT, aggregated across modes, 2005 DKK/hr

Step 2:

The income aggregated Spanish short-distance VOT are shown in *Table 28*. These values are obtained by: *values in Table 23 x 0.67 (i.e. country factor).*

Time components	BA	BU	BI	BF	nUU	EE
Invehicle time (rail)	18.5	20.2	14.3	14.7	12.1	74.3
Invehicle time (bus)	27.8	20.2	14.3	24.9	20.6	111.6
PT access/egress	*14.8*	*15.6*	*15.6*	*31.2*	*17.0*	*80.3*
PT waiting time	*38.3*	*102.7*	*74.2*	*57.0*	*51.7*	*134.0*
PT interchange time	*38.3*	*102.7*	*74.2*	*32.8*	*51.7*	*134.0*
Car free flow time	29.1	20.3	12.9	22.9	21.3	40.6
Car congestion time	60.2	20.3	12.9	62.1	47.2	113.5

Table 28: Spain: VOT, aggregated across income groups, 2005 DKK/hr

Step 3:

If we now aggregate BI, BF and nUU into two purposes only, which is Private and Holiday, as proposed in section 5.3.2 we get now a complete VOT table for Spain originating from the OTM VOT (Table 27).

These values are obtained in the following way:

Commuting: BA values in Table 26.
Private: 0.35*BIvalues + 0.65*nUUvalues (BI and nUU values from Table 28)
Holiday: (Commuting + Private) / 2
Business: EE values in Table 28

Time components	Commuting	Private	Holiday	Business
Invehicle time (rail)	18.5	12.9	15.7	74.3
Invehicle time (bus)	27.8	18.4	23.1	111.6
PT access/egress	14.8	16.5	15.6	80.3
PT waiting time	38.3	59.5	48.9	134.0
PT interchange time	38.3	59.5	48.9	134.0
Car free flow time	29.1	18.3	23.7	40.6
Car congestion time	60.2	35.2	47.7	113.5

Table 29: Spain: VOT, aggregated across income groups, 2005 DKK/hr

Note that when building the model, the VOT need to be presented in Euro/min. So, the figures in *Table 29* must first be divided by 7.5 (i.e. the ratio between the Euro and DKK) and then by 60.

Step 4:

Note that the values presented in Table 28 and Table *29* need to be sensitive towards the GDP (a known variable in the zonal data) in the **TEN**CONNECT model application, e.g. if in the model scenario for 2030 GDP goes up 5% in Portugal and 3% in Denmark, then the baser year VOT need to be projected to 2030 differently for the two countries. The proposed way for doing this to assume the following:

- change of +1% in GDP equals 0,6% increase in VOT for Private and Holiday segments,
- change of +1% in GDP equals 0.8% increase in VOT for Commuter segment, and
- change of +1% in GDP equals 1.0% increase in VOT for Business segment.

If we now assume that Spain's GDP will increase by 1.5% annually up to 2030, then the 2030 VOT adjusted for GDP are presented in *Table 30*.

These values are obtained by:

2030 – 2005 = 25 years
25 years * 1,5% = 38%
An increase of 38% applies on Business segment
An increase of 38% * 0,8 = 30,4% applies on Commuting segment
An increase of 38% * 0,6 = 22,8% applies on Private and Holiday segments

Time components	Commuting	Private	Holiday	Business
Invehicle time (rail)	24.2	15.9	19.4	103.3
Invehicle time (bus)	36.4	22.7	28.5	155.1
PT access/egress	19.4	20.3	19.3	111.7
PT waiting time	50.3	73.5	60.4	186.3
PT interchange time	50.3	73.5	60.4	186.3
Car free flow time	38.1	22.6	29.2	56.4
Car congestion time	79.0	43.4	58.9	157.7

Table 30: Spain 2030: VOT (DKK/hr), aggregated across income groups

5.4 Mode/Destination choice model

We propose to apply a join mode/destination choice model application for the **TEN**CONNECT project's short trips (up to 100 km). The model includes 4 model-segments according to travel purposes, i.e. *commuting, leisure, holiday, business*. Note that the holiday and private segments have the same VOT, e.g. coefficients (the result of the thorough model test). The available modes are *car driver, car passenger, bus and train*.

The model application must have the utilities in the form of coefficients (and not VOT), otherwise we will have a scaling problem in the model and the elasticities will be wrong. In the case of Denmark and the year of 2005, the VOT (aggregated across the income groups) in Euro/min are presented in *Table 31*.

Time Components	**Commuting**	**Private**	**Holiday**	**Business**
Invehicle time (rail)	0.06128	0.05204	0.05204	0.24658
Invehicle time (bus)	0.09207	0.07660	0.07660	0.37001
PT access/egress	0.04911	0.05189	0.05189	0.26644
PT waiting time	0.12711	0.16228	0.16228	0.44443
PT interchange time	0.12711	0.16228	0.16228	0.44443
Car free flow time	0.09639	0.07861	0.07861	0.13465
Car congestion time	0.19968	0.15821	0.15821	0.37634

Table 31: Denmark 2005: VOT (Euro/min), aggregated across income groups

In order to calculate the coefficients we start by applying the OTM 5.0 cost coefficients, as shown in *Table 32*.

Coefficient	**Commuting**	**Private**	**Holiday**	**Business**
Cost	-0.18052	-0.19083	-0.19083	-0.03862

Table 32: Cost coefficient; 1/Euro

Ferry sailing time coefficient is agreed to be 1/7 of the car free flow coefficient, while the ferry waiting time coefficient is agreed to be equal to car congested time coefficients.

Two "Car Ownership" parameters are also introduced in the model; a positive value for Car utility and a negative value for Public Transport modes (bus and train). Car Ownership variable is equal to that in the frequency model, i.e. number of cars per 1.000 inhabitants in the Origin zone.

Finally, in order to model better zone internal trips a positive constant is introduced.

The full set of coefficients for Denmark in 2005 is therefore (time = 1/min, cost = 1/Euro):

Utility components	**Commuting**	**Private**	**Holiday**	**Business**
Invehicle time (rail)	-0.01106	-0.00993	-0.00993	-0.00952
Invehicle time (bus)	-0.01662	-0.01462	-0.01462	-0.01429
PT access/egress	-0.00887	-0.00990	-0.00990	-0.01029
PT waiting time	-0.02295	-0.03097	-0.03097	-0.01716
PT interchange time	-0.02295	-0.03097	-0.03097	-0.01716

Car free flow time	-0.01740	-0.01500	-0.01500	-0.00520
Car congestion time	-0.03605	-0.03019	-0.03019	-0.01453
Cost coefficient	-0.18052	-0.19083	-0.19083	-0.03862
Ferry sailing time	-0.00249	-0.00214	-0.00214	-0.00074
Ferry waiting time	-0.03605	-0.03019	-0.03019	-0.01453
PT car ownership	-1.84	-1.840	-1.840	-2.24
Car car ownership	+1.440	+1.630	+1.630	+1.310
Intra zone time Train/Bus	+0.605	+0.605	+0.605	+1.705
Intra zone time Car Driver/Car Passenger	+1.21	+1.21	+1.21	+3.41

Table 33: Denmark 2005: cost and time coefficients

Therefore, the Danish 2005 utilities for commuter segment are now:

$U_{\text{car driver, ij}}$ = C_CarD – 0.18052*CarCost_{ij} – 0.0174*CarFF_{ij} – 0.03605*CarCNG_{ij} – 0.00249*FerrySailing_{ij} – 0.03605*FerryWaiting_{ij} + 1.44*CarOwnership_i + 1.21*ifeq(O-zone, D-zone) + (DestJ_Cnst + β_j*log(max(jobs_j,1))

$U_{\text{car passenger, ij}}$ = C_CarP – 0.0174*CarFF_{ij} – 0.03605*CarCNG_{ij} – 0.00249*FerrySailing_{ij} – 0.03605*FerryWaiting_{ij} + 1.44*CarOwnership_i + 1.21*ifeq(O-zone, D-zone)+ (DestJ_Cnst + β_j*log(max(jobs_j,1))

$U_{\text{bus, ij}}$ = C_Bus – 0.18052*BusCost_{ij} – 0.01662*(2*(CarFF_{ij} + CarCNG_{ij} + FerrySailing_{ij} + FerryWaiting_{ij})) – 1.84*CarOwnership_i + 0.605*ifeq(O-zone, D-zone) + (DestJ_Cnst + β_j*log(max(jobs_j,1))

$U_{\text{train, ij}}$ = C_Train – 0.18052*TrainCost_{ij} – 0.01106*TrainT_{ij} – 0.00887*TRAINaccegr_{ij} – 1.84*CarOwnership_i + 0.605*ifeq(O-zone, D-zone) + (DestJ_Cnst + β_j*log(max(jobs_j,1))

where:
i is from-zone
j is to-zone

C_CarD, C_CarP, C_Bus and C_Train are mode constants to be estimated in the calibration process.

log(max(jobs(dest(D)),1) is zone attraction variable (found on the basis of zone data) based on number of jobs in the destination zone j.

The logsum parameter β_j is fixed to 1.

DestJ_Cnst is the destination attraction constant to be estimated in the calibration process.

CarCost is LOS car driving cost, in Euro[3].
CarFF is LOS car driving free-flow time, in min
CarCNG is LOS car driving congested time, in min
FerrySailing is LOS ferry sailing time, in min
FerryWaiting is LOS waiting time for ferry, in min
BusCost is LOS bus travel costs, in Euro
BusT is LOS bus driving time, in min
TrainCost is LOS train travel costs, in Euro
TrainT is LOS train driving time, in min
TRAINaccegr is LOS train access/egress time, in min
CarOwnership is Zonal data number of cars per 1.000 inhabitants

These four utilities have to be repeated for all travel purposes and for all countries. For each new year (scenario) where we calculate a set of VOT (according to section 5.3.3), new utilities must be presented. What is constant is the cost coefficients presented in Table 32.

Calibration process guide:
Calibration of the mode/destination choice model a purpose of fitting the model to the observed matrices with respect to mode totals and zone-attraction totals. Once when all constants are calibrated, then with the help of base year LOS we can calculate a number for each mode/destination utilities and that is an input for estimating Zone Logsums. Zone Logsums is a measure of accessibility and it represents a connection between the frequency model and the mode/destination choice model, i.e. the frequency model cannot be estimated before the Zone Logsums are calculated. Opposite to that, the estimation and calibration of the mode/destination choice model is totally independent of the frequency model.

We explain here step-by-step the calibration process for the mode/destination choice model for the commuter segment:

Step 1:
Let us take an example shown on figure 1; i.e. zone 1 attracts 100 trips, zone 2 attracts 110 trips and zone 1400 attracts 95 trips. So **Dest1_Cnst** needs to replicate 100 trips attracted to zone 1, **Dest2_Cnst** needs to replicate 110 trips attracted to zone 2, and **Dest1400_Cnst** needs to replicate 95 trips attracted to zone 1400. The numbers 100, 110 and 95 comes from the model base matrix for commuters. More specifically we can calculate these numbers by summing all the trips in the base year trip matrix by to-zone (i.e. 110 is the sum of all trips to zone 2).

---

[3] CarCost is calculated as Distance*km_cost + TollCost, where km_cost are trip purpose defined (in the OTM km-cost for non-business segments is 0.77 DKK/km and for business segment 2.98 DKK/km. If we apply these values then they need to be divided by 7.5 before application).

zone	*car driver*	*car pass*	*bus*	*tog*		
1	50	10	15	25	100	**Dest1_Cnst**
2	70	15	5	20	110	**Dest2_Cnst**
...	...	...	...	...	...	
...	...	...	...	...	...	
...	...	...	...	...	...	
1400	20	5	30	40	95	**Dest1400_Cnst**
	x1	x2	x3	x4		
	C_CarD	**C_CarP**	**C_Bus**	**C_Train**		

Figure 13: Calibration of the mode/destination choice model

On the modal split side, C_CarD constant need to replicate X_1 commuter trips by car driver mode, C_CarP constant need to replicate X_2 commuter trips by car passenger mode, C_Vus constant need to replicate X_3 commuter trips by bus, and C_Train constant need to replicate X_4 commuter trips by train. X_1, X_2, and $X_{3,}$ and X_4 comes from the model base matrices for commuters by summing *all* trips in the each mode-specific matrix. In other words X_1 is the total numbers of CarDriver trips in the base year for the entire model.

Mode and destination constants are calibrated against the matrix totals and not against the sell values. This is to say, that the totals will fit but on a sell level we will have discrepancies between the matrix values and the model values.

For each origin zone the mode/destination utilities to be included in the calibration process origin in the travel matrices.

<u>Mode and destination constants must be calibrated simultaneously.</u> At no point will be obtained a perfect match to the matrix-outputs, but that is not crucial because we apply a pivot-point procedure at the level of mode/destination choice model.

Step 2:
In order to execute this step we have to be able to run the entire model (calculate utilities, calculate choice probabilities, calculate mode-specific OD matrices). For the first calibration iteration, we will assume:

- All mode constants (C_CarD etc.) = 0
- All **β-**constants = 1

In order to calculate the OD matrices we need $GeneratedTrips_{ik}$, which come from the frequency model. In the calibration we use sums from the base year matrices instead: $GeneratedTrips_{ik}$ is the sum of all trips with purpose **k** from zone **i**.

The calibration process is based on the following formula:

Example of Car Driver mode:
$C\_CarD^{New\_value} = -\ln(Model\_trips / Observed\_trips)$
where:

Observed_trips is equal to X_1 and it comes from base year trip matrices (N-values in *Table 34*)
Model_trips come from the model run (S-values in *Table 34*)
$C\_CarD^{New\_value}$ is the new mode constant to be used in the next calibration iteration.

The same procedure is applied for the destination constants **DestJ_Cnst**.

We propose to build a following calibration table for the modal split part (a similar table should be built for the attraction part):

Mode	**Observed Mode Trips (trip matrices)**	**Observed Mode Shares**	**Model Trips**	**Model Mode Shares**
Rail	N1	N1/N (%)	S1	S1/N (%)
Bus	N2	N2/N (%)	S2	S2/N (%)
Car Driver	N3	N3/N (%)	S3	S3/N (%)
Car Passenger	N4	N4/N (%)	S4	S4/N (%)
Total	N	100 %	N	100 %

Table 34: Calibration table for the model split

The purpose of the table is to help determine whether enough calibration iterations have been conducted by comparing observed mode shares to modeled mode shares.

Step 3:
When new adjustments are calculated (i.e. mode/destination constants), then new mode/destination utilities need to be calculated and after that the new probabilities.

Let us imagine that from zone 1 we can travel only to zones 2 and 3 (all other zones are on distances longer than 100 km), and that we can do that by all 4 modes. This is to say that we need to write 8 mode/destination utilities:

$\mathbf{U}_{\textbf{car driver}}^{\text{zone 1 to 2}}$, $\mathbf{U}_{\textbf{car passenger}}^{\text{zone 1 to 2}}$, $\mathbf{U}_{\textbf{bus}}^{\text{zone 1 to 2}}$, $\mathbf{U}_{\textbf{train}}^{\text{zone 1 to 2}}$, $\mathbf{U}_{\textbf{car driver}}^{\text{zone 1 to 3}}$, $\mathbf{U}_{\textbf{car passenger}}^{\text{zone 1 to 3}}$, $\mathbf{U}_{\textbf{bus}}^{\text{zone 1 to 3}}$, $\mathbf{U}_{\textbf{train}}^{\text{zone 1 to 3}}$

With the known LOS, and the calibration betas and mode constants, all the above utilities are numbers (values).

The probabilities are then to be calculated for all 8 cases as following:

$p_{\text{car driver}}^{\text{zone 1 to 2}} = \exp(U_{\text{car driver}}^{\text{zone 1 to 2}}) / \text{sum}(\exp(\text{all utilities}))$

$p_{\text{car passenger}}^{\text{zone 1 to 2}} = \exp(U_{\text{car passenger}}^{\text{zone 1 to 2}}) / \text{sum}(\exp(\text{all utilities}))$

$p_{\text{bus}}^{\text{zone 1 to 2}} = \exp(U_{\text{bus}}^{\text{zone 1 to 2}}) / \text{sum}(\exp(\text{all utilities}))$

$p_{\text{train}}^{\text{zone 1 to 2}} = \exp(U_{\text{train}}^{\text{zone 1 to 2}}) / \text{sum}(\exp(\text{all utilities}))$

$p_{\text{car driver}}^{\text{zone 1 to 3}} = \exp(U_{\text{car driver}}^{\text{zone 1 to 3}}) / \text{sum}(\exp(\text{all utilities}))$

$p_{\text{car passenger}}^{\text{zone 1 to 3}}$ = exp ($U_{\text{car passenger}}^{\text{zone 1 to 3}}$) / sum (exp (all utilities))

$p_{\text{bus}}^{\text{zone 1 to 3}}$ = exp ($U_{\text{bus}}^{\text{zone 1 to 3}}$) / sum (exp (all utilities))

$p_{\text{train}}^{\text{zone 1 to 3}}$ = exp ($U_{\text{train}}^{\text{zone 1 to 3}}$) / sum (exp (all utilities))

The sum of **p**'s is equal to 1.

With this we can also calculate the final product of the entire model: mode-specific OD-matrices:

$OD_{\text{car driver, ijk}} = P_{\text{car driver, ij}} * \text{GeneratedTrips}_{ik}$

$OD_{\text{car passenger, ijk}} = P_{\text{car passenger, ij}} * \text{GeneratedTrips}_{ik}$

$OD_{\text{bus, ijk}} = P_{\text{bus, ijk}} * \text{GeneratedTrips}_{ik}$

$OD_{\text{train, ijk}} = P_{\text{train, ijk}} * \text{GeneratedTrips}_{ik}$

Step 4:
Repeat ***Step 2*** and ***Step 3*** until the values in the Observed Mode Shares and Destination Attraction Shares are reasonably close to model mode shares and model destination attraction shares respectively, say about 1%.

Step 5:
Once we know the mode constants as well as the destination betas, for the case of Denmark 2005, then we must run the model for 4 scenarios:

a – increase of car driving cost by 10%
b – increase of total car travel time by 10%
c – increase of public transport fare by 10%
d – increase of public transport travel time 10%

Then, demands for car and public transport need to be compared to the base demands for all 4 travel purposes. The calculated elasticities should be close to those presented in appendix.

If not, the whole procedure of model calibration must be repeated for the new set of cost and time coefficients (as shown in tables 13 and 14).

Logsum
Logsum value is to be calculated for each of 1.441 zones per trip purpose segment. That is to say that for one segment (say commuters), one zone has only one Logsum value to be calculated from the mode/destination choice model.

Going back to our example for zone 1, the Zone 1 Logsum (i.e. Logsom $^{zone\ 1}$) is equal to:

$Logsum^{zone\ 1} = Ln(exp(U_{car\ driver}^{zone\ 1\ to\ 2}) + exp(U_{car\ passenger}^{zone\ 1\ to\ 2}) + exp(U_{bus}^{zone\ 1\ to\ 2}), exp(U_{train}^{zone\ 1\ to\ 2}) + exp(U_{car\ driver}^{zone\ 1\ to\ 3}) + exp(U_{car\ passenger}^{zone\ 1\ to\ 3}) + exp(U_{bus}^{zone\ 1\ to\ 3}) + exp(U_{train}^{zone\ 1\ to\ 3}))$

Results of the calibration process for the mode/destination choice model
For each iteration of the model calibration we a) accepted a set of new coefficients values presented in Table 32, b) a new attraction coefficient β_j, c) and calibration coefficients.

The Greater Copenhagen Area (GCA) is produced after each iteration. The elasticises are then compared to those presented in the appendix. Those elasticises which as accepted at the moment are presented in *Table 35* to *Table 38*. The according models have coefficients from Table 32 scaled down to 0.50 for car modes and to 0.40 for PT modes, while the attraction $\beta_j = 0.70$.

Trips	Business	Private	Holiday	Commuting	total
car driver	-0.058	-0.272	-0.217	-0.027	-0.152
car pass	0.002	0.033	0.028	0.204	0.061
train	0.002	0.030	0.027	0.197	0.096
bus	0.002	0.030	0.026	0.195	0.116
total	-0.044	-0.146	-0.106	0.030	-0.072

Table 35: Driving cost elasticity for GCA

Trips	Business	Private	Holiday	Commuting	total
car driver	-0.133	-0.218	-0.206	0.038	-0.108
car pass	-0.136	-0.263	-0.240	-0.001	-0.205
train	0.003	0.055	0.072	0.391	0.189
bus	0.003	0.055	0.071	0.388	0.232
total	-0.129	-0.214	-0.213	0.066	-0.114

Table 36: Driving time elasticity for GCA

Trips	Business	Private	Holiday	Commuting	total
car driver	0.000	0.006	0.002	0.046	0.022
car pass	0.000	0.006	0.002	0.047	0.012
train	-0.183	-0.585	-0.509	-0.453	-0.517
bus	-0.166	-0.539	-0.461	-0.384	-0.444
total	-0.006	-0.034	-0.010	0.003	-0.014

Table 37: Public transport fare elasticity for GCA

Trips	Business	Private	Holiday	Commuting	total
car driver	0.000	0.003	0.002	0.035	0.016
car pass	0.000	0.003	0.002	0.035	0.008
train	-0.222	-0.189	-0.173	-0.143	-0.171

bus	-0.422	-0.327	-0.340	-0.362	-0.348
total	-0.011	-0.017	-0.007	0.001	-0.008

Table 38: Public transport travel time elasticity for GCA

5.5 Frequency model

Beside the zone population, the frequency model must be sensitive to *GDP, car ownership and mode/destination logsums* (i.e. connection to the destination/mode choice model).

It is proposed to try to estimate a disaggregated model based on the TU 2003-2006 choice data.

5.5.1 Disaggregate model

Estimation

On the basis of the TU data we will estimate a MNL model with the following utilities; zero trips, 1 trip, 2 trips, 3 trips, r trips and 5+ trips.

Utilities are defined as follows:

U_{trips} = asc + x_1 * Person_Car_Availability + x_2 * Person_Income + x_3 *ZoneLogsum

where,
asc is trip-constant,
x_1 and **x_2** are coefficients to be estimated,
x_3 is 1 in all segments,
Person_Income and Person_Car_Ownership are known in the TU estimation data, and ZoneLogsum is value calculated in the mode/destination model. It is assumed that the coefficient by the ZoneLogsum variable is equal to 1.

Person_Car_Availability is a number between 0 and 1, i.e. in the TU data we know the person car ownership while in the zonal data this value is calculated via “Cars per 1000 inhabitants” variable (if, for instance, for zone 5 the “Cars per 1000 inhabitants” is 387 then **Person_Car_Availability** for zone 5 is 0.387).

Person_Income is annual gross personal income in ´000 Euro known both in the estimation and zonal data.

A probability that a random person makes, for instance, 2 trips is then:
p_{two} = exp(U_{two}) / sum (exp(all trip utilities))

Based on choice data restrictions we are forced to assume that the **x_1** and **x_2** coefficients are identical across zones (countries), i.e. model sensibility with respect to GDP (income) and car ownership is identical across the geography. So, the number of trips generated by a random zone is dependent of the zonal characteristics (population size, GDP and car ownership), as well as the accessibility.

Estimation results
The frequency model has been estimated based on the 2003-2006 TU data for the Greater Copenhagen Area. The following results are obtained:

- Utilities are made for 0 trips, 1 trip, ….., 5+ trips with trip constants, and income and car ownership parts.
- The logsum part is omitted for the moment.
- The three structure beta-scale is set to 1.0, i.e. we have a MNL model structure.

The estimated coefficients are presented in the table below.

Trip purpose	ZeroCnst	1 Cnst	2 Cnst	3 Cnst	4 Cnst	5+Cnst	Car avail.*	Income*
Commuting	0	-4.346	-2.3	-7.71	-6.54	-9.31	0.85	0.0033108
Business	0	-6.25	-6.32	-7.41	-8.33	-8.43	1.00	0.02
Private	0	-2.65	-1.68	-3.06	-3.09	-3.32	0.85	0.02
Holiday	0	-5.18	-5.51				0.85	0.035

Table 39: Estimated coefficients in the Frequency Mode. * Adjusted values

Where:
- Columns 1 to 6 are trip constants.
- Note that for the “Holiday” segment the maximum number of trips is 2.
- Zero trips is a default value.
- Car availability coefficient for “Holiday” segment is set to the same value as in the “Private” segment.
- Income coefficient for “Holiday” segment is set to the same value as in the “Private” segment.
- Car availability and Income coefficients are positive in sign (as expected) and they are high in t-values (over 2.0).

The estimated values are checked against the zonal data for four random countries, i.e. Denmark, UK, Germany and Bulgarian. Table below shows the obtained trip rate results.

	Cars per person	GDP (in ‘000 Euro)	Obtained trip rate
DK	0.367	38.4	2.06
UK	0.45	30.3	2.17
DE	0.54	27.2	2.32
BG	0.28	2.8	1.69

Table 40: Obtained trip rates for 4 randomly selected countries

In the case of Denmark, the trip rate of 2.06 trips/person/day are split by travel purposes in the following way: Commuter – 0.55, Business – 0.06, Private – 1.42, Holiday – 0.03. For the Greater Copenhagen population of 1.6 mil people that gives 0.88 mil commuting trips, 0.10 mil business trips, 2.27 mil private trips and 0.05 mil holiday trips (all per day), in total 3.3 mil trips per day completed by car, bus and train.

Calibration

Calibration process guide:

We explain here step-by-step the calibration process for the frequency model for the case of the commuter segment (i.e. the same procedure must be applied to other 3 travel purpose segments):

Step 1:

Let us assume that in the observed base-year travel matrices for commuter there are 1000 trips generated.

The calibration utilities for 0, 1, 2, 3, 4, and 5 trips have the following definition:

U_{trips} = asc + x_1 * Person_Income + x_2 * Person_Car_Availability + x_3 * ZoneLogsum + freq_const

From table 20 we can see that for the commuter segment:

asc for one trip is -4.346

asc for two trips is -2.3

asc for three trips is -7.71

asc for four trips is -6.54

asc for five+ trips is -9.31

x_1 is 0.85

x_2 is 0.0033108

The freq_const is at the moment equal to 0.2346, and this is the coefficient to be changed so that the model reproduces 1000 trips generated by the base-year commuter matrix.

So, the trip utilities in the commuter segment are as following:

U_{zero} = 0

$U_{1\ trip}$ = -4.346 + 0.85 * Person_Car_Availability + 0.0033108 * Person_Income + x_3 * ZoneLogsum + 0.2346

$U_{2\ trips}$ = -2.3 + 0.85 * Person_Car_Availability + 0.0033108 * Person_Income + x_3 * ZoneLogsum + 0.2346

$U_{3\ trips}$ = -7.71 + 0.85 * Person_Car_Availability + 0.0033108 * Person_Income + x_3 * ZoneLogsum + 0.2346

$U_{4\ trips}$ = -6.54 + 0.85 * Person_Car_Availability + 0.0033108 * Person_Income + x_3 * ZoneLogsum + 0.2346

$U_{5\ trips}$ = -9.31 + 0.85* Person_Car_Availability + 0.0033108 * Person_Income + x_3 * ZoneLogsum + 0.2346

Again, **Person_Income** is in '000 Euro and **Person_Car_Availability** is between 0 and 1 (arrives from zone variable called "Cars per 1000 inhabitants").

ZoneLogsum is calculated in the mode/destination choice model.

When applying the coefficients for the utilities let us imagine that the model produces 800 commuting trips.

Step 2:
The calibration process is based on the following formula:

Freq_const^{New_value} = - ln (Model_trips / Observed_trips)
where:
Observed_trips come from trip matrices
Model_trips come from the model run

In our case a new value for *Freq_const^{New_value}* is:

0.2346 – ln(800 / 1000) = 0.2346 + 0.2231 = 0.4577

Step 3:
When new adjustments are calculated then new utilities for 0, 1, 2, 3, 4, 5+ trips need to be written, following the new probabilities. These probabilities give new trip rates. New trip rates, for the know population, gives new trips generated by a zone.

The new model zone generated trips suppose to be closer now to those observed in the base matrix.

Step 4:
Repeat ***Step 2*** and ***Step 3*** until the values in the Observed Mode Shares and Model Shares (see columns in Table 31) are reasonably close, say about 1%.

Model scenario forecasts
The model scenario forecasted number of trips generated by a zone, $GeneratedTrips_i^{sc}$, is: $GeneratedTrips_i^{sc}$= Zero_trips + 1trip + 2trips + 3trips + 4trips + 5trips

5.6 Application; pivot point

The forecasted generated trips, $GeneratedTrips_i$, are calculated in a pivot-point procedure: $GeneratedTrips_i = T_{base} * (GeneratedTrips_i^{sc} / T_{sc\_base})$, where T_{base} is known from the base 2005 travel matrices, $GeneratedTrips_i^{sc}$ is calculated as shown above, while T_{sc_base} is the same as T_{sc} but for the base 2005 year.

The pivot point procedure as to be applied in both the frequency model and the mode/destination choice model. In both cases we need a set of base observed matrices and base synthetic matrices. When they are produced they should be kept in the model system without possibilities for the model user to change them. Then, each time we make a model run, the scenario produced matrices need to be pivoted around the ratio of the base observed matrices and base synthetic matrices.

In the case where for a specific zone (generation model) or a zone pair combination (mode/destination model) a cell has a number equal to zero a special rule has to be assigned. These rules are specified in the OTM 5 project and they can be applied also here if necessary.

5.7 Appendix 1; OTM 5.0 elasticities

Cost elasticity, all travel purposes

Transport Mode	Car	PT	Bicycle	Walk
Car	-0,10	+0,09	+0,07	+0,06
Public Transport	+0,06	-0,42	+0,09	+0,07

Travel time elasticity, all purposes

Transport Mode	Car	PT	Bicycle	Walk
Car	-0,15	+0,18	+0,13	+0,08
Public Transport	+0,04	-0,26	+0,06	+0,03

Separate purposes

Cost elasticity; commuters

Transport Mode	Car	PT	Bicycle	Walk
Car	-0,13	+0,11	+0,06	+0,02
Public Transport	+0,08	-0,33	+0,11	+0,05

Travel time elasticity; commuters

Transport Mode	Car	PT	Bicycle	Walk
Car	-0,24	+0,21	+0,14	+0,06
Public Transport	+0,06	-0,27	+0,08	+0,03

Cost elasticity; private trips

Transport Mode	Car	PT	Bicycle	Walk
Car	-0,10	+0,08	+0,08	+0,06
Public Transport	+0,05	-0,52	+0,08	+0,08

Travel time elasticity; private trips

Transport Mode	Car	PT	Bicycle	Walk
Car	-0,13	+0,14	+0,11	+0,08
Public Transport	+0,03	-0,26	+0,04	+0,03

Cost elasticity; business trips

Transport Mode	Car	PT	Bicycle	Walk
Car	-0,03	+0,06	+0,09	+0,12
Public Transport	+0,02	-0,10	+0,13	+0,05

Travel time elasticity; business trips

Transport Mode	Car	PT	Bicycle	Walk
Car	-0,09	+0,24	+0,30	+0,36
Public Transport	+0,03	-0,26	+0,02	+0,02

6 Description of the Passenger Demand Model for Long distance passenger Trips

6.1 Introduction

The following document describes the structure of the long-distance (above 100 km.) passenger model for the TENCONNECT project. The objective of the model is to forecast the level and distribution of long-distance European passenger transport on the basis of Level-of-Service (LoS) variables and zone data.

The main dimensions of the model are outlined below in Table 41;

Choice dimension	Number of alternatives	Description
Trip purpose	4	Holiday Private Commuting Business
Mode	5	Car as driver Car as passenger Bus Train Airplane
Distance	2	Short Long
Destination	1441	NUTS3 level

Table 41: Dimensions of the model.

More specifically, the models will be aimed at policy experiments with LoS variables measuring the impact from changes in infrastructure, and regional indicators such as Gross Domestic Product (GDP), population, and jobs.

The overall model structure is based on a random utility framework and decomposed in a trip generation part and a trip distribution part. The structure is illustrated below in Figure 14.

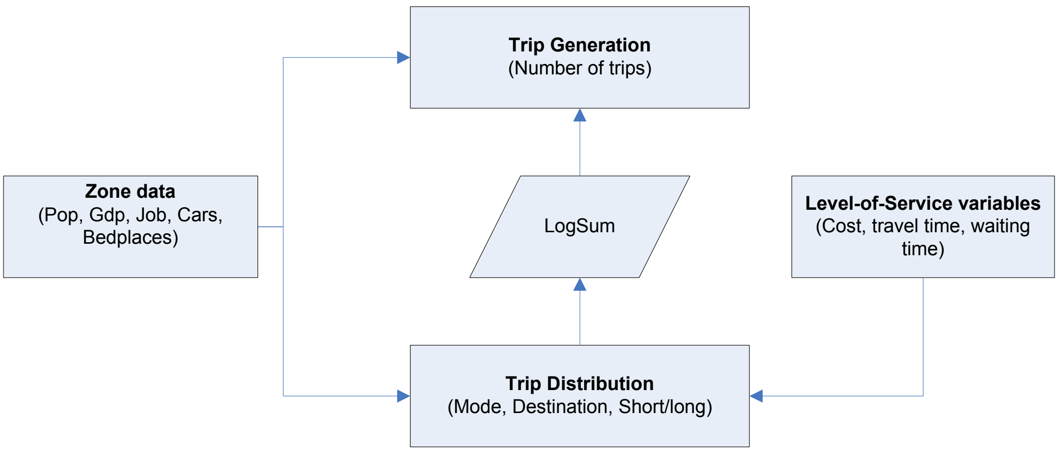

Figure 14: Overall model structure.

In order to consistently link the generation part with the distribution part, logsums are feed from the distribution part to the generation part. This accounts for accessibility effects in the generation modelling.

As illustrated, both model components are dependent of zone data, whereas only the distribution part is directly dependent on LoS variables, although, the generation part is indirectly dependent through the logsum inclusion.

6.1.1 Limitations

Due to limitations in data, there are important limitations of the model specification.

- The model only covers trips beyond 100 km. This is due to the fact that the DATELINE survey only covers trips beyond 100 km.

- The modelling area covers 1441 zones including the Far East (Russia, Belarus and Ukraine), however, the DATELINE survey primarily covers EU25 and is mainly covering old member states. Hence, it is assumed that preferences estimated on the basis of DATELINE are valid for the modelling area.

- The DATELINE survey was collected in 2003 but here it is applied to 2005. Especially for high-growth countries such as Bulgaria, Romania, Slovakia, Slovenia, and Turkey this may be seen as a problem.

- Individual income data have not been available. As a result, the only income effect is through zone GDP measures. This affects the size variables as well as the trip generation model.

- Due to the revealed-preference (RP) nature of the DATELINE survey, there has been a problem in identifying both in-vehicle-time and out-of-pocket-costs. As a result, we have applied country-wide value-of-time estimates to produce a generalised in-vehicle-cost measure.

- Due to the 100 kilometre threshold, the DATELINE survey only holds very few commuting trips. As a result it has been impossible to properly estimate a model for this trip purpose. Instead, commuter trips have been pooled with business trips.

- The model has been formulated with a car-passenger alternative although this alternative is not present in the DATELINE. The initial market shares (alternative specific constants) of this alternative have been fitted to conform to European average measures by purpose. The out-of-pocket cost for passengers has been fixed to 0.

- The duration of trips are not considered.

6.2 Definition of tours and trips

The model is based on the DATELINE survey. In DATELINE there are separate files for trips, journeys, and excursions. It is therefore possible to follow the complete trip chain from home to work and eventually combined with a shopping trip before returning to home.

It has been decided to overcome the complexity of the trip chaining by the following simplifications;

- Stratification on purpose
- Journeys are converted into single generation-attraction (GA) trips by attaching a main-mode and a main destination.
    - For business journeys, however, we allow non-home based journeys, with an attached main mode and main destination.
    - For private trips and holiday trips, we only allow home-based trips.

For business trips there may be many trips in a chain, however, all sub-trips (not origin and final destination) are excluded whether the trip is home based or non-home based.

Figure 15 below illustrates two typical examples of reduced trip chains.

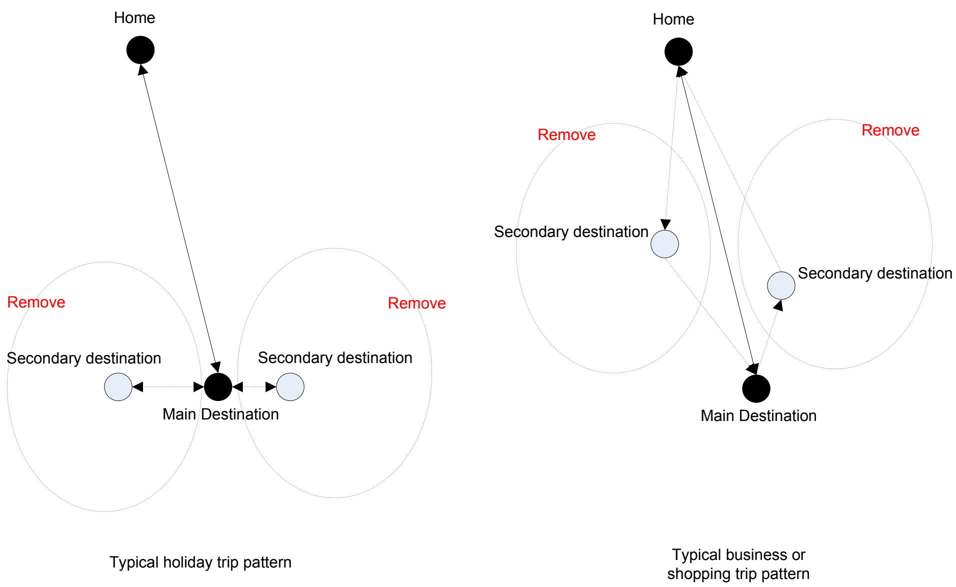

Figure 15: Illustration of how trip chains are changed to simple home-based trips.

To the left in Figure 15, a typical holiday trip pattern is shown. It consists of a long journey (e.g. airplane to the Canary Island) and excursions departing from the main destination. In the model only the trip to the main destination is maintained, whereas trips to the secondary destinations are left out.

To the right in Figure 15, a typical business or private trip pattern is shown. It may consist of a main destination and a number of sub-trips along the way to the main destination. However, all secondary destinations are considered as detours and excluded. As a result, only the trip from the home to the main destination is maintained. Compared to the illustration to the left, this trip chain reduction may well produce a synthetic new trip, which was not in the original set of trips.

The consequences of the trip chain reductions may seem more critical that they are. The point to remember is that the objective of the model is to capture overall differences in preferences rather than precisely mimic the trip pattern of households. In other words, excluded trips will only have impact to the extent preferences differ. In any case, the model will be pivot-point adjusted prior to the assignment. This means that the trip patterns represented in the OD matrix will be the offset for the demand effects of the model.

Moreover, in terms of excluded mileage the chain simplification account for less than 10%.

6.3 Data

In the following section the various data sources are considered in more details. The most important data sources are listed below;

- The DATELINE survey (European long-distance travel survey from 2003).
- Level-of-service (LOS) data.
- OD matrices, split by purpose and mode.
- Zone data primarily based on Eurostat.
- TU data (Danish travel-diary survey conducted from 1992-2006).

6.3.1 Level-of-service data

The model includes five modes: car as driver, car as passenger, bus, rail, and airplane. All of the modes are assigned on their respective network, although, cars and busses are not separated. The set of LoS variables across modes is shown in Table 42 below;

LoS component	**Car / Bus**	**Rail**	**Air**
Out-of-pocket costs	X	X	X
In-vehicle-time	X	X	X
Congestion time	X		
Ferry time	X	X	
Access-Eggress time		X	X
Frequency		X	
HeadWay time			X
TransferTime			X

Table 42: Outline-of-level of service variables.

All of the LoS variables except for the out-off-pocket costs for rail have been described in the documentation of the various assignment models. The rail cost, however, has not been available prior to the model exercise. As a result is has been necessary to estimate a separate cost function on the basis of a sample for rail ticket costs. The sample has been compiled by visiting web-sites of rail operators and then subsequently, estimating a model for cost as a function of length and country specific dummies.

The cost sample covers 500 observations and is intentionally sampled for different countries and for different ticket types. Two models have been estimated, a model for business trips and a model for all remaining trips.

The base model is a log-linear model;

(1)
$$Cost_i = k + k_{c|i} + Length_i^{\alpha} + \varepsilon_i$$
$$\Leftrightarrow$$
$$\log(Cost_i) = k + k_{c|i} + \alpha \cdot \log(Length_i) +$$

It is found that $\alpha = 0.95$, which indicate a decreasing marginal effect of length on the ticket price. $k_{c|i}$ define country-wise dummies. Result of the model estimation is shown in appendix A.

6.3.2 Zone data

The zone data to be used in the model are primarily based on Eurostat data. However, there have been problems in that the new zone structure of the model does not conform to the official NUTS3 codes as used by Eurostat. In these cases, we have applied a distribution of variables due to population which was available for the new zone structure.

An important issue in the model specification is that the model is to be used as a forecast module. It means that a variable which cannot be forecasted is not considered as input.

6.3.2.1 Population

The population is taken from the REG_D3AVG database in EUROSTAT. It measures the annual average population by gender at NUTS3 level. However, the distinction on gender is not used, only the total population.

For countries not included in the NUTS3 codex, e.g. zones in Russia, Belarus, and Ukraine, the population has been compiled from national statistics.

6.3.2.2 Hotel capacity

An important issue in destination modelling is the capacity of tourists. The EUROSTAT database includes a range of regional statistical indicators which may be used. Part of the data is included in the tourist database and contains information on nights spend by non-residents as well as residents. The data, however, do not distinguish between business travellers and tourists.

- TOUR_CAP_NUTS3: Number of establishments, bedrooms and bed places (NUTS-3 annual data).
- TOUR_OCC_NINRN2: Nights spent by non-residents (NUTS2 annual data).
- TOUR_OCC_NIRN2: Nights spent by residents (NUTS2 annual data).

The TOUR_CAP_NUTS3 database includes the following activity types;

Activity type	Description
A100	Hotels and similar establishments
B010	Tourist campsites
B020	Holiday dwellings
B040	Other collective accommodation n.e.s.
B100	Other collective accommodation establishments

Table 43: Hotel capacity activity types in Eurostat.

Each of the activity types is divided into three indicator types: Establishments (A001), bedrooms (A002), and bed-places (A003).

As a mean to describe capacity effects in the destination model, a total “bed-place” capacity across all activity types has been used. Clearly, one of the challenges is to forecast the activity level, especially in parts of Turkey, Romania, and Bulgaria.

For the countries not in the Eurostat bedplace/population rations for comparable countries has been applied.

6.3.2.3 Jobs

In the model, only the total number of jobs has been used. However, it has been considered whether part of the employment statistics should be used. The REG_LFE2ENACE database includes information on jobs in various sectors. It has been considered to distinguish between NACE classifications for;

- Hotels and restaurants
- Agriculture, hunting, forestry and fishing
- Total industry (excluding construction)
- Industry
- Construction
- Services
- Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods; hotels and restaurants; transport, storage and communication
- Financial intermediation; real estate, renting and business activities
- Public administration and defence, compulsory social security; education; health and social work; other community, social and personal service activities; private households with employed persons; extra-territorial organizations and bodies

However, the database only includes information at the NUTS2 level, which would necessitate further distribution onto NUTS3 zones. Moreover, the NUTS2 information only covers EU27. In other words, for many East European countries there would be no information.

Finally, from a forecast perspective, it would be very hard to do forecasting at a sector/branch level. As a result, we have considered only the total number of jobs.

6.3.2.4 GDP

The GDP measure is taken from the REG_E3GDP database. It measures the Gross domestic product (GDP) at current market prices at NUTS3 level.

GDP is measured in Mill. Euros per year.

For countries not included in the Eurostat database, national statistics has been used. Moreover, where the Eurostat have obvious wrong values a calculation based on the population has been used (e.g. Aberdeen).

6.3.3 The DATELINE survey

The DATELINE survey is a revealed-preference (RP) dataset which includes information on actual observed behaviour for the interviewed persons.

As discussed in the introduction, DATELINE represents a “diary type” survey in the sense that people were asked to provide their past travel history. The various trips are then divided into main destinations and secondary destinations. Secondary destinations may be “excursions” or simply secondary destinations.

The “past” in this context differs from purpose to purpose. The interview periods are summarised in Table 44 below.

Purpose	Period of record	Weights
Business	3 months	4
Holiday	1 Year	1
Private	3 month	4
Commuters	4 weeks	10,5

Table 44: Duration of interview periods and corresponding "weights".

The "weights" is the naïve weight that brings the survey to an annual basis.

The division on trip purposes are somewhat important because the models are stratified on trip purpose. E.g. different models are applied to different trip purposes. The purpose split is illustrated in Table 45 below.

Purpose	Frequency	Percentage
Business	10855	9,7%
Holiday	73326	66,55%
Private	27221	24,33%
Commuters	465	0,42%
Sum of trips	111867	

Table 45: Distribution of trip purpose.

Commuter trips account for only 0.42%, which is not sufficient for a valid estimation.

The mode choice variable is another central endogenous variable. The split is seen below in Table 46.

Mode	Frequency	Percentage
Air	16588	14,83%
Bus	10419	9,31%
Car	74608	66,69%
Train	10252	9,16%
Sum of trips	111867	

Table 46: Distribution of trip mode.

Car has a dominating market share of 67%, however, all of the included modes are sufficiently well represented from an estimation point of view. In order to properly represent the distribution of trip purpose and modes on annual basis, the weights in Table 44 can be applied. This adjusted distribution is shown in Table 47 below.

Purpose	Frequency	Percentage	Mode	Frequency	%
Business	43420	29,17%	Air	22597	15,18%
Holiday	73326	49,26%	Bus	11900	7,99%
Private	27221	18,29%	Car	97917,5	65,78%
Commuters	4882,5	3,28%	Train	16435	11,04%
Sum of trips	148849,5		Sum of trips	148849,5	

Table 47: Distribution of trip purpose and trip mode corrected to a year base.

Clearly, the trip purpose is not independent of the choice of mode. Table 48 illustrates the cross-tabulation of the two variables.

Mode	Business		Holiday		Private		Commuters	
	Freq	%	Freq	%	Freq	%	Freq	%
Air	31136	17,93%	13971	19,05%	2612	2,40%	189	3,87%
Bus	6784	3,91%	7319	9,98%	10616	9,75%	231	4,73%
Car	108560	62,51%	46489	63,40%	84092	77,23%	3265,5	66,88%
Train	27200	15,66%	5547	7,56%	11564	10,62%	1197	24,52%
Sum of trips	173680	100,00%	73326	100,00%	108884	100,00%	4882,5	100,00%

Table 48: Relation between trip mode and trip purpose (weighted).

It can be seen that the air alternative is more frequently used for business and holiday trips. However, for private trips, air trips only account for 2.4%. Business and commute trips also use rail more often than the average.

However, the important issue with respect to Table 48 is to identify possible coverage problems in the trip purpose stratification. Do all entries in the purpose×mode matrix have sufficient elements for estimation? This is in fact the case except for the commuting alternative which is pooled with business trips in the estimation.

As discussed in the introduction the model does not consider trip durations endogenously. However, it is clear that there are significant differences among mode and trip purpose. These differences are shown in Table 49 and Table 50 below.

Mode	Average (days) journey duration	Std.Dev.	Max. (days) journey duration	Min. (days) journey duration
Air	12,37	13,92	365	0
Bus	7,7	9,32	150	0
Car	7,93	10,22	365	0
Train	7,35	9,86	153	0

Table 49: Trip-duration by trip-mode.

Purpose	Average (days) journey duration	Std.Dev.	Max. (days) journey duration	Min. (days) journey duration
Business	2,2	5,42	180	0
Holiday	11,84	11,65	365	2
Private	2,07	4,41	365	0
Commuters	7,32	7,1	40	0,5

Table 50: Trip-duration by trip-purpose.

The consequences of leaving out durations are difficult to assess, however, it is not considered to be critical. However, in the long run there may be duration effects which would impact the demand assessment.

There may be substitution effects from travel expenses to accommodation expenses. E.g. if the cost of travelling increased dramatically – eventually as the result of increasing oil prices – it could cause durations to go up in order to bring down the cost of travel relative to accommodation. However, the opposite effect could happen as well if new flexible modes were introduced.

6.3.3.1 Geographical representativity of DATELINE

The representativity of the survey is of some interest to the model estimation. If there is an extremely bad coverage of the countries included in the final model (42 countries in total) it may weaken the validity of the model. The following section is concerned with the country-wise coverage of DATELINE.

The first measure is to consider information about the country of departure and the country of destination of the trip. This is shown in Figure 16. The number of flights used to create this figure is not weighted to the annual total.

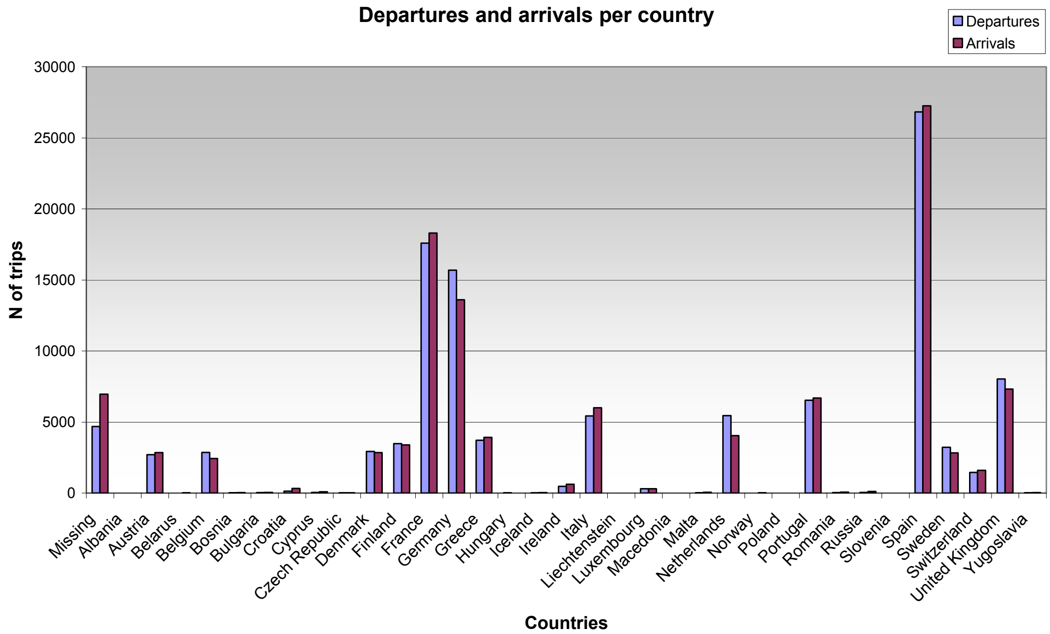

Figure 16: Ingoing and outgoing trips for each country.

It appears that there are some countries that are more represented than others, e.g. Spain tends to be relative over-represented compared other countries. It is also strongly visualised that many countries are not included or strongly under-represented. Unfortunately, this tends to apply to most East European countries.

The average number of outgoing and ingoing trips is quite similar as expected.

In order to further investigate the geographical representation ingoing and outgoing trips has been divided according to purpose. The results are illustrated in Figure 17 to Figure 19.

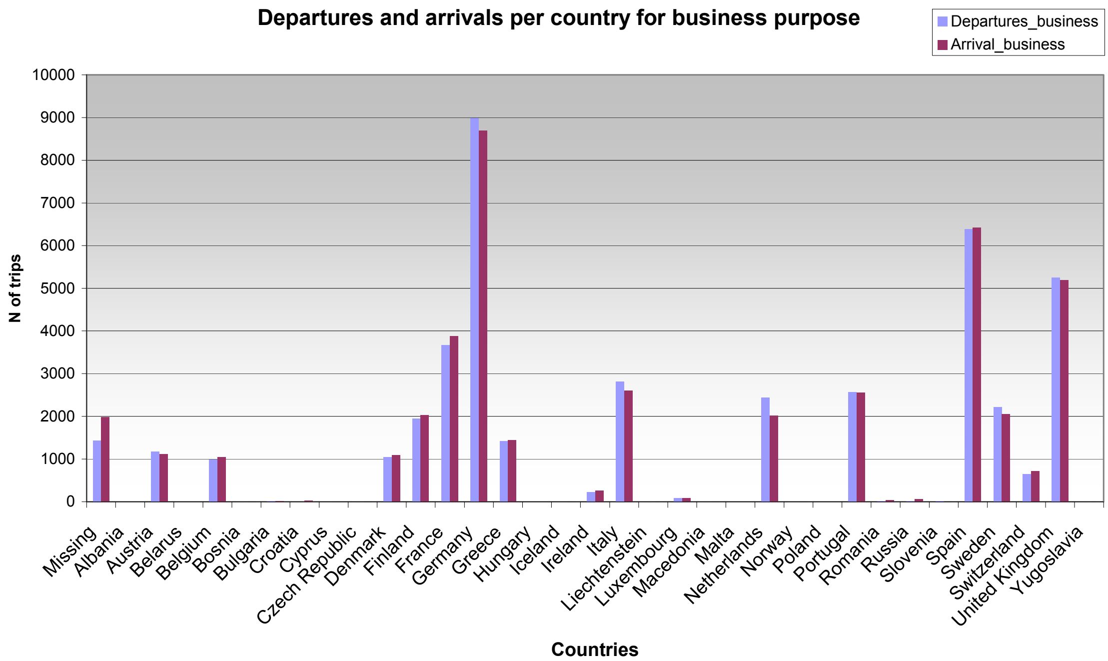

Figure 17: Ingoing and outgoing trips for business purpose.

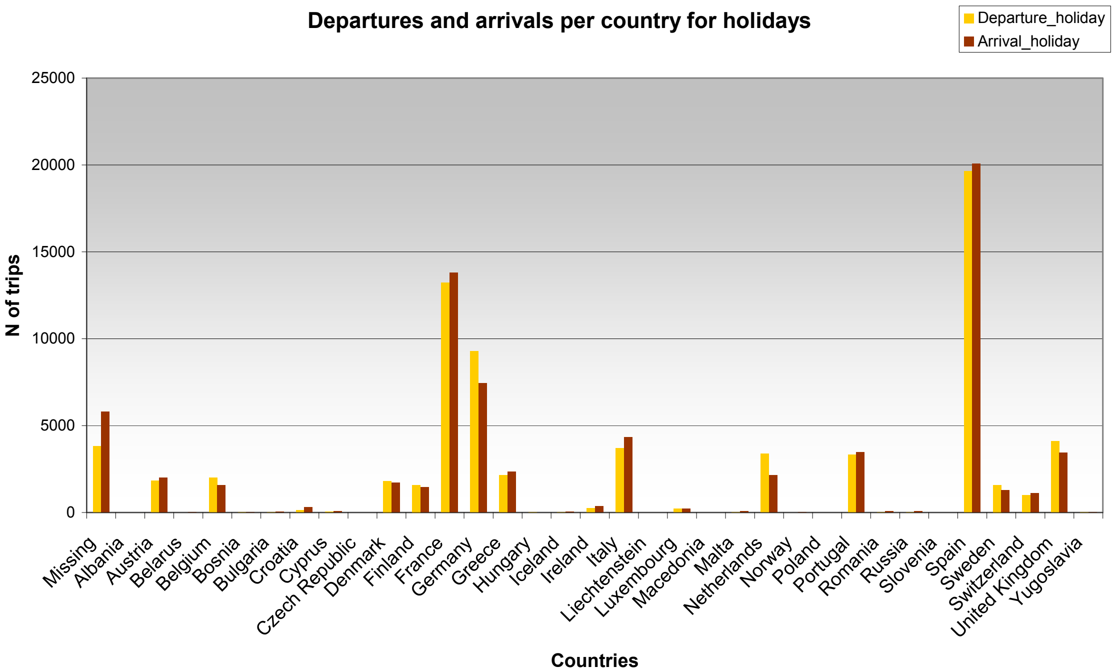

Figure 18: Ingoing and outgoing trips for holiday purpose.

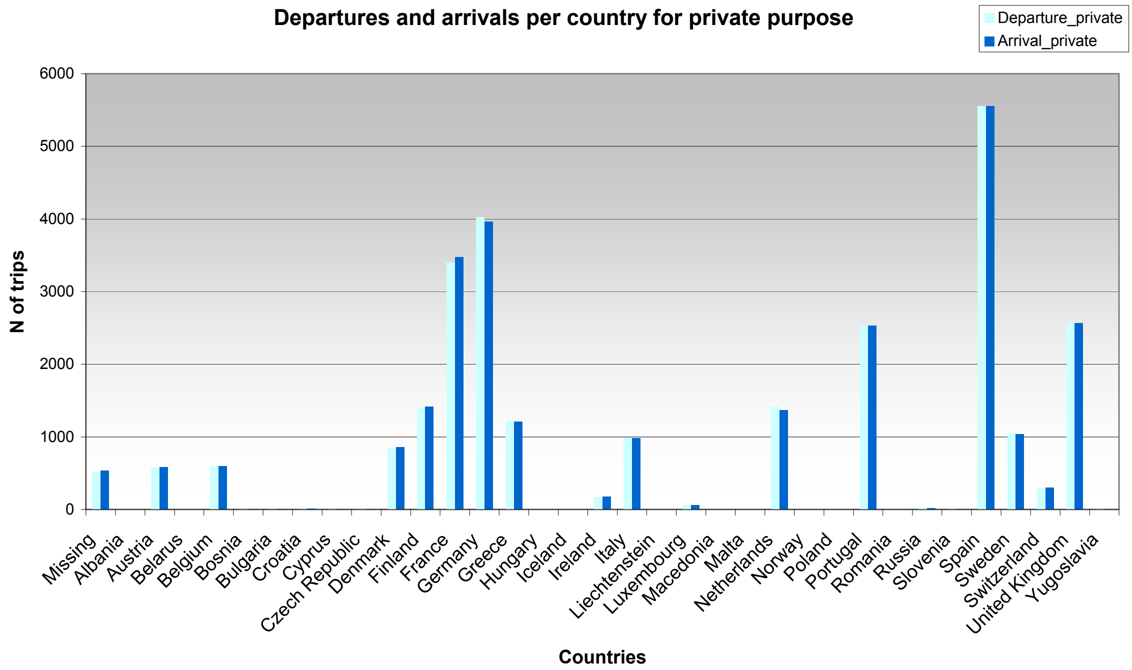

Figure 19: Incoming and outgoing trips for private purpose.

The complete table of country-wise departures and arrivals is shown in Table 85 in the Appendix

6.3.3.2 *Value-of-time estimates*

As discussed in section 6.4.1 below it has not been possible to properly estimate value-of-time measures from the DATELINE survey.

Moreover, even if it was possible, the weak coverage for large parts of Europe would force an external value-of-time estimate anyhow.

There have been two options;

- To specify an aggregated country-wise value-of-time estimate divided on trip purpose.
- To apply VOT based on a NUTS3 division.

Clearly, the NUTS3 division is appealing from a preference point of view in that – presumably - there are great differences in the regional value-of-time levels. The question, however, is if GDP per capita is a reasonable scaling instrument. The conclusion we have drawn is that this is probably not the case. One reason is that the GDP measure does not consider tax issues. Another issue is that there may be considerable regional differences in the consumer prices and the income distribution, which will make GDP a bad instrument for the scaling.

Rather the Purchasing Power Parity (PPP) index seems more appropriate for VOT scaling. The fundamental argument is that time is a commodity as any other commodity and the price of time should be proportional (for average people) to the PPP derived on the basis of the pool of all goods.

The derivation of country-wise value-of-time estimates is based on VOT studies in Switzerland, UK, the Netherlands, and Denmark. Based on these studies we find ratios between commute, other (private and holiday) and business trips.

The ratios are 0.65 for other and 1.33 for business. These calculations are for an average trip, i.e. we assume that the ratios are the same for short and long journeys.

Subsequently, the VOT for long-distance commuter trips is calculated on the basis of studies from Sweden (Algers et al. 1995), Switzerland (Axhausen et al. 2003), and Danish value-of-time estimates carried out in this project based on the DATIV (refer to Fosgerau et al. (2006) for a discussion of these data). As these studies are the only available studies in which long-distance trips are separated in the value-of-time analysis these are applied to scale long versus short trips.

The (un-weighted) average value is given by 13.98 Euros per hour with SE = 16.04, CH = 13.79, and DK = 12.10.

To arrive at country-wise VOTs we apply a table of purchasing power parity (PPP) from Eurostat. The average PPP for CH, DK, and SE is 1.28. It is assumed that the VOT follows PPP linearly and that a PPP of 1.28 relates to a commute VOT of 13.98 Euros per hour.

This is used to find the commute VOT for each country. Finally, the ratios of 0.65 and 1.33 are used to find VOTs for the remaining two other purposes. The VOT table is shown in Table 84 appendix.

The only exception to the country-wise VOT estimates is that we have allowed a different value-of-time for air travelers. The problem is that, since we use a generalised cost, the inherited weighting between time and cost elasticities becomes crucially dependent on the scales. Moreover, it may be argued that the VOT for air travelers is higher due to higher income and longer trips. As a result it has been assumed that the VOT for air travelers is the double of other modes.

6.4 Model structure – Distribution model

In the following we describe the model structure of the distribution demand model. From the point of view of the complete model framework illustrated in Figure 14, this represents the lower part of the demand model.

What is described in the following refers only to estimation and not application. The implementation of the model and the transformation of the estimated model to a forecast module are described separately in section 6.7.

The model is formulated as a nested discrete choice model. The “gross” nested structure has been spanned by three choice dimensions; short/long, mode, and destination. The choice tree is shown in Figure 20 below.

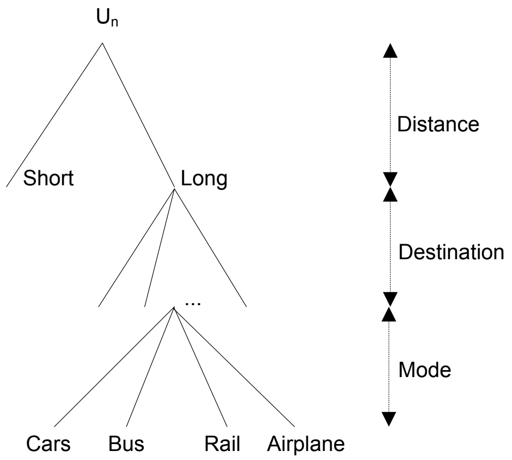

Figure 20: Choice decision tree for the passenger model.

After numerical testing, it has been decided to fix the upper logsum parameter (connecting short/long with destination) to unity. As a result the model structure collapse to a 2-level nested choice model.

The main reason for this restriction has to do with identification. Testing has clearly demonstrated that there is parameter instability depending on the long/short dummy. E.g. time preferences are different for long and short trips. In case of different time and cost parameters for long/short trips, the corresponding identification of a logsum modelling the split between long/short becomes problematic.

The motivation for a destination and mode nest can be backed by empirical findings. In fact, the structure with destination before mode (counting from the top) is evidenced in several studies including Fox and Sivakumar (2006).

6.4.1 Mode-choice alternatives

The choice of mode is explained by LoS differences as well as the number of available cars.

As discussed in the introduction of this chapter the optimal nested tree structure is found to be mode below destination. In the following section we will focus on the functional form at this level.

6.4.1.1 Function form

The issue of functional form is one of the most important aspects of the model specification, especially with very long trips as in the present survey.

In the estimation we have consider the functional form issue along two dimensions;

- Distance dependent parameter split (under/over 600 KM)
- Linear versus logarithmic specification of the generalised travel cost variable (GTC)

The first action has been applied to all models and to all time and cost components. There is very strong evidence that the hypothesis of equal parameters for long and short distances fails.

The second issue has also turned out to be important. For all of the trip purposes it was found that a linear time and cost component for short trips and a logarithmic time and cost for long trips performed best in terms of goodness of fit.

6.4.1.2 Utility functions

Each of the three trip purposes has been tested separately

Let m represent the index for mode, d the index for destination, and q the short/long indicator.

(2)

$$V_{m|d,q} = k_m + Size_d + Adj_{d,q}$$
$$+ \sum_{q=1}^{2} \varphi_{q,TC} f\left(GTC_{m|d,q}\right) + \varphi_{q,AE} AccEgg_{m|d,q} + \varphi_{q,F} Freq_{m|d,q} + \varphi_{q,FT} FerryTime_{m|d,q}$$
$$+ \varphi_{q,HW} HeadWayTime_{m|d,q} + \varphi_{q,TT} TransferTime + \varphi_{q,CA} CarAv_i$$

With the different variables described as;

Variable name	**Description**
$Size_d$	The attraction variable that varies over destinations (refer to 6.4.1.3).
$Adj_{d,q}$	Sampling correction factor (refer to 6.4.1.4).
$f\left(GTC_{m\|d,q}\right)$	Generalised travel cost on the basis of In-Vehicle-Time and out-of-pocket costs (see below)
$AccEgg_{m\|d,q}$	Access-Egress time. This variable is only valid for the rail and air mode.
$Freq_{m\|d,q}$	Rail frequencies.
$FerryTime_{m\|d,q}$	Gross ferry time including on-board ferry time and waiting time. For rail, only onboard time exist.
$HeadWayTime_{m\|d,q}$	Headway time for the air mode.
$TransferTime_{m\|d,q}$	Transfer time for the air mode.
$CarAv_i$	Car availability based on the number of private cars in the households (recorded from the DATELINE survey).

Table 51: Description of model variables.

The definition of $f\left(GTC_{m|d,q}\right)$ is as follows;

(3)

$$GTC_{m|d,q} = Cost_{m|d,q} + \gamma_{nm}\left(OnBoardTime_{m|d,q} + \kappa_n CongestionTime_{m=1\wedge 2|d,q}\right)$$

Where $Cost$ define variable car costs, γ_{nm} is a general value-of-time (VOT) measure for countries and modes (refer to section 6.3.3.2), and κ_n is a mark-up used to further scale congestion time. In the project results from a recent Danish project has been applied (refer to Table 52 below).

Purpose	**Average value of time (γ_n)**	**"Mark-up" ration for congestion (κ_n)**
Business	11.46	1.60
Private	5.60	1.57
Holiday	5.60	1.57
Commuting	8.61	1.57

Table 52: Value-of-time components – average values – in Euro/Hour.

The $f(\cdot)$ function refers to the outer functional form of the GTC variable. Two specifications have been used, $f(\cdot)$ equal to the identity function (the linear specification) and $f(\cdot)$ equal to the log function (log space).

For all models, the following configuration has been used;

$q = 1$ (short trips): $f(\cdot)$ = linear
$q = 2$ (long trips): $f(\cdot)$ = logarithmic

The remaining time components have been described in the documentation for the LoS variables. All of these follow a linear form.

6.4.1.3 Destination alternatives

The destination alternatives introduce two non-trivial issues. These are;

- Measurement of attractions
- Sampling of alternatives

The correct way of estimating size variables has been described by Daly (1982), however, this approach has not been possible in the present estimation. Instead, the form of attraction variables has been estimated prior to the discrete model. The primary reason for this is excessive correlation between the different size-variable candidates, in particular between Population, Jobs, regional Gross Product, and hotel capacity.

For forecasting purposes the model will need to incorporate all four candidates. E.g. if job changes occurs we will require the distribution model to react to this change. Ideally, in a position with much correlated variables, a principal component approach might be used to form new alternative variables as a linear combination of the correlated input variables. However, principal components are not suitable for forecasting, and we need to formulate the model based on core variables. A way to go has been to estimate a prior size-relationship between trips entering a zone (taken from the DATELINE survey) and the various size variables.

The construction of size-variables is accomplished by three successive regressions.

In the first regression GDP and hotel Capacity is regressed onto the trip vector T_i. The regression is a standard log-linear Poisson regression in order to explain size-effects in the number of trips as a function of GDP and CAP, e.g.

$$T_i = \rho \prod_k \left(GDP_{k,i}\right)^{\alpha_k} \left(CAP_{k,i}\right)^{\beta_k} \Leftrightarrow \tag{4}$$
$$\log(T_i) = \rho' + \sum_k \alpha_k \log\left(GDP_{k,j}\right) + \beta_k \log\left(CAP_{k,i}\right)$$

Where k defines trip purpose with 1=Business, 2=Private, and 3=Holiday trips. Moreover, and also due to scaling issues, the trip vector T_i, has been scaled to the population level by the use of expansion factors in the DATELINE survey. The average expansion factor is 4796.

The ideal approach would have been to include pop and job as well in the first stage estimation, however, due to correlation, the model estimate negative parameter values.

The second estimation is a simple ordinary linear square (OLS) in which Jobs is regressed onto the population, e.g.

$$POP_i = \lambda_k JOB_i \Rightarrow S_i^* = \log\left(POP_i + \hat{\gamma}_k JOB_i\right) \tag{5}$$

This regression defines the internal scaling between jobs and populations all other things equal. The R-square for this regression is as high as 0.986 which underline the correlation issue for these two variables.
The third and final regression is a log-linear Poisson regression where S_i^* is regressed onto the predicted trip vector $\hat{T}_i$. from the first regression, e.g.

$$\hat{T}_i = \theta_k S_{i,k} \tag{6}$$

The regression gives the relative scaling of size-variables contained in the first regression and the size-measure made up of jobs and population. The regression is without intercept. The final parameter values are given in Table 53 below.

Parameter name	**Purpose**	**Estimate**	**Lower 95% Conf.**	**Upper 95% Conf.**
α1 (LogCap)	Business	0.0662	0.0658	0.0666
α2	Private	0.3475	0.3472	0.3477
α3	Holiday	0.9073	0.9072	0.9075
β1 (Log(GDP*100))	Business	0.5966	0.5963	0.5968
β2	Private	0.4884	0.4882	0.4886
β3	Holiday	0.2072	0.2071	0.2074
γ1 (JOB)	Business	1.5793	1.5542	1.5944
γ2	Private	1.5798	1.5592	1.5993
γ3	Holiday	1.5828	1.5665	1.6067
θ1 (S)	Business	0.8645	0.8645	0.8645

θ2	Private	0.9271	0.9271	0.9271
θ3	Holiday	1.0017	1.0017	1.0017

Table 53: Estimation of size-variable components.

From the estimates in Table 53, three size variables can be formulated, one for each trip purpose, e.g.

(7)
$$Size_{i1} = 0.0662 LogCap_i + 0.5966 Log(GDP_i * 100) + 0.8645 Log\left(POP_i / 1.5793 + JOB_i\right)$$

(8)
$$Size_{i2} = 0.3475 LogCap_i + 0.4884 Log(GDP_i * 100) + 0.9271 Log\left(POP_i / 1.5798 + JOB_i\right)$$

(9)
$$Size_{i3} = 0.9073 LogCap_i + 0.2072 Log(GDP_i * 100) + 1.0017 Log\left(POP_i / 1.5828 + JOB_i\right)$$

Note, that for scaling purposes, the GDP has been multiplied by 100. Hence, from the point of view of implementation, the above form should be applied.

These three size variables are then applied directly in the following discrete choice estimations with the parameter fixed to unity. The fact that we fix the parameter to unity force the size variable to be proportional to the estimated probabilities.

6.4.1.4 Sampling of alternatives

The model operates on a zone structure of 1441 zones at the NUTS3 level. For destination choice modelling sampling of alternatives are required in order to reduce the memory consumption of the model during estimation. At present, the memory consumption for the largest model segment (holidays) is above 800 MB. A full-scale estimation without sampling would require in the range of 60-120 GB of swap space and would not be computationally feasible.

In order to reduce the number of destination alternatives, an importance sampling strategy has been applied. The idea of importance sampling in the context of destination choice modelling is to over-sample destinations close to the home and under-sample destinations away from home (refer to Table 54 below).

This sampling scheme is more efficient in terms of its ability to reduce parameter variance as compared to a simple random sampling approach where all alternatives are sampled with equal weight.

In the present project, we have applied a sampling strategy with the following configuration;

Short/Long	**Distance bands**	**Number of alternatives**	**Average selection prob**
Below 600 KM.	0-200	10	0.5633
	200-400	5	0.3460
	400-600	5	0.1683
Above 600 KM.	600-1200	8	0.0517
	1200-1800	7	0.0347
	1800-	5	0.0249

Table 54: Importance sampling scheme.

Clearly, the sampling scheme for the longer trips is the most critical as seen from the selection probability. On the other hand, it seems as if trips below 600 KM, have been covered well.

There have been various exceptions to the sampling pattern illustrated above.

- If it has not been possible to sample enough zones in the inner-band, the remaining zones has been sampled from the medium-band
    - If in this case, there are not enough alternatives in the medium-band, the remaining zones have been sampled from the outer-band (and example where this is necessary would be the Canary Islands).
- If there is enough zones in the inner-band, but not enough in the medium band, sampling of alternatives for the medium-band has used inner-band zones
    - If in this case, the inner-band "runs empty" we have sampled the remaining zones from the outer-band.
- If there are not enough zones in the outer-band, we have sampled first from the medium-band and then subsequently from the inner-band.

For trips below 600 KM and due to the exceptions listed above, it may happen that the number of sampled alternative destinations is lower than 20. In that case we apply the sampled number rather than excluding them from the estimation. For the trips above 600 KM there are always 20 alternatives.

As evidenced by McFadden (1978), sampling of alternatives may be used in a MNL model. If an importance sampling is applied the MNL needs to be adjusted in order to compensate for the fact that the sample selection is not random.

It can be shown (Ben-Akiva and Lerman, 1985) that the sample correction term $\Omega_{n,b}$ for individual n and distance band b is given by

(10) $$\Omega_{n,b} = -\log(q_{n,b})$$

Where $q_{n,b}$ is the selection probability.

For nested logit models, the proof for consistency of the MNL estimator as provided by McFadden does not apply.

However, given the fact that importance sampling is needed due to feasibility of the computations, we instead examine the sensitivity of the estimated parameters with respect to the sampling scheme. In Figure 21 and Figure 22 the percentage deviation of the parameters GTC, FerryTime, and AccEggTime are tested as a function of sampled alternatives.

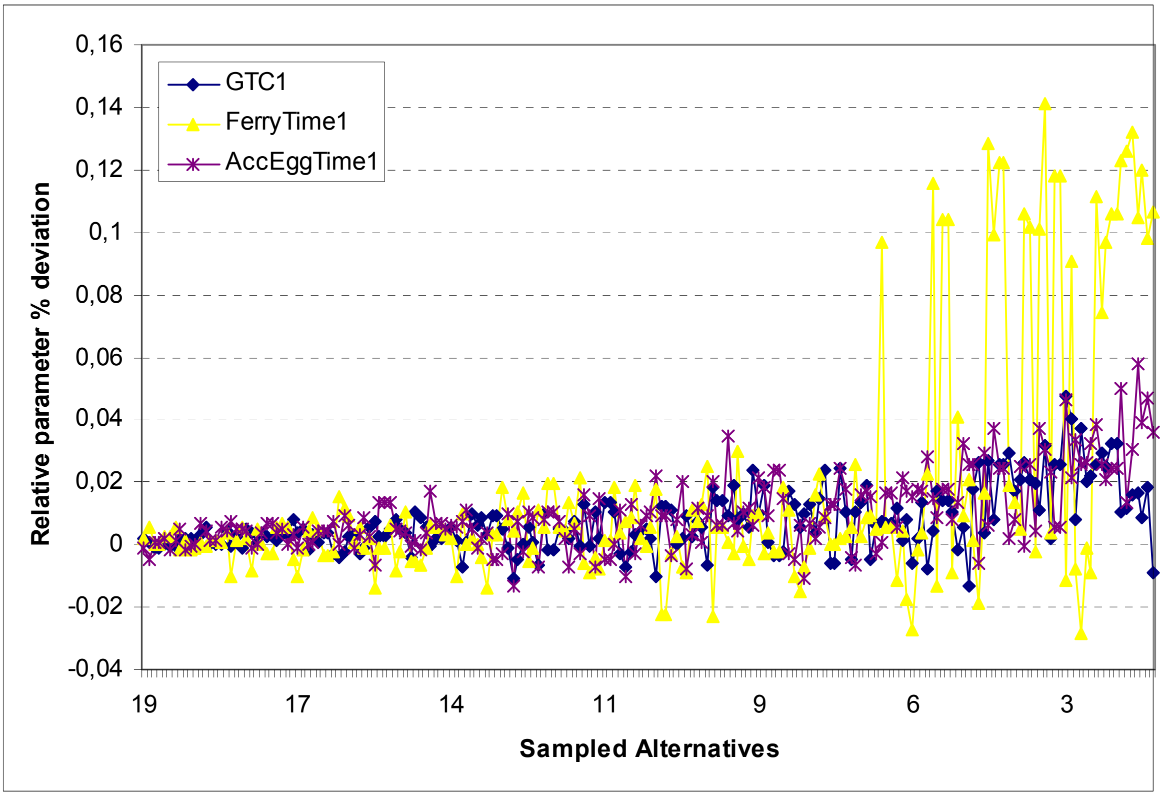

Figure 21: Parameter sensitivity analysis due to sampling of short-distance destinations.

The first sensitivity analysis in Figure 21 has been carried out by first running the model with all 20 destination alternatives. Hereafter, we have analysed parameters after removing one alternative randomly. This process is repeated 10 times for each number of alternatives. E.g. for 10 alternatives, we analyse the effect of 10 different random combinations of alternatives. Each minor tick-mark on the x-axis represents a simulation.

Since the model is split on two-distance bands, the most variation due to sampling of short-distance destinations is in the short-distance parameters.

Figure 21 indicate that the sampling strategy is sufficient. Clearly there are signs of some parameter instability in the sense that parameters “drift” upward or downward as we reduce the number of sampled alternatives. However, the relative deviation tends to be quite low. Moreover, it seems as if the parameters stability beyond 10 sampled alternatives is fairly good with a maximum deviation of less than 2%. The maximum deviation for 15 or more alternatives is only 1%.

It is interesting to see that even without destination sampling, e.g. when only one alternative is included, the model performs fairly well for the GTC1 and the AccEggTime1 with a maximum deviation of 5%. The reason is that most of the travel time preferences explained by the model is in fact explained by the mode-choice. However, for FerryTime1 the destination choice seems to play a role.

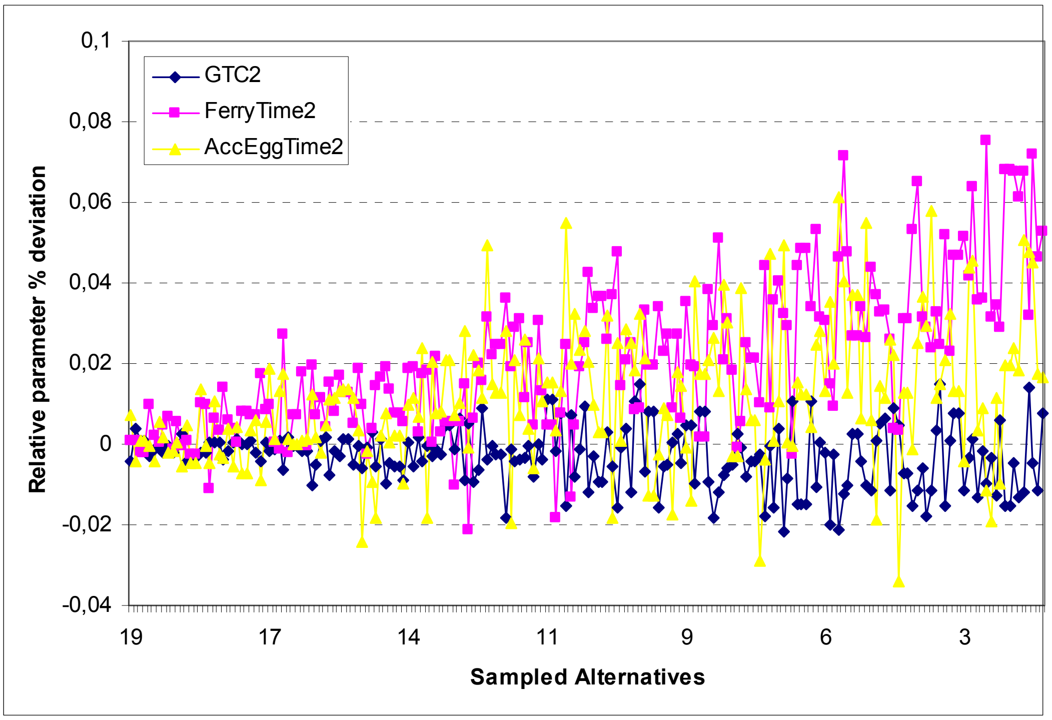

Figure 22: Parameter sensitivity analysis due to sampling of long-distance destinations.

As expected, there is more variation in the long-distance parameters as seen in Figure 22. This is due to a lower sample selection probability as shown in Table 54. Still, however, the parameter stability from 17 sampled alternatives and up seems quite reasonable.

6.4.2 Tree-structure

Initially, 6 possible nesting structures were considered.

- Short/Long, Mode, and Destination
- Short/Long, Destination, and Mode
- Mode, Short/Long, and Destination
- Mode, Destination, and Short/Long
- Destination, Short/Long, Mode
- Destination, Mode, and Short/Long

The type of nesting has no behavioural interpretation in the sense that some decisions are causal to others. The nesting structure only represents a statistical specification of the error-structure in the model.

It was decided to neutralise the short/long nest due to identification problems and the fact that we have non-generic time/cost parameters with respect to the short/long division.

In the choice among the different nesting, we have considered;

- Consistency of logsum parameters (within the unit interval)
- Overall likelihood value (see below)

For estimation of models, a “same-scale” constraint has been applied to all nests. It implies that the logsum parameters are constrained to be identical across the active nests (the scaling of the lower nest is per default = 1). The result of the different nesting structures is given below;

Purpose	Tree structure	LogL(0)	LogL(β)
Private	Mode, Dest	-97254	-57018
	Dest, Mode		-56687
Business	Mode, Dest	-49015	-25086
	Dest, Mode		-24619
Holiday	Mode, Dest	-519999	-177.236
	Dest, Mode		-166.822

Table 55: Test of nesting structures.

Horowitz (1983) provide at test for comparison of non-nested hypothesis aimed at discrete choice models. In the following we adopt the approximation described in Ben-Akiva and Lerman (1985).

The pseudo rho-square measure is given by

$$\bar{\rho}^2 = 1 - \frac{\ell(\hat{\beta}) - K}{\ell(0)} \qquad (11)$$

Where K defines the number of estimated parameters, whereas $\ell(\hat{\beta})$ and $\ell(0)$ are the log-likelihood value evaluated at $\hat{\beta}$ and 0 respectively.

Nest structure	Private	Business	Holiday
Mode. Dest	0.4138	0.48850	0.65918
Dest. Mode	0.4172	0.49803	0.67922

Table 56: $\bar{\rho}^2$ values for different nesting structures and by purpose.

As stated in Ben-Akiva and Lerman (1985):

"*for more than 250 observations with two or more alternatives and models having the same number of parameters, if the $\bar{\rho}^2$ of the two models differ by 0.01 or more, the model with the lower $\bar{\rho}^2$ is almost certainly incorrect.*"

In other words, without further examination, we may conclude that the nesting structure illustrated in Table 56 is optimal from a statistical point of view.

6.5 Estimation results

The estimation results will be presented in three sections, one for each purpose.

6.5.1 Business trips

6.5.1.1 Functional form

The choice of functional form of the GTC-variable has been investigated. The likelihood value for six models are shown in Table 57.

MNL(Linear): Multinomial logit, linear GTC1 (short) , linear GTC2 (long)
MNL(Log): Multinomial logit, logarithmic GTC1 (short) , logarithmic GTC2 (long)

MNL(Linear/Logit): Multinomial logit, linear GTC1 (short) , logarithmic GTC2 (long)

Model Specification	Log-Likelihood	$\bar{\rho}^2$
MNL L(0)	-49015	
MNL (Linear)	-25710	0.4573
MNL (Log)	-25388	0.4641
NMNL (Linear)	NA	NA
NMNL (Log)	-23979	0.4973
NMNL (Linear/Log)	-23837	0.4980

Table 57: Test for functional form[4].

Table 57 suggests that the combined linear/log-specification is the better model.

However, the log-likelihood value may be a poor basis for judgment of functional form. To further test the specification, we looked at the prediction ability for the two models.

Mode	Long/short split	Obs	Linear/Log	Log
1	Long	282	316.45	314.07
1	Short	3698.00	3775.95	3751.88
2	Long	42	36.94	31.22
2	Short	200	234.49	219.81
3	Long	82	80.69	82.02
3	Short	889	818.34	846.34
4	Long	544	459.03	470.53
4	Short	541	556.06	562.10
Std.Dev.			54,66	41,58

Table 58: Observed and predicted number of trips by long/short split.

Overall, the prediction ability for long and short trips respectively is good. The log-specification tends to produce a slightly lower standard deviation and is therefore slightly better in this sense.

Model estimates is presented below.

```
Parameter      DF Estimate Std.Error  t Value Pr > |t|
  M1            1   -0.6889    0.0958   -7.19  <.0001
  M2            1   -3.5810    0.1139  -31.44  <.0001
  M3            1   -3.1346    0.1334  -23.50  <.0001
  Size2         0    1.0000         0
  Adj           0    1.0000         0
  GTC_1         1 -0.002893 0.0000984  -29.40  <.0001
  LOG_GTC2      1   -0.7229  0.009011  -80.23  <.0001
  FerryTime_1   1   -0.2538    0.0326   -7.78  <.0001
  FerryTime_2   1   -0.1363    0.0181   -7.51  <.0001
  AccEggTime_1  1   -0.3206    0.0168  -19.04  <.0001
  AccEggTime_2  1   -0.1475    0.0142  -10.35  <.0001
  HeadWayTime_1 1   -0.2696    0.0235  -11.46  <.0001
  HeadWayTime_2 1   -0.1332    0.0136   -9.80  <.0001
  Freq_1        1    0.0113  0.000827   13.72  <.0001
  Freq_2        1  0.005775  0.001540    3.75  0.0002
```

[4] Refer to the appendix for detailed goodness-of-fits reports.

Restrict1	1	2029	99.9903	20.29	<.0001*	Linear EC
Restrict2	1	3939	49.4857	79.59	<.0001*	Linear EC

Table 59: Multinomial logit estimates for the Business model with the Linear/Log specification.

Parameter	DF	Estimate	Std.Error	t Value	Pr > |t|	
M1_L1	1	-1.5735	0.0865	-18.20	<.0001	
M2_L1	1	-3.5681	0.1070	-33.34	<.0001	
M3_L1	1	-3.1708	0.1632	-19.43	<.0001	
Size1_L1	0	1.0000	0			
Adj_L1	0	1.0000	0			
CarAv_1_L1	0	0.3695	0			
CarAv_2_L1	0	0.3695	0			
GTC_1_L1	1	-0.002644	0.000156	-16.90	<.0001	
LOG_GTC2_L1	1	-0.8455	0.0160	-52.94	<.0001	
FerryTime_1_L1	1	-0.002296	0.000138	-16.62	<.0001	
FerryTime_2_L1	1	-0.001254	0.0000648	-19.33	<.0001	
AccEggTime_1_L1	1	-0.005916	0.000195	-30.40	<.0001	
AccEggTime_2_L1	1	-0.002694	0.000157	-17.16	<.0001	
HeadWayTime_1_L1	1	-0.001997	0.000397	-5.03	<.0001	
HeadWayTime_2_L1	1	-0.002280	0.000287	-7.94	<.0001	
Freq_1_L1	1	0.0208	0.002363	8.82	<.0001	
Freq_2_L1	1	0.002092	0.003291	0.64	0.5251	
INC_L2G1	1	0.5620	0.008450	66.51	<.0001	
Restrict1	1	-1905	60.1080	-31.68	<.0001*	Linear
Restrict2	1	2731	31.3428	87.14	<.0001*	Linear
Restrict3	1	200.3211	27.6738	7.24	<.0001*	Linear
Restrict4	1	120.5128	16.3595	7.37	<.0001*	Linear

Table 60: Nested logit estimates for the Business model with the Linear/Log specification.

Mode **DistID**	**1** **1**	**1** **2**	**2** **1**	**2** **2**	**3** **1**	**3** **2**	**4** **1**	**4** **2**
Base	3777	316	235	37	818	81	556	459
Car: GTC +25%	3552	292	266	41	919	89	620	499
Bus: GTC +25%	3806	319	192	31	825	81	562	463
Rail: GTC +25%	3857	322	240	38	715	68	571	467
Air: GTC +25%	3855	331	241	39	837	85	483	407
AccEgg: AccEggTime +25%	3951	341	248	41	778	82	445	392
HeadWay: HeadWayTime +25%	3850	328	240	39	836	85	481	419
Freq: Freq +25%	3502	301	216	35	1171	102	514	437
Ferry: FerryTime + 25%	3771	309	236	37	821	80	561	463

Table 61: Sensitivity tests with a 25% increase in selected variables for the Linear/log model.

Mode **DistID**	**1** **1**	**1** **2**	**2** **1**	**2** **2**	**3** **1**	**3** **2**	**4** **1**	**4** **2**
Mode	-0.238	-0.306	0.542	0.404	0.498	0.407	0.462	0.345
DistID	0.031	0.030	-0.725	-0.687	0.034	0.040	0.040	0.032
GTC(Car)	0.085	0.072	0.099	0.088	-0.503	-0.645	0.105	0.072
GTC(Bus)	0.083	0.179	0.102	0.253	0.095	0.245	-0.527	-0.455
GTC(Rail)	0.185	0.306	0.230	0.478	-0.193	0.078	-0.799	-0.587
GTC(Air)	0.078	0.141	0.099	0.225	0.089	0.213	-0.538	-0.348
AccEggTime	-0.291	-0.192	-0.313	-0.205	1.731	1.065	-0.305	-0.197
HeadWayTime	-0.006	-0.096	0.023	0.048	0.014	-0.017	0.033	0.036

Table 62: Approximate elasticities for selected variables[5].

[5] Calculated on the basis of a 25% change.

These elasticities conform to compensated elasticities without income effects.

6.5.2 Private trips

6.5.2.1 Functional form

The choice of functional form of the GTC-variable has been investigated. The likelihood values are shown in table.

Model	LogLikelihood	$\bar{\rho}^2$
MNL(0)	-97254	
MNL (Linear)	-65136	0.3304
MNL (Log)	-66201	0.3151
NMNL (Linear)	-57301	0.4109
NMNL (Log)	-56971	0.4144
NMNL (Linear/Log)	-56687	0.4173

Table 63: Log-likelihood performance for different GTC specifications for the private.

From the point of view of the Horowitz test, the linear/log formulation is the better model.

To further test the specification, we looked at the prediction ability for the two models.

Mode	Long/short split	Obs	Linear/Log	Log
1	Long	464	384.98	412.52
1	Short	11193	11465.70	11424.29
2	Long	105	110.22	88.43
2	Short	1388	1289.07	1360.95
3	Long	147	163.61	166.72
3	Short	1402	1217.35	1205.95
4	Long	225	228.48	228.40
4	Short	213	277.57	249.70
	Std.Dev.		235.57	108.78

Table 64: Observed and predicted number of trips by long/short split.

As before, it seems as if the log-specification deals with distance slightly better than the corresponding linear specification. However, the Linear/Log specification brings out more explanation power in the remaining LoS variables.

The estimation results for the different models is shown below;

```
Parameter        DF  Estimate Std.Error t Value  Pr > |t|
  M1              1    2.9393    0.1151   25.54  <.0001
  M2              1    1.1322    0.1177    9.62  <.0001
  M3              1    0.3560    0.1295    2.75  0.0060
  Size1           0    1.0000         0
  Adj             0    1.0000         0
  GTC_1           1 -0.008607  0.000126  -68.26  <.0001
  LOG_GTC2        1   -0.9669  0.008092 -119.50  <.0001
  FerryTime_1     1   -0.0459  0.006404   -7.16  <.0001
  FerryTime_2     1   -0.0718  0.004888  -14.70  <.0001
```

AccEggTime_1	1	-0.1298	0.0161	-8.06	<.0001	
AccEggTime_2	1	0.0266	0.006756	3.94	<.0001	
HeadWayTime_1	1	-0.2440	0.0336	-7.27	<.0001	
HeadWayTime_2	1	-0.1667	0.0194	-8.58	<.0001	
Freq_1	1	0.008609	0.000584	14.74	<.0001	
Freq_2	1	0.004702	0.001204	3.90	<.0001	
Restrict1	1	19048	157.4848	120.95	<.0001*	Linear EC
Restrict2	1	4943	63.9715	77.27	<.0001*	Linear EC

Table 65: Multinomial logit estimates for the Private model with the Linear/Log specification.

Parameter	DF	Estimate	Std.Error	t Value	Pr > |t|	
M1_L1	1	1.1585	0.1235	9.38	<.0001	
M2_L1	1	0.3914	0.1230	3.18	0.0015	
M3_L1	1	-0.4388	0.1550	-2.83	0.0047	
Size2_L1	0	1.0000	0			
Adj_L1	0	1.0000	0			
CarAv_1_L1	1	0.7383	0.0198	37.28	<.0001	
CarAv_2_L1	1	0.7344	0.0470	15.61	<.0001	
GTC_1_L1	1	-0.007960	0.000133	-59.63	<.0001	
LOG_GTC2_L1	1	-1.7268	0.0249	-69.44	<.0001	
FerryTime_1_L1	1	-0.003315	0.000222	-14.91	<.0001	
FerryTime_2_L1	1	-0.000967	0.0000747	-12.94	<.0001	
AccEggTime_1_L1	1	-0.003052	0.000156	-19.54	<.0001	
AccEggTime_2_L1	1	0.0000768	0.000136	0.57	0.5718	
HeadWayTime_1_L1	1	-0.000778	0.000435	-1.79	0.0739	
HeadWayTime_2_L1	1	-0.000213	0.000365	-0.59	0.5584	
Freq_1_L1	1	0.0108	0.001836	5.86	<.0001	
Freq_2_L1	1	0.0137	0.002724	5.02	<.0001	
INC_L2G1	1	0.3748	0.004898	76.53	<.0001	
Restrict1	1	-249.0811	66.1122	-3.77	0.0002*	Linear
Restrict2	1	2642	36.4161	72.56	<.0001*	Linear

Table 66: Nested estimates for the Private model with the Linear/Log specification. Latest re-est (4 sep.).

As before, model estimates are fairly encouraging in that most of the LoS variables are properly identified.

Again, the nested logit model out-perform the MNL model, and the logsum parameter is in the unit interval.

Mode **DistID**	**1** **1**	**1** **2**	**2** **1**	**2** **2**	**3** **1**	**3** **2**	**4** **1**	**4** **2**
Base	11462	385	1288	110	1222	164	277	228
Car: GTC +25%	10913	333	1533	126	1453	186	338	256
Bus: GTC +25%	11726	395	971	81	1264	171	292	236
Rail: GTC +25%	11693	397	1325	117	954	123	289	238
Air: GTC +25%	11500	396	1294	116	1228	172	252	177
AccEgg: AccEggTime +25%	11581	393	1309	115	1141	161	231	207
HeadWay: HeadWayTime +25%	11529	397	1301	118	1234	175	213	170
Freq: Freq +25%	11174	373	1248	104	1550	202	267	219
Ferry: FerryTime + 25%	11449	364	1297	114	1228	167	284	233

Table 67: Sensitivity test with an increase in 25% for selected variables.

Mode **DistID**	**1** **1**	**1** **2**	**2** **1**	**2** **2**	**3** **1**	**3** **2**	**4** **1**	**4** **2**
GTC(Car)	-0.192	-0.541	0.759	0.587	0.753	0.534	0.865	0.491
GTC(Bus)	0.092	0.110	-0.986	-1.072	0.136	0.189	0.215	0.128
GTC(Rail)	0.081	0.122	0.115	0.256	-0.877	-0.996	0.173	0.175
GTC(Air)	0.014	0.119	0.019	0.204	0.019	0.212	-0.367	-0.893
AccEggTime	0.042	0.087	0.064	0.171	-0.267	-0.074	-0.675	-0.376
HeadWayTime	0.024	0.127	0.038	0.280	0.037	0.284	-0.929	-1.023
Freq	-0.100	-0.127	-0.127	-0.221	1.072	0.944	-0.146	-0.171
FerryTime	-0.004	-0.217	0.028	0.133	0.019	0.090	0.094	0.084

Table 68: Approximate elasticities for selected variables[6].

6.5.3 Holiday trips

For holiday trips, the issue of functional form has been slightly mixed. It has not been possible to estimate separate parameters for the GTC variable split on distance.

To account for this, we have introduced the following constraint in the estimation

(12) TC_1 * a = TC_2

In other words, we apply the same internal scaling between long and short distance GTC as for the private trip segment, whereas the level is set by the model itself.

This is most likely due to a weak attraction description for the holiday segment. In fact, this is not surprising at all, in that many short holiday trips is bound for summer houses and rural areas, which is very difficult to describe by aggregate attraction variables.

	MNL	
Variables	**Estimate**	**T-test**
ModeConst1	0.4282	15.68
ModeConst2	-1.0234	-41.90
ModeConst3	-1.8962	-80.70
CarAv_1	0.5343	37.98
CarAv_2	0.3911	23.26
Size1	1.0000	Fixed
GTC_1	-0.2334	190.11
GTC_2	-0.6909	190.11
FerryTime_1	-0.0893	-40.99
FerryTime_2	-0.1116	-55.63
AccEggTime_1	-0.0575	-24.23
AccEggTime_2	-0.003676	-2.38
Freq_1	0.0800	Fixed
Freq_2	0.1000	Fixed
HeadWayTime1	0	Fixed
HeadWayTime2	0	Fixed
Adj	1.0000	Fixed
Logsum	1.0000	Fixed

Table 69: parameter estimates for the holiday trips.* Estimated under the constraint that GTC_1 * 2.96 = GTC_2.

[6] Calculated on the basis of a 25% change.

Parameter	DF	Estimate	Error	t Value	Pr > |t|	
M1_L1	1	-0.0965	0.0272	-3.55	0.0004	
M2_L1	1	-1.1642	0.0248	-46.95	<.0001	
M3_L1	1	-1.4419	0.0263	-54.83	<.0001	
Size3_L1	0	1.0000	0			
Adj_L1	0	1.0000	0			
CarAv_1_L1	1	0.7262	0.0172	42.11	<.0001	
CarAv_2_L1	1	0.8611	0.0168	51.24	<.0001	
GTC_1_L1	1	-0.003078	0.0000348	-88.40	<.0001	
LOG_GTC2_L1	1	-0.6402	0.007243	-88.40	<.0001	
FerryTime_1_L1	1	-0.000331	0.0000444	-7.45	<.0001	
FerryTime_2_L1	1	-0.001642	0.0000307	-53.54	<.0001	
AccEggTime_1_L1	0	-0.001877	0			
AccEggTime_2_L1	0	-0.000622	0			
HeadWayTime_1_L1	0	-0.002440	0			
HeadWayTime_2_L1	0	-0.000951	0			
INC_L2G1	1	0.3414	0.002828	120.71	<.0001	
Restrict1	1	-9042	155.5674	-58.13	<.0001*	Linear
Restrict2	1	23656	86.1518	274.58	<.0001*	Linear
Restrict3	1	-3145	42.6657	-73.71	<.0001*	Linear
Restrict4	1	-286727	6337	-45.25	<.0001*	Linear
Restrict5	1	106134	1146	92.60	<.0001*	Linear
Restrict6	1	-410243	5607	-73.17	<.0001*	Linear
Restrict7	1	42846	4015	10.67	<.0001*	Linear

Table 70: Parameter estimates for the holiday purpose, Llinear/log form. Final re-estimate (4. sept).

In the following we shall briefly consider the demand sensitivity of the model with respect to other LoS variables than the TC-variable.

The derived elasticity structure has been calculated as a marginal increase by 25% in the respective variables and only the business trip models has been tested. In fact, there are quite good accordance in terms of these variables for the three trip purposes, e.g. we expect the business trip purpose to be fairly representative.

Variable	Car	Bus	Rail	Airplane
Air AccEggTime	0.1189	0.1217	0.1253	-0.5652
Bus FerryTime	0.0029	-0.0790	0.0030	0.0042
Car FerryTime	-0.0230	0.0319	0.0332	0.0466
Rail AccEggTime	0.0376	0.0384	-0.2025	0.0354
Rail FerryTime	0.0008	0.0009	-0.0053	0.0021
Air HeadwayTime	0.016073	0.017211	0.019573	-0.0789

Table 71: Elasticities for other time components.

As seen, there are significant responses to changes in these LoS variables. Although fairly small for the ferry time component it is large for the Access-Egress component.

6.5.4 Calibration

The various models have been calibrated in various respects. Firstly, we have calibrated mode constants according to the OD matrices as seen in Table 72.

Mode	Parameter name	Private	Business	Holiday
Car driver	M1_L1	1.809122856	-1.084637845	-0.735597967
Car passenger	M2_L1	0.604128709	-2.79989712	-1.07108248
Bus	M3_L1	1.357672992	-2.805965386	0.220565109
Rail	M4_L1	-1.042150646	-3.130247393	0.132256341
Air	M5_L1	-0.87528367	-1.466191871	0.526118442

Table 72: Alternative calibrated mode constants.

In addition, the frequency model has been calibrated as well, to fit levels at the zone-level. In both cases a Lerman and Manski (1977) approach has been used.

6.6 Trip generation

The long distance generation model is formulated as a random utility model, where probabilities are assigned to the generation of 0, 1, 2, 3, and 4 or more trips. The model is aimed at forecasting and only variables that may be forecast are included.

6.6.1 Data

DATELINE was used for the estimation. The survey includes information on households, journeys, and trips. The household database includes number of long journeys split by purpose (THJL=holidays trips in a year, TPJL=private trips within a 3 months period, and TBJL=business trips within a 3 months period).

The data include the geographical location of each household at NUTS3 level and the number of cars (private) of each household. GDP and population statistics are available at NUTS3 level from EUROSTAT.

The data consist of 55544 observations. Because of inconsistent NUTS3 codes and some missing data, the final data have a total of 44666 observations.

In the tables the +weight column indicates the average number of trips made by households in the 6+ group for business and private trips, and 4+ trips for holiday trips.

Purpose	0	1	2	3	4	5	6+	+weight
Business	92.92	4.15	1.37	0.67	0.33	0.34	0.21	6.00
Holiday	46.81	30.60	13.90	8.65	0.05			4.52
Private	79.13	15.52	3.33	1.09	0.45	0.39	0.09	6.09

Table 73: Generation by trip purpose before cleaning.

There is no great difference between the frequencies in Table 73 and Table 74.

Purpose	0	1	2	3	4	5	6+	+weight
Business	92.98	4.03	1.41	0.66	0.35	0.34	0.22	6.00
Holiday	44.12	31.33	14.81	9.69	0.05			4.46
Private	78.97	15.63	3.36	1.08	0.49	0.39	0.09	6.10

Table 74: Generation by trip purpose.

The DATELINE survey includes information about household expansion factors, however, more than half of these are missing and as a result no weighting has been applied.

6.6.2 Model

A multinomial logit (MNL) model is used to model trip generation preferences, i.e. the probability of k trips is given by

(13) $$P_i(k) = \frac{e^{V_{ik}}}{\sum_{k \in K} e^{V_{ik}}}.$$

The dependent variable is the number of trips carried out by a household in a year. The utility function is given by

(14) $$\mathrm{V_{ik}} = \mathrm{a_k} + \mathrm{b_k} \cdot \ln(\mathrm{GDPper_i}) + \mathrm{g_k} \cdot \ln(\mathrm{ncars_i} + 1) + \mathrm{q_k} \cdot \mathrm{logsum_i}$$

where GDPper_i is GDP per capita in the zone (in 1000 Euros) of the household, ncars is the number of private cars in the household (in the aggregate model this may be approximated by the average zonal car ownership), and logsum is the logsum for zone i from the destination/mode choice model.

The logarithmic transformation has been applied in order to avoid scale-effects, e.g. due to GDP changes.

A simple model approach has been chosen to assure that higher GDP and car ownership gives rise to more trips. The estimation results are given below. For the business purpose, nine coefficients were estimated.

Parameter	Estimate	Std. dev.
a_1	-6.9319	0.53
a_2	-7.8490	0.83
a_3	-7.8490	0.83
a_{4+}	-8.6267	0.83
g_1	1.3357	0.07
g_{2+}	1.6875	0.08
b_{1+}	0.4166	0.05
q_1	0.0870	0.03
q_{2+}	0.2049	0.06

Table 75: Estimates for business trip generation.

For the private purpose, nine coefficients were estimated.

Parameter	Estimate	Std. dev.
a_1	-2.8751	0.25
a_2	-3.9204	0.40
a_3	-5.0543	0.40
a_{4+}	-6.2691	0.40
g_1	0.6839	0.03
g_{2+}	1.0010	0.06
b_1	0.0992	0.03
b_{2+}	0.3875	0.05
q_{1+}	0.0307	0.02

Table 76: Estimates for private trip generation.

For the holiday purpose. twelve coefficients were estimated.

Parameter	Estimate	Std. dev.
a_1	-9.0801	0.92
a_2	-11.5499	1.01
a_3	-12.8207	1.02
a_{4+}	-14.0150	1.04
g_1	1.0551	0.03
g_2	1.3587	0.04
g_{3+}	1.4682	0.05
b_1	0.6324	0.03
b_2	0.9416	0.04
b_{3+}	1.1726	0.04
q_1	0.4370	0.06
q_{2+}	0.6663	0.07

Table 77: Estimates for holiday trip generation.

The models outlined in the tables conform to a random-utility set-up with utility functions for the different alternatives. As an example the utility functions for the holiday purpose is given by.

(15)
$$\begin{aligned}
V_{i0} &= 0 \\
V_{i1} &= a_1 + b_1 \cdot \ln(\mathrm{GDPper_i}) + g_1 \cdot \ln(\mathrm{ncars_i} + 1) + q_1 \cdot \log sum_i \\
V_{i2} &= a_2 + b_2 \cdot \ln(\mathrm{GDPper_i}) + g_2 \cdot \ln(\mathrm{ncars_i} + 1) + q_{2+} \cdot \log sum_i \\
V_{i3} &= a_3 + b_{3+} \cdot \ln(\mathrm{GDPper_i}) + g_{3+} \cdot \ln(\mathrm{ncars_i} + 1) + q_{2+} \cdot \log sum_i \\
V_{i4+} &= a_{4+} + b_{3+} \cdot \ln(\mathrm{GDPper_i}) + g_{3+} \cdot \ln(\mathrm{ncars_i} + 1) + q_{2+} \cdot \log sum_i
\end{aligned}$$

Note that coefficients with a + also apply to all higher cases, e.g. for business $a_{6+} = a_5 = a_4 = a_{4+}$.

The DATELINE survey covers varying periods depending on the trip purpose (Table 44). In order to expand the model output to an annual basis, the output should be scaled accordingly.

6.7 Implementation and forecasting

The estimation and application of the model framework are two different exercises. In the estimation part, which is based on the DATELINE survey, the aim is to reflect the preference structure of the survey in the best possible way. In the application part, model structure and preferences are held constant (reflected by the parameters from the estimation) and applied to reference data and scenario data.

The implementation of the model differs from the estimation in the sense that, whereas estimation is carried out on the basis of the DATELINE survey, the reference of the implementation is pure LoS matrices, zone data, and OD matrices.

The “cook-book” following the implementation of the model is outlined below in 15 successive steps;

Step 1: Compile an aggregate data structure

This first step involves setting up the complete “gross” data-structure. The structure should be in a form identical to Table 78 below.

FromZoneID	ToZoneID	Purpose	Var1	Var2…
1	1	1	X	g
1	2	2	Y	h
1	3	3	Z	j
…	…	…	…	…

Table 78: Data structure.

More specifically, the variables included in the table should be as described in Table 79 below;

Variable	Description	Unit
FromZoneID	The TENCONNECT 1441 zone coding, from zone	
ToZoneID	The TENCONNECT 1441 zone coding, to zone	
Purpose	1 = Business, 2 = Private, 3 = Holiday	
Distance	Distance matrix (Euclidian distances) in km	km
DistanceMedium	1 if Distance <= 600; 0 else where	0/1
VOT	Value-of-time	Euro/Hour
POP	Population	People
JOB	Jobs	Jobs
GDP	Gross Domestic Product	Mill. Euros
CAP	Bed place capacity	Bed places
CarAv	Average car propensity for households	Cars/household
Cost1	Variable km cost – car	Euro
Cost3	Variable km cost – bus	Euro
Cost4	Estimated ticket price costs – rail	Euro
Cost5	Estimated ticket price costs – air	Euro
TollCost1	Toll cost – car	Euro
TollCost3	Toll cost – bus	Euro
In-Vehicle-time1	FreeFlow time – car	Minutes
TravelTime3	“Fixed” pre calculated bus travel time (see below) – bus	Hours
In-Vehicle-time4	Onboard time – rail	Minutes
In-Vehicle-time5	Onboard time – air	Minutes
CongestionTime1	Congestion time – car	Minutes
Access-EgressTime4	Access-Egress time – rail	Minutes
Access-EgressTime5	Access-Egress time – air	Minutes
FerryTime1	Ferry onboard time – car	Minutes
FerryTime3	Ferry onboard time – bus	Minutes
FerryTime4	Ferry onboard time – rail	Minutes
FerryWTime1	Ferry waiting time – car	Minutes
FerryWTime3	Ferry waiting time – bus	Minutes
FerryWTime4	Ferry waiting time – rail	Minutes
HeadWayTime5	HeadWay time – air	Minutes
Freq4	Frequency variable – rail	Freq per 24 hour
Transfer5	Transfer time – air	Minutes

Table 79: Gross variable list.

If in the above table, variables do not vary across purpose (e.g. POP), then values are just repeated. This makes the implementation simpler.

In the table above, TravelTime3 is a “fixed” variable measuring the travel time for bus. It is fixed in the sense that it is not allowed to vary from the base-line data to scenario data.

Step 2: Defining variables to be used in the utility functions

Given the data structure described in Table 78 and Table 79, we are in a position to derive all of the variables to be used in the specification of the utility model.

In addition to the 4 initial modes, we add a “car passenger mode”. The numbering of modes is given by;

1: Car
2: Car passenger
3: Bus
4: Rail
5: Air

Add the following variables to the above data structure;

GTC’s

GTC1 = Cost1 + TollCost1 + VOT*(In-Vehicle-time1 + 1.5*CongestionTime1 + 1.5*FerryWTime1)

GTC2 = VOT*(In-Vehicle-time1 + 1.5*CongestionTime1 + 1.5*FerryWTime1)

GTC3 = Cost4 + VOT*TravelTime3

GTC4 = Cost4 + VOT*(In-Vehicle-time4 + 1.5*FerryWTime4)

GTC5 = Cost5 + q_{vot}*VOT*(In-Vehicle-time5)

Distance dependent GTC’s

GTC1_1 = GTC1 * DistanceShort

GTC1_2 = GTC1 * (1-DistanceShort)

GTC2_1 = GTC2 * DistanceShort

GTC2_2 = GTC2 * (1-DistanceShort)

GTC3_1 = GTC3 * DistanceShort

GTC3_2 = GTC3 * (1-DistanceShort)

GTC4_1 = GTC4 * DistanceShort

GTC4_2 = GTC4 * (1-DistanceShort)

GTC5_1 = GTC5 * DistanceShort

GTC5_2 = GTC5 * (1-DistanceShort)

Log of GTC’s

LogGTC1_1 = ln(GTC1_1 + 1)

LogGTC1_2 = ln(GTC1_2 + 1)

LogGTC2_1 = ln(GTC2_1 + 1)

LogGTC2_2 = ln(GTC2_2 + 1)

LogGTC3_1 = ln(GTC3_1 + 1)

LogGTC3_2 = ln(GTC3_2 + 1)

LogGTC4_1 = ln(GTC4_1 + 1)

LogGTC4_2 = ln(GTC4_2 + 1)

LogGTC5_1 = ln(GTC5_1 + 1)

LogGTC5_2 = ln(GTC5_2 + 1)

Distance dependent – other LoS

AccEggTime4_1 = Access-EgressTime4 * (DistanceShort)

AccEggTime4_2 = Access-EgressTime4 * (1-DistanceShort)

AccEggTime5_1 = Access-EgressTime5 * (DistanceShort)

AccEggTime5_2 = Access-EgressTime5 * (1-DistanceShort)

FerryTimes1_1 = FerryTime1 * (DistanceShort)

FerryTimes1_2 = FerryTime1 * (1-DistanceShort)

FerryTimes2_1 = FerryTime1 * (DistanceShort)

FerryTimes2_2 = FerryTime1 * (1-DistanceShort)

FerryTimes3_1 = FerryTime3 * (DistanceShort)

FerryTimes3_2 = FerryTime3 * (1-DistanceShort)

FerryTimes4_1 = FerryTime4 * (DistanceShort)

FerryTimes4_2 = FerryTime4 * (1-DistanceShort)

HeadWayTime5_1 = HeadWayTime5 * (DistanceShort)

HeadWayTime5_2 = HeadWayTime5 * (1-DistanceShort)

Freq4_1 = Freq1 * (DistanceShort)

Freq4_2 = Freq1 * (1- DistanceShort)

Transfer5_1 = Transfer5 * (DistanceShort)

Transfer5_2 = Transfer5 * (1- DistanceShort)

Distance dependent – car availability

CarAv1_1 = CarAv * (DistanceShort)

CarAv1_2 = CarAv * (1- DistanceShort)

CarAv2_1 = CarAv1_1

CarAv2_2 = CarAv1_2

Attraction variables

Size =
(Purpose=1) * [0.0662*Ln(Cap+1) + 0.5966*Ln(GDP*100+1) + 0.8645*Ln(POP/1.5793 + JOB)]
+
(Purpose=2) * [0.3475*Ln(Cap+1) + 0.4884*Ln(GDP*100+1) + 0.9271*Ln(POP/1.5798 + JOB)]
+
(Purpose=3) * [0.9073*Ln(Cap+1) + 0.2072*Ln(GDP*100+1) + 1.0017*Ln(POP/1.5828 + JOB)]

Note3: Since each record is coded with separate indicators for trip purpose, the "size" variable will vary with purpose.

Handling non-availables

For some modes it may be that the present destination is non-available. In the LoS data "non-availability" is coded as 999999 or 0.

In the implementation of the model, we suggest to simply retain the 999999 coding because it will produce a probability of 0 for all practical purposes. For LoS entries having the value of 0, we suggest a re-coding accordingly, i.e. insert the following coding:

If GTC1 = 0 then GTC1 = 999999;
If GTC2 = 0 then GTC2 = 999999;
If GTC3 = 0 then GTC3 = 999999;
If GTC4 = 0 then GTC4 = 999999;
If GTC5 = 0 then GTC5 = 999999;

Step 3: Calculating utility functions V

The general form of the utility functions, leaving out an index for trip purpose, is given by;

If Purpose=1:

$$V_{m=1,d} = Size_d + k_{m=1} + \varphi_{1,GTC} * GTC_{m=1,d,1} + \varphi_{2,GTC} * LogGTC_{m=1,d,2} + \varphi_{1,FT} * FerryTime_{m=1,d,1} + \varphi_{2,FT} * FerryTime_{m=1,d,2} + \varphi_{1,CA} * CarAv_{m=1,d,1} + \varphi_{2,CA} * CarAv_{m=1,d,2}$$

$$V_{m=2,d} = Size_d + k_{m=2} + \varphi_{1,GTC} * GTC_{m=2,d,1} + \varphi_{2,GTC} * LogGTC_{m=2,d,2} + \varphi_{1,FT} * FerryTime_{m=2,d,1} + \varphi_{2,FT} * FerryTime_{m=2,d,2} + \varphi_{1,CA} * CarAv_{m=2,d,1} + \varphi_{2,CA} * CarAv_{m=2,d,2}$$

$$V_{m=3,d} = Size_d + k_{m=3} + \varphi_{1,GTC} * GTC_{m=3,d,1} + \varphi_{2,GTC} * LogGTC_{m=3,d,2} + \varphi_{1,FT} * FerryTime_{m=3,d,1} + \varphi_{2,FT} * FerryTime_{m=3,d,2}$$

$$V_{m=4,d} = Size_d + k_{m=4} + \varphi_{1,GTC} * GTC_{m=4,d,1} + \varphi_{2,GTC} * LogGTC_{m=4,d,2} + \varphi_{1,AE} * AccEgg_{m=4,d,1} + \varphi_{2,AE} * AccEgg_{m=4,d,2} + \varphi_{1,FT} * FerryTime_{m=4,d,1} + \varphi_{2,FT} * FerryTime_{m=4,d,2} + \varphi_{1,FREQ} * Freq_{m=4,d,1} + \varphi_{2,FREQ} * Freq_{m=4,d,2}$$

$$V_{m=5,d} = Size_d + k_{m=5} + q_{a1} * \varphi_{1,GTC} * GTC_{m=5,d,1} + q_{a2} * \varphi_{2,GTC} * LogGTC_{m=5,d,2} + \varphi_{1,AE} * AccEgg_{m=5,d,1} + \varphi_{2,AE} * AccEgg_{m=5,d,2} + \varphi_{1,HT} * HeadwayTime_{m=5,d,1} + \varphi_{2,HT} * HeadwayTime_{m=5,d,2} + \varphi_{1,TR} * Transfer_{m=5,d,1} + \varphi_{2,TR} * Transfer_{m=5,d,2}$$

If Purpose=2:

$$V_{m=1,d} = Size_d + k_{m=1} + \varphi_{1,GTC} * GTC_{m=1,d,1} + \varphi_{2,GTC} * LogGTC_{m=1,d,2} + \varphi_{1,FT} * FerryTime_{m=1,d,1} + \varphi_{2,FT} * FerryTime_{m=1,d,2} + \varphi_{1,CA} * CarAv_{m=1,d,1} + \varphi_{2,CA} * CarAv_{m=1,d,2}$$

$$V_{m=2,d} = Size_d + k_{m=2} + \varphi_{1,GTC} * GTC_{m=2,d,1} + \varphi_{2,GTC} * LogGTC_{m=2,d,2} + \varphi_{1,FT} * FerryTime_{m=2,d,1} + \varphi_{2,FT} * FerryTime_{m=2,d,2} + \varphi_{1,CA} * CarAv_{m=2,d,1} + \varphi_{2,CA} * CarAv_{m=2,d,2}$$

$$V_{m=3,d} = Size_d + k_{m=3} + \varphi_{1,GTC} * GTC_{m=3,d,1} + \varphi_{2,GTC} * LogGTC_{m=3,d,2} + \varphi_{1,FT} * FerryTime_{m=3,d,1} + \varphi_{2,FT} * FerryTime_{m=3,d,2}$$

$$V_{m=4,d} = Size_d + k_{m=4} + \varphi_{1,GTC} * GTC_{m=4,d,1} + \varphi_{2,GTC} * LogGTC_{m=4,d,2} + \varphi_{1,AE} * AccEgg_{m=4,d,1} + \varphi_{2,AE} * AccEgg_{m=4,d,2} + \varphi_{1,FT} * FerryTime_{m=4,d,1} + \varphi_{2,FT} * FerryTime_{m=4,d,2} + \varphi_{1,FREQ} * Freq_{m=4,d,1} + \varphi_{2,FREQ} * Freq_{m=4,d,2}$$

$$V_{m=5,d} = Size_d + k_{m=5} + q_{a1} * \varphi_{1,GTC} * GTC_{m=5,d,1} + q_{a2} * \varphi_{2,GTC} * LogGTC_{m=5,d,2} + \varphi_{1,AE} * AccEgg_{m=5,d,1} + \varphi_{2,AE} * AccEgg_{m=5,d,2} + \varphi_{1,HT} * HeadwayTime_{m=5,d,1} + \varphi_{2,HT} * HeadwayTime_{m=5,d,2} + \varphi_{1,TR} * Transfer_{m=5,d,1} + \varphi_{2,TR} * Transfer_{m=5,d,2}$$

If Purpose=3:

$$V_{m=1,d} = Size_d + k_{m=1} + \varphi_{1,GTC} * GTC_{m=1,d,1} + \varphi_{2,GTC} * LogGTC_{m=1,d,2} + \varphi_{1,FT} * FerryTime_{m=1,d,1} + \varphi_{2,FT} * FerryTime_{m=1,d,2} + \varphi_{1,CA} * CarAv_{m=1,d,1} + \varphi_{2,CA} * CarAv_{m=1,d,2}$$

$$V_{m=2,d} = Size_d + k_{m=2} + \varphi_{1,GTC} * GTC_{m=2,d,1} + \varphi_{2,GTC} * LogGTC_{m=2,d,2} + \varphi_{1,FT} * FerryTime_{m=2,d,1} + \varphi_{2,FT} * FerryTime_{m=2,d,2} + \varphi_{1,CA} * CarAv_{m=2,d,1} + \varphi_{2,CA} * CarAv_{m=2,d,2}$$

$$V_{m=3,d} = Size_d + k_{m=3} + \varphi_{1,GTC} * GTC_{m=3,d,1} + \varphi_{2,GTC} * LogGTC_{m=3,d,2} + \varphi_{1,FT} * FerryTime_{m=3,d,1} + \varphi_{2,FT} * FerryTime_{m=3,d,2}$$

$$V_{m=4,d} = Size_d + k_{m=4} + \varphi_{1,GTC} * GTC_{m=4,d,1} + \varphi_{2,GTC} * LogGTC_{m=4,d,2} + \varphi_{1,AE} * AccEgg_{m=4,d,1} + \varphi_{2,AE} * AccEgg_{m=4,d,2} + \varphi_{1,FT} * FerryTime_{m=4,d,1} + \varphi_{2,FT} * FerryTime_{m=4,d,2} + \varphi_{1,FREQ} * Freq_{m=4,d,1} + \varphi_{2,FREQ} * Freq_{m=4,d,2}$$

$$V_{m=5,d} = Size_d + k_{m=5} + q_{a1} * \varphi_{1,GTC} * GTC_{m=5,d,1} + q_{a2} * \varphi_{2,GTC} * LogGTC_{m=5,d,2}$$
$$+ \varphi_{1,AE} * AccEgg_{m=5,d,1} + \varphi_{2,AE} * AccEgg_{m=5,d,2}$$
$$+ \varphi_{1,HT} * HeadwayTime_{m=5,d,1} + \varphi_{2,HT} * HeadwayTime_{m=5,d,2}$$
$$+ \varphi_{1,TR} * Transfer_{m=5,d,1} + \varphi_{2,TR} * Transfer_{m=5,d,2}$$

Note that the V's for purpose 1, 2, and 3 are identical in form in the use of GTC and LogGTC.

Parameter	**Business**	**Private**	**Holiday**
"$Size_d$"	1	1	1
$k_{m=1}$	-1.2391	1.796338	-1.11107
$k_{m=2}$	-2.90673	0.452256	-1.6186
$k_{m=3}$	-2.34376	1.77229	0.635542
$k_{m=4}$	-2.55259	-0.033	-0.19483
$k_{m=5}$	1.879349	1.651778	2.789086
$CarAv_{m=1,d,1}$	0.3695	0.7383	0.7262
$CarAv_{m=1,d,2}$	0.3695	0.7344	0.8611
$\varphi_{1,GTC}$	-0.002644	-0.007960	-0.003078
$\varphi_{2,GTC}$	-0.8455	-1.7268	-0.6402
q_{a1}	2	2	2
q_{a2}	2	2	2
q_{vot}	2	2	2
$\varphi_{1,FT}$	-0.002296	-0.003315	-0.000331
$\varphi_{2,FT}$	-0.001254	-0.000967	-0.001642
$\varphi_{1,AE}$	-0.005916	-0.003052	-0.001877
$\varphi_{2,AE}$	-0.002694	0	-0.000622
$\varphi_{1,FREQ}$	0.0208	0.0108	0
$\varphi_{2,FREQ}$	0.002092	0.0137	0
$\varphi_{1,HT}$	-0.001997	-0.000778	-0.002440
$\varphi_{2,HT}$	-0.002280	-0.000213	-0.000951
$\varphi_{1,TR}$	-0.001997	-0.000778	-0.002440
$\varphi_{2,TR}$	-0.002280	-0.000213	-0.000951
μ	0.5620	0.3748	0.3414

Table 80: Final parameter-values for the distribution model.

Step 4: Calculating lower-level probabilities

Calculate all of the following variables;

SumM = exp(V1) + exp(V2) + exp(V3) + exp(V4) + exp(V5)

ProbL1 = exp(V1) / SumM

ProbL2 = exp(V2) / SumM

ProbL3 = exp(V3) / SumM

ProbL4 = exp(V4) / SumM

ProbL5 = exp(V5) / SumM

LogsumM = μ * Ln(SumM)

EVd = exp(LogsumM)

Note: μ is the logsum parameter coming from the estimation of the nested logit model (μ differs across purpose).

Step 5: Calculate summation of EVd across destinations

First calculate, in a secondary step, the sum of EVd across all ToZoneID's. Call this variable SumD.

Join this table onto the master table from step 4.

Step 6: Calculate upper-level probabilities

Calculate all of the following variables;

ProbU = EVd / SumD

Step 7: Calculate final probabilities

Prob1 = ProbL1 * ProbU

Prob2 = ProbL2 * ProbU

Prob3 = ProbL3 * ProbU

Prob4 = ProbL4 * ProbU

Prob5 = ProbL5 * ProbU

Step 8: Output final logsum

For each FromZoneID and purpose, calculate the final logsum;

LogSumD = ln(SumD)

Export to a data structure: {FromZoneID, Purpose, LogsumD}

Step 9: Clean distribution model data

After exporting the logsum, retain the following variables from the outcome of Step 7;

{FromZoneID, ToZoneID, PurposeID, Prob1, Prob2, Prob3, Prob4, Prob5}

Step 10: Set up data for the trip frequency model

First join the zone data with the logsum data from step 8. This should yield a table with the following variables;

Variable	Description
ZoneID	The TENCONNECT 1441 zone coding
GDP	GDP data in million Euro

JOB	Job data
POP	Pop data in 1000s
CAP	Bed place capacity
NCARS	Number of cars per household
PH	Number of persons per household
Logsum	Logsum coming from the distribution model

Table 81: Gross variable list for the trip frequency model.

Step 11: Define utility functions for the trip frequency model

The coefficients a, b, g and q may be found in Tables 42. To calibrate the estimated model to observed matrices 9 constants are added. The calibration constants may be found in the excel file calibration6, sheet Coefficients as a 9 by 1441 matrix.

If Purpose = 1

$V_{f=0,i} = -c_{f=10,i} - c_{f=11,i}$
$V_{f=1,i} = a_{f=1} - c_{f=11,i} + b_{f\geq1} * \ln(GDP_i/POP_i) + g_{f=1} * \ln(NCARS_i+1) + q_{f=1} * logsum_i$
$V_{f=2,i} = a_{f=2} + b_{f\geq1} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq2} * logsum_i$
$V_{f=3,i} = a_{f=3} + b_{f\geq1} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq2} * logsum_i$
$V_{f=4,i} = a_{f\geq4} + b_{f\geq1} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq2} * logsum_i$
$V_{f=5,i} = a_{f\geq4} + b_{f\geq1} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq2} * logsum_i$
$V_{f\geq6,i} = a_{f\geq4} + b_{f\geq1} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq2} * logsum_i$

If Purpose = 2

$V_{f=0,i} = -c_{f=20,i} - c_{f=21,i} - c_{f=22,i}$
$V_{f=1,i} = a_{f=1} - c_{f=21,i} - c_{f=22,i} + b_{f=1} * \ln(GDP_i/POP_i) + g_{f=1} * \ln(NCARS_i+1) + q_{f\geq1} * logsum_i$
$V_{f=2,i} = a_{f=2} - c_{f=22,i} + b_{f\geq2} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq1} * logsum_i$
$V_{f=3,i} = a_{f=3} + b_{f\geq2} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq1} * logsum_i$
$V_{f=4,i} = a_{f\geq4} + b_{f\geq2} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq1} * logsum_i$
$V_{f=5,i} = a_{f\geq4} + b_{f\geq2} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq1} * logsum_i$
$V_{f\geq6,i} = a_{f\geq4} + b_{f\geq2} * \ln(GDP_i/POP_i) + g_{f\geq2} * \ln(NCARS_i+1) + q_{f\geq1} * logsum_i$

If purpose = 3

$V_{f=0,i} = -c_{f=30,i} - c_{f=31,i} - c_{f=32,i}$
$V_{f=1,i} = a_{f=1} - c_{f=31,i} - c_{f=32,i} + b_{f=1} * \ln(GDP_i/POP_i) + g_{f=1} * \ln(NCARS_i+1) + q_{f=1} * logsum_i$
$V_{f=2,i} = a_{f=2} - c_{f=32,i} + b_{f=2} * \ln(GDP_i/POP_i) + g_{f=2} * \ln(NCARS_i+1) + q_{f\geq2} * logsum_i$
$V_{f=3,i} = a_{f=3} + b_{f\geq3} * \ln(GDP_i/POP_i) + g_{f\geq3} * \ln(NCARS_i+1) + q_{f\geq2} * logsum_i$
$V_{f\geq4,i} = a_{f\geq4} + c_{f=34,i} + b_{f\geq3} * \ln(GDP_i/POP_i) + g_{f\geq3} * \ln(NCARS_i+1) + q_{f\geq2} * logsum_i$

Parameter	Business	Private	Holiday
$a_{f=1}$	-6,9319	-2,8751	-9,0801
$a_{f=2}$	-7,8490	-3,9204	-11,5499
$a_{f=3}$	-7,8490	-5,0543	-12,8207
$a_{f\geq4}$	-8,6267	-6,2691	-14,0150
$b_{f=1}$		0,0992	0,6324
$b_{f\geq1}$	0,4166		
$b_{f=2}$			0,9416
$b_{f\geq2}$		0,3875	
$b_{f\geq3}$			1,1726
$g_{f=1}$	1,3357	0,6839	1,0551

$g_{f=2}$			1,3587
$g_{f\geq 2}$	1,6875	1,0010	
$g_{f\geq 3}$			1,4682
$q_{f=1}$	0,0870		0,4370
$q_{f\geq 1}$		0,0307	
$q_{f\geq 2}$	0,2049		0,6663

Table 82: Trip frequency model parameters.

Step 12: Calculate trip frequency probabilities and the number of trips

Calculate all of the following variables by purpose;

For business and private

$SumF = \exp(V_{f=0}) + \exp(V_{f=1}) + \exp(V_{f=2}) + \exp(V_{f=3}) + \exp(V_{f=4}) + \exp(V_{f=5}) + \exp(V_{f\geq 6})$

For holiday

$SumF = \exp(V_{f=0}) + \exp(V_{f=1}) + \exp(V_{f=2}) + \exp(V_{f=3}) + \exp(V_{f\geq 4})$

$ProbF0 = \exp(V_{f=0}) / SumF$

$ProbF1 = \exp(V_{f=1}) / SumF$

$ProbF2 = \exp(V_{f=2}) / SumF$

$ProbF3 = \exp(V_{f=3}) / SumF$

$ProbF4 = \exp(V_{f=4}) / SumF$ or (for holiday) $ProbF4^{+} = \exp(V_{f\geq 4}) / SumF$

$ProbF5 = \exp(V_{f=5}) / SumF$

$ProbF6^{+} = \exp(V_{f\geq 6}) / SumF$

For business and private;

$Tp_i = 1*ProbF1 + 2*ProbF2 + 3*ProbF3 + 4*ProbF4 + 5*ProbF5 + 6*ProbF6^{+}$

Trips per person are given by;

$T_i = 4 * Tp_i$

We have to multiply with 4 to get the number of trips per year as the estimation is based on a 3 month period.

For holiday;

$Tp_i = 1*ProbF1 + 2*ProbF2 + 3*ProbF3 + 4*ProbF4^{+}$

Trips per person are given by;

$T_i = Tp_i$

The total number of trips per year for any purpose is then given by;

$Trips_i = T_i * POP_i * 1000$

This output is per year. If it is needed per day it should be divided by 365.

Step 13: Clean trip frequency output

Retain in the above dataset resulting from step 12 the following variables;

{ZoneID, PurposeID, Trips}

Step 14: Final model evaluation and calculation of GA matrices

Join the data from step 13 with the data resulting from step 9. Join the ZoneID onto the FromZoneID.

The data will include the following variables;

{FromZoneID, ToZoneID, PurposeID, Prob1, Prob2, Prob3, Prob4, Prob5, Trips}

Calculate the GA matrices by purpose as;

$T_{ijm=1}$ = Trips * Prob1
$T_{ijm=2}$ = Trips * Prob2
$T_{ijm=3}$ = Trips * Prob3
$T_{ijm=4}$ = Trips * Prob4
$T_{ijm=5}$ = Trips * Prob5

Step 15: Pivot point processing

The GA matrices are compared relative to the base matrix.

6.7.1 Summary and discussion

This document describes the model structure for the long distance passenger model.

The model is concerned with trips above 100 km and has been estimated on the basis of the DATELINE survey.

The model structure is decomposed into two sequentially linked model parts; a lower-level distribution model and an upper level trip generation model. These models are linked with a logsum variable in order to make trip generation a function of accessibility changes.

6.7.1.1 Distribution model

In the distribution model we have applied a standard random utility framework. The model is formed as a discrete two-level nested logit model with destination in the upper nest and mode in the lower nest.

The model applies importance sampling to cope with the destination dimension. The degree of sampling bias on model parameters has been simulated and it is demonstrated that the sampling scheme is sufficient.

A range of model specifications has been tested. The main issue has been to test a linear versus a logarithmic form of the in-vehicle-time variable. It is found that the logarithmic form performs best for business and private trips, whereas for holiday trips, the linear specification is superior.

6.7.1.2 Generation model

The generation model is a standard multinomial logit model where we explain the generation of trips at the household level using GDP per person and car ownership as explanatory variables. The model specification is kept simple to assure that variable may be forecast. Both variables are significant and their specification assures that generation rises with both GDP per person and car ownership.

6.7.1.3 Interpretation of preliminary results

The presented results are preliminary in that LoS variables have not been finalized at the time of estimation.

However, generally speaking, the model structure is encouraging in several respects;

- It is possible to properly identify a wide range of different time components
- The nested logit formulation out-perform the more simple MNL specification and is consistent with the requirement of the logsum parameters
- Analysis of functional has significantly improved the model
- The elasticity-structure is reasonable compared to international findings

Remaining issues include;

- Link the two model structures in order to have accessibility effects in the generation model
- Simulate demand responses with the final nested logit structure

6.8 References

Algers, S., Lindqvist-Dillén, J. and Widlert, S. (1995), 'The National Swedish Value of Time Study', Proceedings of Seminar F, PTRC, 1995.

Axhausen K. et al. (2003), Swiss value of travel time savings, working paper

Daly A. (1982), Estimating choice models containing attraction variables. *Transportation Research*, (1):5-15, 1982.

Ben-Akiva M., Lerman S.R. (1985), Discrete Choice Analysis. MIT Press Cambridge, Massachusetts, 1985.

Fosgerau et al. (2006), 'An integrated approach to the estimation of the value of travel time', European Transport Conference, Strasbourg.

Fox J. (2005), OTM5 Demand Model Estimation: Frequency Modelling. Report for The Danish Transport Research Institute, no. PM-2139-DTF.

Fox J., Sivakumar A. (2005), OTM5 Demand Model Estimation: Mode-Destination Modelling. Report for The Danish Transport Research Institute, no. PM-2140-DTF.

6.9 Appendix – Rail costs

6.9.1 Dummy allocation

The different countries have different dummies attached.

```
AT: c24
AL: c14
BA: c12
BG: c2
BY: c14
BE: c1
CH: c23
CZ: c18
DE: c20
DK: c3
EE: c9
ES: c16
FI: c4
FR: c26
GR: c5
HR: c6
HU: c19
IS: c3
IE: c25
IT: c21
LT: c9
LU: c3
LV: c9
MD: c12
MK: c12
NL: c22
NO: c8
PL: c9
PT: c10
RO: c11
RU: c14
SE: c17
SL: c15
SK: c14
TR: c5
UA: c14
UK: c13
UK: c13
RS: c12
```

6.9.2 Model

Two cost functions have been estimated as a log-linear model. The following functions define the rail cost as a function of length measured in kilometres (Dist) and country specific dummies.

CostHigh = exp(0.95*ln(Dist+1) +
-0.93108 + 0*C1 -0.74418*C2 + 0.06593*C3 - 0.22144*C4 - 0.95589*C5 + 0.03612*C6 + 0*C8 - 0.41980*C9 -0.47972*C10 - 0.55009*C12 + 0.48881*C13 - 1.19080*C14 -

0.19391*C15 - 0.2142*C16 - 0.00703*C17 -0.11524*C18 - 0.34698*C19 + 0.04221*C20 - 0.39612*C21 +
-0.01109*C22 + 0.04080*C23 - 0.10707*C24 + 0.47769*C25 - 0.26382*C26);

CostLow = exp(0.95*ln(Dist+1) +
-0.91620 + 0*C1 -1.15403*C2 - 0.29528*C3 - 0.61373*C4 - 1.37624*C5 - 0.27513*C6
-0.28542*C8 - 0.85736*C9 -0.59806*C10 - 0.99843*C12 + 0.15211*C13 - 1.51442*C14
-0.63591*C15 - 0.47132*C16 - 1.01490*C17 -0.38506*C18 - 0.52098*C19 - 0.31215*C20
-0.64760*C21 -0.22395*C22 - 0.37790*C23 - 0.35616*C24 + 0.12880*C25 - 0.52021*C26);

CostLow = max(2, CostLow) and CostHigh = max(3, CostHigh);

In the estimation, there is a tendency of a marginal falling kilometre cost. The country dummies have been included according to originating country.

The models have been validated due to distribution of residuals, goodness-of-fit, and prediction capabilities. The test for normality is accepted and the R^2 is high. Prior to the final estimation a small set of outliers has been removed. These outliers have impact primarily on the country dummies and to a very little extent on the estimates of α, β, and γ.

For countries not included in the above sample, we have applied the constant from neighbouring countries or for countries with a similar economic profile as seen in the dummy list. The following rules have been applied;

Country (estimation)	**Country 1 (application)**	**Country 2 (application)**	**Country 3 (application)**
Serbia	Moldovia	Macedonia	Bosnia
Poland	Lithuania	Latvia	Estonia
Bulgaria	Belarus	Russia	Albania
Greece	Turkey	Malta	
Scotland	Ireland		
Switzerland	Luxembourg	Liechtenstein	

Table 83: Attachment of country constants for countries not in the sample.

If for a country, a purpose constant has not been identified, whereas this is the case for another purpose, the latter constant will apply for both purposes.

6.10 Appendix – VOT table

CountryCode	PPP	VOTcom	VOToth	VOThol	VOTbus
AL	0.458	5.002	3.252	3.252	6.653
AT	1.038	11.336	7.368	7.368	15.077
BA	0.439	4.798	3.119	3.119	6.382
BE	1.064	11.623	7.555	7.555	15.458
BG	0.350	3.825	2.486	2.486	5.087
BY	0.394	4.299	2.794	2.794	5.717
CH	1.345	14.686	9.546	9.546	19.532
CY	0.888	9.695	6.301	6.301	12.894
CZ	0.570	6.221	4.043	4.043	8.274
DE	1.064	11.623	7.555	7.555	15.459
DK	1.341	14.644	9.519	9.519	19.476
EE	0.577	6.299	4.094	4.094	8.378
ES	0.901	9.842	6.397	6.397	13.090
FI	1.158	12.643	8.218	8.218	16.816
FR	1.116	12.186	7.921	7.921	16.207
GR	0.826	9.017	5.861	5.861	11.993
HR	0.619	6.761	4.395	4.395	8.993
HU	0.604	6.601	4.291	4.291	8.780
IE	1.194	13.042	8.477	8.477	17.346
IS	1.430	15.618	10.152	10.152	20.772
IT	1.036	11.315	7.355	7.355	15.049
LI	1.345	14.686	9.546	9.546	19.532
LT	0.484	5.284	3.434	3.434	7.027
LU	1.095	11.962	7.775	7.775	15.909
LV	0.467	5.102	3.317	3.317	6.786
MD	0.350	3.825	2.486	2.486	5.087
ME	0.433	4.724	3.070	3.070	6.283
MK	0.370	4.046	2.630	2.630	5.381
MT	0.671	7.331	4.765	4.765	9.750
NL	1.079	11.785	7.660	7.660	15.674
NO	1.332	14.548	9.456	9.456	19.349
PL	0.549	5.998	3.899	3.899	7.977
PT	0.850	9.284	6.034	6.034	12.347
RO	0.427	4.659	3.028	3.028	6.197
RS	0.394	4.299	2.794	2.794	5.717
RU	0.427	4.659	3.028	3.028	6.197
SE	1.164	12.717	8.266	8.266	16.913
SI	0.725	7.918	5.147	5.147	10.531
SK	0.530	5.794	3.766	3.766	7.705
TR	0.575	6.278	4.081	4.081	8.350
UA	0.350	3.825	2.486	2.486	5.087
UK	1.098	11.992	7.795	7.795	15.949

Table 84: Country-wise VOT measures in Euro/Hour.

6.10.1 Appendix – DATELINE country-wise tabulation

	Business		Holiday		Private		Commuters	
Country	Dep.	Arr.	Dep.	Arr.	Dep.	Arr.	Dep.	Arr.
Missing	1428	1988	3825	5792	513	535	10,5	1459,5
Albania	0	0	2	2	1	1	0	0
Austria	1172	1112	1828	1993	576	579	94,5	94,5
Belarus	0	0	5	9	0	0	0	0
Belgium	992	1048	2003	1564	597	592	178,5	178,5
Bosnia	4	4	15	32	3	3	0	0
Bulgaria	20	20	30	47	2	3	0	0
Croatia	4	28	138	303	6	13	0	0
Cyprus	0	4	44	81	2	2	0	0
Czech Republic	4	4	4	5	2	3	0	0
Denmark	1048	1088	1816	1716	843	858	105	105
Finland	1948	2032	1591	1456	1398	1412	199,5	178,5
France	3672	3888	13222	13792	3397	3476	693	672
Germany	8992	8700	9294	7430	4022	3959	1260	409,5
Greece	1424	1444	2138	2350	1216	1205	63	63
Hungary	0	0	13	2	0	0	0	0
Iceland	0	4	8	40	0	0	0	0
Ireland	228	264	250	373	167	175	21	0
Italy	2816	2608	3716	4340	985	979	325,5	325,5
Liechtenstein	8	8	2	2	0	0	0	0
Luxembourg	80	84	214	216	61	61	42	0
Macedonia	8	8	0	1	0	0	0	0
Malta	0	0	31	61	0	0	0	0
Netherlands	2440	2012	3382	2149	1415	1363	409,5	283,5
Norway	0	4	2	9	0	0	0	0
Poland	0	0	2	3	0	0	0	0
Portugal	2568	2560	3322	3478	2523	2528	472,5	451,5
Romania	12	44	35	62	0	1	0	0
Russia	24	56	32	73	17	21	0	10,5
Slovenia	12	4	1	0	2	2	0	0
Spain	6388	6428	19651	20072	5553	5549	231	283,5
Sweden	2220	2060	1588	1280	1054	1032	231	105
Switzerland	652	716	991	1118	289	297	136,5	105
United Kingdom	5252	5196	4109	3446	2573	2564	409,5	157,5
Yugoslavia	4	4	22	29	4	8	0	0

Table 85: Dateline stat - outgoing trips divided by purpose and scaled to annual total.

7 Traffic forecast – The TEN-Connect trade prediction model (TPM)

7.1 Introduction

The model building in the TEN-Connect study deals with both, passenger travel and freight. Therefore both, passenger flows and freight flows have to be predicted for the study years 2020 and 2030. This section describes the methodology used to make the prediction of freight flows.

Freight flow predictions start from the production-consumption (PC) flow matrix, provided by the ETIS-BASE dataset for the reference year 2005. It shows flows in tons, differentiated according to 11 NSTR commodity groups, for each PC pair of regions within the study area as well as for each pairing of regions in the study area (Europe, North Africa, Middle East, and Caucasus) with a country or group of countries in the rest of the world.

We apply a top-down procedure, starting from global trends in economic development (economic growth, decline of national barriers to trade, and changing composition of trade flows), going on with predicting trade within and between countries, stepping down further to flows between regions, and then finally updating the detailed ETIS-BASE tables such that they perfectly fit with the predicted flows within and between countries. The steps are roughly described in Figure 23 .

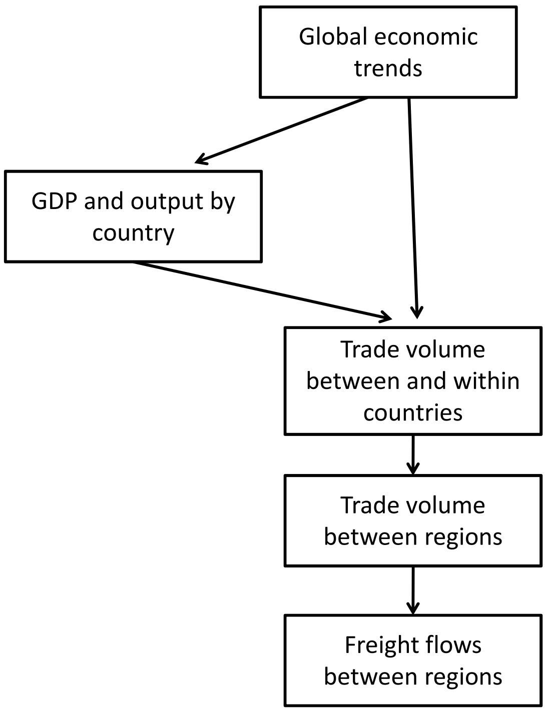

Figure 23: Freight flow prediction

7.2 The gravity formulation

As mentioned above, we predict freight flows in a top-down approach, starting from predicting international trade volumes on country basis, including within-country trade. Trade volumes are monetary trade figures in constant (year 2005) prices.

For the core of the trade prediction model, we rely on the *modern gravity model*, leading to the so-called *doubly constrained approach*:

$$\begin{cases} T_{ijt} = A_{it} B_{jt} T_{ij0} f_{ijt} \\ \sum_j T_{ijt} = S_{it} \ \forall i, t \\ \sum_i T_{ijt} = D_{jt} \ \forall j, t \end{cases}$$

where T_{ijt} is trade volume from country i to country j at time t; $t = 0$ is the reference year 2005. It is well known that even the simplest gravity equation has a high explanatory power, such that using this up-to-date formulation as a predictive tool seems attractive.

A_{it} and B_{jt} are fixed effects covering all effects specific to the export and import country, respectively, for each year. Resources, productivities, and prices are all covered by these factors. They are obtained from the two sets of constraints on the trade flows, where S_{it} is total supply (not just value added) and D_{jt} is total demand. Both cover trade within as well as between regions. These are predicted by updating with the help of the externally prescribed GDPs.

f_{ijt} represents the trade impediment changes from 0 until $t > 0$, specified as:

$$f_{ijt} = \begin{cases} 1 & if \quad i = j \\ \exp(g_{ij0} - g_{ijt}) & else. \end{cases}$$

g_{ijt} is a *country-pair specific border barrier*, which is a function of trade resistance variables. The most important trade resistance variable is geographic distance. Others are dummies for joint membership/non-membership in trade associations, for common language, for a common border, variables of cultural dissimilarity, and the like. Predictions of g_{ijt} are obtained by extrapolating the trends of the respective estimates in the trade panel 1993 to 2004, taking account that barriers of country-pairs having already low barrier levels will approach a lower plafond. We will explain the underlying procedure in more detail in the next subsection.

Note that our gravity formulation implies that a uniform rate of growth of output in all regions and constant trade impediments lead to a uniform rate of growth of trade, equal to the uniform rate of output growth, as it should be. If we observe trade to grow faster than GDP, this is primarily due to a decline of g_{ijt}. Another reason that, according to this model, global trade may grow faster than global GDP is faster growth of smaller (in terms of supply and demand) countries. For the theory behind this model, see e.g. Anderson and van Wincoop (2003).

7.3 The prediction of international trade impediments

In order to come up with the values of trade impediment change factor for the prediction years (starting from 2005), we first estimate the values of underlying parameters for years 1993-2004 using the gravity regression, and later apply an extrapolation procedure.

Our regression analysis starts from the well-known gravity formulation for trade flow from country i to country j at time t:

$$T_{ijt} = \tilde{A}_{it}\tilde{B}_{jt}\tilde{f}_{ijt}$$

A common method is to express $\tilde{f}_{ijt}$ as a function of trade resistance variables: distance, membership in FTAs, other measures of geographical and cultural proximity between the trading partners. We apply the following formulation:

$$\tilde{f}_{ijt} = \exp\left[-\lambda_{jt} \cdot C_j + \phi_t \cdot CB_{ij} + \varphi_t \cdot LA_{ij} + \sum_k \mu_t^k \cdot FTA_{ij}^k - \gamma_t \cdot d(Dist_{ijt})\right].$$

Here, parameter λ_{jt}, premultiplying the importer-country dummy C_j, delivers the size of the *border barrier* of an importing country. The border barrier summarises all factors impeding trade with other countries, but not trade within the country (C_j=0 if $i = j$). Recent research brought about ample evidence that these barriers are high. This was for the first time demonstrated by Bröcker (1984), using within country and between country transport flows for the EU6. The most often cited reference nowadays is McCullum (1995). Helliwell (1998) and Helliwell and Schembri (2005) review lots of literature, showing that recent estimates of border barriers turn out to be somewhat lower than McCullum's estimates, but similar to Bröcker's estimates.

CB_{ij} and LA_{ij} are the dummies for, respectively, existence of a common border (CB_{ij}=1), and the close affinity of languages between the trading partners (LA_{ij}=1). Parameters ϕ_t and φ_t thus show the effects of these proximity indicators on trade. By our definition, $CB = LA = 1$ for country-internal flows.

Parameters μ_t^k give the size of the trade-facilitating effect of the *k* free trade agreements (specifically, we take account of the different stages of EU enlargement and the NAFTA agreement). FTA_{ijt}^k is a dummy indicating whether the respective preferential effect applies (FTA_{ijt}^k=1) or not.

Finally, $d(Dist_{ijt})$ is the representation of the distance function. An underlying assumption is that this term should increase with distance, but at a diminishing rate. Bröcker (1998) suggests the Box-Cox form of this function:

$$d(Dist_{ij}) = \frac{(Dist_{ij})^{\omega} - 1}{\omega},$$

where $0 < \omega < 1$ is a curvature parameter. This parameter can be estimated together with the rest of the trade resistance parameters using the maximum-likelihood techniques. From our estimations, however, it followed that ω is quite close to zero, suggesting that values of d would be very close to $\ln(Dist_{ij})$. That is why we eventually used this log-representation in the regression analysis.

Distances are population weighted great-circle distances. For population weighting, we use the CIESIN (2005) database, offering population of the world, geographically assigned to a very fine grid (2.5 x 2.5 minutes) of the land map of the world. Note, that this method implies that the distances are time varying.

We estimate the following equation for trade between 187 countries of the world for each year between 1993-2004:

$$T_{ijt} = \exp\left[a_{it} \cdot P_i + b_{jt} \cdot C_j + \phi_t \cdot CB_{ij} + \varphi_t \cdot LA_{ij} + \sum_k \mu_t^k \cdot FTA_{ij}^k - \gamma_t \cdot \ln(Dist_{ijt}) \right] + v_{ijt},$$

for $i \neq j$,

where $a_{it} = \ln(\tilde{A}_{it})$, $b_{jt} = \ln(\tilde{B}_{jt}) - \lambda_{jt}$, P_i is the exporter-country dummy, and v_{ijt} is an error term. We apply the method of Bröcker (1985) (see also Bröcker and Rohweder, 1990). The approach has been reinvented by Silva and Tenreyro (2006) under the name “Poisson pseudo-maximum-likelihood” (PPML). The original source of trade data is the UN COMTRADE database.

This produces the estimates:

$$\hat{T}_{ijt} = \exp\left[\hat{a}_{it} \cdot P_i + \hat{b}_{jt} \cdot C_j + \hat{\phi}_t \cdot CB_{ij} + \hat{\varphi}_t \cdot LA_{ij} + \sum_k \hat{\mu}_t^k \cdot FTA_{ij}^k - \hat{\gamma}_t \cdot \ln(Dist_{ijt}) \right].$$

The purpose of this estimation is to calculate the time series of border barriers. Note, that for the country-internal trade, the expression for T_{ijt} becomes:

$$T_{iit} = \exp\left[(a_{it} + b_{it} + \lambda_{it}) \cdot P_i + \phi_t \cdot CB_{ii} + \varphi_t \cdot LA_{ii} - \gamma_t \cdot \ln(Dist_{iit}) \right]$$

Having the results of the estimation above, the only unknown parameter in this equation of internal trade is the border barrier λ_{it}. Thus, we simply calculate its values for every year and country as:

$$\hat{\lambda}_{it} = \ln(T_{iit}) - (\hat{a}_{it} + \hat{b}_{it}) \cdot P_i - \hat{\phi}_t \cdot CB_{ii} - \hat{\varphi}_t \cdot LA_{ii} + \hat{\gamma}_t \cdot \ln(Dist_{iit}).$$

Since we are interested in the development of the estimated border barriers in the future, we look at their time trend in order to get an idea of their future values. Thus, we estimate a simple linear trend regression:

$$\hat{\lambda}_{it} = c_i + tr_i \cdot time + \varepsilon_{it}, \text{ where}$$

$time$ is the time period indicator, with 1993 entering with value 1 and 2004 with value 12. The resulting trend coefficients are very heterogeneous across countries, with many extreme values, which is of course to be expected, given the small number of observations we have. Our idea is to use these coefficients to extrapolate the values of $\hat{\lambda}_{it}$ into the future. For this we need reliable values of the time trends. To rule out the outliers, we perform the aggregation of the country-specific trends into trends for country groups. We group together countries with similar institutional, historical, and geographical properties. For each of the resulting 15 groups (for example, EU15, NMS12, FSU, Balkan, Central America, South America etc), we calculate a weighted average value of the trend coefficient, with weights given by the GDP shares of the respective countries in 2005:

$$tr_g = \sum_{i \in g} \hat{tr}_i \frac{GDP_i}{GDP_g}, \text{ for any group } g.$$

As expected, all of these aggregate trend coefficients are negative and correspond to the average decline rate of the border barrier $\hat{\lambda}_{it}$ equal to 3% per annum in year 2004.

Now we come to the forecasting step. From our estimation it follows that the parameters $\hat{\phi}_t$, $\hat{\varphi}_t$ and $\hat{\gamma}_t$ are quite stable over time. We therefore assume that the time-varying trade impediment consists of only two components: the country-specific border barrier and the FTA effects. Thus, we define g_{ij0}, the *country-pair specific border barrier* in the benchmark year 2004 (our time 0 for the impediment-forecasting part), to be given by:

$$g_{ij0} = \hat{\lambda}_{j0} - \sum_k \hat{\mu}_t^k \cdot FTA_{ij}^k$$

Correspondingly, the trade impediment changes from $t = 0$ until $t > 0$, are specified as:

$$f_{ijt} = \begin{cases} 1 & if \quad i = j \\ \exp(g_{ij0} - g_{ijt}) & else. \end{cases}$$

How does g_{ijt} develop over time? We assume the trade impediments observed in the benchmark year 2004 to "melt" over time down to a lower bound. In other words, *the barrier curve* (the graph of g_{ijt} against time) smoothly approaches the lower bound in the course of time. One possibility is to assume that the trade impediment goes to zero by the year 2030. We, however, take a more conservative approach and assume that the lower bound for g_{ijt} is given by the lowest g_{ij0} level in the starting period (in the year 2004). The rate of decline is chosen such that the curve is smooth in the sense that, at reference year 2004, its slope is like it was in the historical trend.

Depending on the country pair, the curve starts at different levels. This level depends first on the importing country, having each their respective specific import barrier $\hat{\lambda}_{j0}$ at reference year 0. Second, for a given importing country, it may vary with respect to trade partner, in case of FTA membership. If e.g. we take French imports, then $\hat{\lambda}_{j0}$ applies to trade with non-FTA partners (USA, say), while the starting point of the barrier curve is $\hat{\lambda}_{i0} - \mu_0^{EU1991}$ for French imports from Germany.

Formally, the formula we use is as follows:

$$g_{ijt} = bound + (g_{ij0} - bound) * \exp(\beta_{ij} * t),$$

with $\beta_{ij} = trend_g / (g_{ij0} - bound)$ for all $i \in g$.

As a *bound,* we take the lowest value of g_{ji0}, which in our case corresponds to the flows from the new member states (NMS12) to the Netherlands.

7.4 Predicting freight flows between regions

The next step is to use international trade as well as GDP predictions to predict freight flows between regions in the study area. As freight is in tonnes, while trade is in values, we first need a tonnes-to-value conversion for the reference year 2005. This is done in two steps. First, a preliminary value estimate is obtained by applying the *value/weight ratio* from the UN COMTRADE database to international flows, and by applying an estimate for the within-country value/weight ratio to within-country flows. The latter is estimated as a sum of country of origin and country of destination fixed effects in a regression explaining value/weight ratios. The value/weight ratio for commodity ℓ, origin r and destination s obtained this way is called $p_{\ell rs}$. Let $x_{\ell rs}$ denote the corresponding freight flow in tonnes. The preliminary value estimate $x_{\ell rs}p_{\ell rs}$ is then further corrected as follows to adjust it to the known totals of regional production values, regional values of intermediate and final demand, and values of international trade:

$$\begin{aligned} v^o_{\ell rs} &= x^o_{\ell rs} p^o_{\ell rs} a^o_r b^o_s c^o_{ij} \\ &= x^o_{\ell rs} \pi^o_{\ell rs}, \quad r \in \{i\}, s \in \{j\}. \end{aligned}$$

$v^o_{\ell rs}$ is the estimated trade value, $\pi^o_{\ell rs}$ is the factor converting tonnes to values (i.e. the corrected value/weight ratio). Superscript o indicates the reference year. The multipliers a^o_r, b^o_s and c^o_{ij} are obtained from the restrictions

$$\sum_\ell \sum_r v^o_{\ell rs} = D_{0s} \ \forall r,$$

$$\sum_\ell \sum_s v^o_{\ell rs} = S_{0r} \ \forall r,$$

$$\sum_\ell \sum_{r \in \{i\}} \sum_{s \in \{j\}} v^o_{\ell rs} = T_{0ij} \ \forall i, j.$$

$\{i\}$ and $\{j\}$ denote the sets of regions belonging to countries i and j, respectively.

Putting it differently, $v^o_{\ell rs}$ is the minimum information estimate under the restrictions (3) to (5), given the prior $x_{\ell rs}p_{\ell rs}$. It makes best use of the information supplied by the constraints that macro values have to add up to the known totals of sales und purchases in each region and of international trade.

Finally, freight flows are predicted for the future years according to the following formula. It combines the elasticity approach with the adjustment to macro totals for the future year:

$$x^t_{\ell rs} = x^o_{\ell rs} \left(\frac{y^t_r}{y^o_r} \right)^{\eta_\ell} \left(\frac{y^t_s}{y^o_s} \right)^{\zeta_\ell} h_r g_s d_{ij}.$$

There are five updating factors in the formula:

1. $\left(y^t_r / y^o_r \right)^{\eta_\ell}$ represents *the impact of GDP on the commodity structure of the ori-*

gin.

2. $\left(y_s^t / y_s^o\right)^{\zeta_\ell}$ represents *the impact of GDP on the commodity structure of the destination*.
3. h_r represents all determinants related to the origin, and which are not commodity specific.
4. g_s represents all determinants related to the destination, and which are not commodity specific.
5. d_{ij} represents all determinants related to the respective country pair, and which are not commodity specific.

The multipliers h_r, g_s and d_{ij} are obtained by adjusting predicted values $v_{\ell rs}^t = x_{\ell rs}^t \pi_{\ell rs}^t$ to the predicted macro totals, as in (3) to (5),

$$\sum_\ell \sum_r v_{\ell rs}^t = D_{ts} \ \forall r,$$

$$\sum_\ell \sum_s v_{\ell rs}^t = S_{tr} \ \forall r,$$

$$\sum_\ell \sum_{r \in \{i\}} \sum_{s \in \{j\}} v_{\ell rs}^t = T_{tij} \ \forall r, s.$$

The commodity specific elasticities η_ℓ and ζ_ℓ are estimated in a regression explaining the time series of international trade flows in tonnes by NSTR, taken from Eurostat COMEXT database. The regression uses fixed effects for controlling all effects except the impact of GDP on the commodity composition of trade. η and ζ are identified only up to an arbitrary constant in that regression. Note however, that any constant added to them does not change the prediction, because it would be exactly compensated by a corresponding change of h_r and g_s. The conversion factor $\pi_{\ell rs}^t$ is obtained by updating $\pi_{\ell rs}^o$ using commodity specific trends in the value/weight ratio.

7.5 Updating the ETIS-BASE

The original data in the ETIS-BASE contains the freight flows in tons between the regions in the study area that are split not only by NSTR commodity group, but also by the route taken (with the possibility of mode change on the way). The trade prediction model uses the same updating factor to forecast the future flows on all of the routes connecting a given production-consumption pair. The output of the trade prediction model is then passed on to the Trans-Tools modal split model, that reallocates the flows between routes and modes based on the transport capacity and minimal transport cost considerations.

Not all flows inside the study area of the TEN-Connect project are covered by the version of the ETIS-BASE we had access to by the end of September, 2008. The important gaps include the internal trade in a wide range of countries: Cyprus, Malta, Albania, Bosnia, Belarus, Switzerland, Liechtenstein, FYR Macedonia, Croatia, Iceland, Moldova, Russia, Turkey, Ukraine, Serbia. Most importantly, the commodity-specific flows between regions in Switzerland, Belarus, Croatia, Russia, Turkey, Ukraine, and Serbia are not available. That is why we do not provide a commodity-specific forecast of these flows. However, the effects of infrastructure projects on the regional GDP can still be calculated using the techniques of trade impact model (to be presented next).

7.6 References

Anderson, J. E. van Wincoop, E. (2003), “Gravity with Gravitas: A Solution to the Border Puzzle”, *American Economic Review, 93(1), 170–192*.

Bröcker J. Rohweder, H. (1990), “Barriers to International Trade: Methods of Measurement and Empirical Evidence”, *Annals of Regional Science, 4, 289-305*.

Bröcker, J. (1984) “How do International Trade Barriers Affect Interregional Trade?”, *Regional and Industrial Development Theories: Models and Empirical Evidence*, *219-239*

Bröcker, J. (1998/2000), “How would an EU-membership of the Visegrád countries affect Europe’s economic geography?” *Annals of Regional Science, 32 (1998), 91-114; Erratum 34 (2000), 469-471*.

Bröcker, J. (2002), “Spatial effects of European transport policy: a CGE approach”, *Trade, Networks and Hierarchies*, edited by G.J.D. Hewings et al. New York, Springer.

Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World Version 3 (GPWv3). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Internet: http://sedac.ciesin.columbia.edu/gpw.

Helliwell, J. (1998), “How Much Do National Borders Matter?”, *Brookings Institution Press*, Washington, D.C

Helliwell, J. F. Schembri, L. L. (2006), “Borders, Common Currencies, Trade, and Welfare: What Can We Learn from the Evidence?”, *Bank of Canada Review*, Internet: http://www.bankofcanada.ca/en/review/rev_spring2005.html.

McCallum, J. (1995), “National Borders Matter: Canada-U.S. Regional Trade Patterns.”, *American Economic Review 85, 615–23*.

Silva, J. M. Santo, C. Tenreyro, S. (2006), “The Log of Gravity”, *The Review of Economics and Statistics, 88, 641–658*.

8 Traffic forecast – The TEN-Connect trade impact model (TIM)

8.1 Introduction

The trade impact model is essentially the *CGEurope model* (Bröcker, 1998/2000, 2002) frequently applied to transport policy evaluation before, with an important modification. In the model version used so far, interregional trade was estimated as a by-product of the model calibration, assuming CES composition functions to be symmetrical and identical for all users. The strength of the approach was its applicability even without explicit interregional trade information. In the current project we have to keep consistency with the trade estimates of the TPM as far as possible. We therefore prefer to introduce so called Armington preferences in trade allowing calibrating the model such that it exactly reproduces the trade values of the TPM described before. This makes the model theoretically less elegant, but brings it closer to the data.

We briefly repeat the description of the basic structure of CGEurope. It is a static general equilibrium model for a closed system of regions covering the whole world. The regions are the 1441 regions on the NUTS3 level of the TEN-Connect project and 19 external zones. In each region reside identical immobile households owning the regional stock of production factors that are immobile as well. Their incomes stem from regional factor returns as well as from an interregional income transfer that can have a positive or negative sign. Income transfers are exogenous (in real terms) and add up to zero for the entire world. A further income source is introduced if we simulate charges to be paid by users of infrastructure. Revenues from these charges are redistributed to households. The redistribution shares of regions are imported from the transport model.

Households spend their income for buying goods and services partly produced in their own regions and partly produced in other regions. Households' demand represents total final demand, which means, private as well as public consumption, and investment. There is no separate public sector in the model; that is households have to be regarded as an aggregate of private and public households, their budget constraint is the consolidated budget constraint of private and public households in the region.

Households are price takers on all markets. They maximize a Cobb-Douglas utility depending on the quantity of local goods and the quantity of an index of diversified tradable goods. Hence, they spend fixed shares ε and $1-\varepsilon$ of their income for local and tradable goods, respectively. Utility changes of households, measured in monetary terms by Hicks' equivalent variation concept, are our measure of regional welfare effects of any change in the level of services brought about by transport policy.

The production sector is represented by identical immobile firms. There are two types of firms: 1) firms producing local goods and 2) firms producing tradable product varieties. There is no further sectoral differentiation. Local goods are produced under constant returns to scale and, as the name says, can only be used within the region itself. Tradable goods, however, are produced by a "Dixit-Stiglitz-Industry", which is now standard in empirical trade modelling. This assumption implies economies of scales in production of tradables.

For the sake of simplicity local as well as tradable goods are assumed to be produced by a Cobb-Douglas-technology with cost shares α, β and γ for primary factors, local goods, and tradable goods that are used as inputs, respectively. Primary factors are modelled as a single homogeneous factor. One may also regard them as a composite of an arbitrary number of factors combined by a linear homogeneous technology. As we do not distinguish between sectors having different factor intensities, this would be formally equivalent.

Analogous to household consumption, firms use tradable goods as a composite index that is composed of all variants produced anywhere in the world. The same index is used for final demand as for intermediate inputs: as usual, varieties are composed by a CES index with an elasticity of substitution controlling how sensitive demand responds to changes of relative prices for goods from different origins.

The decisive assumption for transport policy evaluation is that there are transaction costs for goods delivered from one region to another. Transport policy is evaluated by comparing two equilibria for the same simulation year (2020 or 2030, say), one representing a benchmark without the policy instruments under study in place, another representing an alternative with the policy implemented. All costs of spatial interaction in the reference situation are incorporated in the Armington preference parameters, while the cost changes characterizing the transition from the reference to the alternative are explicitly taken into account in the simulation.

The explained assumptions imply the equilibrium to consist of the following system of equations. Note that the respective equations for different regions are interdependent, such that they can be solved only simultaneously for all the 296 regions of the system.

Production P_r must equal use for non-tradables input βP_r, plus non-tradables final demand εE_r, plus transport service T_r, plus supply of tradables S_r, all in values (E_r is income, which equals expenditure),

$$P_r = \beta P_r + \varepsilon E_r + T_r + S_r.$$

Income is factor income $w_r F_r$ plus exogenous received transfer G_r, plus redistributed charges C_r,

$$E_r = w_r F_r + G_r + C_r,$$

with factor price w_r and factor stock F_r. The price of non-tradables output is the minimal unit cost, which by the Cobb-Douglas assumption yields

$$p_r = w_r^{\alpha} p_r^{\beta} q_r^{\gamma}.$$

q_r denotes the price of one unit of the CES composite used for production input consumption. It fulfills the equation

$$q_r^{1-\sigma} = \sum_s \theta_{sr} p_{sr}^{1-\sigma}.$$

σ is the Armington elasticity of substitution. p_{sr} is the inclusive price of goods from s to be paid in destination r, which is composite price of varieties produced at origin s, $\tilde{p}_s$, plus transport cost per unit τ_{sr} plus charges per unit c_{sr},

$$p_{sr} = \tilde{p}_s + \tau_{sr} + c_{sr}.$$

It can be shown that under the Dixit-Stiglitz assumption $\tilde{p}_s$ is

$$\tilde{p}_s = (\varphi_s p_s^{\rho} / S_s)^{\frac{1}{\rho-1}}.$$

$\rho > 1$ is the elasticity of substitution between brands, φ_s is a scaling parameter fixing units of measurement for the composite. Note that this just introduces a scale effect for tradables supply, which is the smaller, the bigger is ρ. It vanishes for ρ tending to infinity. One has to restrict this elasticity to $\rho \geq \sigma$ to preserve stability of the solution.

The equilibrium on the factor market is,

$$w_r F_r = \alpha P_r.$$

On the market for tradables it is

$$S_r = \tilde{p}_r \sum_s t_{rs}.$$

t_{rs} is the trade flow in real terms, which by the CES assumption is

$$t_{rs} = \theta_{rs} \frac{p_{rs}^{-\sigma}}{q_s^{1-\sigma}} D_s.$$

D_r, the demand for tradables (in values, here for region r) is production input of tradables βP_r plus final demand for tradables εE_r,

$$D_s = \gamma P_r + (1-\varepsilon) E_r.$$

Finally, the value of transport services is

$$T_r = \sum_s t_{sr} \tau_{sr},$$

and total charge revenue equals total redistribution,

$$\sum_r \sum_s c_{rs} t_{rs} = \sum_r C_r.$$

The model is parameterized as follows:

- Cost shares α, β and γ are taken from national accounts. We allow them to differ between countries, but not between regions, as there is no information on that.
- The parameters θ_{rs} are trade value shares of the reference situation. They are obtained from the values that are predicted by the TPM, aggregated over commodities.
- As already said, the parameter φ_s just fixes units and is thus arbitrary. We choose it to be equal to the reference value of tradables supply. This implies $\tilde{p}_s = 1$ for $p_s = 1$.
- The factor stock F_r is calibrated such that the benchmark equilibrium GDP equals the benchmark data. Choosing $w_r = 1$ for the benchmark implies that F_r

is just the benchmark GDP for the region.

- Finally, elasticities cannot be calibrated. We have to rely on past experience and econometric estimates taken from our own and others' research.

The calibration just described implies that all reference prices equal unity.

8.2 Data inputs and outputs

The external **data inputs for the TEN-CONNECT trade model** for each scenario (referred to as ScenarieX below) are the following:

1. P/C-pair based generalized freight costs by NSTR commodity group, aggregated over modes, with separated payments to the public budget, in euros per ton.
2. Corresponding freight flows in tons per year.

   This information (1 and 2) is to be found in the file N3FreightGeneralisedCosts.csv located in the folder TransToolsData\Matrices\ScenarieX.
3. G/A-pair based generalized passenger cost by purpose (business, private), with separated payments to the public budget, in euros per trip.
4. Corresponding passenger flows in trips per year.

   This information (3 and 4) is to be found in the file N3PassengerGeneralisedCosts.csv located in the folder TransToolsData\Matrices\ScenarieX.
5. Scenario year GDP by NUTS3 region, in mln. euros

   This information is to be found in the file N3GdpFutureYear.csv located in the folder TransToolsData\Zones\ScenarieX.
6. Scenario description containing the year that GDP figures refer to (needed to forecast international impediments and value-to-weight ratios).

This information is to be found in the file Scenario.csv located in the folder TransToolsData\Scenario\ScenarieX.

The **output of the TEN-CONNECT trade model** is the matrix of predicted trade flows between P/C NUTS2 pairs, by NSTR commodity group, in tons.

This information will be placed in the file N2FreightBeforeMSTripMatrix.csv located in the folder TransToolsData\Matrices\ScenarieX.

The **additional data input for the CGEurope model** is the revenue from pricing schemes, tolls and taxes, collected by NUTS3 region, in mln. euros.

This information is to be found in the file N3Revenue.csv located in the folder TransToolsData\Matrices\ScenarieX.

The **output of the CGEurope model** is the updated GDP by NUTS3 region in mln. 2005 euros, and in % changes to the reference situation.

This information will be placed in the file N3GdpFromCGEurope.csv located in the folder TransToolsData\Zones\ScenarieX.

The **output of the CGEurope model** is also the updated matrix of predicted trade flows between P/C NUTS2 pairs, by NSTR commodity group, in tons.

This information will be placed in the file N2FreightAfterShock.csv located in the folder TransToolsData\Matrices\ScenarieX.

8.3 References

Bröcker, J. (1998/2000), “How would an EU-membership of the Visegrád countries affect Europe’s economic geography?” *Annals of Regional Science, 32 (1998), 91-114; Erratum 34 (2000), 469-471*.

Bröcker, J. (2002), “Spatial effects of European transport policy: a CGE approach”, *Trade, Networks and Hierarchies*, edited by G.J.D. Hewings et al. New York, Springer.