SKEMA

Grant Agreement No. TREN/FP7/TR/218565/"SKEMA"

D2.1.1.2 Trade and Transport Flows

Project Start Date:	15th June 2008		
Project End Date:	14th June 2011		
Co-ordinator:	Prof. Yannis Katsoulakos, Athens University of Economics & Business		
Deliverable No:	D2.1.1.2	**WP Leader:**	Antti Permala, VTT
WP No:	2	**Dissemination Level:**	Public
Submission date:	12th May'09		

A Project supported by the European Commission.

Directorate-General for Energy and Transport.

Document summary information

Role	Persons	Organisation	Project Role
Authors	James Kehoe	Nautical Enterprise Centre Ltd	Participant
Contributors	Gerry Trant	Nautical Enterprise Centre Ltd	Participant
Contributors	Y Koliousis	AUEB	Participant
Reviewer	Linda Ramstedt	UGOT	Participant

Revision history

Rev.	Who	Date	Comment
1	Linda Ramstedt	11th May'09	Accepted
2	T Katsoulakos	24 Aug09	Integration of NECL NEAC study with AUEB study on trade and transport models

Quality Control

	Who	Date
Checked by WP Leader	Antti Permala	
Reviewed by	Linda Ramstedt	11th May'09
Approved by Coordinator	Dr. Takis Katsoulakis	24th Aug 09

Contents

1 Objectives

Forecasts on transport flows are a major input and critical element for all planning and investment actions. This study will review and elaborate on the transport flow models used by public and private stakeholders in conjunction with various databases to assist them in their decision making processes.

Specific objectives include:

1. Provide description of general principles in forecasting interregional and international flow of goods and an overview of related EU projects
2. Review modal split models with specific reference to 'NEAC modal split'

It is anticipated that there will be a substantial increase in the amount of information available on trade-flows as the adoption of e-Maritime techniques and technology increases. This will bring significant developments in the field of transport modelling, including mode choice modelling.

1.1 Target Stakeholders

- Transport operators and freight forwarders
- Maritime industry business analysts
- Maritime & Logistics research institutes
- Policy generation and evaluation bodies

1.2 Approach

The approach adopted in respect to the study objectives is as follows:

- *Trade models.* The diversity of trade models and databases create difficulties for users to estimate trade flows in order to design efficiently transport services. Among the issues of concern, reliability, consistency, timeliness and flexibility in terms of data/updates are the main ones. Key areas to be addressed:
    - a) Principles behind the various economic models that are currently in use.
    - b) the Development of a European Transport policy Information System (ETIS) as a basis for transport planning and policy formulation
    - c) survey of projects
- *NEAC modal-split model* The approach of this study was to initially review existing literature on the NEAC modal-split model. The review then moved to explore aspects of the NEAC model in more detail within the general field of transport modelling. Based on an understanding of the NEAC model from the previous work the various aspects of the approach/techniques used within the NEAC model were explored further to determine there strengths and weakness. Alternative approaches were also investigated to frame the choices made by the NEAC modellers. To aid understanding no prior knowledge of mathematical modelling techniques or transport simulation was assumed in the composition of this study.

2 Trade models

2.1 General principles

Trade and transport models. The diversity of trade models is associated with economic, geographic, network deployment aspects (Bergstrand (1985), Deardorff (1995), Egger (2000), Head, (2003) Koo et al (2006).

Currently there are a number of databases that are used for producing forecasts regarding interregional and international flow of goods. The common practice is usually the following: most end users attempt to link the outputs from traditional economic measures and indicators (e.g. Gross Domestic Product, Trade Balance) with transport economics and cost data, in order to come up with a prediction as to what the expected volume of goods will be. However, there are a number of inadequacies in deriving results through the use of these databases.

Some of the key problems are:

a) Use ambiguous economic models to estimate flows as well as forecasts
b) Reporting organizations are reluctant to provide accurate figures
c) Organizations provide data / figures that may not be harmonized. Particularly, each port uses its internal methods in calculating specific inputs and outputs.
d) Use outdated measurement units (e.g. volume of goods moved for the unaccompanied trailers in the port industry)
e) Use trivial methods to extrapolate the forecasts, i.e. simple moving averages without considering relevant / extraordinary events (e.g. the Olympic Games in 2004 in Greece, strikes in European Ports, etc.).

As witnessed by members of this research group, in almost all engagements with organizations (airport and port authorities, railway operators, etc) the officials refer to the "data-gap" and almost always suggest using their internal figures. Especially for the maritime sector, there is a need at the moment for disambiguation of the forecasts, since a number of stakeholders (ranging from vessel operators to port authorities to logistics companies) are evaluating options to transfer freight flows from traditional means (e.g. road, etc) to more sustainable means (intermodal transport, short sea shipping) .

2.2 The ETIS Database

The European Commission has supported in the Fifth Framework Program the Development of a European Transport policy Information System (ETIS) as a basis for transport planning and policy formulation. It was envisaged that once the ETIS-BASE was set up it could be easily used by the various stakeholders to support them in their planning and policy making decisions at the EU level. However, practice has shown that the original goals of the ETIS Project may not have been fully achieved, since a number of deficiencies and hindrances restrict its practical value.

ETIS related projects include BASE, AGENT and LINK.

Difficulties in using the ETIS database include:

- *usability,* it is not easy is for users to retrieve specific sets of data;
- *integration* with already established databases (e.g. EuroSTAT, ESPO, etc);

2.3 EUNET – SASI

The EUNET/SASI project, (June 1996 -June 1999), developed methods to assess the wider socioeconomic and spatial impacts of transport initiatives, and to demonstrate that these methods can be made fully operational. , EUNET took a regional/corridor view and focused on the demonstration of methodology, while SASI took a more global view of impacts across Europe

The EUNET modelling system developed in the form of prototype software to support policy decisions included:

- methods for valuing socio-economic effects;
- a tool for estimating spatially-resolved indicators of regional accessibility and social cohesion, including effects on national output and employment;
- a new approach to regional economic modelling, where the sources of potential travel growth are identified separately as a function of specific social and economic activities;
- a database for estimating the costs of vehicle and infrastructure operation, by country, through to 2020;
- an assessment framework combining both cost-benefit and multi-criteria analysis methods.
- The model was demonstrated for the Trans-Pennine corridor in the UK and for long-distance freight movements in Finland.

SASI divided the Member States into 201 regions), and provides forecasts to the year 2016. Innovative attributes included:

- the prediction of regional unemployment;
- the estimation of the spatial redistribution effects of the TEN-T;
- the calculation of accessibility taking account of proximity to nodes of the transport network;
- the calculation of indicators of cohesion for the European Union;
- dynamic modelling of the development of the transport network and socio-economic impacts over time.

2.4 Joint project on developing Euro-Asian transport links

The aim of this project is to assist Member States of the United Nations Economic Commission for Europe (UNECE) and United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) to develop efficient, safe, sustainable and secure Euro-Asian land and land-cum-sea transport links and promote cooperation in the field of transport so as to facilitate interregional trade and tourism between Europe and Asia.

Until now the project has accomplished the following results:

- the identification of the main Euro-Asian road, rail, inland water transport routes and transhipment points,

- an initial analysis of the main physical and non-physical obstacles along the selected routes,
- evaluation and prioritization of infrastructure projects on the basis of an agreed methodology,
- collection and processing of a large volume of technical and operational data using Geographic Information System,
- organization of a number advisory workshops and interregional meetings dealing with transport infrastructure and facilitation issues.

2.5 MC-ICAM[1]

Project MC-ICAM (2001[-2003) addresses issues related to the implementation of efficient or marginal costbased pricing in transport. The relevant marginal costs are social marginal costs comprising of infrastructure costs, congestion costs, environmental costs and accident costs..

Key dimensions and key policy issues The five key dimensions and relates policy issues are as follows:

- (1) Scope or coverage of the pricing system. What should be priced and who should pay? The answers in part depend on constraints regarding the number of market segments that can be priced distinctly e.g. by: geographical or spatial coverage, modal coverage, user groups covered and externalities covered. An important question related to these issues is about priority: in what order? This is affected by potential existence of synergy benefits.
- (2) Level and composition of pricing measures. What should be the level and composition (i.e. which pricing instruments to use) of prices? Allowance may need to be made for maximum tolls or price caps, budget constraints determining minimum or maximum total revenues etc, and also possible distortions in other links, modes, regions and sectors. Also these issues are affected by synergy benefits discussed above.
- (3) Degree of differentiation across vehicles / infrastructure users, over time and spatially. The desired extent of differentiation depends on the additional welfare benefits associated.
- (4) Use of revenues. It has been argued, and also shown in various studies, that the way revenues are used can have great welfare impacts – often much greater than the direct impacts of pricing. And it is demonstrated in many occasions to be a critical question from the acceptability viewpoint (public and political).
- (5) Supplementary measures and actions. It is similarly argued that the implementation of marginal cost pricing should be considered within broader policy packages including use of revenues and other supplementary non-price measures. We can consider here also policies or actions that aim to directly promote technological development, and improve other practical preconditions or enablers such as estimates of relevant marginal costs and quality of relevant impact data.

2.6 SCENARIOS project

The main aim of the SCENARIOS project (09/1996 – 12/1998) was to develop a reference scenario for the European transport sector for the year 2020, covering:

Socio-economic variables;

Regional dynamics and spatial elements;

Factors affecting transport supply and demand;

---

[1] http://www.transport-research.info/web/projects/project_details.cfm?id=6931&backlink=%2Fweb%2Fcommon%2Fsearch.cfm&referer=y*12|searchstring*MC-ICAM|x*11

Transport policy trends (e.g. European liberalization and harmonization policies).

The project also sought to develop projections of the evolution of technologies, not based on trends, but rather on conditions for their entry to the market.

SCENARIOS concluded that, in the current political climate in Europe, the most likely policy measures for transport involve demand regulation and pricing to alter modal shares. Support for public transport infrastructure seems less likely. SCENARIOS also defined a "European trend policy scenario" assessing the effects of current policies on liberalization and harmonization. The policy scenario was applied in the European Commission's pilot Strategic Environmental Assessment of the trans-European transport network. For the road sector, the scenario showed a decrease or no change in costs under liberalization policy, followed by an increase in costs due to harmonization measures. For rail, costs will increase under liberalization but remain stable under harmonization.

The SCENARIOS project has produced an external reference scenario for European transport in 2020. Using a forecasting methodology, and a range of projection tools, medium and long term trends were identified and analysed for the key scenario variables: external factors, determinants of demand and supply and the transport policy environment. In addition a review of strategic transport modelling was undertaken to provide recommendations for further research. The complexity of European and national decision levels is also discussed for the definition of a policy scenario.

2.7 The European Economic Shipping Impact Study

The European Economic Impact Study (EIS) (97-99) provides policy makers with a tool to quantify the economic significance of the shipping sector to both individual countries in Europe and at European Union level. The project is based on Input-Output Analysis. This instrument allows for the qualification of both direct and indirect effects of a sector, in terms of value added, employment and backflow to the government. By closing the Input-Output model for labour income, induced spending effects can be calculated as well. Included the quantitative evolution of the shipping sector in Belgium, Germany, Italy, the Netherlands and the United Kingdom. The relation between the shipping sector and other sectors of the economy will be analysed, using the E-EIS methodology, as well as the total economic impact of the shipping sector in terms of value added, employment and backflow to the government, as well as induced spending effects;

2.8 The CAPOEIRA project

The CAPOEIRA project (05/2006 – 04/2008) concentrates on maximizing the opportunities for successful R&D and innovation in the field of freight transport activities ports. Included method of evaluation of innovation projects under the form of a "project opportunity assessment grid" aggregating indicators defined so as to minimize the risks of investments in innovation (validated by port actors, technology platforms and advisory councils) and to guarantee that research will pass the commercialization threshold;

3 NEAC modal-split model

3.1 Introduction

The NEAC modal-split model determines the division of freight between modes within Europe, including the short sea shipping (SSS) mode. The model uses an aggregate 'multi-nominal logit' approach, which requires less data than an Input/Output model. This approach is common in the field of modal-split modelling but does come with known limitations, which have been reduced as much as practical by modelling individual commodity flows on specific Origin-Destination (OD) relationships.

In the following sections the theory behind the 'NEAC modal-split' model is examined and briefly the 'NEAC trade flow' is reviewd model as it directly impacts the modal-split model. The strengths and weaknesses of the approach are assessed and recommendations made where relevant.

The primary documents used to review the NEAC modal-split model were:

- E van der Leest, M Duijnisveld and P Hilferink, (2006), 'Update of the NEAC modal-split model', NEA, NL.
- TRANS-TOOLS (2006), Deliverable 3: Report on model specification and calibration results, TNO Inro, NL.

3.2 History of NEAC Modal-Split Model

In the 1980's NEA, a transportation consultancy based in the Netherlands, built a European transportation model (NEAC) to predict the impact of proposed European transport policy. The NEAC model is based on the standard transportation model referred to as the 'Traditional (four-step) Model', predominately used in the passenger market.

This model consists of the 'Trip Generation' stage which aims to predict the total number of trips generated and attracted to each defined region. The 'Trip Distribution' stage which aims to allocate the trips specified in the 'Trip Generation' stage to particular destinations resulting in a trip matrix, Ortúzar and Willumsen (2008). The 'Modal-Split' stage which assigns the trade flows to specific modes of transport. The NEAC Modal-Split model is one of the only European transportation models that includes short sea shipping (SSS) as a mode of transport. Lastly the 'Assignment' stage which assigns the distributed trade flows to specific routes. The NEAC model combines 'Trip Generation' and 'Trip Distribution' into one step entitled Trade Flows.

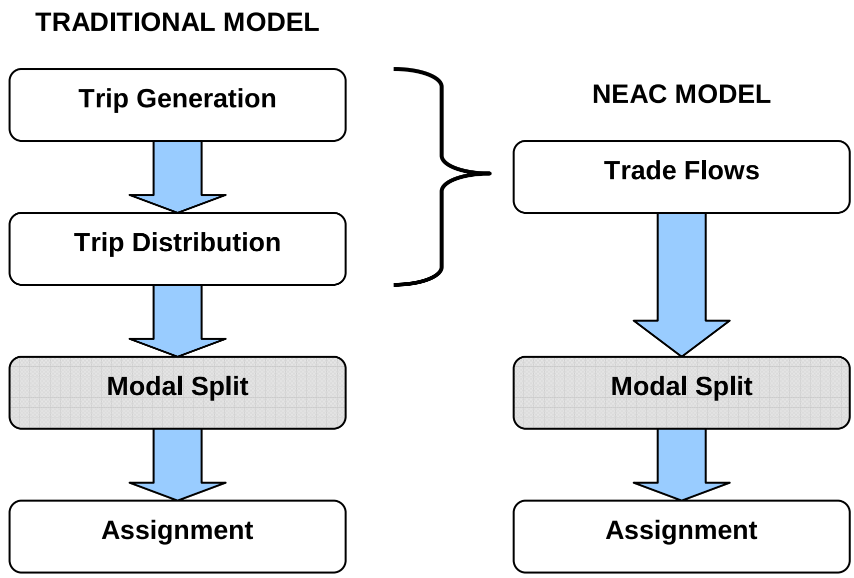

Figure 1: NEAC Transportation Model Functional View

Using the 'Traditional Model' assumes that large, sudden, changes in trade flows do not normally occur and future trade flows can be predicted by either increasing or decreasing trade flows according to economic indicators. For example an increase in GDP in one region will result in higher imports resulting in increased transportation requirements.

To build this type of model all trade flows, including modal-split, for a reference year (referred to as the 'base year') must be known. Economic models are then used to predict the change in trade flows from the base year to the forecast year. The modal-split model then takes into account the new trade flows and predicts the change in modal-split from the base year

The NEAC model uses an 'aggregate modal-split model' to determine forecasted modal split. Aggregate modal-split models usually use binomial or multinomial logit models, which make mode estimates based on the shares of different modes for a number of regions, de Jong and Walker (2004). These models provide the forecasted market share of a mode, not the absolute amount of transport. Consequently, the elasticities from such models are conditional elasticities[2] (conditional on the quantity of demand) Beuthe et al. (2001). The 'aggregate modal-split model' can be based on the theory of individual utility maximisation, but only under very restrictive assumptions.

The 'NEAC modal-split model' determines the market share of a given mode in the forecast year based on changes in costs and time of the different transport modes available. Mode cost and time factors are affected by changes in the organisation of the transport mode, changes in the characteristics of the transport mode, or changes in the modal infrastructure.

[2] Elasticity is the ratio of the percent change in one variable to the percent change in another variable.

3.3 Model Inputs

As mentioned in the previous section, for the modal-split model to make forecasts the following information must be inputted:

- Base Year Data, specifying:
    - Origin, Transhipment & Destination Regions
    - Transport Mode at Origin, Transhipment & Destination Regions
    - Commodity Group
    - Tonnes
- Change in trade flows from base year to forecast year
- Explanatory variables for each mode of transport on each Origin-Destination (OD) route
    - Time per mode
        - Journey time
        - Waiting time
    - Cost per mode
        - Fixed cost per hour
        - Waiting cost per hour
        - Variable cost per km
        - Fuel cost per km
        - Toll cost per km
    - Existence of service per mode
    - Border resistance per mode (dummy variables)
    - Direct line distance

The main source of the input data for the NEAC model is the 1997 EUROSTAT and other national data sources. This database was further updated as part of the European funded 'TransTools' project in 2000 to include Central & Eastern Europe, the Russian Federation and the option of short sea shipping (SSS). Within this database the trade flows are categorised by commodity (11 commodities) and transport mode (5 modes). These are elaborated in the following tables:

COMMODITIES

0	Agricultural Products
1	Foodstuffs
2	Solid Mineral Fuels
3	Crude Oil
4	Ores, Metal Waste
5	Metal Products
6	Building Minerals & Materials
7	Fertilisers
8	Chemicals
9	Machinery & Other Manufacturing
10	Petroleum Products

TRANSPORT MODES

1	Road
2	Rail
3	Inland Waterway
4	Sea
0	Other (Air, Pipeline, etc.)

3.3.1 Constructing the Base Year Trade Flow Database

The data required for the base-year database is not readily available therefore the data that is available must be interrogated in depth such that missing data can be inferred from know trade flows. For example; all trade into a country is well documented at the borders, however, this data does not specify the final destination of the goods. This presents a challenge if the goods are just passing through (transhipment) as they could be miss-interrupted as trade flow to the transhipment country.

To minimise the impact of the lack of sufficiently detailed data the NEAC model follows a 'transport chain approach', which is based on the macro-economic theory that transport flows are caused by trade flows. This approach attempts to build a full picture of the trades between countries including modes at all transhipment points. To facilitate this approach the NEAC database was constructed using the following three steps:

a) Construction of Country to Country Trade Flow Matrix

EUROSTAT data details international trade at the export and import side. Typically only the mode of transport at the border crossing is registered.

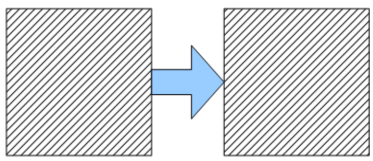

Figure 2: Country to Country Flow

b) Identification of Transhipment Regions

This is particular relevant for countries containing hub ports such as the Netherlands and Belgium. National statistics are crucial to distinguishing between import/export and transhipment trade flows.

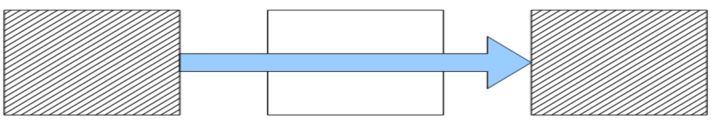

Figure 3: Transhipment Regions

c) Assigning Country Trade Flow to Regions

Country trade flows are divided among regions, preferentially utilising national statistics data and EUROSTAT data elsewhere.

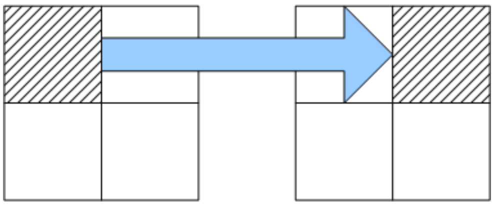

Figure 4: Region to Region Trade Flow

d) Inclusion of Country's Domestic Trade Flow

Unlike international trade flows, which are measured on a continuous basis at borders, domestic trade flows statistics are determined through occasional transport surveys.

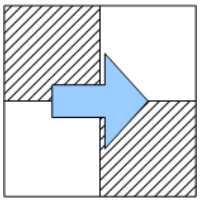

Figure 5: Domestic Trade Flow

3.3.2 Forecasting Future Trade Flows

Once the database is constructed an economic model describing trade flow between regions is generated using an inductive approach. This implies that the data is analysed and general laws, expressed as formulas, are inferred. Built into these formulas (models) are factors that influence trade between countries/regions such as:

- supply factors in the exporting region
- the demand factors in the importing region
- resistance to trade.

These models are similar in mathematical structure to Sir Isaac Newton's gravity equation. Newton's equation describes the attractive force between two large bodies and as such economic models in this format are referred to as gravity models.

The basis for using gravity models is the 'comparative advantage' theory; implying that trade is a result of the specialisation of countries or regions. This allows particular countries/regions to manufacture and exports specific goods cheaper than other countries can, thereby increasing the GDP of the exporting country. The gravity model takes into account these push and pull factors as well as resistance factors; such as distance between exporting and importing country/region. See Appendix-1 for further information and a sample mathematical derivation of the gravity model formula.

3.3.3 Explanatory Variables

These variables are determined for each mode on each Origin-Destination (OD) route and used in the equation that determines the modal-split for each commodity on each OD route. These variables are calculated based on hard data such as fuel prices, wage rates etc. How these ‘explanatory variables’ are used in the mathematical model will be examined in section 4.1..

3.4 Model Structure

The NEAC modal split model has two aspects to its structure; how the data is analysed to predict the modal split and how the raw data, gathered from European and national databases, is organised to improve the predictability of the model. The data analysis aspect assigns relative probabilities to the selection of a specific mode by an individual. Probability is usually expressed as the frequency of an expected event divided by the sum of the total possible outcomes. This would equate to:

Probability of choosing truck mode = $\frac{Truck}{Train + ShortSeaShipping + Inlandwaterway}$

The calculation of these probabilities can only be based on observed and quantifiable factors therefore the attractiveness of each mode is calculated. This approach is called ‘Random Utility Theory’ and will be examined further in the next section.

The second aspect of the NEAC model structure is how the data is organised to improve predictability. It can not be expected that a formula derived to predict the modal split of fresh fish transported from the Baltic Sea states to the rest of Europe, could accurately predict the modal split of office furniture transported from the Iberian Peninsula to the rest of Europe. For this reason, the raw data is segmented into similar groups and different formula derived for each group. This greatly improves the accuracy of the model. Both of these structural aspects are further in the next two sections.

3.4.1 Random Utility Theory

Random utility theory assumes that the decision maker has perfect discrimination capability. Based on this principle, the decision maker will choose the available mode that yields the most utility. The utility of a mode is a function of the characteristics of the mode and the characteristics of the individual decision maker. Some of these characteristics are directly observable and are termed ‘systematic utility’, other characteristics which cannot be directly observed are accounted for with a random utility component. Therefore, the total utility of a mode is defined as the sum of all systematic and random utility components:

Equation-1: $$U_{m,n} = V_{m,n} + \varepsilon_{m,n}$$

$U_{m,n}$ = Total utility of mode m for individual n

$V_{m,n}$ = Systematic utility of mode m for individual n

$\varepsilon_{m,n}$ = Random component of utility of mode m for individual n

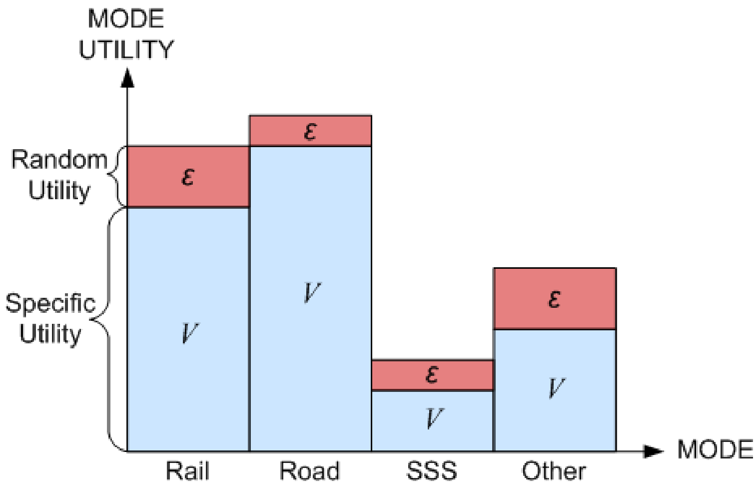

Figure-7: Indicative graphical representation of Equation-1

The individual choosing the mode of transport maximises utility and thus chooses the alternative with the highest total utility $U_{m,n}$, but in the model only the $V_{m,n}$ can be observed (measured). So now it is possible to define the probability $P_{m,n}$ that an individual n chooses mode m with this equation:

Equation-2: $$P_{m,n} = \frac{e^{V_{m,n}}}{\sum_{l \in M} e^{V_{l,n}}} = \frac{e^{\beta' x_{m,n}}}{\sum_{l \in M} e^{\beta' x_{l,n}}}$$

Where $V_{m,n} = \beta' x_{m,n}$

$x_{m,n}$ = Observed characteristics of mode m and individual n

β = Logit parameter

The function for the systematic utility $V_{m,n}$ is defined as the multiplication of the vectors β and $x_{m,n}$. The value of β is determined separately for each segment. Using this number the choice probabilities of the available modes per commodity group for every OD relation are determined by the following formula:

Equation-3: $$P_{m|cij} = \frac{e^{V_{m|cij}}}{\sum_{l \in M} e^{V_{l|cij}}}$$

Where: $V_{m|cij} = \beta_{m0} + \sum_k \beta_{mk} x_{cijmk}$

$P_{m|cij}$ = Choice probability of mode m given commodity group c and OD relation ij

$V_{m|cij}$ = Systematic utility of mode m given commodity group c and OD relation ij

x_{cijmk} = Explanatory variable k for mode m given commodity group c and OD relation ij

β_{mk} = Logit parameter for mode m and level of service k.

e = Unique real number (2.7182818284)

Different modal-split shares of the market are estimated for every OD relation (ij) and commodity group (c). The explanatory variables or observed characteristics used in the model are:

x_1	relative total cost per ton	x_5	border resistance dummies
x_2	relative total time	x_6	domestic transport dummies
x_3	distance	x_7	intercontinental transport dummies
x_4	total annual transport volume	x_8	port region dummies

Therefore the total utility for mode (m) carrying commodity (c) on OD route (ij) can be described as:

$$V = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8$$

Since the modal split and explanatory variables are known for the base year it is possible to calculate the values of $\beta_1 - \beta_8$ for each commodity group (c) on OD route (ij). The mode choice probability is calculated by substituting the explanatory variable for the forecast year and the difference between this result and the base year mode split is applied to the forecast year tonnage from the Trade-Flow model.

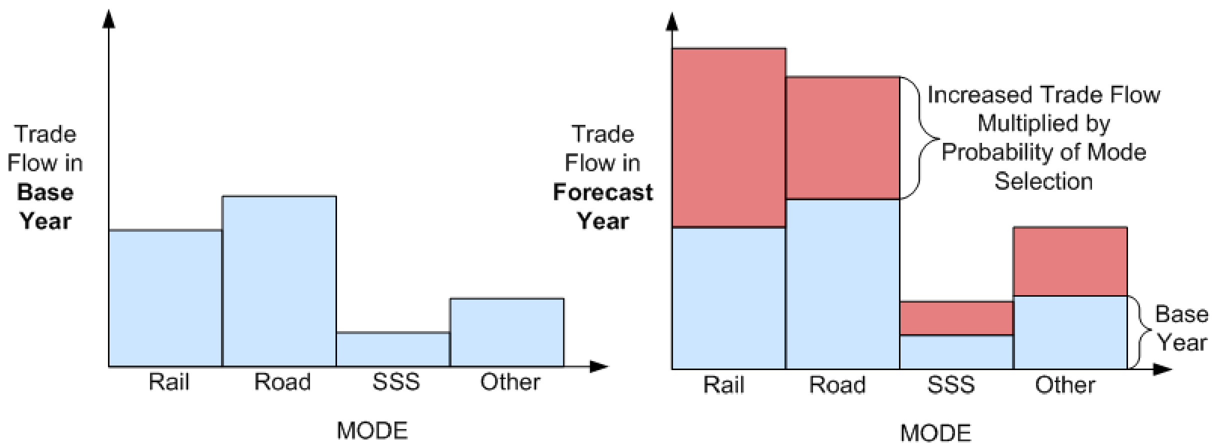

Figure-8:Indicative calculation of Modal-Split in Forecast Year for OD (ij) and Commodity (c)

This approach assumes that the main determinants of mode choice are cost and time of transport. To improve the accuracy of this approach the data needs to be organised into homogeneous segments.

3.4.2 Segmentation

The information within the database is organised by mode and commodity, however it is very heterogeneous and formulas derived from this data are not very sensitive to variations in 'cost' or 'time'. This condition conflicts with assumptions made in section 4.1, therefore the data is separated into more homogeneous segments with similar characteristics. These segments consist of groups of transport markets (each OD & Commodity relation is considered a separate market).

For the segmentation analysis the CHAID (Chi-Squared Automatic Interaction Detection) technique is used to split the data in homogeneous groups. CHAID is a criterion-based segmentation tool limited to nominal and ordinal categorical variables. It is an evolution of AID (automatic interaction detection), which uses a hierarchical binary splitting algorithm on a set of categorical variables. One explanatory variable is chosen as predictor for the division of the data set at every stage. The choice of the predictor is based on the minimization of the residual sum of squares of the dependent variable (criterion) of the divided data set. The splitting of the data set in two mutually exclusive and exhaustive subsets results from the division of the categories of the predictor variable in two groups. So observations in the data set with a value of the predictor in the first category go in the first subset and observations with a predictor value in the second category go in the second subset.

This process is repeated creating subsets of subsets until no further improvement, that is no lower residual sum of squares of the criterion, can be attained by dividing the data set in more subsets. Some drawbacks of AID are a bias for the selection of variables with more than two categories as the best predictor and the possibility that the choice of an optimal split in an early stage can lead to a sub-optimal final solution: it may be possible to reach a lower minimum residual sum of squares for the final division of the data set by choosing a predictor that does not minimize the residual sum of squares at an early stage. CHAID eliminates these drawbacks. Furthermore, CHAID considers the whole distribution of the dependent variable, it is not restricted to binary splits, and it makes no assumptions of normality. It utilizes the chi-square test for independence to assess statistical significance. CHAID is able to automatically identify significant interaction effects between categories of dependent/criterion variables.

The first step is merging. For each pair of categories eligible to be merged of each predictor, chi-squared tests are computed to test for independence in the subset of data formed by the two categories of the predictor variable and the dependent variable. Among those pairs found to be non-significant, the most similar are merged. These pairs have the smallest chi-square value. In the algorithm, for any subset containing three or more categories, a test of the significance for any predictor associated with a category against the other categories in that subset is performed to see if any predictor should be unmerged. If more than one category is eligible to be unmerged, the one having the highest chi-square is selected to be unmerged. These procedures are repeated until only significant pairs remain and thus no categories are eligible to be unmerged. Then the probability p that the observed sample relationship between the predictor and the dependent variable would occur if the two variables were in fact statistically independent is calculated.

The second step is splitting. The predictor with the lowest significant probability p is selected and the group is split on this predictor. If no predictor has a significant probability p, the group is not split. These steps are repeated until all subgroups have been analyzed or contain too few observations.

The result is a tree where the data set is the root and the branches are the more homogeneous subsets.

OD relations with a different availability of modes of transport have different markets for mode choice. The modal-split model is estimated separately for these markets. There are five possible combinations of available modes:

Mode Combination
Road & Rail
Road & Short Sea
Road, Rail & Inland Waterway
Road, Rail & Short Sea
Road, Rail, Inland Waterway & Short Sea

For each of these five markets a segmentation analysis is carried out. Based on preliminary test results and expert opinion; commodity group, the location of the origin and destination region, the distance and the tonnage are used as explanatory variables in the segmentation analysis. Therefore the modal-split in the base year is the dependent variable and the respective explanatory variable categories used are as follows:

Commodity		OD Regions	Distance	Tonnage
0	Agricultural Products	Scandinavia	0 – 400km	<5000
1	Foodstuffs	UK & Ireland	400 – 700km	5000 – 50,000
2	Solid Mineral Fuels	Western Europe	700 – 1000km	50,000 – 500,000
3	Crude Oil	Eastern Europe	1000 – 1500km	>500,000
4	Ores, Metal Waste		>1500km	
5	Metal Products			
6	Building Minerals & Materials			
7	Fertilisers			
8	Chemicals			
9	Machinery & Other Manufacturing			
10	Petroleum Products			

3.5 Model Limitations

The multinominal logit approach used in the NEAC modal-split model does have draw backs, such as the independence of irrelevant alternatives (IIA). This property declares that the relative choice probability for any pair of alternatives is independent of the absence or presence of other alternatives. This results in failure of the model when correlated alternatives are present. A well known example of this property is the blue bus-red bus problem[3]. The other property is the identical distribution of the

[3] See Appendix 2

random components. This means that the variance is equal across alternatives. Thus, a given absolute difference between the characteristics of two modes has the same impact on the share of the modes for every level. So for instance a difference of 15 minutes has the same impact on a trip with a transport time of one hour as it has on a trip of two days.

4 Other Modal-Split Models

4.1 Demand Model

Direct demand models combine three of the traditional model steps, Generation, Distribution & Modal Split. These models predict the number of trips (or kilometres) for each mode, De Jong et al (2004). This approach has the advantage of minimising errors caused by breaking up each of the fours steps and approaching them sequentially. Direct demand models are a form of econometric modelling and determine modal split solely on the basis of cost, time and reliability.

4.2 Disaggregate Modal Split Models

In disaggregate modal-split models mode choice is modelled at the level of individual shipments, as apposed to at the regional level. Such models use data from surveys of shippers, commodity surveys and so on. Usually a combination of Revealed and Stated Preference data is used. Using Stated Preference data allows the impact of policy variables on modal-split to be incorporated in the modelling, and in particular the relative importance of variables other than cost and time can be identified, such as reliability. Such models tend to be multinominal or nested logit models, relying on the random utilisation theory.

4.3 Multi-modal Network Models

Multi-modal network models predict mode and route choice simultaneously. To determine the optimal mode-route path through the network, cost minimisation algorithms are used. The optimal path through the network may involve one or more changes in modes, and therefore multi-modal network approaches can model intermodal movements directly. The costs of each route can contain cost and travel time components by mode, and costs associated with transhipment when multi-mode paths are considered. These models require less information than demand models and have a theoretical basis.

4.4 Probit Model

The probit model is a discrete choice model and is very similar to the logit model. This model also relies on the assumption of random utility theory. The probit model assumes that the random variable ($\varepsilon_{m,n}$) is normally distributed and the model has been proven to predict with greater generality. This benefit comes at the cost of large computation requirements as the equations that need to be solved are significantly more complex than those of the logit model. For this reason the logit model has been more popular among modellers in the past.

4.5 Artificial Neural Networks Model

This is a relatively new area within the field of modal-split modelling, and started to gain acceptance in the early 21st century. Artificial neural networks use pattern association and error correction as the underlying mechanism to represent a problem. These models function similarly to the human brain and are capable of learning and self-correcting; in fact this attribute is essential as the model must be trained so as to improve its accuracy. The model is built with two building blocks, neurons and synapses. Neurons have one output each which is related to the state of the neuron, its activation. This output is sent to other neurons via the synapses, these synapses are weighted, which form the memory of the model. A neuron functions similarly to an AND gate in electronics. All of the inputs to a neuron are inputted to a threshold (usually non-linear) function and if the function output value is greater than a threshold level a neuron output is generated which propagates to the next layer of neurons.

Neurons are typically organised in three layers, the input layer which accepts model inputs, the output layer which output the final answer (in this case modal split) and the hidden layer which functions as the intermediary between the input and output layers. These elements are represented in the figure 9 and 10.

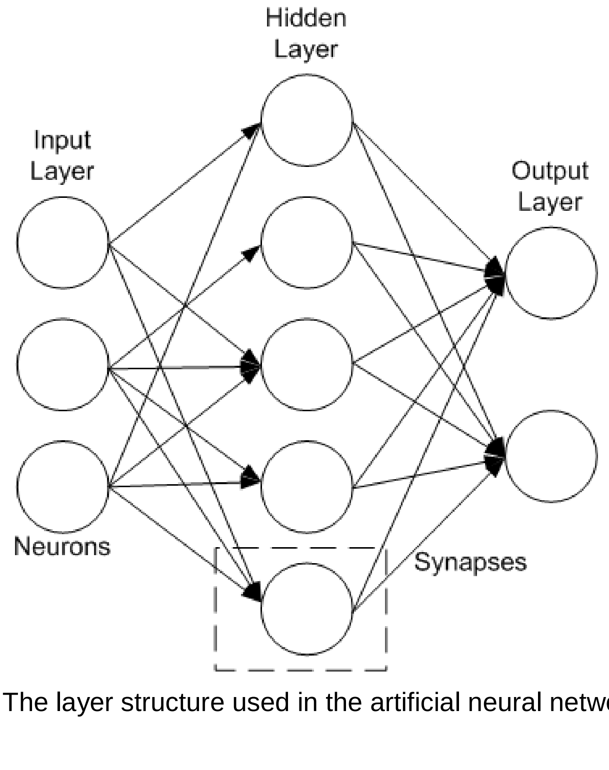

Figure-9: The layer structure used in the artificial neural network model

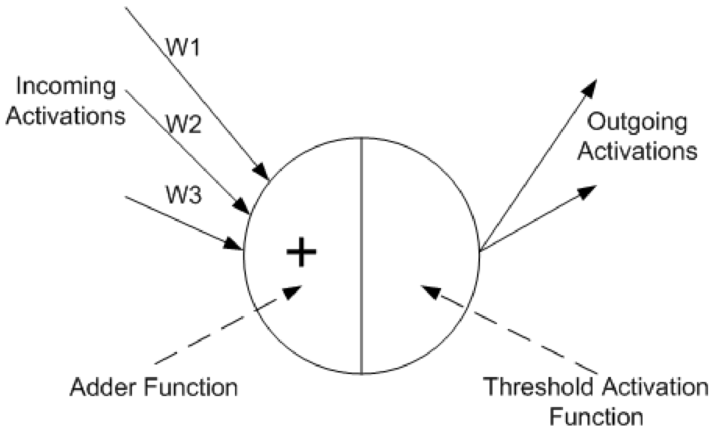

Figure-10: Functioning of a neuron (expanded view of the dashed box area in Figure-9)

The model input patterns in this system consist of attributes of the modes. A disadvantage to this model is the lack of transparency within the model and the absence of a along proven history, Rand Europe, (2002).

5 Policy Implications

The policy implications are:

- The diversity of trade models and databases create difficulties for users to estimate trade flows in order to plan infrastructure investments or design efficiently transport services. Among the issues of concern: reliability, consistency, timeliness, flexibility in terms of data/updates and interoperability of EU databases.
- Many projects address specific regional trade flows but results are not used to establish a coherent picture of the key trade corridors in Europe and their potential development.
- The NEAC modal-split model has been used to carry out transport flow analysis for the EU, "Transport forecasting goods and passengers for the year 2020". If this model is to provide an input to policy, and, in the light of recent world market changes it is recommended that the 2020 study should be re-run with updated economic data, E van der Leest, et al (2006). According to reviewed literature the NEAC model has a higher probability of accurately forecasting modal split on established trade routes within the EU-15 member states. However, due to the limited data on the use of SSS in the base year the model may not specify modal-shift opportunities, TRANS-TOOLS (2006). In tandem with this limitation, new SSS services can start with a relatively low investment and lead time, unlike road and rail. This allows the SSS market to be more flexible and reactive and hence more difficult to model or predict.

6 Recommendations

- Transport flow databases including ETIS need to be integrated with Eurostat and to facilitate easy access by users to require datasets.
- A coherent view of key EU trade corridors from a maritime transport perspective is needed to facilitate planning of 'green corridors' in the context of sustainable transport
- The status of modal-split modelling should be reviewed annually as it is a continuously developing field and will potentially experience exponential growth as various e-Maritime initiatives make information more available in electronic format. Accurate and timely information is a key obstacle to model development and the increased information availability will increase the accuracy and usefulness of these models.

7 Conclusions

- Further work is needed to establish a library of useful trade and transport models and databases as well as guides for their use
- As the NEAC modal-split model is a proprietary software package details on the constants and assumptions used in the creation of the model were not available, therefore, the approach used in this paper was to evaluate the theory behind the model. The approached

used by the model has been tried, tested and proven effective within the field of passenger modal split on land and in this instance has been adapted to apply to freight.

Further validation of the model and an assessment of the impact of base year data on the SSS option must be carried out. Once completed this would provide greater confidence in the model with respect to SSS. This can be achieved by testing predicted outcomes against actual outcomes and assessing the deviation if any.

8 References

1. Anderson, J. (1979) "A Theoretical Foundation of the Gravity Model." American Economic Review 69(1), 106-16
2. Tinbergen, J. (1962) "Shaping the World Economy: Suggestions for an International Economic Policy". The Twentieth Century Fund. New York
3. Bergstrand, J. (1985) "The Gravity Model in International Trade: Some Microeconomic Foundations and Empirical Evidence." The Review of Economics and Statistics 67, 474-81
4. Deardorff, A.V. (1995) "Determinants of Bilateral Trade: Does Gravity Work In a Neoclassical World?" NBER Working Paper No. 5377. National Bureau of Economic Research. New York.
5. Egger, P. (2000) "An Econometric View on the Estimation of Gravity Models and the Calculation of Trade Potentials." The World Economy. 27(2). 297-312(16)
6. Head, Keith. (2003) "Gravity for Beginners." Working Paper. Faculty of Commerce. University of British Columbia
7. Linneman, H. (1966) "An Econometric Study of International Trade Flows" Amsterdam. North Holland Publishing Co.
8. Carrillo, C. and Li, C. A. (2002) "Trade Blocks and the Gravity Model: Evidence from Latin American Countries". Economics Discussion Papers 542, University of Essex, Department of Economics.
9. Koo, W., Kennedy, P.L., Skripnitchenko, A. (2006) "Regional Preferential Trade Agreements: Trade Creation and Diversion Effects." *Review of Agricultural Economics.* 28(3)
10. Deardorff, A.V. (1995) "Determinants of Bilateral Trade: Does Gravity Work In a Neoclassical World?" NBER Working Paper No. 5377. National Bureau of Economic Research. New York.
11. Fosgerau, M., Kveiborg O. (2004) "A Review of some Critical Assumptions in the Relationship Between Economic Activity and Freight Transport" International Journal of Transport Economics, 31 (2) 247-261.
12. Egger, P. (2000) "A note on the proper econometric specification of the gravity equation" Economics Letter, 66 25-31.
13. Stéphanie Vincent Lyk-Jensen, "The future flow of goods in the enlarged EU", Danish Transport Research Institute

14. Tavasszy. L. A. (1996). "Modelling European Freight Transport Flows" Ph.D. Thesis. Delft University of Technology.
15. "The Swedish model system for goods transport – SAMGODS" SAMPLAN rapport 2001
16. "Computable General Equilibrium Model for studying Economy-Energy-Environment Interactions for Europe and the World" Final Report, DG Research, 5th Framework Programme

NEAC Model

1. Ben-Akiva, M., Meersman, H. and de Voorde, E. V. (ed.) (2008) *Recent Developments in Transport Modelling: Lessons for the Freight Sector*, Bingley, Emerald.
2. Beuthe, M., Jourquin, B., Geerts, J., Ndjang Ha, C. (2001) *'Freight transportation demand elasticities: a geographical multimodal transportation network analysis'*, Transportation Research Part E, Vol. 37 : p.253-266
3. De Jong, G., Walker W. (2004) 'National and Internation Freight Transport Models: Overview and Ideas for Future Development', Transport Reviews, Vol 24, No. 1 : p.103-124
4. Kennedy, P. (2008) '*A Guide to Econometrics*', 6th Edition, Oxford, Blackwell Publishing
5. Konings, R., Priemus, H. and Nijkamp, P. (ed.) (2008) 'The future of Intermodal Freight Transport', Chletenham, Edward Elgar Publishing.
6. Ortúzar, J. D., Willumsen, L. G. (2008) *Modelling Transport*, Sussex: Wiley.
7. Rand Europe, (2002) 'Review of Freight Modelling', Report B2 – *Review of Models in Continential Europe and Elsewhere*, Department of Transport and Economic Appraisal,
8. Stopford, M. (2008) 'Maritime Economics', 3rd Edition, Oxon, Routledge.
9. Vido, E. and Prentice, B. (2003) '*The Use of Proxy variable in Economic Gravity models: A Cautionary Note'*, Journal of the Transport Research Forum, Vol. 57, No. 1 : p.123-137
10. E van der Leest, M Duijnisveld and P Hilferink, (2006), 'Update of the NEAC modal-split model', NEA, NL.
11. TRANS-TOOLS (2006), Deliverable 3: Report on model specification and calibration results, TNO Inro, NL.

Key projects

TRANS-TOOLS, FP7 Project, completed in 2006

Key web sites

http://www.inro.tno.nl/transtools/index.html

http://site.nea.nl/neac/neac_modules.htm#Modal-Split%20Model

http://www.skematransport.eu/

http://www.etcproceedings.org

9 Appendix-1 Derivation of the Gravity Model

Whilst the NEAC 'modal-split model' does have its limitations it is in line with common practice within the field of freight modelling. Additionally, it is one of only a few European models that include the SSS mode. However, this modal-split model can not be examined in isolation as the results from such a model are only as good as the quality of its inputs; in this case the forecast trade flows from the 'Trade Flow' model.

As already mentioned, the 'Trade Flow' model is based on the gravity model and this has inherent limitations. Gravity models assume that the 'Law of One Price' applies. This implies that in an efficient market all identical goods must have only one price and differences in price are accounted for by the cost of moving goods.

The above assumption means that the geographic information data requirements are less onerous for gravity models. Traditional trade equilibrium models require commodity prices in different regions. Economic gravity models require an accurate measure of commodity flows and transportation cost between origins and destinations. Due to the disparate sources of information on transportation costs, distance and other variables are used as a proxy for transportation costs, Vido and Prentice (2003). This approach then implicitly assumes that there is only one distance between regions and that freight rates are the same on the front and back haul, neither of these assumptions stand up to scrutiny.

Gravity models account for the economic attractive forces between trade regions and utilise cross-section base year data (i.e. one specific point in time) to determine relationships. The disadvantage of this approach is that a region or a mode that may not have been a significant contributor to trade in the base year may become a large contributor in ensuing years. Unless the model is updated it will overlook this discrepancy. The impact of these structural weakness of gravity models have been reduced within the NEAC modal-split model through the use of segmentation and generation of specific formula for each commodity type on each OD relation.

The gravity model can be derived from Sir Isaac Newton's gravitational law as follows.
Newton's gravitational law:

Equation.-5: $$F = GM_i M_j D^{-2}$$

F = Force of attraction between object (i) and (j)

G = Gravitational constant

M_i = Mass of object (i)

M_j = Mass of object (j)

D = Distance between (i) and (j).

This formula is seen as analogous to trade flow between countries and the components are relabelled and one new term introduced (η) as follows:

Equation.-6: $$T_{ij} = \alpha 1 M_i M_j D^{-1} \eta$$

T_{ij} = Flow of trade between region (i) and (j)

M_i = Attractive characteristics of region (i), (typically GDP)

M_j = Attractive characteristics of region (j), (typically GDP)

D = Resistance to trade between (i) and (j), usually Distance

$\alpha 1$ = Proportionality Constant

η = Error correction term

As the formula may only use quantitative values therefore:

- M_i becomes P_i, the added value of the sector that supplies the commodity in country/region i.
- M_j becomes A_j the added value of the sector that consumes the commodity in country/region j.
- η becomes e^{DUMMY} where the dummy variable captures economic co-operation between countries/regions, such as free trade zones etc.
- Estimators $\alpha 2$, $\alpha 3$, $\alpha 4$ & $\alpha 5$ are introduced to weight the equation parameters and distinct values are estimated for 10 of the 11 commodity groups (crude oil was excluded).

This results in the following equation:

Equation.-7: $$T_{ij} = \alpha 1 P_i^{\alpha 2} A_j^{\alpha 3} D_{ij}^{\alpha 4} e^{\alpha 5 DUMMY}$$

The equation is then put in log form and is referred to as the log-linear regression equation:

Equation.-8: $$Log T_{ij} = Log \alpha 1 + \alpha 2 Log P_i + \alpha 3 Log A_j + \alpha 4 Log D_{ij} + \alpha 5 DUMMY$$

The constants ($\alpha 1$- $\alpha 5$) are derived from plotting known variables from the above equation and matching the above equation form to the data. The method used is referred to as 'Ordinary Least Squares' (OLS) which minimises the difference between the derived equation and the plotted data. The equation is examined for consistency and the signs of the constants should be as follows for a consistent formula.

The estimators should have appropriate signage, as follows:

- ($\alpha 2 > 0$) a larger added value within the producing sector in the exporting country should have a larger effect on the trade.
- ($\alpha 3 > 0$) a larger added value within the attracting sector in the importing country should have a larger effect on the trade.
- ($\alpha 4 < 0$) a larger distance between the exporting and the importing country should have a negative effect on trade.
- (DUMMY ±) the dummy variable can either be positive or negative

10 Appendix-2 Blue bus, red bus paradox

Suppose an individual must reach a given destination and the individual has the same probability of using their car or of taking the bus:

P(car) = P(bus)

=) P(car)/P(bus) = 1

Suppose now that two buses can be used, identical except for their colour, red or blue. The choice set is given by:

A = {car, blue bus, red bus}

Assume that the individual pays no attention to colour so that:

P(blue bus) = P(red bus)

=) P(blue bus)/P(red bus) = 1

Intuitively one would think that the respective choice probabilities would be:

P(car) = ½

and P(blue bus) = P(red bus) = ¼

However, the choice axiom[4] of the logit model implies that:

P(car)/P(blue bus) = P(red bus)/ P(blue bus) = 1

=) P(car) = P(red bus) = P(blue bus)

The sum of all the probabilities must equal to one, therefore:

P(car) + P(blue bus) + P(red bus) = 1

The only probabilities that allow this equation to be correct are as follows:

P(car) = P(red bus) = P(blue bus) = $^1/_3$

This would imply that the probability of an individual choosing a bus would be equal to:

P(bus) = P(red bus) + P(blue bus)

P(bus) = $^1/_3$ + $^1/_3$

P(bus) = $^2/_3$

This would imply that the simple introduction a new coloured bus would increase the usage of buses from 50% to 66%. As this is clearly not the case this paradox must be accounted for in logit modelling.

---

[4] The choice axiom implies that the ratio of choice probabilities is constant