European Transport \ Trasporti Europei

European Transport \ Trasporti Europei
International Journal of Transport Economics, Engineering and Law
Rivista Internazionale di Economia, Ingegneria e Diritto dei Trasporti

ISSN 1825-3997

YEAR XIII NUMBER 37, December 2007

Publisher/Direzione, redazione e amministrazione

I.S.T.I.E.E.
Istituto per lo Studio dei Trasporti nell'Integrazione Economica Europea

Via Lazzaretto Vecchio, 13
34123 TRIESTE
Tel. +39.040.311464
Fax. +39.040.311465
Mail: istiee@units.it

www.istiee.org/te

Print/Stampa
ARTIGRAFICHERIVA SRL TRIESTE

Autorizzazione del Tribunale di Trieste, n. 915 del 31 ottobre 1995

Spedizione in A.O. 70% DC TS
Tassa Pagata/Taxe Perçue

Index

YEAR XIII, NUMBER 37, December 2007

European Transport \ Trasporti Europei n. 37 (2007): 1-15

The liberalisation of maritime transport and the island regions in EU. Evidence from Greece

C. I. Chlomoudis [1], P. L. Pallis [1], S. Papadimitriou [1*], E. S. Tzannatos [1]

[1] *Department of Maritime Studies, University of Piraeus*

Abstract

The liberalisation of maritime transport constitutes a substantial and vital progress not only in the maritime transport market but also in influencing the development of distant island regions. In the European Union (EU), the establishment of the legal framework and policy regarding the liberalisation is standing from the year of 1992. In Greece, the first attempt to harmonize the inevitable political and state framework was in 2001 (2932/2001 Act). Prevailing factors encumbered the liberalisation planning and implementing, mainly because of the incoherent procedures applied and the partial cover of principal issues regarding impingement of public interests and goods relative to island development.

In our paper, we provide strongly support to the argument that the major initiative considering the development of island regions throughout European territory is that European Policy for the liberalisation of maritime transport should be oriented to the distinctive needs of island regions. Moreover, we provide a "road map" for completion of the liberalisation procedures. Hence, we address specific proposals and measures towards the healing of the inadequate regional development.

Keywords: Maritime transport; Liberalisation; EU policy; Island regions.

---

1. Introduction

The general aim of this paper is to discuss the EU maritime policy and its relation to the islands regions across EU. The Greek Institute for Local Authorities (GILA) report for 'The Inter-Island Transport problem, counter measures and the role of local Authorities' ((2006), Chlomoudis, C. et al.) composed a substantial resource for our study.

As more than 90% of the EU's external trade and some 43% of its internal trade is transported by sea, international maritime transport ought to be a liberalised activity. If it were not, nobody would benefit from the role this mode of transport plays in international trade[1]. However, it is only since 1st January 1993 that cabotage by sea[2] has

* Corresponding author: Stratos Papadimitriou (stratos@unipi.gr)

[1] http://europa.eu/scadplus/leg/en/lvb/124040.htm

started to be phased in, as agreed in 1992. The introduction of cabotage and the need for the Community to help improve the conditions for international maritime transport have resulted in the adoption of measures relating to the competition policy, to the prevention of unfair pricing practices, to standards for ships engaged in the transport of dangerous goods and to the working conditions[3]. Furthermore, maritime transport policy is specifically required to take into account the overriding EU principles of social solidarity and cohesion between central and peripheral regions (CEC, 1994a, 1996a), which are concerned primarily with measures to reduce regional inequalities in economic and social people's well-being factors.

The physical characteristics and natural conditions of island communities, are a drawback to their economical and social development. Maritime liberalisation adapted to these regions could serve as a basis for specific policies designed to offset these natural handicaps providing the islands with adequate supply and quality of services which could serve to foster territory continuity and the well being of the island population. The aim is to provide the proper mix of measures and policies adjusting them to the island peculiarities, that would offer the islands a real chance of competing, in the near future, with the mainland and in particular with the most accessible and most developed Regions (Eurisles, 1998).

In this paper, the main features of liberalisation on maritime transport are outlined as essential preliminaries to the understanding of how cabotage constitutes a substantial and vital progress not only in the maritime transport market but also in influencing the development of distant island regions. Various models of cabotage adjustment across the countries of Europe are then examined, with particular attention to address how each country considers sea transport as a public service obligation. The adequacy of the current situation of Greece is analysed in terms of the maritime liberalisation adjustment, the domestic market conditions and the necessity of considering sea transport as public service. The most notable feature of our study is that even though EU established the legal framework and the policy regarding the liberalisation, EU countries – including Greece – have not applied all the necessary procedures and policies in order to be completely harmonized with the EU directives. Finally, we propose some significant measures and characteristics that all the island regions across Europe should bear in mind as a common basis towards a rational and effective development.

2. The liberalisation of maritime transport in the European Union

In general, liberalisation involves the exposure of the transport market to laissez-faire, or free market, achieved through the removal of most regulatory controls over pricing, while permitting carriers to enter and leave the markets at will.

Council Regulation (EEC), No 3577/92, of 7th December 1992, applied the principle of freedom to provide services to maritime transport within Members States (maritime cabotage)[4]. The regulation grants freedom to provide maritime transport services within

[2] Council Regulation (EEC) No 3577/92 of December 1992 applying the principle of freedom to provide services to maritime transport within Member States (maritime cabotage) [Official Journal L 364 of 12.12.1992]

[3] http://europa.eu/scadplus/leg/en/lvb/124040.htm

[4] http://europa.eu/scadplus/leg/en/lvb/124040.htm

a Member State for Community ship-owners operating ships registered in a Member State and flying the flag of a Member State. A substantial article of the regulation deals with the right that Member States have to provide transport services subject to public service obligations (PSO) in the interest of maintaining adequate cabotage services between the mainland and its islands and between the islands themselves.

Under European Commission guidelines, a PSO in ferry services may only be declared to provide services needed to address problems of peripherality, insularity and/or economic disadvantage that would not normally be addressed without public intervention. The definition of a PSO is set out as: "any obligation imposed upon a carrier to ensure the provision of a service satisfying fixed standards of continuity, regularity, capacity and pricing, which standards the carrier would not assume if it were solely considering its economic interest".

Consequently, there is a time-lag between the implementation of domestic and international liberalisation, the former being much easier to implement, either at the scale of the individual country or in a trading bloc such as the EU. For this reason, safeguard measures may be taken by the Commission where the internal market would seriously be disrupted by the liberalisation of cabotage or where countries whose domestic island ferry cabotage rules and contracting procedures are in violation[5] of Council Regulation EEC 3577/92. In the case of France, Italy, Greece, Portugal and Spain mainland cabotage was gradually liberalised according to a specific timetable for each type of transport service. Mainland-island and inter-island cabotage for these countries was liberalised in 1999[6].

3. Island regions in European Union and ferry services

Our aim is to demonstrate that the island space has inherent characteristics which affect the economic development of the European island regions. Matters of size, dependence and distance imply different consequences, as do historical issues or economic systems, but overall, there are specific island characteristics which determine the type of economic development of a territory.

---

[5] (Spain) In December 1998, the European Commission opened an investigation into state financing made available to the publicly-owned Spanish company Compania Trasmediteerranea S.A. (Transmed). Following the investigation Spain was required to re-tender the ferry services. Spain C-205/99 Royal Decree No 1466 was inconsistent with Community law, in particular Articles 1, 2 and 4 of Council Regulation (EEC) No 3577/92. (Portugal) An infringement procedure was started against Portuguese decree-law No 194/98 of 10 July 1998 on maritime cabotage. The Commission sent a reasoned opinion to the Portuguese authorities inviting them to amend the text to bring it into line with Council Regulation (EEC) No 3577/92. (France) The European Commission initiated a formal investigation procedure into the proposed restructuring aid to Société Nationale Corse-Méditerranée (SNCM). (Greece) The Greek legislation of 2000 did not appropriately reflect the EU basic principle according to which public service obligations can be imposed on ship-owners performing island cabotage in the event of market failure to provide adequate services. (Italy) In 1999, the European Commission opened a State aid investigation against Italian ferry company Gruppo Tirrenia di Navigazione with regard to subsidies paid by the Italian authorities to the six companies of the ferry group Gruppo Tirrenia di Navigazione.

[6] This exemption was prolonged until 2004 for scheduled passengers and lighter services and services involving vessels of less than 650 gross tonnages in the case of Greece.

The island system[7] has the following characteristics *«One fact which always holds true is that the islands are smaller that the mainland areas. While this seems to be stating the obvious, it nevertheless leads to a very simple first level of reasoning. Because of its endemism to the islands, smallness implies rarity. While this rarity manifests itself in various ways, it is measured mainly by the scarcity and paucity of resources (raw materials, infrastructures, human potentialities, etc.). In order to manage these rare resources, the island economies adopt a specific management system. This is based on the need to have access to a major network of exchange with the outside. From colonialism to today's trade deficits, this phenomenon is a constant factor of the economic history of all the island economies. Rarity and openness to the outside are factors which lead to a high degree of dependence, due to a virtual single-product export activity and a high level of imports. This dependence, which is due to the weakness of the domestic economy and the dominant role of external trade, is significantly aggravated if, in addition, the island is remote and located far from its markets.»*

The “Regional disparity indicators in the European Union” report identifies three sources of disparities which affect inter-regional cohesion within the European Union:

- Disparities which are linked to the natural conditions and physical characteristics of the regions;
- Disparities which are indicative of an unequal development of production potential;
- Disparities which highlight the great differences in income and living standards from one region of the European Union to another.

While not denying the importance of the second and third category, the first type of disparity is obviously the one which is of primary concern to the islands. Moreover, this type of disparity is concerned with the insularity as a phenomenon which weights most socio-economic data and the phenomenon of physical discontinuity. (Eurisles, (1998))

Nevertheless, island communities have a special inheritance, a way of life that is to be cherished and valued. Their peripheral location can make it difficult to share equally in the economic and social life of the nation, and they face particular barriers in ensuring full participation arising from their location (Malachy Walsh and Partners in association with Posford Haskoning, Raymond Burke Consulting, McCaig Watson and Seosamh Mac Donnacha).

Our study aims to point out those islands communities that face special difficulties that need to be addressed and, where possible, sustainable solutions must be designed and implemented in the light of available resources.

It has been observed that due to their geographical isolation, inadequate transport infrastructure and irregular transport services, island communities have greater difficulty in traveling and must bear relatively higher transport costs. It is recommended that the provision of a socially desirable minimum level of transport be a priority for all islands, and that there should be all-year round services to ensure the importation of essential supplies, to provide reliable means of exporting island produce and to gain access to the mainland for social, employment and/or education purposes. It also noted that every effort should be made to ensure that the services take into account the needs

[7] Eurisles, 1998, regional disparities: statistical indicators linked to insularity and ultra-peripherality.

of people with disabilities in terms of easy access to vessels and subsequent disembarkation (Malachy Walsh et al).

Therefore, the importance of Ferry Services to the Islands has been substantially investigated[8]. Ferry services are the lifeblood of the islands and the need for ferry services to the islands has been confirmed under numerous studies and EU regulations (Malachy Walsh et al.). Moreover, from a socio-economic perspective, the availability of regular and guaranteed services provides substantial improvement to the quality of life for islanders, reduction of relative remoteness and enrichment of the island populations which remain together and sustain the community identity. Finally, the availability of a ferry service also fosters a vibrant tourism industry which contributes to local economic well-being, and to an enhanced awareness and positive appreciation of local traditions, way of life and cultural heritage (Malachy Walsh et al.). We summarize the services offered in the EU countries[9] as follows.

The contracted services on the sea routes are operated by domestic private shipping companies or local government companies, but in the most cases are not exclusive and the authorities of each country are at liberty to procure other services on the same routes. Particularly, contracted services on the 14 routes in Ireland, are operated by domestic private shipping companies. Moreover in Denmark, at Bornholm, there are three maritime lines: a subsidized direct line with Copenhagen, one with Sweden (Ystad) which has recently become entitled to a subsidy and one with Germany (Sassnitz) the subsidy of which was abolished on foot of a complaint by a private company.

In Netherlands, there are six domestic Dutch ferry services currently operated by private operators. Finally, in Scotland, passenger, vehicle and shipping services are provided by Caledonian MacBrayne (CalMac) to twenty-two islands and four peninsulas spread over the West Coast of Scotland and in the Clyde Estuary.

On the other hand, in Sweden, inter-island maritime traffic managed by the local authorities consists essentially of ferries belonging to local government companies.

In Finland, all domestic ferry services are supported by the government either through procurement or direct ownership.

Moreover, ferry services are regulated by state (Laender) governments in Germany. The two states with island ferry services are Lower Saxony and Schleswig Holstein each with their own sets of laws governing ferry operations.

In addition, in France, links between Corsica and the French mainland are subsidized on the principle of territorial continuity. This gives preference to national companies.

In Spain, responsibility for inter-island shipping is regional, while responsibility for Balearic Islands/Mainland sea transport lies with the national authorities.

In Malta, ferry services between Malta and Gozo are subsidized by the Malta Ministry of Communications on the basis that they would not have otherwise been commercially viable. These consist of roll-on/ roll-off passenger service for foot passengers, car and car passengers as well as cargo vehicles, including hazardous cargo. Gozo Channel

[8] Luis, J. A. H. (2002), Robert Y Canava (2004), Rutz, W. O. A. and Coull, J. R. (1996), Werner, O., Rutz, A. and Coull, J. R. (1996).

[9] Steer Davies Gleeve, (2005) "Research and advice on risk management in relation to the subsidy of Ferry services", The Scottish executive, Enterprise, transport & lifelong Learning & Malachy Walsh and Partners In association with Posford Haskoning, Raymond Burke Consulting, McCaig Watson and Seosamh Mac Donnacha, (2004), "Review of Certain Subsidized Ferry Services to the Islands" Department of Community, Rural and Gaeltacht Affairs.

Shipping Line, a wholly owned subsidiary of Malta Government Investments Limited, provides the services.

In general, subsidized ferry contracts are granted to operators on the condition of compliance with the marine regulations, health and safety rules, and by-laws set by the relevant local authorities. The subsidized services are not set up as PSOs in all the cases. This is because some of the islands generate a small number of trips per year; hence they are not bound by the European Union's public procedures.

The operators, on the other hand, must meet minimum levels of service, in light of the following:

- Determine the frequency (and regularity/continuity) of the services on offer, with the minimum of weekly services for continental transport. For some services, the itinerary must include a fortnightly inter-island stop;
- Operators could also be obliged to undertake a stop at all islands, or provide intra-island services, if it is found that the islands are not regularly and sufficiently supplied;
- Demonstrate that they are able to run the services;
- Run services and itineraries with the scheduled regularity, for at least one year
- Comply with EU legislation;
- Utilize exclusively crew formed by nationals or from a European member state.

Additional services, however, are provided according to demand. Passenger and cargo charges are often capped, with further discounts for island residents. To avoid the problem of defining "residents", the principle of a direction-based subsidy is applied in France.

4. The case of Greece

Law 2932 of 2001 was brought in to open up the Greek maritime market and to bring ferry services in Greece in line with the Regulation. This legislation, however, was deemed to conflict with the spirit of the Regulation. A number of disputes have emerged and been resolved with regards to the interpretation of the Regulation and its application in the Greek context. Following the opening of the maritime cabotage market, services are open to any operator that complies with the EU requirements and provides some basic documentation confirming their capacity to provide services. No requirements are set with regards to the service provided apart from the fact that the vessel has to comply with the safety requirements of the ports they serve, and be in service for a minimum of 10 months per year. Domestic ferry services in Greece continue to be provided by domestic operators who keep serving their traditional routes, albeit in some cases also expanding to others.

Greece has a significant number of islands to which services would not be commercially viable, especially in the winter months, due to the low demand. Currently, lifeline[10] services operate on some 70 island routes. Islands for which applications to provide free market services are not made in January of each year are those islands for which services need to be provided in a Public Service Contract (from now on, PSC) framework under annual contracts. The requirements for the provision of the PSC services are as follows:

- Frequency: determined number of crossings per week.
- Period of operation: the ferry has to be in operation for a minimum of 10 months a year.
- Back-up: A replacement ferry has to be made available for the remaining two months a year. The details of this have to be made available at the stage of submitting the bid.
- Ferry standards: passenger capacity, car parking surface, total length. A standard type of service is required (i.e. conventional passenger-car ferry).
- Safety: The ferry has to be suitable for safe use of the port facilities and safe service provision.
- Fares: A maximum fare is set for standard/ economy class. Also discounts are imposed for concessionary travelers.
- Vehicle age: Needs to be less than 35 years.

The national budget supplies the funding for the lifeline services. In addition, a 3% surcharge is imposed on all other ferry services (not subsidized) in order to subsidize the thin routes (Malachy Walsh et al).

Even though, the liberalisation has been in effect for many years, some components should be rationalized. A notable characteristic, in Greek coastal market, is the vessel age. The findings of the GILA report ((2006), Chlomoudis, C. et al.) pinpoint the fact that during the previous decade (1996-2005), as shown in the Table 1, the average age of the vessels, has been substantially decreased and as a consequence the average speed of ships has constantly increased. Particularly, according to the Figure 1, 35,7% of ships are 12 years old or less, 20,8% of ships are between 13-24 years old and the 43,5% of ships are between 25 and 35 years old.

Table 1: ship characteristics of Greek coastal market.

CHARACTERISTICS	**YEAR**				
	1996	**1998**	**2000**	**2002**	**2005**
Average Speed	14.9	15.3	15.8	17.4	19.1
Average Age	23.6	21.7	20.4	18.4	18.1

(The GILA report for ‘The Inter-Island Transport problem, counter measures and the role of local Authorities’ ((2006), Chlomoudis, C. et al.)

[10] Lifeline services are defined as those that would not have been provided by the free market.

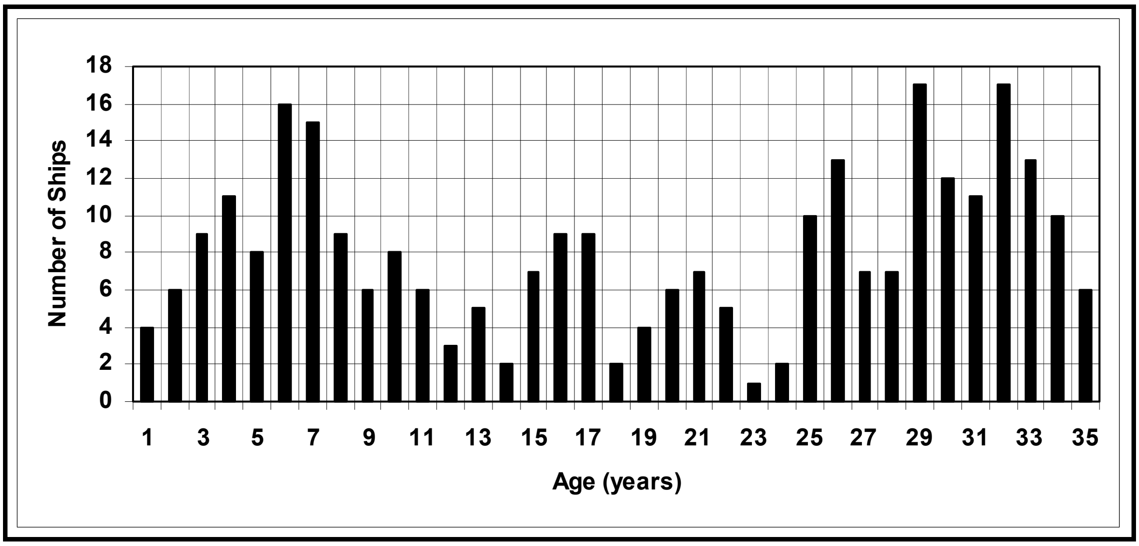

Figure 1: age distribution of Greek coastal markets' ships.
(The GILA report for 'The Inter-Island Transport problem, counter measures and the role of local Authorities' ((2006), Chlomoudis, C. et al.)

National legislation imposes the withdrawal of ships over 35 years old (Figure 2), Act 187/1973 (30 years old from the year of 2008, Act 2932/2001). This legislation is instituted only by Greece and creates several problems at the Greek Coastal market, which entitles 350 ships of all modes and is operated through main and secondary sea routes in a total number of 1500 connections between 40 and 100 mainland and island ports respectively.

Figure 2: ships withdrawal.
(The GILA report for 'The Inter-Island Transport problem, counter measures and the role of local Authorities' ((2006), Chlomoudis, C. et al.)

Moreover, as highlighted in the table 2, the demand for coastal services, during the winter, decreases over 80% as regards to summer period. Approximately 40% of inter-island connections for the summer period do not operate during the winter period. This situation is aggravated by the fact that because a lot of scheduled trips are cancelled in the winter due to the bad weather conditions. For this reason, the Greek Government should provide PSO for all the sea routes during the year and particularly for these sea routes which are not commercially viable.

Table 2: ferry services frequency (August vs January).

Service Frequency	*August*	*January*
At least once a day	50,5%	40 %
2-6 times per week	40 %	47 %
Once a week	9,5 %	13 %
TOTAL	100 %	100 %

(The GILA report for 'The Inter-Island Transport problem, counter measures and the role of local Authorities' ((2006), Chlomoudis, C. et al.)

The government subsidizes PSO routes under the following financial sources, the national budget and the 3% increase of the ticket price.

The maximum amount of subsidy is calculated on a basis of route distance. As shown in the Figure 3, the amount of the subsidy for the year 2003-2004 was 37,1 million Euros. The overall amount of subsidy between the years 2001-2006 was 100 million Euros with an annual growth rate of 74%.

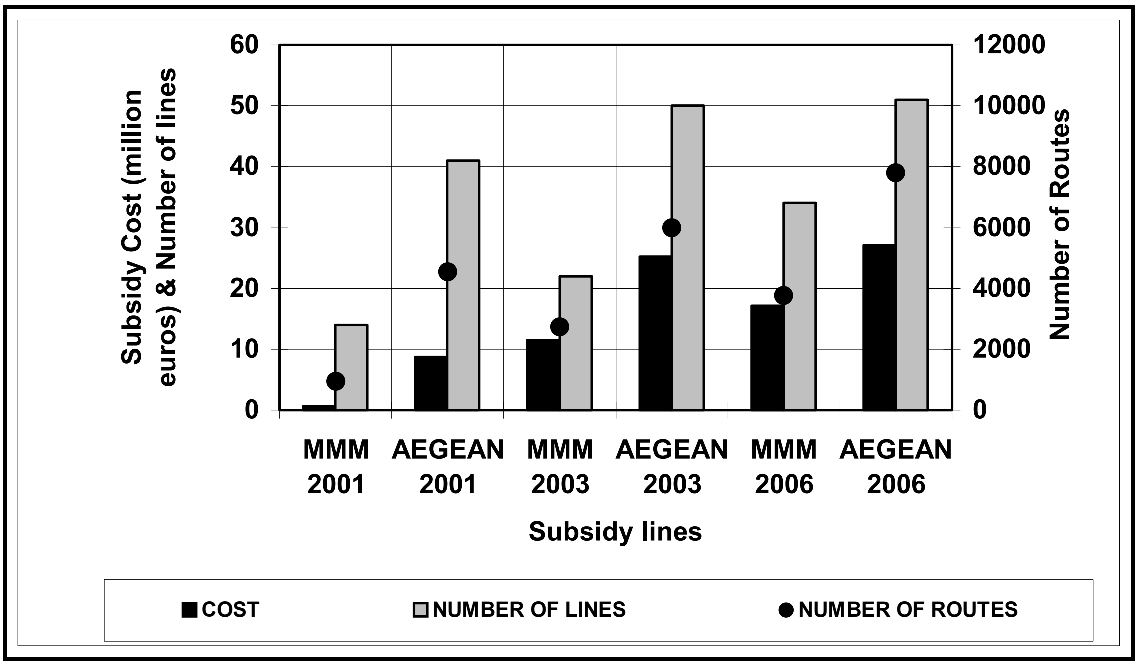

Figure 3: subsidy cost and routes (Ministry of Mercantile Marine Vs. Ministry of Aegean).
(The GILA report for 'The Inter-Island Transport problem, counter measures and the role of local Authorities' ((2006), Chlomoudis, C. et al.)

5. A "road map" towards liberalisation completion

The government undertakes the responsibility to fulfil the needs of the society in areas of public goods such as transportation, energy provision, water supply and telecommunications and the concept of Universal Service is closely attached to that mission. In particular, Universal Service is a defined set of services at a certain quality, available to any user without geographical constraints with an affordable price.

Potential aims of the Universal Service, as they are set at the GILA report ((2006), Chlomoudis, C. et al.) are as follows:

- Availability and adequacy of the provided services
- Basic access to goods and services
- Affordability of the prices
- Provision for quality of service
- Adequate public service obligations routes.

According to the international experience (Malachy Walsh et al.), in network industry markets, universal services obligations were established in order to facilitate the smooth transfer to liberalisation. In particular, EU determined Universal Service Obligations for all the key infrastructure networks as telecommunications, energy and surface transport systems.

For the Central and Western Europe, where island regions are absent, Universal service obligations addressed the needs for the rail and road networks and inland waterways.

On the contrary, EU has not yet proceeded to determine Universal service obligations in sea passenger transport. However, the necessity for such a provision covering the needs of the Greek territory is obvious since its implementation would safeguard an adequate level of public service provision for the islands and address a fully defined negotiation platform with the stakeholders (maritime companies, local authorities, proper ministries).

The solution to the transportation problem in the island regions and the implementation of an effective and efficient strategy requires an integrated approach which while considering the political and social environment at the European and national levels would search for solutions parameters on the basis of their impact on the economic development of the islands and social welfare of the islanders.

The two basic parameters - economic development and political aiming- define generally the external or socio-economic environment. In order to investigate further the relationship between the external environment and the transportation system, the external environment has to be categorized into several dimensions, such as, territorial, institutional, economical and social.

As Chlomoudis, C. et al. mention, there are many alternative scenarios which address concrete lines of prerequisites and which highlight different financially viable solutions. In particular, these scenarios address a 'road map' with distinct choices under specific circumstances. However, due to the absence of a clearly formulated development national policy but also due to the lack of a clear political mandate from EU, the determination of alternative scenarios has to be formulated through the acknowledgement of two different measures and their corresponding indicators:

- Economic development using the GDP indicator and
- The social state status applying the social state principles indicator or the global market indicator.

Regarding the first measure, two scenarios are considered: one characterized by low and one with a high rate of development. Particularly, high development is considered when the GDP indicator is 2.5% and more, while low development is considered when the GDP indicator is less than the European average, of 2.5%.

Regarding the second measure, it reflects on the choices and decisions in the European level that affect significantly the national development policies. The two basic hypotheses are:

- "Reinforcement of social state". In the EU, the dominating opinion is that the reinforcement of the social state constitutes the primary strategy. Therefore, the policy of enhancement of the European *aquis communautaire* is supported and thus particular emphasis is provided to persons with disabilities, isolated regions, and reduction of insularity and improvement of social cohesion.
- "Emphasis on globalisation". In the EU, the liberal approach is dominating. This means that extra measures for the enhancement of competitiveness of the European economy in the demanding new global market would be required.

According to the Table 3, there are four major scenarios that emerge for EU development:

Table 3: "Road map" scenarios.

	High Development	***Low Development***
Reinforcement of Social State	High Development - Social State Scenario	Low Development - Social State Scenario
Emphasis on globalization	Globalization - High Development Scenario	Globalization - Low Development Scenario

(The GILA report for 'The Inter-Island Transport problem, counter measures and the role of local Authorities' ((2006), Chlomoudis, C. et al.)

(a) *High Development - Social State Scenario*

This scenario is characterized by high economic growth while placing strong emphasis in the social state. Its basic hypothesis considers that under specific circumstances (adequacy of means), the state (government) supports financially selective interventions targeting equal development throughout the European regions.

(b) *Low Development - Social State Scenario*

It is a scenario which is also characterized with high emphasis in the social state however it has to deal with low development. In this case, the state cannot support the desirable interventions for the equivalent multi central development.

(c) *Globalization - High Development Scenario*

It is a scenario characterised by high development with emphasis on competitiveness and globalization. Its central idea is that due to the shifts in the priorities of the European policy, there will be a substantial deceleration in the regional development.

(d) *Globalization - Low Development Scenario*

It is a scenario characterised by low development with emphasis on globalization.

It is concerned as the most probable to be followed. In particular, this scenario consists of following:

- *Territorial dimension*: It will sustain the urban centralized pattern. The existent infrastructure is enforced through an adequate inter modal transportation system.
- *Institutional dimension*: Due to the dominant role of the market and the dependence of the entrepreneurs for boosting competitiveness, the market is deregulated and the services are provided from the private companies.
- *Economical dimension*: In order to cover the accessibility of all possible passengers, even the residents of small and isolated islands, a significant program of subsidies is applied. The cost of that program is disproportional large for the government, since the whole environment is market oriented.
- *Social dimension*: The state attempts to intervene in the market by allocating resources fairly among social groups and geographical regions. Additionally, there is limited intervention in the fairs charged (only in PSO routes) and there is no consideration for equal distance pricing with other modes of transport based on the principle of territorial continuity.

6. Our proposals towards island regions development

Therefore, in order to highlight the need to establish distinct measures and policies towards islands regions development, we state the following proposals:

Insularity

The provision of sea transport services should be treated in the same manner for all the islands. The extent of insularity in Greece, as well as in other countries, differs from island to island, since it is based on the size of an island, the distance from the mainland and various others characteristics.

Moreover, efforts towards the creation of a favourable and positive climate towards the islands should be enhanced, aiming at the issuance of a decision regarding “priority to the disparity island regions”. This decision should be incorporated to the principles and policies of EU and the normative and administrative simplification be encouraged. In addition, special financial programs should be provided for the island regions towards the period of 2007-2013.

For that reason, the development of an “Observatory for the Islands of Europe” is required, which will systematically record data regarding the transportation system, the quantity and quality of the services offered and the problems encountered. In general, the observatory will have the responsibility for tracking, monitoring and evaluating the prevailed conditions in the island transport market and providing corrective and effective solutions.

Transport equivalent

One of the most significant tasks and objective in the field of transportation policy is the issue of 'equity' or 'justice', regarding the proper distribution of both positive and negative impacts (benefits and costs) to the citizens. Decisions taken in the field of transportation have a high degree of correspondence to the 'equal' treatment of citizens:

- The quality of services affects the quality of life and the opportunities of people for travelling, finding a job etc.
- Transportation expenses cover a significant percentage of the family budget.
- Prices charged should be consisted with the cost of providing the specific service and the level of service provided.
- Transportation investments are a vital instrument for economic development. Their distribution throughout the country should be fair.

Analysing and safeguarding the appropriate 'equality' to all citizens, mainly regarding the quantity of the services offered to them and the price to be paid, is a complex task since there are different types of 'equality' and different ways of classifying citizens into certain categories (i.e. on the basis of residency, family income, other demographic characteristics, mode of transport in use).

For the islanders, it has been examined and accepted that the previous principles have not been applied in various degrees and that their "unequal" treatment is a fact. For this reason it has been proposed that based on the principle of territorial continuation, the cost for travelling to the islands should equivalent to the cost by road and rail transportation for travelling the same distance on the mainland, taking also into account the time duration and other quality provisions of the offered services. Therefore, the principle of "transport equivalent" should be adopted by the State with the aim of harmonizing the cost of travelling by sea with that of travelling by road or rail, offering the same, adequate quality of service.

In addition, there should be a policy shift in the planning and tendering of services, away from the unimodal nature of the ship, to one which considers the development of an integrated transport system with the coordination of all other modes of transport (boats, airplanes, hydroplanes and helicopters).

Public service obligations

The current policies in Greece, regarding the establishment and operation of subsidized lines and the EU policies for PSOs have not yet attributed to deliver the appropriate accessibility to the residents of the islands, even though the public funds for subsidizing the PSO lines have been increased substantially. It should also be noted though, that a significant portion of this increase has been provided by other island travellers through a ticket surcharge. In particular, the cost for all subsidized lines for the period 2005-2006 exceeded 44 million euros while from that amount 25 millions euros were covered though a 3% surcharge which was imposed on the price of the tickets. Therefore, only 20 millions euros were paid directly through the national budget, in order to safeguard the viability of the Greek islands. The Greek government has a valid point in requesting further funds from the EU in order to cover a substantial portion of this cost, since Greece is the only European country with so many islands to be served under the PSO framework.

Ship Financing

It is obvious that in the near future more ships will be required to cover the increased demand for coastal services. Therefore, the Greek Government should find the proper measure to increase the supply of ships. One noticeable instrument is that of the SPVs. A Special Purpose Vehicle would manage and coordinate all the necessary resources and assets available for providing PSOs. There are many ways to develop the SPV model. The two most acknowledged methods are:

- PPPs (Public Private Partnerships),
- PFIs (Private Finance Initiatives).

Segregation of passenger and freight transportation

Almost all of sea transport is provided by Ro-Ro ships, thus combining passenger and freight transport. The option to segregate passenger and freight transport services through the use of different types of vessels, provided there is sufficient demand, could be a realistic prospect in finding a solution to the reduced boat supply.

6. Conclusions

The European Union Regulations regarding sea transport appear to have adequate provisions to safeguard the islands interests. The key issue today does not seem to be European Union legislation itself, but the amount of influence an island community may exert in its own country over the way transport operators servicing the island are licensed, over the way their public service obligations are drafted and over the way funds are used to subsidize transport costs.

In our study, we present the various adjustment procedures and models of sea transport servicing across the EU countries. Moreover, we examined the case of Greece and pointed out the delay of the harmonisation with the EU directives. Consequently, we set a number of substantial measures and policies, regarding insularity, transport equivalent, public service obligations and distinction between passenger and freight transportation that each country in the EU, should bare in mind and apply towards its harmonisation to maritime cabotage. This set of policies, examined in Greece, could implement throughout European island regions as a common “road map”.

References

Canava, R. Y. (2004) "A qualitative of reintroducing cabotage onto New Zealand's coasts, *Maritime Policy and Management,* 31, pp. 179-198.

Chlomoudis, C., Lekakou, M., Panou, C., Papadimitriou, E., Siriopoulos, T. and Tzannatos, E. (2006) "The Inter-Island Transport problem, counter measures and the role of local Authorities", Greek Institute for Local Authorities, *Report*.

Commission of the European Communities (1994a) *Europe 2000+ Co-operation for European territorial development*, EC, Brussels/Luxembourg.

Commission of the European Communities (1996a) *First report on social and economic cohesion,* EC, Brussels/Luxembourg.

Council Regulation (EEC) No 3577/92 of December 1992 applying the principle of freedom to provide services to maritime transport within Member States (maritime cabotage) [Official Journal L 364 of 12.12.1992].

Eurisles (1998)"Regional disparities: statistical indicators linked to insularity and ultra-Peripherality", *Summary report*.

Eurisles (1999) "The Island regions and the price of intra-EU. transport of goods" *Study realized under the initiative of CPMR Islands Commission*.

Eurisles (2002) "European construction and the problem of the islands", *Study realized under the initiative of CPMR Islands Commission.*

Eurisles (2003) "Island transport and the EU. Situation in 2003", *Study realized under the initiative of CPMR Islands Commission.*

European Commission (2001) "White Paper, European Transport Policy for 2010: Time to Decide".

Giannopoulos, G. A. (2005) "Main transport challenges in South Eastern Europe, after enlargement", *European Transport,* 30, pp. 3-13.

Giannopoulos, G. A. and Aifandopoulou-Klimis, G. (2004) "Inland maritime transport in Greece after the lifting of cabotage and full liberalisation", *Transport Reviews,* 24, pp. 65-78.

Gooverden, C. V., Rietvielt P., Koelemeijer J. and Peeters P. (2006) "Subsidies in public transport", *European Transport,* 32, pp. 5-25.

Luis, J. A. H. (2002) "Temporal accessibility in archipelagos: inter-island shipping in the Canary Islands", *Journal of Transport Geography,* 10, pp. 231-239.

Malachy Walsh and Partners In association with Posford Haskoning, Raymond Burke Consulting, McCaig Watson and Seosamh Mac Donnacha (2004) *Review of Certain Subsidised Ferry Services to the Islands*, Department of Community, Rural and Gaeltacht Affairs.

Romanos, M. J (2005) "Shipping, the state and the market. The evolving role of the European Union in international and Greek shipping politics*". LSE PhD Symposium on Modern Greece*, Proceedings from the 2nd LSE PhD Symposium on Modern Greece: "Current Social Science Research on Greece", LSE, London, UK.

Rutz, W. O. A., and Coull, J. R. (1996) "Inter-Island passenger shipping in Indonesia: development of the system. Presents characteristics and future requirements", *Journal of Transport Geography* 4: 275-286.

Scottish executive, Enterprise, transport & lifelong Learning, (2005),"Invitation to tender for Northern isles Ferry services".

Steer Davies Gleeve (2005) "Research and advice on risk management in relation to the subsidy of Ferry services", *The Scottish executive, Enterprise, transport & lifelong Learning.*

Theodoropoulos, S., Lekakou, M. V. and Pallis, A. A. (2006) *"European Policies for Marittime Industry",* TIPOTHITO – George Dardanos, Athens.

Tzannatos, E. S. (2004) "Technical reliability of the Greek coastal passenger fleet", *Marine Policy,* 29, pp. 85-92.

Werner, O., Rutz, A. and Coull, J. R. (1996) "Inter-island passenger shipping in Indonesia: Development of the system", *Journal of Transport Geography,* 4, pp. 275-286.

Air traffic management in the South East European countries. Current situation and prospects

E. Matsoukis [1*], S. Poulimenakos [1]

[1] *Transport and Traffic Eng. Study Unit, University of Patras*

Abstract

This paper deals with the Air Traffic Management (ATM) situation in the South East Europe (SEE) countries. The current status indicates that the SEE air space users require increased capacity, higher flexibility and improved cost efficiency. On the other hand, some of these countries were in a war situation a few years ago, and now they are seeking a regional cooperation to ensure regional stability and prosperity towards their integration to the European Union.

The paper starts with a description of the cooperation schemes, already in operation in the area. Then the attempt to converge the ATM of the region towards the EU Single European Sky is investigated. The FABA (Functional Airspace Block Approach), a European Union (EU) initiative, applied in the SEE is presented in detail.

In the continuation, the paper reports on the definition phase of the SEE FABA initiative, which is to provide a well founded basis for the reorganisation of the SEE air space into FAB's.

It is concluded that the FAB solution stemming from the EU Single European Sky policy is feasible, highly beneficial and necessary for the SEE region development.

Keywords: Air transport; Air traffic management; Single European sky; Functional airspace blocks; South East European countries.

1. Current situation and challenges

Air Traffic forecasts indicate that the SEE (South East European) region, as elsewhere in Europe, does face a continued challenge in meeting the needs of the airspace users and the Air Navigation Service Providers (ANSPs) in the mid-to long term, unless fundamental changes are implemented in the fragmented airspaces and ATM (Air Traffic Management).

The airspace users require increased capacity, higher flexibility and improved cost efficiency which can only be achieved by a more integrated regional ATM system (see EUROCONTROL SES, Regulatory Framework, 2004).

Consequently, the SEE States should use a coherent approach in their plans, within the context of the overall development of European airspace, for developing the future

* Corresponding author: Evangelos Matsoukis (Emats.glyfada@tee.gr)

airspace structure and route network and that the approach is based on jointly accepted traffic forecasts.

Considering the different legislation of the participating States, the ECAA agreement (see ECAA, 2006) with SES provisions appears as the appropriate legal instrument to create a legal framework for enhanced and consolidated Air Traffic Management cooperation in the SEE region.

The challenge for the SEE region is in developing an adequate, common regulatory framework. This challenge is for many of the participating States to get the sufficient number of skilled human resources. Therefore, high priority should be given to the development of certified human resources required for regulatory authorities and service provision.

The provision of air navigation services must be provided by Air Traffic Controllers Officers (ATCOs) (see ESARR3, 2003) operating under standardised rules, regulations and procedures. This will necessitate a harmonisation in ATCOs working practices to all Air Traffic Control Centres (ACCs) within the region through a coordinated and standardised approach to training and licencing.

The current cooperation between some SEE States should be used as a major foundation stone for the establishment of an enhanced regional cooperation.

2. The EU cooperation schemes for the Balkans

Fostering regional cooperation is a key element in EU transport policy for the Balkans as, firstly, it opens the way for regional stability and prosperity. Secondly, it acts as a vehicle for rapprochement and, in the long term, for the integration of these countries into EU "acquis communautaire".

The cooperation initiative schemes operating in the region, in conjunction with the institutional mechanism of the EU Stability and Association Process, and the S.E. Europe Stability Pact, are best equipped for acting as catalysts to reinforce regional cohesion.

The principal inter-state regional cooperation initiatives in the area of South-Eastern Europe are:

i. the South East European Cooperation Process (SEECP).The SEECP constitutes an indigenous cooperation format that stems exclusively from the countries of S.E. Europe. Albania, FYROM, Serbia, Montenegro, Bulgaria, Romania, Turkey, Bosnia-Herzegovina, Croatia and Greece, all participate in the SEECP. Since October 2004 Moldova has participated in the SEECP, having being granted observer status. The SEECP was revived in 1996, pursuant to a Bulgarian and Greek initiative (MFA meeting in Sofia, 6-7.07.96), which led to the Crete Summit, (3-4.11.1997).

ii. the Stability Pact, SE Europe (SPSEE).The Stability Pact was established in Cologne in June 1999, and adopted by the Heads of State and Government in Sarajevo on 30 July 1999. The pact stemmed from an older French proposal, which at the end of the Kosovo war was once more put forward by Germany, with the support of the US. The aim was to create a zone of stability around F.R. of Yugoslavia as well as a means to coordinate the allocation of economic development aid to the countries of South Eastern Europe. The participating countries were divided into beneficiaries or recipients (Albania, FYROM, Bulgaria, Romania, Bosnia & Herzegovina and, since 2001, Moldova) and facilitators or donors. The Stability Pact aimed to stabilise S. E. Europe through programmes of a regional nature that have been established for this purpose.

iii. the South East Europe Cooperation Initiative (SECI). SECI comprises a consultation framework for addressing economic and environmental problems with a regional dimension. The role of SECI was to complement and strengthen existing regional initiatives and actions for transferring know-how, realising private investment and harmonising the trade legislation and policies of the countries in the region. The SECI Initiative cooperates closely with the UN Economic Commission for Europe (UNECE) and the Organisation for Security and Cooperation in Europe (OSCE). The following countries are SECI members: Greece, Albania, FYROM, Serbia and Montenegro, Bulgaria, Romania, Turkey, Bosnia-Herzegovina, Croatia, Slovenia, Hungary and Moldova. The following countries are SECI partners: US, Italy, Austria, Switzerland, Liechtenstein and the Czech Republic.

iv. the Adriatic - Ionian Initiative (AII) The AII comprises a forum for debate and cooperation between its member states (Greece, Italy, Albania, Croatia, Bosnia Herzegovina, Serbia & Montenegro) in the following sectors: economy and tourism; sustainable development and environmental protection; education and culture; combating organised crime.

The AII was instituted at the Conference on Development and Security in the Adriatic and the Ionian (Ancona, 19 & 20 May 2000), which led to the Ancona Declaration that acts more or less as the Initiatives charter.

The different regional cooperation schemes in the Balkan area, except from a general political effect, have also specific objectives to achieve, especially in the fields of institutional reforms, finance , communication and transport. It should be remembered that the EU has gone through extensive planning exercises resulting in transeuropean networks (TEN) for the European Union and the accession countries. It comes out of these exercises that there is a need for further planning in the region of SEE. It is evident that any development of a regional nature has to take full account of links with neighbouring countries, both EU Member States and candidate countries and the paneuropean policies currently under implementation. In particular for the transport sector, the results of the work undertaken in the framework of the Transport Infrastructure Needs Assessment (TINA) process will be considered as a given for the development of infrastructure in these candidate countries of the region. <u>One of the very first targets set by EU in the area of air transport is the de-fragmentation of air traffic according to the existing EU regulations, and this certainly applies for the SEE region.</u>

Taken under consideration the above mentioned initiatives for regional cooperation in the South Eastern Europe (SEE) it is evident that an attempt to re-organise the Air Traffic Management (ATM) of the region is at least an interesting and challenging exercise.

3. Air Traffic Management(ATM) situation in the EU member states

3.1 Background

In the late 1990s, the high levels of European air traffic growth, combined with the liberalisation of the air transport industry within the EU raised concerns on the ability of the Air Traffic Management (ATM) sector to meet the projected capacity requirements, needed to support increasing air traffic demand. There has been a sharp rise in delays to aircrafts.

This had major repercussions for users and placed a substantial financial burden on airlines. Delays cost to airlines in Europe a substantial amount of money, between €1.3 and €1.9 billion per year (Van Houtte B., 2004). These delays are due to a combination of factors: insufficient capacity of the air traffic control system, adverse weather, problems of airports infrastructure or problems within airline operations.

It has been estimated that air traffic will grow by 4% a year over the next 15 years, leading to a nearby doubling of figures by 2020 (see EUROCONTROL/CFMU Reports).

In response to the challenges outlined above, the European Commission published a Communication Mandate on the creation of the Single European Sky which concluded that: irrespective of the legal and economic structure of air navigation service providers, there is a need to establish an adequate overall European regulatory framework. This will ensure that services meet the necessary levels of safety, interoperability and performance, particularly if they are to continue being provided on a monopolistic basis.

Clearly, a thorough structural reform and development of the necessary regulatory framework in Pan European level would require high-level political support and the development of the necessary political and legislative control mechanisms in order to proceed towards a Single European Sky.

3.2 The High Level Group

With the support of the European Council, the European Commission constituted in 1999, a High Level Group (HLG), bringing together civilian and military representatives of Community Member States, together with representatives of Norway and Switzerland, in order to:

- Define the modalities of functioning of the Single European Sky within conditions of efficient delivery of services and with respect to public service obligations, responsibilities and safety objectives for the benefit of civil and military users;
- Examine the technical issues, implementation decisions and restructuring measures to be considered at national or European level in order to achieve such a reorganisation of routes, airspace structures and their operational usage;
- Propose harmonisation of national systems along a coherent Community approach implying central decision making processes and solidarity mechanisms, and Indicate how the Community framework can be supported by the use of the EUROCONTROL organisation in the implementation of its conclusions.

The HLG developed a final report in 2000 with the objective of undertaking a genuine reform of air traffic management. Primarily, this involved adopting a more coherent organisational role at the Community level, while at the same time accommodating expected traffic growth with more efficient use and organisation of airspace.

The HLG concluded that the main deficiencies that ATM was facing were the following:

i. Air traffic is growing and will continue to grow in the future .Unprecedented delays can only be resolved by effective measures at European level ATM in Europe is fragmented, and this results in an inefficient use of available capacity;
ii. The current system is not able to keep pace with demand;
iii. There is a shortfall of qualified controllers;
iv. EUROCONTROL does not at present have the necessary decision-making process and enforcement powers to ensure rapid improvement of the situation.

The HLG finally suggested a reform process to:

- Reinforce mechanisms to optimise the performance of European ATM as a whole;
- Establish a 'European' airspace as a single continuum, managed for overall system efficiency , ensuring sufficient access to airspace for both civil and military purposes, while respecting national security and defence requirements for the use of airspace;
- Develop a coherent ATM system design across Europe;
- Establish high-level rules at the European level for safety and system performance;
- Establish strong and independent regulators and develop a process that ensures implementation backed up by effective enforcement;
- Be consistent with the international framework and comply with the basic requirements of the EU Treaty.

3.3 The Single European Sky

The Single European Sky is a tool to re-organise airspace and ATM system, according to HLG findings, at a European rather than at a local level. The Single European Sky is essentially a regulatory initiative. It brings to the existing traditional engineering inspired and consensus-driven approach that prevails in the ATM industry a set of clear rules that define the rights but also the responsibilities of the different actors.

The main principles of the Single Sky legislation are:

- to reduce fragmentation between States, Civil and military ATM organisations;
- to introduce new technology at a Pan- European level;
- to improve Synergies between EU and Eurocontrol.

The four EC regulations which consist the founder stones for the Single European Sky are:

- Regulation (EC) No 549/2004 of the European Parliament and of the Council of 10 March 2004 **laying down the framework for the creation of the single European sky** (the framework Regulation);
- Regulation (EC) No 550/2004 of the European Parliament and of the Council of

10 March 2004 on **the provision of air navigation services in the single European sky** (the service provision Regulation);
- Regulation (EC) No 551/2004 of the European Parliament and of the Council of 10 March 2004 on **the organisation and use of the airspace in the single European sky** (the airspace Regulation);
- Regulation (EC) No 552/2004 of the European Parliament and of the Council of 10 March 2004 on **the interoperability of the European Air Traffic Management network** (the interoperability Regulation).

The Regulations provide for the development by the European Commission of more detailed implementing rules, with the assistance of EUROCONTROL and the assistance of the Single Sky Committee.

As a result, in the field of the organisation and management of the European airspace, an effort would be made to move from the traditional national approach, to a more integrated European perspective.

Clearly Member States retain responsibility for their airspace under the Chicago Convention (ICAO, 1944), but nothing prevents them from exercising this responsibility collectively under the EU umbrella and from agreeing uniform rules.

This implies the definition of a harmonised classification of airspace with a reduced number of airspace categories, together with harmonised rules on airways and sector design.

The latter will make the whole process of airspace design and management much more straightforward and open it to the creation of cross-border airspace structures.

As a more seamless airspace is being developed, this should be recognised by the establishment of the single European Upper Information Region and by the preparation of comprehensive aeronautical information to be managed and made accessible centrally.

It has long been acknowledged, technically speaking, that national boundaries are not necessarily the optimal basis for optimal air traffic management architecture.

4. The FAB methodology

The Single Sky legislation obliged Member States to analyse the way in which air traffic is managed over their territories in the light of a number of technical criteria, and to organise the provision of the relevant services on the basis of optimised zones which are called "Functional Airspace Blocks" (FAB).

Since Airspace is a common resource, the key to a more rational organisation of airspace is an integration across borders through functional airspace blocks (FABs) in order to improve capacity, enhance security and lower costs of air traffic services. These FABs should be based on operational requirements rather than existing national borders. This means that for the purpose of optimizing ATM in Europe, national boundaries will become less relevant than operational parameters such as traffic flows and cost-benefit analysis; and that there is a scope for greater efficiency and integration, thereby also reducing a number of safety hazards.

Over time it is expected to record the creation of a number of cross-border functional airspace blocks, drawing on the experience of Maastricht, NUAC and CEATS.

The establishment of functional airspace blocks is considered as a "bottom up" process, driven by the Member States concerned and by their service providers. They know best where there are opportunities for operational improvements.

There is no intention to direct this process towards a specific result, but technical and financial support is to make sure that all opportunities are being pursued diligently and that a forum for discussion is provided, if difficulties arise between States. It is also fair to say that we are putting a lot of faith in the ability and willingness of States and industry to come up with good proposals is being put.

The airspace regulation (Regulation (EC) No 551/2004); foresees in its article 5, that FABs shall respect the following criteria:

(a) be supported by a safety case;
(b) enable optimum use of airspace, taking into account air traffic flows;
(c) be justified by their overall added value, including optimal use of technical and human resources, on the basis of cost-benefit analyses;
(d) ensure a fluent and flexible transfer of responsibility for air traffic control between air traffic service units;
(e) ensure compatibility between the configurations of upper and lower airspace;
(f) comply with conditions stemming from regional agreements concluded within the ICAO, and
(g) respect regional agreements in existence on the date of entry into force of this Regulation, in particular those involving European third countries.

FABs will only be established following consultation with all interested parties, including other Member States and the Commission. Whilst mutual agreement between the Member States concerned is fundamental to the establishment and running of a FAB, in the event of any dispute, parties may ask the Single Sky Committee for advice. Up to now, based on the strong political willingness of European Commission to implement SES, numerous initiatives have started to be examined by different states according to the bottom up approach .

In this procedure, a number of critical issues will be examined in a case by case procedure. A feasibility study for each initiative will support the decision making of the participating States. Up to now the starting point for all economic evaluations is <u>the current ATM fragmentation cost</u>. All on going initiatives intend to reduce this cost, but it remains to each feasibility study to prove how much ,where and when this cost will be reduced.

5. The SEE FABA initiative

At the European Councils in Lisbon and Feira in the first half of 2000, it was confirmed that the Stabilisation and Association Process is the centrepiece of the Unions' policy towards the region and that the countries concerned are potential candidates for membership in the European Union (Matsoukis, E. and Poulimenakos, S. 2001)

In the Thessaloniki Summit in June 2003, the EU reiterated its unequivocal support to the European perspective of the Western Balkan countries and to prepare them for integration into European structures and ultimately for membership into the European Union.

In 2005 the Western Balkan countries, along with Bulgaria, Romania and the United Nations Interim Administration Mission in Kosovo (herewith referred to as UNMIK) agreed to create with the European Union, a European Common Aviation Area (ECAA) and thereby accepted to align their national aviation legislation to the complete set of EU legislation in the area of Civil Aviation .

The Air Traffic Management (ATM) part of the ECAA agreement extends the Single European Sky (SES) Regulations to the contracting parties. With regard to the airspace it shall be reconfigured into Functional Airspace Blocks (FABs) with a view to achieving maximum capacity and efficiency of the ATM system while maintaining a high level of safety.

On 1st April, 2005 the Directors General of Civil Aviation of Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Romania, Serbia and Montenegro, The Former Yugoslav Republic of Macedonia and UNMIK ("the Directors General") decided to examine opportunities with a view to implementing the Functional Airspace Block Approach (FABA) in South East Europe in consistency with the rest of European Air Traffic Management network. Consequently, a Working Group was created to support this task: the South East Europe Functional Airspace Block Approach Working Group (see EUROCONTROL SEE FABA Final Report, 2006).

In February 2006 the report of this Working Group, "The Opportunities for the Application of the Functional Airspace Block Approach in South East Europe", was presented to the Directors General of Civil Aviation of the SEE-FABA States. In accordance with the conclusions and recommendations of this report, the Directors General approved the commencement of the Definition Phase (SEE FABA Initiative).

The SEE FABA States have also stressed the possibility to extend the participation on an equal basis in the Definition Phase to other States of the Region that express interest and commitment to be part of the FAB approach.

During the working process representatives from the European Transport Workers´ Federation (ETF) and from Civil Air Navigation Services Organisation (CANSO) were consulted on appropriate subjects in the light of their role as social dialogue partners for the European Commission. The authors of this paper participated in the SEE FABA Working Group.

The Working Group identified a number of key issues from the EUROCONTROL report that need to be addressed and studied more thoroughly in a Definition Phase of SEE FABA. These major areas of interest are: operational, technical, military, legal/institutional/ organizational, economic/financial and social.

The regional concept embedded in the implementation of FAB(s) supports:

i. A regional approach to the airspace solving some of the restrictions put on the airspace structure due to national considerations. The FAB(s) will – on a regional basis – ensure that the airspace structure follows the traffic flows and meets the airspace users' requirements;
ii. The development and implementation of common operational concepts and procedures creating a harmonised operational environment and harmonised safety management in the region covered by the FAB(s);
iii. The planning, deployment and operation of a CNS/ATM infrastructure from a regional perspective, based on common functionality requirements and harmonised (or integrated) operational procedures in the ATM system;
iv. The enhancement of capacities through seamless operation within flexible use of the airspace;

v. Practical use of the existing infrastructure in the region and of the ongoing initiatives.

The SEE FAB approach should be based on operational requirements. With an Operationally Driven Approach, an optimal ATS Route Network can be structured with the SEE Region from which the Sector Families can be identified and grouped appropriately into a number of potential FAB scenarios on the basis of which the economical, social, military, institutional and other elements need to be considered.

Although the SES Regulation called for FAB(s) to be established above FL 285 (28.500 feet), this should not be seen as a limiting boundary. The totality of the airspace should be considered to ensure that not only are the benefits extended to all areas but also that an arbitrary division level does not then become another artificial boundary (see also Willis, P., Matsoukis, E. and Poulimenakos, S. 2001).

In order to identify the best suitable scenario for the SEE region, the operational requirements should be developed with the active participation of all SEE FABA States and taking into account neighbouring FAB(s) and cooperation with neighbouring States.

The institutional aspects related to the establishment of the FAB(s) are very complex and must be addressed very early in the establishment of the FAB(s). Basically, the SEE FAB(s) institutional model(s) will then have to be developed and agreed during the Definition Phase (2006-2009).

The economic and charging mechanism related to the establishment of the FAB(s) touches upon the State sovereignty (fiscal regime) and economic interest of participating ANSPs and the air space users while it should support the operational improvements created by the FAB(s). Therefore, the economic model will have to be developed and agreed during the Definition Phase.

The harmonisation and integration process will mainly be driven by common Operational Concepts to be developed before actual activities are performed. At the same time, it will be important to safeguard the present investments in CNS/ATM infrastructure in order to reduce the cost base of providing Air Traffic Services (ATS).

Some planning and implementation activities will also be initiated, such as:

- Development of a common harmonised and integrated CNS/ATM infrastructure to support the provision of ANS;
- Identification of other areas where a common approach should entail early benefits.

As an example the implementation of regional ground-to-ground backbone network is an area that was recommended by the SEE FABA WG as a candidate for early implementation. Implementing a regional harmonised and integrated approach towards safety supporting the implementation of FAB(s) in the region should provide considerable benefits to the airspace users and to the organisations regardless of the chosen institutional framework for the future provision of services by reducing the existing fragmentation in the ATM environment.

Social dialogue in FAB environment shall be more “international minded” with the main goal to foster cooperation instead of competition. Therefore, it is of the highest importance to initiate coordinated bottom up social dialogue at national, at regional and at European level.

Human resource planning and training are important supporting elements of FAB(s). Clear and visible short term benefits can be gained in these areas if training programmes are targeted towards FAB(s) implementation and at the same time promoting the best ATM practices and sharing scarce instructor resources. The training focus on FAB(s) will bring a very much needed fresh input to the current ATC training and it will also pave the way for more harmonized ATCO performance matched with appropriate social benefits.

6. The proposed strategy for implementation-the definition phase (2006-2009)

The main objective of the SEE-FABA Definition Phase is to deliver all the relevant implementation proposals in order to provide by March 2009 the SEE-FABA States with a well founded basis for the reorganisation of their airspace into FAB(s), in line with the strategic objectives identified by the SEE FABA WG and detailed in the previous paragraph.

The Definition Phase will consider the operational, technical, legal and institutional requirements for the development of Functional Airspace Block Proposals and apply the FAB Key Issues in order to make implementation proposals for the establishment of Functional Airspace Blocks within the airspace of the SEE-FABA States.

To ensure coherency and harmonisation, the Definition Phase will take into consideration Regional and pan-European developments. In order to identify the best suitable scenario for the SEE region, the operational requirements shall be developed with the active participation of all SEE FABA States and taking into account neighbouring FAB(s) and cooperation with neighbouring States, and will fully associate all the relevant actors concerned. In that process, the cooperative approach will not interfere with questions of national sovereignty, in conformity with the principles applied to the Single European Sky, and will let full responsibility to the States for the endorsement of any agreement.

In line with the Chicago Convention, the obligations pertaining to each individual State will be taken into account when establishing FAB(s) as they will have to be addressed in the State level FAB Agreement (or at lower level) to clearly identify the rules applicable to the FAB airspace.

The SEE FAB approach will be based on operational requirements regardless of existing boundaries, which are the primary criteria of a FAB. Therefore the operationally driven concept should be based on the delineation of Air Traffic Control Sector Families based on the air traffic flow complexity which will then be put together in FAB(s) based on secondary criteria such as economic, legal, institutional, and technical.

The key objectives that will be considered during the definition process should be:

i. The identification of domains for which early benefits can be obtained and propose implementation solutions;
ii. The development of operationally driven FAB scenarios;
iii. The address of the FAB Key Issues and the development of solutions for application within the region;
iv. The conduct of a Business Case;
v. The conduct of a Safety Case;
vi. The development of SEE-FABA implementation scenarios with a proposal of a preferred option;
vii. The development of a SEE-FABA Implementation Plan;

viii. The Definition Phase of the SEE-FABA project shall be based on the most suitable working arrangements combining the political and financial support of the European Commission and the Stability Pact for South Eastern Europe with the technical expertise of EUROCONTROL under the governance of the SEE-FABA States.

The implementation should be based on:

- A step-by-step process that will combine an operationally driven development period with political decision milestones. The implementation should be composed of a Definition Phase and an Implementation Phase;
- Initiation of projects leading to enhanced cooperation in the SEE region;
- An evolutionary and incremental process for the transition from the existing situation to the establishment of FAB(s);
- Practical use of ongoing regional initiatives combined with a pan-European perspective;
- National skills and competencies.

The following initiatives would provide clear benefits for all parties on short and mid-term:

- By initiating the process of setting up a Common Interconnected Regional Network (CIRN);
- By harmonising of ATC procedures and ATCO operational training;
- By developing the cooperation between the National Regulatory Authorities.

The key factor for success of such a project can only be guaranteed if all the relevant actors of the ATM domain are fully involved in the various works to be conducted. In this regard, the SEE FABA definition phase will ensure participation of the following stakeholders:

- SEE-FABA States: Albania, Bosnia & Herzegovina, Bulgaria, Croatia, Romania, Serbia, Montenegro, The Former Yugoslav Republic of Macedonia and UNMIK
- SEE-FABA Supporting States: Greece and Italy;
- SEE-FABA Supporting Institutions: European Commission – Directorate General for Transport and Energy (TREN) Stability Pact, EUROCONTROL, North Atlantic Treaty Organisation (NATO), International Civil Aviation Organisation (ICAO);
- National Authorities (civil and military);
- Air Navigation Service Providers (civil and military);
- Airspace Users (civil and military);
- Airports (see Matsoukis, E. and Poulimenakos, S. 1999);
- Social Partners and Industrial Representatives;
- Airspace User Organisations(such as IATA, IACA, ERA, EBAA and IAOPA;
- Social Partners;
- Other interested and ready to comit States;

7. What is at stake/benefits-costs of the European ATM/CNS system due to fragmentation

Given the lack of consistent and complete data at national level, this work intends to provide only a way to make a first estimate of the orders of magnitude. Up to now, we can only talk about the cost of fragmentation to the existing ATM Organization and the strong political will of EU and air space users to rationalize it through the SES. Actual financial data and benefits for each of the future FAB Initiatives will be part of a case-by-case feasibility and business study.

The results of this work comes from the Performance Review Committee/EUROCONTROL Study "The Impact of Fragmentation in European ATM/CNS" prepared by Helios Economics and Policy Services (April 2006).

The adverse impact of fragmentation in the European Air Traffic Management and Communications, Navigation and Surveillance (ATM/CNS) system has long been a concern of all stakeholders. The recent Single European Sky legislation is intended to have a major impact on fragmentation; in particular it will foster airspace rationalization and restructuring, consolidation of facilities, and harmonization of systems and procedures.

"Fragmentation" in the ATM system is defined as referring to the division of air navigation service provision into smaller decision-making or operational units than would result from considerations of optimum scale. In Europe, this has mainly arisen from the organization of ANS at the state level.

However, fragmentation also arises through smaller than optimal operational units within national ANSPs. These units may have become sub-optimal, for example, as changes in the technology of service provision have raised the optimum size of a centre upwards.

In the current European ATM system, there are 69 ACCs, of which 47 operate with 10 sectors or fewer at maximum configuration. In Europe, the average centre operates 9 sectors at maximum configuration; the average US centre 37 sectors. Furthermore, the FAA is currently examining the possibility of further consolidation among its existing 21 centres.

The first step in assessing the costs of fragmentation in the European ATM system were to understand what the costs of that system were and how they arose.

Two major resources are required to run an ATM/CNS system: the capital assets, and the staff, with associated non-staff operating costs.

This work is using all available resources to estimate the order of magnitude both of the total costs of the existing system in each of these categories, and how those total costs arose in terms of the areas of focus of the study: ACCs and ATM systems; CNS, and associated support.

The task of assessing the total costs of the system and how these were divided up in this way was a surprisingly difficult one. ACE data, CRCO data and ANSPs' Annual Reports gave an incomplete picture, particularly concerning how costs and asset values were split between ACCs and ATM systems, CNS and support. We estimated costs based on physical infrastructure and industry estimates of unit capital replacement and operating costs, with validation against Annual Reports (EUROCONTROL, 2006, The Impact of Fragmentation etc).

We assessed the capital replacement costs for the enroute European ATM/CNS system as some €10 billion (2003 data), as shown in the first column of Table 2 below; annual operating costs amounted to some €3500m.

Total annual costs (around €4340m) were obtained by annualising the capital costs over eleven years (a typical ratio for ANSPs between asset acquisition costs and the sum of depreciation and finance costs).

Table 1: Costs for European enroute ATM/CNS (2003 data).

	Capital replacement costs	***Annual operating costs***	***Total annual costs***	
COM	€560 m	€60 m	€110 m	2.5%
NAV	€230 m	€10 m	€30 m	0.7%
SUR	€3,000 m	€210 m	€500 m	11.5%
ACCs & ATM systems	€4,900 m	€2,100 m	€2,500 m	57.6%
Associated support	€1,000 m	€1»100m	€1,200 m	27.7%
Total	**€9,690 m**	**€3,480 m**	**€4,340 rn**	**100%**

It is evident that the costs of NAV and, to a lesser extent, COM, are rather small compared with other aspects of the system. The costs of fragmentation in these areas are therefore likely to be small compared with those in SUR and ATM operations. As a consequence, efforts in trying to identify the adverse impact of fragmentation were concentrated on the latter areas.

The major areas where fragmentation was expected to have an adverse impact are summarised in Table 2 , along with the associated order of magnitudes of costs:

Table 2: Summary of fragmentation costs.

	Cause of fragmentation	***Annualised costs***	***% of cost of fragmentation***
Common issues	Piecemeal procurement (mainly ATM systems)	€30m-€70m	**14%**
	Sub-optimal scale in maintenance and in-service development (mainly CNS)	€10m-€15m	
	Fragmented planning	€60m-€120m	
ACCs	Economies of scale in ACCs (operating costs)	€37Qm-€460m	**53%**
	Economies of scale in ACCs (capital cost)	€105m-€140m	
	Constrained sector design (flight efficiency benefits)	€50m-6100m	
ATM systems	Lack of common systems (operating costs)	€150m-€215m	**23%**
	Lack of common systems (capital costs)	€30m-€90m	
	Increased coordination at interfaces	€IGm-€20m	
CNS infrastructure	Optimum location of en-route navaids	€3m-€7m	**4%**
	Overprovision of secondary radar	€15m-€60 m	
Associated support	Economies of scale in training, administrative costs and R&D	€40m-€100m	**6%**
	Total costs of fragmentation	€880m-€1400m	**100%**

The overall order of magnitude of the costs of fragmentation in the European enroute ATM/CNS system was estimated at some €880m - €1,400m. Although there is inevitably some uncertainty around such estimates, this is undoubtedly a significant amount; it represents around 20-30% of the annual en-route costs (see Table 2).

The main components of the cost of fragmentation are (1) many ACCs are below the optimum economic size, (2) duplication of bespoke ATM systems (including piecemeal procurement and sub-optimal scale in maintenance and in service development), and (3) duplication of associated support (training, administration, and R&D).

A number of other issues were identified as important but were not associated with. Fragmentation costs arising from sub-optimal ACC size could arise both through the economies of scale in the centres themselves, and through lost flight-efficiency as airspace and route design is organised in a fragmented way.

The typical size of a European ACC is much smaller than those in the US, as shown in Figure 1. In the current European ATM system, there are 69 ACCs, of which 47 operate with 10 sectors or fewer at maximum configuration. In Europe, the average centre operates 9 sectors at maximum configuration; the average US centre 37 sectors. Furthermore, the FAA is currently examining the possibility of consolidation among its existing 21 centres.

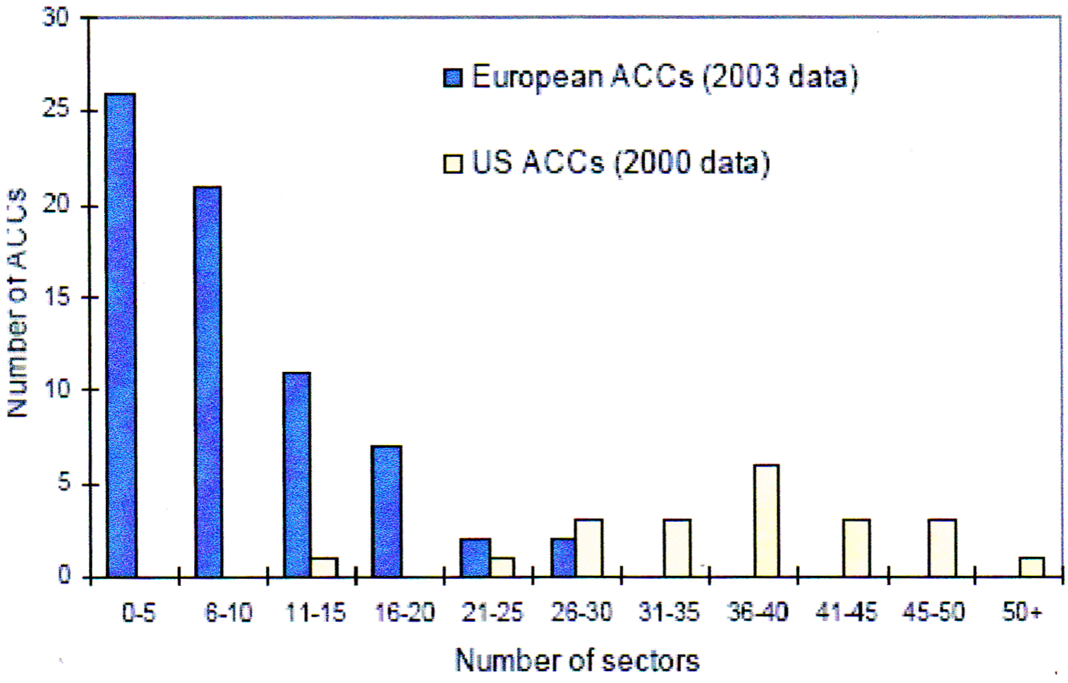

Figure 1: Centre sizes in Europe and the US.

Recent consolidation exercises in the UK and Germany have produced or will produce ACCs that can accommodate 50-60 sectors, making them comparable with the scale of the larger US ACCs. There is a general view that consolidation to scales greater than this may not necessarily lead to further economies of scale – difficulties in coordination and in providing contingency backup may start to outweigh fixed cost savings.

Consequently, in order to estimate the cost savings associated with economies of scale in ACCs in a defragmented system, the costs of current ACCs, estimated according to our assumptions concerning which were fixed and which variable, were compared to the costs estimated for a system in which ACCs operate with around 25 sectors at maximum configuration. This figure is well short of the maximum scale currently operating in Europe. However, such larger scales may only be practically feasible in the densest regions; 25 sectors is a conservative assumption for a non-fragmented system in a FAB environment, operating under SES regulations.

It is important to note that these savings, while associated with ACC fragmentation, do not necessarily require ACCs to be consolidated to achieve them. They could be achieved, for example, by collaborative route and sector design, as should be achieved in a Functional Airspace Block (FAB).

8. ATM systems

An important cost of fragmentation could arise from each centre having different software from others. If a number of ACCs use common ATM systems (identical or very similar systems), further economies can be made on the software element of those systems. Clearly, the lack of such common systems becomes less of a problem as the number of centres is reduced, and therefore cannot be treated in isolation from the costs arising from fragmentation of centres.

The cost arising from this fragmentation was estimated as €5m-€15m capital costs per centre, and €2.2m-€3.2m a year operating costs per centre.

This cost could be applied either to the existing number of centres, or to the number of centres in a defragmented system, as shown in Figure 4.

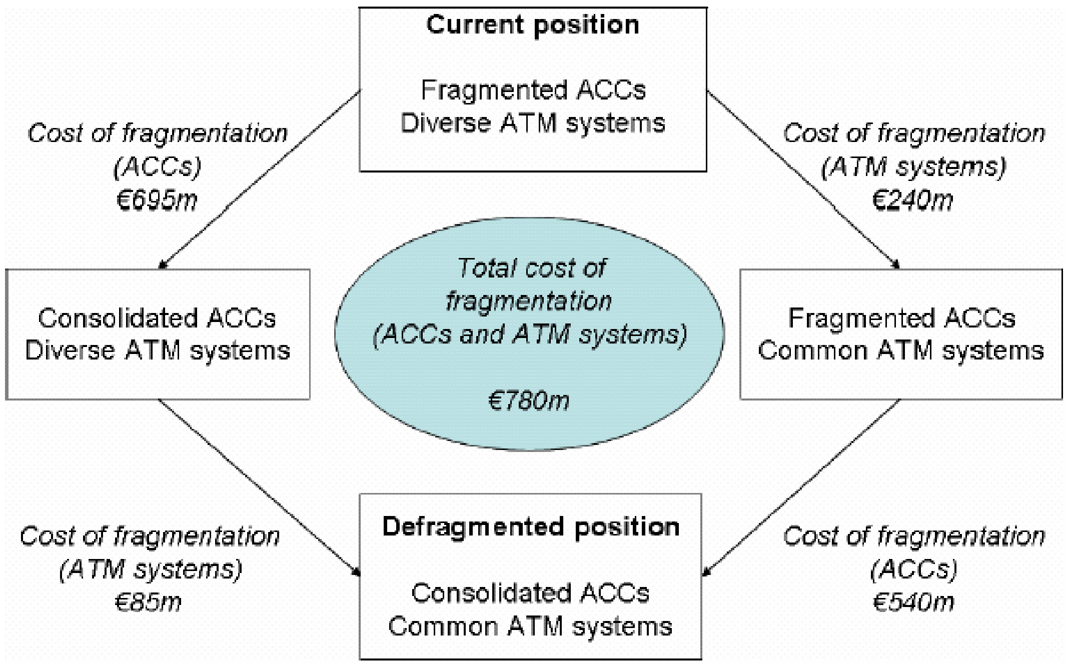

Figure 2: Fragmentation costs of ACCs and ATM systems.

The total cost of fragmentation in ACCs and ATM systems taken together is €780m a year (for clarity, the midpoint of the ranges is shown). This results from reduction of the costs of purchasing and developing the systems, and from the reduction of the number of centres. The top left of the diagram shows the costs of fragmentation obtained by considering ACC consolidation without convergent ATM systems (€695m). This is the figure obtained from the arguments in the previous paragraphs .

The costs of retaining diverse ATM systems in this smaller number of ACCs is relatively small (€85m). However, we considered it clearer to apply the costs of fragmentation in ATM systems to the existing number of ACCs, as in the top right of the figure; a move towards common systems is likely to happen on shorter timescales and with lower transition costs than consolidation of centres. This assigns higher cost to fragmentation of ATM systems (€240m) and a correspondingly lower amount (€540m) to fragmentation of ACCs.

This is an area of fragmentation where costs could rise substantially if no action is taken, because of new requirements for certification of ATM systems. Although a number of initiatives to reduce fragmentation in this area are being undertaken, there is a feeling amongst stakeholders that more needs to be done.

An additional benefit of common systems would be improved coordination at interfaces between centres. The PRU US-Europe study identified inter-centre coordination as possible reason for lower ATCO productivity in Europe. There is higher workload for hand-over in European centres than US centres. In a nonfragmented system, with increased interoperability, the inter-centre coordination time required was assumed to be half that required at present.

9. Conclusion

The coming years will require a sustained effort to improve safety,increase capacity and enhance the efficiency of the European ATM system by reducing fragmentation.

Fragmentation of European en-route ATM/CNS carries a high cost – around €900m - €1,400m annually, perhaps 20-30% of annual European ATM costs; the main causes of this cost were the fragmentation of ACCs, and of ATM systems; other important costs of fragmentation resulted from fragmented planning, piecemeal procurement, and duplication of support activities.

The Single European Sky provides the tools to meet that challenge in a Paneuropean perspective. The EC provides a strong political will to implement the Single European Sky, both to member states and accession states, by using among others the FAB approach for ATM organisation.

This paper demonstrated how the FAB approach stemming from the EU Single European Sky Regulations could also be beneficial, and that the definition and implementation of SEE FAB(s) should be considered necessary for the SEE region.

There are both operational and institutional benefits:

- The operationally driven concept could be based on the delineation of air traffic control sector families based on the air traffic flows complexity and bringing them together, rationalising the airspace organisation and management in the area. Fragmentation of ACCs also has an impact through constrained sector design. Sector design is currently constrained by national and, to a lesser extent, ACC boundaries. A defragmented system would allow improved sector design by removing the constraints of national boundaries. This would allow improved routing through the defragmented airspace and hence greater flight efficiency.

- The costs of procurement of ATM/CNS systems could be reduced by procurement specifications common across ANSPs and the reduced need for adaptation of systems to bespoke designs, as systems become interoperable.
- A highly fragmented ATM/CNS system, where there are many different types of equipment at several locations, requires more maintenance and development staff than a limited number of equipment types.
- The Communications/ Navigation/Surveillance systems harmonisation and possible integration process will mainly be driven by common operational concepts while safeguarding the present investments in CNS/ATM infrastructure in order to increase the cost base providing services. The implementation of regional ground-to-ground backbone network called Common Interconnected Regional Network (CIRN project) is recommended as a candidate for early implementation.
- Fragmentation costs arising from sub-optimal ACC size could arise both through the economies of scale in the centres themselves, and through lost flight-efficiency as airspace and route design is organized in a fragmented way.

The institutional aspects related to the establishment of the FAB(s) are very complex and must be addressed very early in the establishment of the FAB(s).

Regarding unit charges, a more stepwise approach is expected: first a transitional phase with different national unit rates and then a single unit rate after 3/5 years or in accordance to the incoming EC Charging Scheme Regulation.

Implementing a regional harmonised and integrated approach towards safety supporting the implementation of FAB(s) in the region provides benefits regardless of the chosen institutional framework for the future provision of services. Discussing the safety issues the SEE FABA could initiate an initial project concerning the establishment of cooperation between NSAs in the region.

In this context, conditions for improvement would include:

- rapid progress in the reform process at the national and regional level;
- introduction of new market mechanisms;
- modern operating practices and sound infrastructure financing and management;
- strengthening of institutions through improved legislative and regulatory frameworks;
- restructuring of the sectors;
- introduction of new technologies;
- development, within a regional framework, of a sound investment programme based on the co-operation among the countries of the region and with the Member States and Candidate Countries of the European Union.

References

ECAA (2006) European Common Aviation Area Agreement, Annex 1, part B ("Air traffic Management").
ESARR 3 (2003) "Use of Safety Management Systems by ATM Service Providers", Ed. 1.0.
ESARR 4 (2003) "Risk Assessment and Mitigation in ATM", Ed. 1.0.
EUROCONTROL (2000-2004) CFMU Reports.
EUROCONTROL (2004) SES Regulatory Framework.
EUROCONTROL (2005) "Medium-Term Forecast, Flight Movements 2005-2011", Vol. 1, Ed. 2.
EUROCONTROL (2006) "SEE FABA Final Report".
EUROCONTROL (2006) "Report on the SEE FABA Definition Phase".
EUROCONTROL(2006) "The Impact of Fragmentation in European ATM/CNS", commissioned by the PRC/ EUROCONTROL and prepared by the Helios Economics and Policy Services.
EUROPEAN PARLIAMENT, REGULATION (EC) No 2096 (2005) "Common Requirements for the Provision of Air Navigation Services".
EUROPEAN PARLIAMENT, REGULATION (EC) No 549 (2004) "Framework for the Creation of the Single European Sky (the Framework Regulation)".
EUROPEAN PARLIAMENT, REGULATION (EC) No 550 (2004) "Provision of Air Navigation Services in the Single European Sky (the service provision Regulation)".
EUROPEAN PARLIAMENT, REGULATION (EC) No 551(2004) "Organization and Use of the Airspace in the Single European Sky (the airspace Regulation)".
EUROPEAN PARLIAMENT, REGULATION (EC) No 552 (2004) "Interoperability of the European Air Traffic Management Network (the Interoperability Regulation)".
EUROPEAN UNION (2004) SES Regulatory Framework.
ICAO (1944) Convention on the International Civil Aviation, Chicago, USA.
Matsoukis, E. and Poulimenakos, S. (1999) "Determining the realistic prospects for private financing and developing Airports", *3rd A.C.I. Europe Airport Economic Conference* "Challenges at the Turn of the Century", Pisa, Italy.
Matsoukis, E. and Poulimenakos, S. (2001) "Air transport at the MEDA area" in *Proceedings of the International Conference on Airports and Air Transport* " Evolutions in the 21st century", Patras, Greece.
Van Houtte, B. (2004) Head of ATM UNIT/DG TREN "THE SES".
Willis, P., Matsoukis, E. and Poulimenakos, S. (2001) "RNAV GNSS Instrument Procedures for Airports" in *Proceedings of the International Conference on Airports and Air Transport* "Evolutions in the 21st century", Patras, Greece.

Abbreviations and Acronyms

ACC=Area Control Centres
AII=Adriatic - Ionian Initiative
ANSPs=Air Navigation Service Providers
ATM=Air Traffic Management
ATM/CNS=Air Traffic Management and Communications, Navigation and Surveillance
ATCOs=Air Traffic Controllers Officers
CANSO=Civil Air Navigation Services Organisation
CEATS=Central European Air Traffic Services
CRCO=Central Route Charging Office
EBAA=European Business Aviation Association
ECAA=European Common Aviation Area
ERA=European Regions Airline Association
ETF=European Transport Workers´ Federation
EU=European Union
FAA=Federal Aviation Administration
FABA=Functional Airspace Block Approach
FAB=Functional Airspace Blocks
HLG=High Level Group

ICAO=International Civil Aviation Organisation
IATA=International Air Transport Association
IACA=International Air Carrier Association
IAOPA=International Council of Aircraft Owner and Pilot Association
NATO=North Atlantic Treaty Organisation
NUAC= Nordic Upper Area Control Centre
OSCE=Organisation for Security and Cooperation in Europe
SECI=South East Europe Cooperation Initiative
SEE=South East European
SEECP=South East European Cooperation Process
SES=Single European Sky
SPSEE=Stability Pact of the South East Europe
TEN=TransEuropean Networks
TINA=Transport Infrastructure Needs Assessment
TREN=European Commission – Directorate General for Transport and Energy
UN=United Nations
UNECE=UN Economic Commission for Europe
UNMIK=United Nations Interim Administration Mission in Kosovo

European Transport \ Trasporti Europei n. 37 (2007): 35-61

A systematic comparison of continuous and discrete mixture models

S. Hess [1], M. Bierlaire [2*], J. W. Polak [1]

[1] *Centre for Transport Studies, Imperial College London*

[2] *Transport and Mobility Laboratory, School of Civil and Environmental Engineering*
Ecole Polytechnique Fédérale de Lausanne

Abstract

Modellers are increasingly relying on the use of continuous random coefficients models, such as Mixed Logit, for the representation of variations in tastes across individuals. In this paper, we provide an in-depth comparison of the performance of the Mixed Logit model with that of its far less commonly used discrete mixture counterpart, making use of a combination of real and simulated datasets. The results not only show significant computational advantages for the discrete mixture approach, but also highlight greater flexibility, and show that, across a host of scenarios, the discrete mixture models are able to offer comparable or indeed superior model performance.

Keywords: Discrete choice; Mixture models; Random parameters; Discrete and continuous distributions.

1. Introduction and context

Allowing for variations in behaviour across decision makers is one of the most fundamental principles in discrete choice modelling, given that the assumption of a purely homogeneous population cannot in general be seen to be valid.

The typical way of allowing for such variation is through a deterministic approach, linking the taste heterogeneity to variations in socio-demographic factors such as income or trip purpose. While appealing from the point of view of interpretation (and

[*] Corresponding author: Michel Bierlaire (michel.bierlaire@epfl.ch)

especially for forecasting), it is often not possible to represent all variations in tastes in a deterministic fashion, for reasons of data quality, but also due to inherent randomness in choice behaviour. For this reason, random coefficient structures, such as the Mixed Multinomial Logit (MMNL) model, which allow for random variations in behaviour across respondents, have an important advantage in terms of flexibility. In general, such models have the disadvantage that their choice probabilities take on the form of integrals that do not possess a closed form solution, such that numerical processes, typically simulation, are required during estimation and application of the models. This greatly limited the use of these structures for many years after their initial developments. Over recent years, gains in computer speed and the efficiency of simulation based estimation processes (see for example Hess et al. 2006) have however led to increased interest in the MMNL model in particular, by researchers and, to a lesser degree, also practitioners.

Despite the improvements in estimation capability, the cost of using the MMNL model remains high. While this might be acceptable in many cases, another important issue remains, namely the choice of distribution to be used for representing the random variations in tastes across respondents. Here, there is a major risk of producing misleading results when making an inappropriate choice of distribution, as discussed by Hess et al. (2005).

In this paper, we explore an alternative approach, based on the idea of replacing the continuous distribution functions by discrete distributions, spreading the mass among several discrete values. Mathematically, the model structure of a discrete mixture (DM) model is a special case of a latent class model (cf. Kamakura and Russell, 1989; Chintagunta et al., 1991), assigning different coefficient values to different parts of the population of respondents, a concept discussed in the field of transport studies for example by Greene and Hensher (2003) and Lee et al. (2003). Latent class approaches make use of two sub-models, one for class allocation, and one for within class choice. The former models the probability of an individual being assigned to a specific class as a function of attributes of the respondent and possibly of the alternatives in the choice set. The within class model is then used to compute the class-specific choice probabilities for the different alternatives, conditional on the tastes within that class. The actual choice probability for individual n and alternative i is given by a sum of the class-specific choice probabilities, weighted by the class allocation choice probabilities for that specific individual.

The latent class approach is appealing from the point of view that it allows for differences in sensitivities across population groups, where the group allocation can be related to socio-demographic characteristics. However, in practice, it may not always be possible to explain group allocation with the help of a probabilistic model relating the outcome to observed variables. This situation is similar to the case where taste heterogeneity cannot be explained deterministically, leading to a requirement for using random coefficients models. As such, in this paper, we explore the use of models in which the class allocation probabilities are independent of explanatory variables, and are simply given by constants that are to be estimated during model calibration. As such, the resulting model exploits the class membership concept in the context of random coefficients models, with a limited set of possible values for the coefficients.

Thus far, there have seemingly been only two main applications of this approach in the area of transport research, by Gopinath (1995), in the context of mode choice for freight shippers, and by Dong and Koppelman (2003), who made use of discrete mixtures of MNL models in the analysis of mode choice for work trips in New York, referring to the resulting model as the "Mass Point Mixed Logit model". Although the properties of DM models have been discussed by several other authors (e.g. Wedel et al., 1999), the model structure does not seem to have received widespread exposure or application, despite its many appealing characteristics. Given this observation, part of the aim of this paper is to re-explore the potential advantages of DM models, with the hope of encouraging their more widespread use. Additionally, the paper aims to offer a systematic comparison of the performance of discrete and continuous mixture models across a host of situations, making use of simulated data.

The remainder of this paper is organised as follows. The next section sets out the theory behind DM models. Section 3 presents a case study using real data, while Section

4 uses four different simulated datasets in a systematic comparison of discrete and continuous mixture models. Finally, Section 5 presents the conclusions of the paper.

2. Methodology

Let x_{in} be a vector defining the attributes of alternative i as faced by respondent n, and let β be a vector defining the tastes of the decision maker, where, in purely deterministic models, β is constant across respondents. Furthermore, let x_n be a vector grouping together the individual vectors x_{jn} across the alternatives contained in the choice set of respondent n. We can then define $P_n\left(i \mid x_n, C_n, \beta\right)$ to give the choice probability of alternative i for individual n, with a choice set C_n, conditional on the observed vector x_n, and for given values for the vector of parameters β (to be estimated).

In a discrete mixture context, the number of possible values for the taste coefficients β is finite. Here, we divide the set of parameters β into two sets; $\overline{\beta}$ represents a part of β containing deterministic parameters, while $\widehat{\beta}$ is a set of K random parameters that have a discrete distribution. Within this set, the parameter $\widehat{\beta}_k$ has m_k mass points $\widehat{\beta}_k^j$, $j = 1,...,m_k$, each of them associated with a probability π_k^j, where we impose the conditions that[1]

$$0 \leq \pi_k^j \leq 1, \quad k = 1,\ldots,K; j = 1,\ldots,m_k, \tag{1}$$

and

$$\sum_{j=1}^{m_k} \pi_k^j = 1, \quad k = 1,\ldots,K. \tag{2}$$

[1]These constraints can be avoided by setting $\pi_i = \frac{e^{\alpha_i}}{\sum_{j=1}^{J} e^{\alpha_j}}$, where α_j with $j = 1,\ldots,J$ are estimated without constraints. While avoiding the need for constraints, this formulation becomes highly non-linear and difficult to handle in estimation.

For each realisation $\widehat{\beta}_1^{j_1},...,\widehat{\beta}_K^{j_K}$ of $\widehat{\beta}$, the choice probability is given by

$$P_n\left(i \mid x_n, C_n, \beta = \langle \overline{\beta}, \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle\right), \quad (3)$$

where the deterministic part of $\overline{\beta}$ stays constant across realisations of the vector $\widehat{\beta}$.

The unconditional choice probability for alternative i and decision maker n can now be written straightforwardly as a mixture over the discrete distributions of the various elements contained in $\widehat{\beta}$ as:

$$\begin{aligned} & P_n\left(i \mid x_n, C_n, \overline{\beta}, \widehat{\beta}, \pi\right) \\ & = \sum_{j_1=1}^{m_1} \cdots \sum_{j_K=1}^{m_K} P_n\left(i \mid x_n, C_n, \beta = \langle \overline{\beta}, \widehat{\beta}_1^{j_1}, \ldots, \widehat{\beta}_K^{j_K} \rangle\right) \pi_1^{j_1} \cdot \ldots \cdot \pi_K^{j_K}, \end{aligned} \quad (4)$$

where $\overline{\beta}$, $\widehat{\beta}$ and π ($\pi = \langle \pi_1^1, \ldots, \pi_1^{m_1}, \ldots, \pi_K^1, \ldots, \pi_K^{m_K} \rangle$) are vectors of parameters to be estimated in a regular maximum likelihood estimation procedure. An obvious advantage of this approach is that, if the model (3) used inside the mixture has a closed form, then so does the DM itself.

In this paper, we focus on the simple case where the underlying choice model is of MNL form; however, the form given in equation (4) is appropriate for any underlying model (e.g. Nested Logit). The approach can easily be extended to the case of combined discrete and continuous random taste heterogeneity, by partitioning β into three parts; the above defined parts $\overline{\beta}$ and $\widehat{\beta}$, and an additional part $\widetilde{\beta}$, whose elements follow continuous distributions[2]. This however leads to a requirement to use simulation, as with all continuous mixture models.

Finally, a treatment of repeated choice observations analogous to the standard continuous mixture treatment, with tastes varying across individuals, but not across observations for the same individual, is made possible by replacing the conditional choice probabilities for individual observations in equation (4) by probabilities for sequences of choices, and by using the resulting DM term inside the log-likelihood function.

[2]This approach can then also be used to include error components for correlation or heteroscedasticity.

Several issues arise in the estimation of DM models. Firstly, the non-concavity of the log likelihood function does not allow the identification of a global maximum, even for discrete mixtures of MNL. Given the potential presence of a high number of local maxima, performing several estimations from various starting points is advisable. Also, it is good practice to use starting values other than 0 or 1 for the π_k^j parameters. Secondly, constrained maximum likelihood must be used to account for constraints (1) and (2). Thirdly, clustering of mass points (for example around the mode of the true distribution) is a frequent phenomenon with DM models, and the use of additional bounds on the mass points can be useful, based on the definition of (potentially mutually exclusive) a priori intervals for the individual mass points. In this context, a heuristic is needed to determine the optimal number of support points in actual applications. Some of these issues have caused problems in past applications of DM models, see for example Dong and Koppelman (2003). Given these problems in past research, the results of our analysis should ideally be reconfirmed in future work.

For the purpose of this analysis, the model was coded into BIOGEME (Bierlaire, 2003), where various constraints on the parameters can be imposed to address the issues described above. This also allows modellers to test the validity of specific assumptions, such as a mass at zero for the VTTS, a concept discussed for example by Cirillo and Axhausen (2006).

3. VTTS case study

In this section, we present the findings of an analysis making use of real world data. We first give a brief description of the data in Section 1, before looking at model specification in Section 2. The estimation results are presented in Section 3.

3.1 Data

The study presented here makes use of Stated Preference (SP) data collected as part of a recent value of time study undertaken in Denmark (Burge and Rohr, 2004).

Specifically, we make use of data describing a binary choice process for car travellers, with alternatives described only in terms of travel cost and travel time. This may be seen as a simple design for a SP survey. However, this is now a standard SP format for VTTS studies in Europe (e.g. Mackie et al., 2003). In any case, the design of the survey should have no direct impact on the comparison between DM and MMNL models in this analysis.

Each respondent was presented with 9 choice situations, including one with a dominating alternative. After eliminating the observations with a dominating alternative, as well as additional data cleaning (removing non-traders and respondents who did not choose the dominating alternative[3]), a sample of 13,386 observations from 1,723 respondents was obtained. This equates to 3,037 observations from 392 commuters, 1,081 observations from 142 respondents travelling for education purposes, 1,767 observations from 230 people on shopping trips, 3,155 observations from 404 people travelling to visit friends or relatives, 1,752 observations from 224 general leisure travellers and 2,594 observations from 331 respondents travelling for other purposes.

To allow us to gauge the stability of the results, multiple random subsamples of around 80% of the original sample size were generated for each of the above listed six purpose segments[4].

3.2 Model specification

The models used in this paper were estimated in log-WTP (willingness to pay) space, avoiding the effect of heterogenous scale (cf. Fosgerau and Bierlaire, 2006), while allowing us to represent random variations in the VTTS without the issue of calculating

[3]The dominating alternative was both cheaper and faster. Respondents observed to choose the slower and more expensive alternative were deemed not to have correctly understood the survey, and were removed from the analysis. The number of non-traders in the data was fairly low, but their removal was in line with previous studies using the same data (cf. Fosgerau, 2006).

[4]The selection was performed at the individual-specific level, rather than the observation-specific level.

the VTTS on the basis of separate randomly distributed coefficients for travel time and travel cost. This was found to be the best specification for this dataset by Fosgerau (2006).

Details of the specification are given by Fosgerau (2006). In summary, we let T_i and C_i define the time and cost attributes of alternative i, and rearrange the data such that $T_1 > T_2$ and $C_1 < C_2$, i.e., the first alternative is slower but cheaper than the second alternative. By further setting $\alpha_{LV} = \ln(VTTS)$, we get the following utility functions:

$$U_1 = \lambda \ln\left(-\frac{\Delta_C}{\Delta_T}\right) + \varepsilon_1 \tag{5}$$

and

$$U_2 = \lambda \alpha_{LV} + \varepsilon_2, \tag{6}$$

where $\Delta_C = C_1 - C_2$ and $\Delta_T = T_1 - T_2$, while ε_1 and ε_2 give the usual type I *iid* extreme value terms. The scale λ is estimated in addition to α_{LV}, and, with travel costs given in Danish Krona (DKK) and travel times given in minutes, the actual VTTS in DKK per hour is obtained by $60 \cdot \exp(\alpha_{LV})$. The specification set out above can now be used in a standard discrete choice framework, with either a fixed estimate for α_{LV}, or with random variation across respondents.

3.3 Model results

During the analysis, four different types of model were estimated on the data; a simple MNL model, a MMNL model using a Normal distribution, and two DM specifications, one with two support points, DM(2), and one with three support points, DM(3)[5]. In the MMNL and DM models, the repeated choice nature of the data was taken into account by specifying the likelihood function with the integration

[5]Models with more than three support points collapsed back to the more basic specifications.

(respectively summation in the DM models) outside the product over replications for the same respondent.

Each of these models was estimated across the six population segments, with 10 different random subsamples for each segment. Given this wealth of results, we presented detailed results only for a single subsample for shopping trips (Section 1), and give summary results for the remaining five population segments (Section 2). It should be said that, across segments and models, the results were very stable across subsamples, where a similar observation was also made with slightly smaller subsamples, allowing for smaller overlap.

3.3.1 Detailed results for shopping trips

The results for the various models estimated on the data for shopping trips are summarised in Table 1. Several differences arise across models in the presentation of the results. As such, for the MNL model, only α_{LV} and λ are estimated. For the MMNL model, α_{LV} follows a Normal distribution, with mean $\alpha_{LV,\mu}$ and standard deviation $\alpha_{LV,\sigma}$. For the two DM models, the value of α_{LV} is spread across several support points $\alpha_{LV,k}$ with associated probabilities $0 \le \pi_k \le 1$, such that $\sum_{k=1}^{K} \pi_k = 1$, with $K = 2$ and $K = 3$ in DM(2) and DM(3) respectively.

The table also shows the calculated VTTS. For the MNL model, the mean VTTS is simply obtained through $60 \cdot \exp(\alpha_{LV})$. However, for the three mixture models, the non-linearity in the exponential means that a different approach is required. With $\alpha_{LV} \sim N(\mu_\alpha, \sigma_\alpha)$ in the MMNL model, the actual VTTS follows a log-normal distribution with mean $\mu_{VTTS} = \exp\left(\mu_\alpha + \frac{\sigma_\alpha^2}{2}\right)$ and standard deviation $\sigma_{VTTS} = \mu\sqrt{\exp\left(\sigma_\alpha^2\right) - 1}$. Both μ_{VTTS} and σ_{VTTS} can then be multiplied by 60 to obtain hourly values. For the DM models, a slightly different approach was used. As such, with K support points $\alpha_{LV,k}$ and associated probabilities π_k, a sequence of draws was generated that contained $\pi_k \cdot N$ points with a value equal to $\exp\left(\alpha_{LV,k}\right)$, with $k = 1,...,K$. The sample mean and standard deviation from this sequence were then used as estimates of the mean and standard deviation for the actual VTTS. For the results presented here, the value of N was set to $100,000$, beyond which no visible differences were observed for σ_{VTTS}. Finally, along with the results for individual subsamples, the table also shows some overall measures, namely the average of the adjusted ρ^2 measure, the average estimation time, and the average for μ_{VTTS} and σ_{VTTS} (together with a standard deviation of this mean across subsamples).

Table 1: Estimation results on Danish shopping data.

	Model	***MNL***	***MMNL***	***DM(2)***	***DM(3)***
	Final LL	-880.96	-880.96	-849.65	-845.40
	adj. ρ^2	0.1036	0.1036	0.13433	0.136613
	Estimation time (s)	1	1	75	1
$\alpha_{LV,\mu}$	est.	-1.1100	-1.0800	-	-
	asy. t-ratio	-14.20	-11.30	-	-
$\alpha_{LV,\sigma}$	est.	-	0.8950	-	-
	asy. t-ratio	-	8.75	-	-
$\alpha_{LV,1}$	est.	-	-	0.5410	0.7770
	asy. t-ratio	-	-	1.89	2.26
π_1	est.	-	-	0.2130	0.1560
	asy. t-ratio $^{(i)}$	-	-	3.73	2.69
$\alpha_{LV,2}$	est.	-	-	-1.4700	-1.7800
	asy. t-ratio	-	-	-13.40	-5.40
π_2	est.	-	-	0.7870	0.4550
	asy. t-ratio $^{(i)}$	-	-	13.80	1.53
$\alpha_{LV,3}$	est.	-	-	-	-0.9100
	asy. t-ratio	-	-	-	-2.09
π_3	est.	-	-	-	0.3890
	asy. t-ratio $^{(i)}$	-	-	-	1.35
λ	est.	0.8380	1.0300	1.0100	1.0300
	asy. t-ratio	11.50	12.10	12.30	12.10
	Mean VTTS (DKK/hour)	19.77	30.41	32.81	34.29
	VTTS standard deviation	-	33.70	36.55	41.86

The first observation that can be made from Table 1 is that all three mixture models offer significant improvements in model fit over the base MNL model. Given the structural differences between the continuous and discrete mixture models, the comparison between these models is carried out using the adjusted ρ^2 measure rather than the log-likelihood function. Here, we can see that DM(2) offers the best performance, ahead of DM(3) and the MMNL model. While the model with three support points obtains slightly better model fit than the model with two support points, the gains are not large enough to be significant when taking into account the additional cost in terms of the number of parameters. In other words, the model with three support points is not able to retrieve significant amounts of additional heterogeneity when compared to the model with two support points. This can partly be seen as a reflection of the success of the model with two support points, but is also an illustration of the difficulties of estimating models with more than two support points, as alluded to in Section 2.

The next observation relates to the much lower estimation cost for the DM(2) model, with an average estimation time of one second, compared to seventy-five with the MMNL model. This much lower estimation cost would give the DM models a significant advantage in the case of larger datasets, where the absolute estimation times would be more substantial. Furthermore, the estimation time for the MMNL model was in this case kept low through the use of only 250 Halton draws in the estimation.

In terms of substantive results, the mean VTTS measures obtained by the three mixture models are significantly higher than the point estimate obtained with the MNL model. This is at least partly a result of the asymmetrical distribution of the VTTS in the mixture models. While there are also some differences between the three mixture models in the estimates for μ_{VTTS}, these are much smaller than the difference when compared to the MNL estimates. Finally, the estimate for σ_{VTTS} is much higher in the DM(3) model, while the estimate for the DM(2) model and the MMNL model are very similar.

3.3.2 Other results

Table 2 summarises the results for the various models estimated on the remaining five purpose segments. The results are very similar to those obtained on the data for shopping trips. As such, all three mixture models outperfom the MNL model, where the best performance is consistently obtained by the DM(2) model. Again, the DM(3) model is not able to retrieve significant levels of additional taste heterogeneity to warrant the estimation of two additional parameters. In fact, the estimates for μ_{VTTS} and σ_{VTTS} are almost universally equivalent across the two models[6]. As in the case of shopping trips, the advantages of the DM models in terms of estimation time are again very significant, across all five purpose segments. Finally, while there are almost no differences in the estimates for μ_{VTTS} between the three different mixture models (where the estimates are again significantly higher than those for the MNL models), the estimates for σ_{VTTS} are now lower in the DM models, something that was not the case in the shopping segment.

Table 2: Summary of results for commuters, education trips, leisure trips, other purposes and visits.

		Commuters	*Education*	*Leisure*	*Other*	*Visit*
MNL	adj. ρ^2:	0.1017	0.1282	0.1102	0.0888	0.1007
	estimation time (s):	1	1	1	1	1
	Mean VTTS (DKK/hour):	29.08	29.32	26.40	22.73	23.82
MMNL	adj. ρ^2:	0.1263	0.1599	0.1395	0.1127	0.1294
	estimation time (s):	131	51	74	107	127
	Mean VTTS (DKK/hour):	39.51	37.28	37.62	34.76	35.83
	Std.dev. VTTS	35.90	29.24	38.43	39.85	39.61
DM(2)	adj. ρ^2:	0.1291	0.1609	0.1433	0.1156	0.1337
	estimation time (s):	2	1	1	2	2
	Mean VTTS (DKK/hour):	39.78	36.96	37.03	34.43	37.36
	Std.dev. VTTS	30.50	24.01	28.03	29.17	36.28
DM(3)	adj. ρ^2:	0.1279	0.1576	0.1412	0.1142	0.1326
	estimation time (s):	4	1	2	3	4
	Mean VTTS (DKK/hour):	39.78	37.04	37.03	34.43	37.18
	Std.dev. VTTS	30.50	24.36	28.03	29.17	36.16

[6]It is worth noting that, with the exception of the education segment, the adjusted ρ^2 measure is higher for the DM(3) model than for the MMNL model.

4. Simulated data case studies

The application presented in Section 3 has shown the potential advantages of using a discrete mixture approach. However, it is clearly impossible to generalise these results, which could well be specific to the data at hand. For this, a systematic comparison between discrete and continuous mixture models is required; this is the topic of this section, which presents the findings of four case studies making use of simulated data.

In each of the four case studies, the generation of the data is based on the Danish VOT data used in the case study described in Section 3. Specifically, we use $10,776$ observations from $1,347$ respondents, and generate choices based on the attributes used in the original survey data. For each of the four different true models, ten sets of choices are generated for each observation, allowing us to gauge the stability of results across different samples. With the exception of the first case study, where the MMNL model had slightly higher variation across samples, the results were relatively stable, such that we only present results for the first subsample in each case[7].

Unlike in the case study described in Section 3, we now work in preference space, with separate coefficients for travel time and travel cost. In each case, the travel cost coefficient is kept fixed while some random distribution is used for the travel time coefficient, with distributions chosen so as to give realistic ranges for the VTTS distribution. Finally, the data generation was in each case carried out under the assumption of constant tastes across replications for the same individual, and the same approach was later used in model estimation.

In the first two case studies, the true model is a discrete mixture, while in the final two case studies, the true model is a continuous mixture. This allows us to gauge the relative difficulties of the two types of model in dealing with data for which the other model type is more appropriate.

Before proceeding to the discussion of the results, it should be noted that all MMNL models presented here make use of a Normal distribution. Attempts to use alternative continuous distribution functions, such as Johnson's S_B, did not lead to consistent

[7]Detailed results are available from the first author on request.

results on the data used here. While the findings from this analysis are thus limited to a comparison between a discrete mixture and a normal mixture, it should be remembered that the vast majority of MMNL studies make use specifically of this Normal distribution, such that the results are still relevant.

4.1 Case study 1: discrete mixture with two support points

The first case study makes use of data generated with the help of a discrete mixture model with two mass points for β_T, at -1 and -0.5, with probabilities of 0.25 and 0.75 respectively. The travel cost coefficient is fixed at a value of -1, such that we obtain a mean VTTS of 37.5 DKK per hour with a standard deviation of 13.33 DKK per hour.

Table 3 presents detailed results for the first of the ten subsamples generated for this case study. In addition to a basic MNL model, we estimated a MMNL model using a Normal distribution and a discrete mixture model with two support points on this dataset[8]. In both cases, we allowed for random variations in β_C as well as β_T. Consistent with the true model, no variations were observed for β_C in the discrete mixture model, labelled $\text{DM}(2)_A$, such that a second model, $\text{DM}(2)_B$, was estimated, in which β_C was kept fixed.

In a comparison between the three remaining models, MNL, MMNL and $\text{DM}(2)_B$, we observe that the discrete mixture model outperforms the continuous mixture model, which in turn outperforms the MNL model. In terms of estimation time, $\text{DM}(2)_B$ has clear advantages over the MMNL model, and the higher estimation cost when compared to MNL is well justified on the basis of the improvements in model performance. All three models offer very good performance in retrieving the mean VTTS, while the two mixture models additionally offer good performance in the estimation of the standard deviation.

[8]No further gains in model performance were obtained by allowing for more than two support points.

Table 3. Estimation results for first simulated dataset.

	MNL		*MMNL*		*$DM(2)_A$*		*$DM(2)_B$*	
Final LL	-4565.42		-4122.22		-4007.05		-4007.05	
par.	2		4		8		5	
adj. ρ^2	0.3885		0.4476		0.4625		0.4629	
est.time (s)	2		234		17		6	
	est.	asy.t-rat.	est.	asy.t-rat.	est.	asy.t-rat.	est.	asy.t-rat.
β_T	-0.4081	-36.16	-	-	-	-	-	-
$\beta_{T,\mu}$	-	-	-0.6409	-36.28	-	-	-	-
$\beta_{T,\sigma}$	-	-	0.1553	10.31	-	-	-	-
$\beta_{T,1}$	-	-	-	-	-0.5050	-40.99	-0.5050	-40.99
$\pi_{T,1}$	-	-	-	-	0.7258	50.22	0.7258	50.22
$\beta_{T,2}$	-	-	-	-	-1.0231	-40.81	-1.0231	-40.81
$\pi_{T,2}$	-	-	-	-	0.2742	18.97	0.2742	18.97
β_C	-0.6424	-34.09	-	-	-	-	-1.0083	-42.20
$\beta_{C,\mu}$	-	-	-1.0613	-36.64	-	-	-	-
$\beta_{C,\sigma}$	-	-	0.2071	9.66	-	-	-	-
$\beta_{C,1}$	-	-	-	-	-1.0083	-12.20	-	-
$\pi_{C,1}$	-	-	-	-	0.3035	0.00	-	-
$\beta_{C,2}$	-	-	-	-	-1.0083	-24.03	-	-
$\pi_{C,2}$	-	-	-	-	0.6965	0.00	-	-
μ_{VTTS}	38.11		37.81		38.50		38.50	
σ_{VTTS}	-		12.75		13.75		13.75	

A final point deserves some special attention. As mentioned above, we initially allowed for random variation in β_C as well as β_T. The estimation of the first discrete mixture model, DM(2)_A, offered no evidence of such heterogeneity, such that the model was replaced by DM(2)_B. However, for the continuous mixture model, MMNL, we retrieved significant heterogeneity for β_C as well as for β_T, despite the fact that β_C was kept fixed in the generation of the data. This offers clear evidence of confounding; by being unable to retrieve the correct patterns of heterogeneity for β_T, the MMNL model explains part of the remaining error in the model through heterogeneity in β_C. As such, while the model is able to correctly retrieve the mean and standard deviation of the VTTS, it does so by incorrectly indicating a variation across respondents in the sensitivity to changes in travel cost.

The findings from Table 3 are confirmed by a graphical analysis of the shape for the distribution of β_T in Figure 1, where this comparison is made possible by the fact that the mean estimate for β_C is essentially equal to -1 in all models.

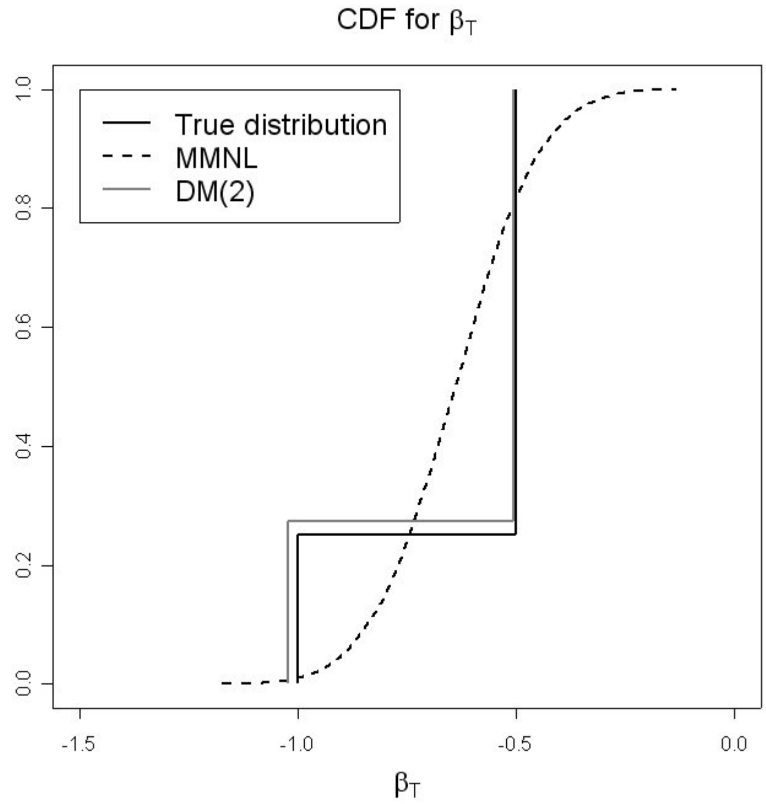

Figure 1: Cumulative distribution function for β_T for first simulated dataset.

4.2 Case study 2: discrete mixture with three support points

In the second case study, the true model is again a discrete mixture of a MNL model, where this time, three support points are used for β_T, at -1, -0.7 and -0.4, with probabilities of 0.3, 0.35 and 0.35. This leads to a true mean VTTS of 41.1 DKK per hour, with a standard deviation of 14.48 DKK per hour. Four different models were estimated on these data; along with the usual MNL and MMNL models, we estimated a

DM with two support points, and a DM with three support points[9]. Again, the DM models were estimated with two different specifications, using a randomly distributed β_C coefficient in DM(2)_A and DM(3)_A, and a fixed β_C coefficient in DM(2)_B and DM(3)_B. The detailed results for the first sample are presented in Table 4.

Table 4: Estimation results for second simulated dataset.

	MNL		*MMNL*		*DM(2)* $_A$		*DM(2)* $_B$		*DM(3)* $_A$		*DM(3)* $_B$	
Final LL	-4721.69		-4155.65		-4126.23		-4227.43		-4120.96		-4120.99	
par.	2		4		8		5		12		7	
adj. ρ^2	0.3676		0.4431		0.4465		0.4334		0.4467		0.4473	
est.time (s)	1		346		16		6		151		13	
	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat
β_T	-0.3925	-33.72	-	-	-	-	-	-	-	-	-	-
$\beta_{T,\mu}$	-	-	-0.6817	-34.78	-	-	-	-	-	-	-	-
$\beta_{T,\sigma}$	-	-	0.2423	25.15	-	-	-	-	-	-	-	-
$\beta_{T,1}$	-	-	-	-	-0.8561	-36.56	-0.4005	-36.62	-0.3930	-29.72	-0.7015	-34.17
$\pi_{T,1}$	-	-	-	-	0.6210	27.38	0.5028	27.72	0.3185	14.88	0.4069	14.02
$\beta_{T,2}$	-	-	-	-	-0.4221	-28.99	-0.8084	-39.33	-0.7031	-32.76	-0.3927	-29.84
$\pi_{T,2}$	-	-	-	-	0.3790	16.71	0.4972	27.41	0.4093	13.50	0.3187	14.89
$\beta_{T,3}$	-	-	-	-	-	-	-	-	-1.0262	-32.13	-1.0234	-34.26
$\pi_{T,3}$	-	-	-	-	-	-	-	-	0.2723	10.94	0.2744	11.64
β_C	-0.5732	-33.39	-	-	-	-	-0.8783	-40.98	-	-	-1.0084	-39.31
$\beta_{C,\mu}$	-	-	-0.9965	-37.48	-	-	-	-	-	-	-	-
$\beta_{C,\sigma}$	-	-	0.0591	4.51	-	-	-	-	-	-	-	-
$\beta_{C,1}$	-	-	-	-	-1.2023	-35.79	-	-	-1.0015	-23.66	-	-
$\pi_{C,1}$	-	-	-	-	0.5357	13.28	-	-	0.8114	0.69	-	-
$\beta_{C,2}$	-	-	-	-	-0.8469	-33.57	-	-	-1.0454	-6.67	-	-
$\pi_{C,2}$	-	-	-	-	0.4643	11.51	-	-	0.1886	0.16	-	-
$\beta_{C,3}$	-	-	-	-	-	-	-	-	-1.2583	0.00	-	-
$\pi_{C,3}$	-	-	-	-	-	-	-	-	0.0000	0.00	-	-
μ_{VTTS}	41.08		41.18		41.22		41.25		41.18		41.09	
σ_{VTTS}	-		14.86		14.68		13.94		14.41		14.44	

[9]No further gains could be made by using more than three support points.

The results show major improvements for the MMNL and various DM models when compared to the MNL model. All six models perform very well in terms of retrieving the mean VTTS, while the five mixture models also obtain a good approximation to the true standard deviation of the VTTS. We now look in more detail at the differences between the various mixture models. As was the case in the case study discussed in Section 1, the MMNL model again falsely recovers some random variation for β_C, where the level of variation is however much lower than was the case in the first case study. When only allowing for two support points, the DM models also retrieve significant variation for β_C, as reflected in the drop in model fit observed from $\text{DM}(2)_A$ to $\text{DM}(2)_B$ when constraining β_C to a fixed value. This is no longer the case when using three support points. Finally, as was the case in Section 1, the DM models again have a significant advantage over the MMNL model in terms of estimation cost.

Figure 2 shows the cumulative distribution functions for β_T in the MMNL model, as well as in $\text{DM}(2)_A$ and $\text{DM}(3)_B$. The advantages of the DM models are again very obvious, especially in the case of the model with three support points.

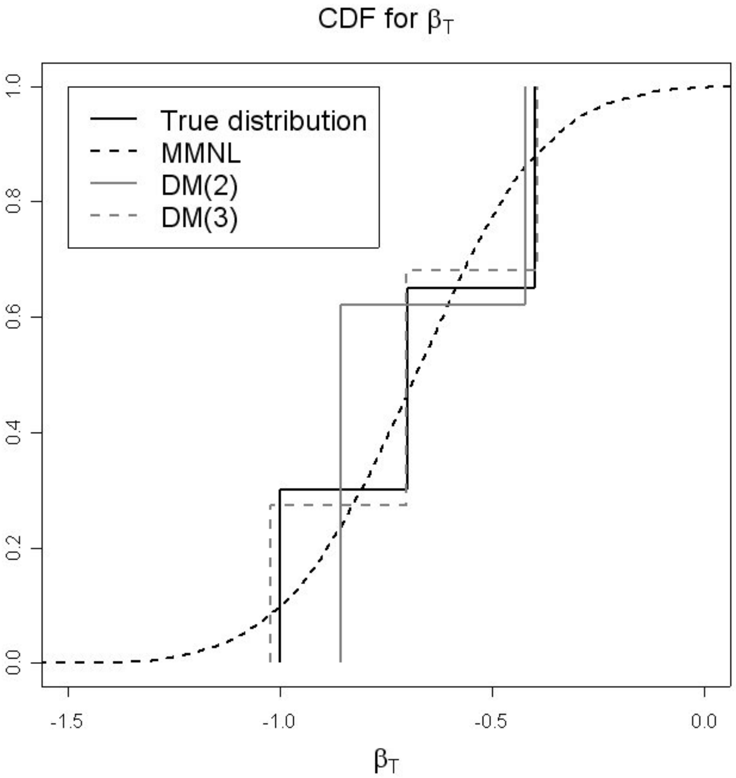

Figure 2: Cumulative distribution function for β_T for second simulated dataset.

4.3 Case study 3: normal mixture

For the third case study, a MMNL model with a normally distributed travel time coefficient was chosen as the true model. Specifically, β_C is still fixed to a value of -1, while β_T now follows a Normal distribution with mean of -0.8 and a standard deviation of 0.3, leading to a mean VTTS of 48 DKK/hour, with a standard deviation of 18 DKK.

The results for the first subsample of the third simulated dataset are summarised in Table 5. A slightly different strategy was employed in the model estimation in this case study. From the experience of the first two case studies, it had to be assumed that some of the distribution of β_T would erroneously be picked up as heterogeneity in β_C. This would apply especially in the discrete mixture models with a low number of support points. As such, alongside the MNL model, two different MMNL models were estimated, one with β_C kept fixed, and one with a randomly distributed β_C. In the discrete mixture models, 2 support points were used for β_C, while the number of support points for β_T was gradually increased up to the point where no heterogeneity was retrieved for β_C, i.e. the random taste heterogeneity in the data is captured correctly by β_T on its own. It was found that this point was reached between five and six support points for β_T. No further gains in model performance could be obtained by increasing the number of support points for β_T any further, independently of the treatment of β_C.

Again, all the different models offer good performance in retrieving the true mean value of the VTTS, while the various mixture models additionally offer a good approximation to the true standard deviation. The six mixture models offer significant improvements in model performance when compared to the MNL model. As in the other examples, the DM models again have computational advantages over the MNL model. Given the results from the other case studies, it is of interest to look at the issue of confounding between the heterogeneity for β_T and β_C. In the MMNL model and the DM model with six support points, the reductions in model fit resulting from using a fixed β_C coefficient are not significant. With only five support points, the drop in model fit is slightly more visible ($DM(5)_A$ *vs* $DM(5)_B$), yet still not significant when taking into account the cost of estimating three additional parameters. However, in earlier models, using fewer than five support points for β_T, this was not the case, and there were significant amounts of confounding[10].

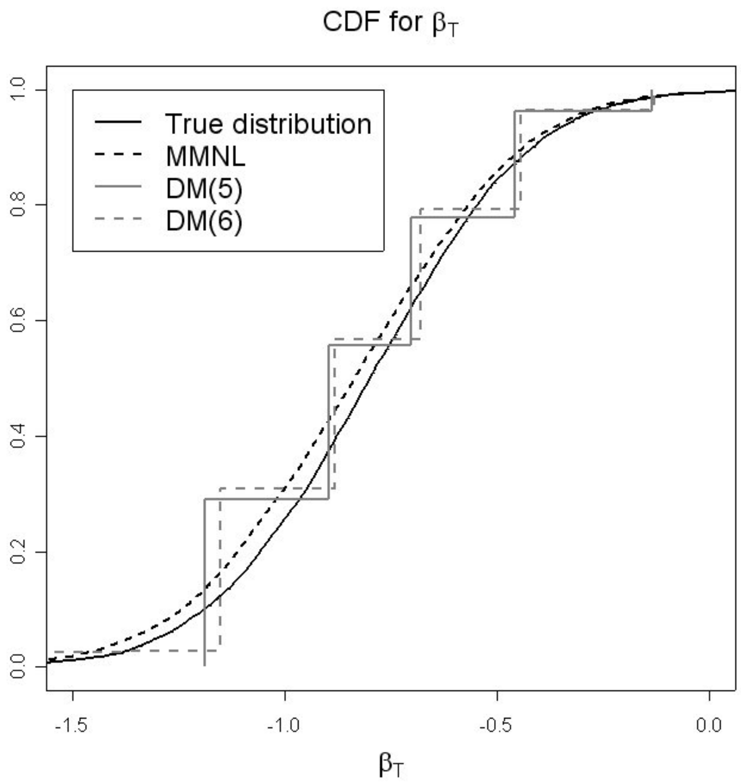

Figure 3: Cumulative distribution function for β_T for third simulated dataset.

Finally, it is of interest to look at the specific patterns of heterogeneity retrieved by the discrete mixture models, where we focus on $MMNL_B$, $DM(5)_A$ and $DM(6)_B$. Here, it can be seen from Figure 3 that the two DM models offer a very good approximation to the Normal distribution.

[10]Detailed results available on request.

Table 5: Estimation results for third simulated dataset.

	MNL		***MMNL*** $_A$		***MMNL*** $_B$		***DM(5)*** $_A$		***DM(5)*** $_B$		***DM(6)*** $_A$		***DM(6)*** $_B$	
Final LL	-4742.06		-3912.57		-3913.9		-3910.2		-3913.54		-3908.43		-3908.61	
par.	2		4		3		14		11		16		13	
adj. ρ^2	0.3649		0.4756		0.4756		0.4746		0.4746		0.4746		0.4750	
est.time (s)	1		341		233		143		41		141		59	
	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat
β_T	-0.4008	-32.11	-	-	-	-	-	-	-	-	-	-	-	-
$\beta_{T,\mu}$	-	-	-0.8359	-36.67	-0.8329	-36.69	-	-	-	-	-	-	-	-
$\beta_{T,\sigma}$	-	-	0.3134	25.27	0.3113	25.03	-	-	-	-	-	-	-	-
$\beta_{T,1}$	-	-	-	-	-	-	-0.1343	-1.99	-0.1867	-4.82	-0.0859	-1.55	-0.1293	-1.88
$\pi_{T,1}$	-	-	-	-	-	-	0.0372	2.34	0.0566	3.82	0.0245	2.55	0.0362	2.32
$\beta_{T,2}$	-	-	-	-	-	-	-0.4585	-13.87	-0.5071	-17.37	-0.3621	-8.79	-0.4449	-14.26
$\pi_{T,2}$	-	-	-	-	-	-	0.1838	6.55	0.2326	9.14	0.0742	1.90	0.1710	6.66
$\beta_{T,3}$	-	-	-	-	-	-	-0.7021	-22.49	-0.7904	-28.13	-0.7102	-20.65	-0.6800	-24.56
$\pi_{T,3}$	-	-	-	-	-	-	0.2231	3.96	0.3529	9.63	0.2026	3.28	0.2247	4.71
$\beta_{T,4}$	-	-	-	-	-	-	-1.1872	-29.66	-1.1208	-26.75	-0.9006	-19.61	-0.8830	-20.63
$\pi_{T,4}$	-	-	-	-	-	-	0.2905	7.70	0.3296	9.67	0.2653	5.11	0.2597	5.79
$\beta_{T,5}$	-	-	-	-	-	-	-0.8964	-19.79	-1.6253	-20.78	-0.5177	-10.19	-1.1502	-29.93
$\pi_{T,5}$	-	-	-	-	-	-	0.2654	5.25	0.0283	2.85	0.1441	3.01	0.2818	7.47
$\beta_{T,6}$	-	-	-	-	-	-	-	-	-	-	-1.1902	-29.58	-1.6423	-21.24
$\pi_{T,6}$	-	-	-	-	-	-	-	-	-	-	0.2893	7.61	0.0267	3.10
β_C	-0.4999	-30.13	-	-	-1.0267	-38.53	-	-	-1.0135	-37.67	-	-	-1.0213	-37.66
$\beta_{C,\mu}$	-	-	-1.0254	-38.58	-	-	-	-	-	-	-	-	-	-
$\beta_{C,\sigma}$	-	-	0.0080	0.47	-	-	-	-	-	-	-	-	-	-
$\beta_{C,1}$	-	-	-	-	-	-	-0.7467	-18.57	-	-	-0.7485	-18.50	-	-
$\pi_{C,1}$	-	-	-	-	-	-	0.0862	2.67	-	-	0.0861	2.68	-	-
$\beta_{C,2}$	-	-	-	-	-	-	-1.0545	-34.46	-	-	-1.0569	-34.41	-	-
$\pi_{C,2}$	-	-	-	-	-	-	0.9138	28.31	-	-	0.9139	28.42	-	-
μ_{VTTS}	48.10		48.93		48.68		48.96		48.72		48.77		48.81	
σ_{VTTS}	-		18.34		18.20		18.08		18.15		18.15		18.15	

4.4 Case study 4: mixture of two normals

For the fourth case study, a more complex mixture was used. As such, the true distribution is now a mixture of two Normal distributions, where $\beta_T = \pi_1 \beta_{T_1} + \pi_2 \beta_{T_2}$, with $\pi_1 = \pi_2 = 0.5$, and with $\beta_{T_1} \sim N(-0.8, 0.2)$ and $\beta_{T_2} \sim N(-0.3, 0.1)$. The cost coefficient β_C was again kept fixed at -1. With this, we obtain a true mean VTTS of 33 DKK/hour, with a standard deviation of 17.76 DKK. In model estimation, the strategy from the third case study was again adopted, gradually increasing the number of support points for β_T in the DM models, while maintaining the number of support points for β_C fixed at 2. Again, the issue of confounding largely disappeared when using five or more support points.

The results for the first subsample are presented in Table 4. Along with the MNL model, two MMNL models were estimated, where MMNL$_A$ and MMNL$_B$ again differ by using a randomly distributed and fixed β_C coefficient respectively. Although the standard deviation for β_C is significantly different from zero in model MMNL$_A$, it is very small compared to the mean value, such that it is no surprise that the effect of using a fixed coefficient is very small, with very similar model performance for MMNL$_B$. In the DM models, we experience a very small, and insignificant drop in model fit when constraining β_C to a single value. Here, two further observations can be made. In model DM(5)$_A$, the difference between $\beta_{C,1}$ and $\beta_{C,2}$ is not significant beyond the 48% level of confidence, while, in model DM(6)$_A$, it is not significant beyond the 50% level of difference. It can also be seen that, on average, when moving from DM(5)$_A$ to DM(5)$_B$ and from DM(6)$_A$ to DM(6)$_B$, the standard errors associated with the various $\pi_{T,k}$ parameters decrease. Finally, model DM(6)$_B$ can be seen to reduce to model DM(5)$_B$; the additional support point, as well as its associated probability, are not significantly different from zero. All seven models again offer good performance in the retrieval of the true mean VTTS, where the six mixture models also perform well for the standard deviation. The DM models maintain their advantages in terms of estimation cost, where these are naturally smaller than before given the higher number of parameters. In terms of model performance, the MMNL models clearly outperform the MNL model, while the various DM models have a small advantage over the MMNL models.

When looking at the retrieval of the true shape for the distribution of β_T, it can be seen that the MMNL models using a single Normal distribution produce a mean that is the weighted average of the mean of the two Normal distributions. The DM models on the other hand do recover the multi-modality of the true distribution[11]. These findings are reflected in the shape of the distributions for β_T in Figure 4, where the DM models ($DM(5)_A$ and $DM(6)_B$) are better able to account for the multi-modality of the true distribution.

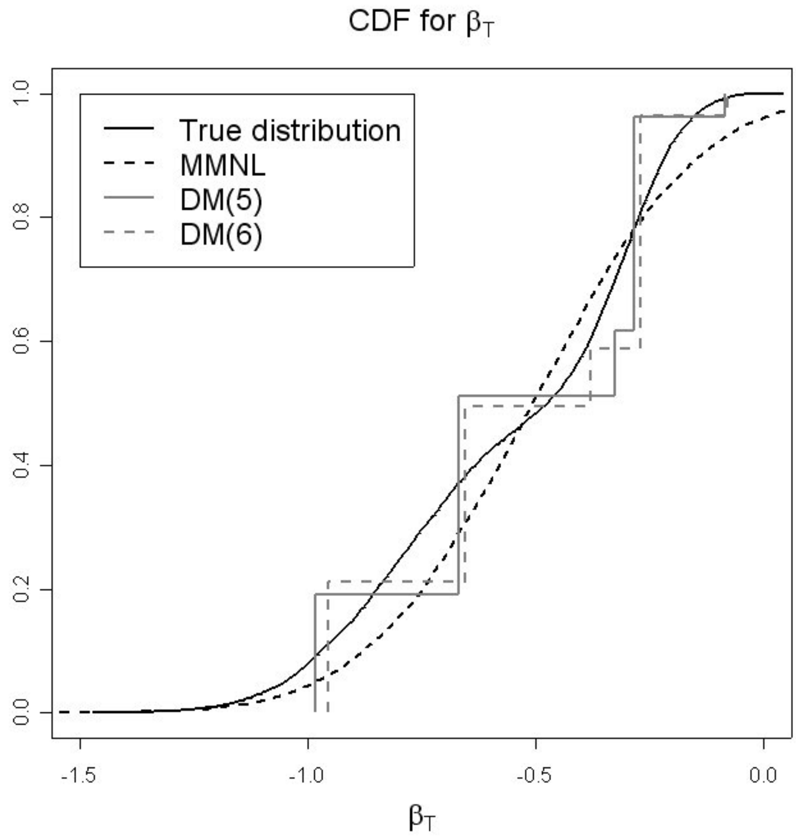

Figure 4: Cumulative distribution function for β_T for fourth simulated dataset.

[11]It should be noted that, in the retained DM model, $DM(5)_B$, two of the probabilities for support points, $\pi_{T,1}$ and $\pi_{T,3}$, are only significant at the 85% level of confidence.

Table 6: Estimation results for fourth simulated dataset.

	MNL		*MMNL $_A$*		*MMNL $_B$*		*DM(5) $_A$*		*DM(5) $_B$*		*DM(6) $_A$*		*DM(6) $_B$*	
Final LL	-5296.84		-4405.54		-4406.11		-4359.07		-4363.23		-4359.03		-4363.23	
par.	2		4		3		14		11		16		13	
adj. ρ^2	0.2906		0.4096		0.4097		0.4145		0.4144		0.4143		0.4141	
est.time (s)	7		341		213		197		33		174		75	
	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat	est.	asy.t-rat
β_T	-0.2153	-27.80	-	-	-	-	-	-	-	-	-	-	-	-
$\beta_{T,\mu}$	-	-	-0.5115	-30.23	-0.5155	-30.62	-	-	-	-	-	-	-	-
$\beta_{T,\sigma}$	-	-	0.2930	29.23	0.2954	29.82	-	-	-	-	-	-	-	-
$\beta_{T,1}$	-	-	-	-	-	-	-0.0844	-1.33	-0.0787	-1.27	-0.0855	-1.36	-0.0787	-1.27
$\pi_{T,1}$	-	-	-	-	-	-	0.0377	1.42	0.0363	1.47	0.0385	1.41	0.0363	1.47
$\beta_{T,2}$	-	-	-	-	-	-	-0.2831	-7.90	-0.2713	-16.19	-0.2835	-7.01	-0.2713	-16.19
$\pi_{T,2}$	-	-	-	-	-	-	0.3461	0.98	0.3761	6.32	0.3395	0.77	0.3761	6.32
$\beta_{T,3}$	-	-	-	-	-	-	-0.9833	-20.76	-0.3788	-11.94	-0.3216	-3.95	-0.3788	-11.94
$\pi_{T,3}$	-	-	-	-	-	-	0.1900	4.71	0.0921	1.46	0.1102	0.25	0.0921	1.46
$\beta_{T,4}$	-	-	-	-	-	-	-0.3253	-4.58	-0.6543	-28.15	-0.7247	-9.82	-0.4676	0.00
$\pi_{T,4}$	-	-	-	-	-	-	0.1047	0.29	0.2848	10.16	0.0749	0.33	0.0000	0.00
$\beta_{T,5}$	-	-	-	-	-	-	-0.6690	-20.61	-0.9546	-30.00	-0.6555	-11.72	-0.6543	-28.15
$\pi_{T,5}$	-	-	-	-	-	-	0.3215	8.36	0.2107	8.94	0.2508	1.13	0.2848	10.16
$\beta_{T,6}$	-	-	-	-	-	-	-	-	-	-	-0.9842	-21.79	-0.9546	-30.00
$\pi_{T,6}$	-	-	-	-	-	-	-	-	-	-	0.1860	4.77	0.2107	8.94
β_C	-0.4194	-31.89	-	-	-0.9721	-37.89	-	-	-0.9737	-37.40	-	-	-0.9737	-37.40
$\beta_{C,\mu}$	-	-	-0.9718	-38.17	-	-	-	-	-	-	-	-	-	-
$\beta_{C,\sigma}$	-	-	0.0352	1.97	-	-	-	-	-	-	-	-	-	-
$\beta_{C,1}$	-	-	-	-	-	-	-1.0781	-22.66	-	-	-0.8725	-15.68	-	-
$\pi_{C,1}$	-	-	-	-	-	-	0.5965	4.03	-	-	0.3572	1.70	-	-
$\beta_{C,2}$	-	-	-	-	-	-	-0.8801	-20.83	-	-	-1.0662	-18.58	-	-
$\pi_{C,2}$	-	-	-	-	-	-	0.4035	2.73	-	-	0.6428	3.06	-	-
μ_{VTTS}	30.80		31.60		31.82		32.60		32.49		32.64		32.49	
σ_{VTTS}	-		18.15		18.23		17.39		17.10		17.38		17.10	

In closing, it should be noted that, in this example, the uni-modal MMNL model still manages to retrieve the true mean and standard deviation of the multi-modal true distribution of the VTTS. This can be explained by the fact that the probabilities for the two Normal distributions were set evenly to 0.5, where the difference in the standard deviation for β_{T_1} and β_{T_2} was also rather small. Different patterns could be expected in a more asymmetrical scenario.

5. Summary and conclusions

With the availability of powerful computers and estimation tools, researchers and practitioners are increasingly making use of continuous mixture structures, such as Mixed Logit, in the representation of random taste heterogeneity across respondents. Despite the gains in estimation power, the cost of using such mixture models remains high, especially in large scale studies. Furthermore, several issues arise due to the models' reliance on specific distribution functions, whose shape is not necessarily consistent with that of the *true*, unobserved distribution.

In this paper, we have discussed an alternative approach for the representation of random taste heterogeneity, making use of discrete mixtures instead of continuous mixtures. Although several issues can also arise in the estimation of such models, they have the advantage of a closed form solution, and can hence be estimated and applied without relying on simulation processes. Furthermore, the models are free from a priori assumptions as to the shape of the *true* distribution.

The paper presents several case studies offering an in-depth comparison of the two modelling approaches, making use of real data as well as four separate simulated datasets. The results of these analyses clearly show the major advantage of the discrete mixture approach in terms of estimation cost. They also show that, across scenarios, the discrete mixture models are able to attain similar or indeed better performance than their continuous counterparts. Finally, they are better able to deal with complicated *true* distributions, such as the presence of multiple modes.

Although further comparisons between the two modelling approaches are required, the results from this paper do suggest that discrete mixture models present a viable alternative, partly thanks to their lower cost in estimation and application, but also due to the absence of a priori shape assumptions, which is of great interest in the context of recent discussions of the issue of the specification of continuous heterogeneity by Hess et al. (2005).

Acknowledgements

Part of the work described in this paper was carried out during a guest stay by the first author in the Institute of Transport and Logistics Studies at the University of Sydney. The authors are grateful to two anonymous referees for comments on an earlier version of this paper.

References

Bierlaire, M. (2003) *BIOGEME: a free package for the estimation of discrete choice models*, Proceedings of the 3rd Swiss Transport Research Conference, Monte Verità, Ascona.

Burge, P. and Rohr, C. (2004) *DATIV: SP Design: Proposed approach for pilot survey*, Tetra-Plan in cooperation with RAND Europe and Gallup A/S.

Chintagunta, P., Jain, D. and Vilcassim, N. (1991) "Investigating heterogeneity in brand preference in logit models for panel data", *Journal of Marketing Research*, 28, pp. 417–428.

Cirillo, C. and Axhausen, K. W. (2006) "Evidence on the distribution of values of travel time savings from a six-week diary", *Transportation Research Part A: Policy and Practice*, 40 (5), pp. 444–457.

Dong, X. and Koppelman, F. S. (2003) "Mass Point Mixed Logit Model: Development and Application", paper presented at *the 10th International Conference on Travel Behaviour Research*, Lucerne.

Fosgerau, M. (2006) "Investigating the distribution of the value of travel time savings", *Transportation Research Part B: Methodological*, 40 (8), pp. 688–707.

Fosgerau, M. and Bierlaire, M. (2006) "Discrete choice models with multiplicative error terms", *Technical Report TRANSP-OR 060831*, Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne.

Gopinath, D. (1995) Modeling heterogeneity in discrete choice processes: Application to travel demand, *Ph.D. Thesis*, MIT, Cambridge, MA.

Greene, W. H. and Hensher, D. A. (2003) "A latent class model for discrete choice analysis: contrasts with mixed logit", *Transportation Research Part B: Methodological*, 37 (8), pp. 681–698.

Hess, S., Bierlaire, M. and Polak, J. W. (2005) "Estimation of value of travel-time savings using mixed logit models", *Transportation Research Part A: Policy and Practice*, 39 (2-3), pp. 221–236.

Hess, S., Train, K. and Polak, J. W. (2006) "On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit model for vehicle choice", *Transportation Research Part B: Methodological*, 40 (2), pp. 147– 163.

Kamakura, W. A. and Russell, G. (1989) "A probabilistic choice model for market segmentation and elasticity structure", *Journal of Marketing Research*, 26, pp. 379– 390.

Lee, B. J., Fujiwara, A., Zhang, J. and Sugie, Y. (2003) "Analysis of Mode Choice Behaviours based on Latent Class Models", paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne.

Mackie, P., Wardman, M., Fowkes, A. S., Whelan, G., Nellthorp, J. and Bates, J. J. (2003) *Valuation of Travel Time Savings in the UK*, Summary Report to the Department for Transport, Institute for Transport Studies, University of Leeds.

Wedel, M., Kamakura, W. A., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S. and Poulsen, C. S. (1999) "Discrete and continuous representations of unobserved heterogeneity in choice modelling", *Marketing Letters*, 10 (3), pp. 219–232.

A mixed logit model for the sensitivity analysis of Greek drivers' behaviour towards enforcement for road safety

G. Yannis [1*], C. Antoniou [1]

[1] *National Technical University of Athens,*
Department of Transportation Planning and Engineering,
Athens, Greece

Abstract

Traffic violations are among the leading causes of road accidents. In this research, the sensitivity of Greek drivers to a hypothetical intensification of police enforcement for speed violations and improper overtaking is analyzed, using stated preference data. Under the assumption of increased police enforcement, drivers were presented with the option to maintain their unsafe driving patterns (and risk getting fined) or comply with the traffic laws (and experience longer trip duration). A parsimonious mixed logit model has been estimated and sensitivity analysis is performed with respect to the main variables. The model explicitly captures the (unobserved) heterogeneity in the sample, and reflects the fixed random parameter across observations from the same respondent. The behaviour of the surveyed drivers depends on socioeconomic characteristics and trip characteristics. Based on the presented sensitivity analysis, it can be argued that while the "typical" Greek driver may not be particularly risk-prone, there are segments of the population that show a tendency to violate traffic laws. This is a useful finding that could be used by policy makers e.g. to develop targeted police enforcement campaigns (or targeted media campaigns, special education initiatives, etc.), aimed at the demographic segments with a higher tendency for traffic violation.

Keywords: Road safety; Police enforcement; Mixed logit model; Panel data; Stated-preference.

1. Introduction

Road safety is one of the most important issues throughout the world. For example, according to the European Commission CARE database for 2002, the number of fatalities from more than 1,250,000 road accidents in the (then fifteen) European countries was 38,637, while another 1,700,000 were injured. In its White Paper on European transport policy (EC, 2001) the Commission has therefore proposed that the

* Corresponding author: George Yannis (geyannis@central.ntua.gr)

European Union (EU) should set itself the target of halving the number of road deaths by 2010 (a proposal that has since been adopted).

Research in the field of road safety has shown that traffic violations constitute one of the most important factors of road accidents. Specifically, it has been shown (Rothengatter and Harper, 1991) that a large proportion of road accidents inside the EU is the result of one or more traffic violations. In the same study it is concluded that the cumulative traffic violations is the most important factor leading to road accidents. Furthermore, according to the European Council of Road Safety (ETSC, 1999), it is calculated that significant improvement in road safety (as high as 50%) could be achieved through measures for the prevention of traffic violations (such as intensification of police enforcement). It is worth mentioning that while drivers commit traffic violations, they believe that police enforcement of road networks should be intensified to the benefit of road safety. According to the SARTRE EU research project, up to 70% of the drivers believe that police enforcement should be intensified in order for traffic violations to be reduced (SARTRE, 1998).

As it has been suggested from various researchers (e.g. Bjornskau and Elvik, 1992, Zaal, 1994, Newstead, Cameron, Mark and Leggett, 2001, Tay, 2005), one way of reducing the number of traffic violations is the intensification of police enforcement. This is supported by empirical results reported by Holland and Conner (1995).

It is clear from the literature that intensification of police enforcement is expected to result in an improvement in road safety. The goal of this paper is to analyze the sensitivity of Greek drivers towards the intensification of police enforcement (targeted primarily at speed limit violations and illegal overtaking). Stated-preference data are collected from a specially designed questionnaire. The collected data includes socioeconomic data, as well as the surveyed drivers' response to a number of hypothetical scenarios. For each scenario, the driver is asked to choose between two alternatives, with different attributes, such as police enforcement intensity, probability of getting fined (and/or being involved in an accident with injury) and trip duration. (For an overview of road safety in Greece, and analysis of factors affecting road safety in Greece c.f. e.g. Matsoukis et al., 1996, Golias et al., 1997, Kanellaidis et al., 1999). Policy makers as well as road safety practitioners could benefit from this research, as they can better support their choices and decisions through the use of the proposed methodologies.

The remainder of this paper is structured as follows. An overview of the techniques that are used in this paper (namely stated-preference surveys and discrete choice analysis using mixed-logit models) is presented in Section 2. The survey design and data collection procedure is described in Section 3. Model specification and estimation results are presented in Section 4, while sensitivity analysis using the estimated model coefficients is presented in Section 5. Conclusions and directions for further research are outlined in Section 6.

2. Background

2.1 Stated-preference techniques

Stated preference techniques are an attractive tool for researching non-existing situations (Louviere et al., 2000). The analysis of stated preference data originated in mathematical psychology with the seminal paper by Luce and Tukey (1964). Stated preference methods were further developed in marketing research in the early seventies and over several decades have had several applications. Stated preference techniques have been used, for example, to examine the effect of travel information on mode choice (Abdel-Aty et al., 1997, Khattak et al., 1996, Polydoropoulou et al., 1996). More recently, Rizzi and Ortuzar (2003) used stated preference techniques in the context of road-safety for the estimation of the value of attributes (travel time, toll and annual accident rate) for the valuation of road accident fatalities.

Stated-preference techniques have also been used specifically for the assessment of drivers' preference with respect to enforcement. Yannis et al. (2005) investigated the behavioral parameters that influence drivers' choices in order to reduce accident risk, using stated-preference techniques and logistic regression models. SARTRE (2004) describes the third wave of a large-scale stated-preference survey across Europe, that was collected data in various aspects of road-safety, including perception and response to enforcement. Kanellaidis et al. (1999) used stated-preference surveys to assess the attitude of Greek drivers towards road safety.

The primary drawback to stated preference data is that they may not be congruent with actual behaviour (for example due to biases). This phenomenon can be critical under certain circumstances, when for example the results are not verified with results from the literature, or revealed-preference data. Additionally, particular attention should be given to the results' interpretation, because respondents show the tendency to exaggerate when they conceive that they take part in some experiment (Lin et al., 1986, van der Hoorn et al., 1984).

2.2 The Mixed-Logit Model for Panel Data

Discrete choice analysis is a well established approach for analyzing individual behavior (Ben-Akiva and Lerman, 1985). In the case of repeated observations (such as the case of stated-preference surveys with multiple responses) one often needs to capture the correlation across observations from the same individual. In general, pooling data across individuals while ignoring correlation across observations and unobserved heterogeneity among responses from different individuals (when it is present) will lead to biased and inconsistent estimates of the effects of pertinent variables (Hsiao, 1986). Several approaches have been developed to incorporate these effects in the model formulation. One is to estimate a constant term for each individual and each choice, which is referred to as a "fixed-effects" approach (Chamberlain, 1980). Perhaps the main drawback to this approach is the large number of parameters (and consequently large number of required observations per individual). A more tractable approach is to replace the fixed term with some probability distribution, which is referred to as a random effects specification (Heckman, 1981, Hsiao, 1986). The most common assumptions for this distribution are the normal and the lognormal. One drawback to this approach, however, is that it does not allow for a closed-form expression for the choice probabilities, thus leading to numerical complications, which will be detailed below.

Mixed logit is a highly flexible model that can approximate any random utility model (McFadden and Train, 2000). It obviates the three limitations of standard logit by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time (e.g. in the case that data from the same individuals are collected at different times). Unlike probit, it is not restricted to normal distributions. Its derivation is straightforward, and simulation of its choice probabilities is computationally simple. Like probit, the mixed logit model has been known for many years but has only become fully applicable since the advent of simulation (Train, 2003). Some indicative applications from the literature follow. Han et al. (2001) develop a mixed logit model to accommodate the random heterogeneity across drivers and to cope with the correlation between repeated choices. Hess et al. (2004) use mixed logit models that allow for random taste heterogeneity for the computation of value-of-time. Bierlaire et al. (2006) present a mixed binary logit model with panel data to analyze the drivers' decisions when traffic information is provided during their trip.

3. Survey design and data collection

The necessary data were collected through a stated preference survey using a specially developed questionnaire. The final sample comprises 251 questionnaires that were completed by drivers familiar with driving in the Greek national road network. The majority of the respondents were from the city of Halkida, where the data-collection took place. Halkida has a population of approximately 100,000 and is located 85 km northeast of Athens, Greece. The surveys were administered at road-side rest areas along the national freeway connecting Halkida with Athens. One potential impact of the survey execution is that interviewed drivers were likely to perceive accident risk related to highway travel, which in principle may be somewhat different than that related to non-highway trips.

The days and hours for the administration of the field survey were chosen so that the sample covers a wide spectrum of driver characteristics (e.g. in terms of age and education). In order to ensure that the sample would be representative and unbiased, further sampling approaches were used. For example only a subset of the drivers stopping at the rest area (1 out of 7) was randomly interviewed to avoid correlation issues. Before starting the questions, the interviewer presented the framework of the survey and made as clear as possible to the interviewees the meaning of the options proposed in the questionnaire; e.g. the current accident risk level was explained allowing to better understand what a 20% risk decrease means.

The questionnaire consisted of three parts and can be completed in approximately five minutes. The first part questions collected demographic characteristics of the subject, such as gender, age, residence area, educational level, occupation and annual income. The second part questions aimed to expose the subject to the road safety problem, and in particular the probability of being involved in an accident. In the third part, a subset of which is shown in Table 1, each respondent was asked about the duration of their usual highway trip, and was subsequently presented with scenarios for this duration. The scenarios were based on the assumption that the current trip duration is shorter than it should, since lack of continuous and effective police enforcement allows for speed violations and illegal overtaking.

Table 1: Subset of part three of the questionnaire.

TRIP DURATION 2 HOURS

Choice	Compliance to increased enforcement is	Trip duration increase (min)	Risk probability reduction (%)
A	unlikely	0	0%
B	likely	+30	-10%

Choice	Compliance to increased enforcement is	Trip duration increase (min)	Risk probability reduction (%)
A	unlikely	0	0%
B	likely	+90	-10%

TRIP DURATION 5 HOURS

Choice	Compliance to increased enforcement is	Trip duration increase (min)	Risk probability reduction (%)
A	unlikely	0	0
B	likely	+ 60	-20%

Choice	Compliance to increased enforcement is	Trip duration increase (min)	Risk probability reduction (%)
A	unlikely	0	0%
B	likely	+180	-10%

If police enforcement was intensified, then the respondents would face the following dilemma: either continue to violate traffic laws and get fined (or get involved in a traffic accident with injury) or comply with traffic laws, not get fined but be subjected to increased trip duration. The assumption is made that in the hypothetical scenario of intensified police enforcement all traffic violations will be recorded and all violators will have to pay a fine (which for the purposes of this research was set at €120 or approximately US$150). The proposed options take into account the fact that higher compliance leads to both longer trip duration and lower risk.

4. Model estimation

Most of the data has been coded as categorical variables. While the order in which the levels are coded (e.g. ascending age groups) follow some logic, assuming that the behavioural patterns of the individuals would follow the same trend is overly restrictive. For example, using a single variable Age in the model, resulting in the estimation of a single coefficient for Age, would imply that the impact of age is a linear function of the age group. To overcome this issue, dummy variables for each level have been introduced. Naturally, for a categorical variable with m levels, only m-1 dummy variables can be defined (while the remaining level serves as the base).

Not all levels of all categorical variables have a significant contribution to the model, however. Therefore, based on formal statistical significance tests, some levels have been grouped together. For example, the two lower levels of the Age factor (18-24 and 25-34) have been grouped together to provide a single level (18-34), which is used as a base for the age groups. Similarly, the two highest levels of the Age factor (55-64 and 65+) have also been grouped together. The specification table for the binary and mixed logit models is shown in Table 2. The available data (including the alternative specific constant) have been used to construct the utility function for the option that the users would be likely to comply with the increased police enforcement and experience a higher travel time (and reduction of their probability of getting involved in an accident with injury) in order not to be fined. As no variables are used for the specification of the alternative option (choosing to violate the speed limit and/or perform illegal overtaking manoeuvres at the risk of getting fined), the utility function of that alternative option is constant and equal to zero. Since only the difference in utilities can be captured in discrete choice models, using one alternative as a reference case in this way does not affect the estimation of the model. More formally, the systematic utility specification for the two alternatives can be expressed as:

$$
\begin{aligned}
V_{Compliance,\,ij} = \; & \beta_{Compliance} * 1 + \beta_{Age\ 35\text{-}44} * X_{age\ dummy\ (35\text{-}44),ij} + \\
& + \beta_{Age\ 45\text{-}54} * X_{age\ dummy\ (45\text{-}54),ij} + \beta_{Age\ 55+} * X_{age\ dummy\ (55+),ij} + \\
& + \beta_{Low\ income} * X_{low\ income\ dummy,ij} + \beta_{High\ income} * X_{high\ income\ dummy,ij} + \\
& + \beta_{Low\ education} * X_{low\ education\ dummy,ij} + \beta_{Trip\ duration} * X_{trip\ duration,ij} + \\
& + \beta_{Trip\ duration\ increase} * X_{trip\ duration\ increase,ij} + \beta_{Risk\ change} * X_{Risk\ change,ij}
\end{aligned}
$$

$$V_{Non\text{-}compliance} = 0$$

The following notation is used in the utility specification:

- $X_{age\ dummy\ (35\text{-}44)}$ Binary dummy variable, taking the value 1 if the age of the individual is between 35 and 44, and 0 otherwise
- $X_{age\ dummy\ (45\text{-}54)}$ Binary dummy variable, taking the value 1 if the age of the individual is between 45 and 54, and 0 otherwise
- $X_{age\ dummy\ (55+)}$ Binary dummy variable, taking the value 1 if the age of the individual is above 55, and 0 otherwise

- $X_{low\ income\ dummy}$ Binary dummy variable, taking the value 1 if the income of the individual is low, and 0 otherwise
- $X_{high\ income\ dummy}$ Binary dummy variable, taking the value 1 if the income of the individual is high, and 0 otherwise
- $X_{low\ education\ dummy}$ Binary dummy variable, taking the value 1 if the education level of the individual is low, and 0 otherwise
- $X_{trip\ duration}$ Numerical explanatory variable, taking a value equal to the trip duration of the individual (in minutes)
- $X_{trip\ duration\ increase}$ Numerical explanatory variable, taking a value equal to the trip duration increase of the individual (in minutes)
- $X_{Risk\ change}$ Numerical explanatory variable, taking a value equal to the percentage of assumed risk change (e.g. if risk change is 20% then this variable is equal to 20).

The utility specification of the binary logit model is given by:

$$U^{binary}_{compliance,ij} = V_{compliance,ij} + \varepsilon_{ij}$$

where ε_{ij} is a zero-mean, random error term that is iid (independently and identically distributed) extreme value.

A random error term has been added in the utility specification of the mixed logit model to account for the presence of serially correlated repeated responses from the same respondent (panel data):

$$U^{mixed}_{compliance,ij} = V_{compliance,ij} + \sigma_{panel}\xi_i + \varepsilon_{ij}$$

where σ_{panel} is an unknown parameter to be estimated, and ξ_i is a standardized normal random parameter $\xi_i \sim N(0,1)$. In the field of transport economics this is often referred to as a compound error term.

Table 2: Model specification table.

Model:	*Binary logit*		*Mixed logit - random effects*	
	Compliance to increased enforcement scenario is		Compliance to increased enforcement scenario is	
	Likely	Unlikely	Likely	Unlikely
$\beta_{Compliance}$	1	0	1	0
$\beta_{Age\ 35\text{-}44}$	Age dummy (35-44)	0	Age dummy (35-44)	0
$\beta_{Age\ 45\text{-}54}$	Age dummy (45-54)	0	Age dummy (45-54)	0
$\beta_{Age\ 55+}$	Age dummy (55+)	0	Age dummy (55+)	0
$\beta_{Low\ income}$	Low income dummy	0	Low income dummy	0
$\beta_{High\ income}$	High income dummy	0	High income dummy	0
$\beta_{Low\ education}$	Low education dummy	0	Low education dummy	0
$\beta_{Trip\ duration}$	Trip duration (min)	0	Trip duration (min)	0
$\beta_{Trip\ duration\ increase}$	Trip duration increase (min)	0	Trip duration increase (min)	0
$\beta_{Risk\ change}$	Risk change (%)	0	Risk change (%)	0

The probability that an individual i chooses the first alternative (comply with the increased enforcement) in experiment j is given by:

$$P_i^{binary}\left(compliance \mid \{compliance, non-compliance\}\right) = \frac{e^{V_{compliance,ij}}}{e^{V_{compliance,ij}} + e^{V_{non-compliance,ij}}}$$

while for the mixed logit model, the same probability is given by:

$$P_i^{mixed}\left(compliance \mid \{compliance, non-compliance\}\right) =$$

$$\int_{\xi_i} \prod_j \frac{e^{V_{compliance,ij}+\sigma_{panel}\xi_i}}{e^{V_{compliance,ij}+\sigma_{panel}\xi_i} + e^{V_{non-compliance,ij}}} f(\xi_i)\, d\xi_i$$

where the product ranges over all experiments j of individual i, σ_{panel} is an unknown parameter to be estimated, and ξ_i is a standardized normal random parameter $\xi_i \sim N(0,1)$, so that

$$f(\xi_i) = \frac{1}{\sqrt{2\pi}} e^{-\xi_i^2/2}$$

Since the sum of the probabilities to choose all alternatives equals to one, the probability that the second (base) alternative is chosen can easily be obtained by subtracting the probability that the first alternative is chosen from one.

The model estimation was performed using the Biogeme software package (Bierlaire, 2003, 2005). Binary logit and mixed-logit models were estimated. The mixed-logit specification differs from the logit in the addition of a zero-mean, normally distributed random component, capturing the unobserved heterogeneity between individuals. The normality assumption for the random component is commonly found in the literature. In the absence of strong evidence suggesting a different distributional assumption (e.g. log-normal), it has been used in this research. Furthermore, the correlation between choices made by the same respondent is explicitly incorporated by recognizing that responses from the same individual are correlated. This is taken into account by estimating the same value of the random parameter for all observations by the same respondent. Unlike logit, mixed-logit model estimation requires simulation, which can be based e.g. on pseudo-random numbers and draws from a Halton sequence (Train, 2003, Sandor and Train, 2004, Sivakumar et al., 2005). For the mixed-logit estimates, draws from a Halton sequence have been used instead of pseudo-random numbers, as they are more efficiently spread over the unit interval. Two hundred draws have been used, which empirically was found to be an adequate number (estimated coefficients had already stabilized well below one hundred Halton draws).

The estimation results for the final models are shown in Table 3. For each parameter, the estimated coefficient value and the robust t-test value are provided. The robust statistics allow for non-severe misspecification errors related with the characteristics of the postulated distributions of the error terms (Bierlaire et al., 2005). For example, the use of robust t-tests alleviates the potential impact of a non-severe misspecification due to the choice of the normal distribution for the random component. Aggregate goodness of fit measures (testing the adequacy of the entire model specification) are presented first (Bierlaire, 2005, Washington et al., 2003). At the individual coefficient level, both informal tests (sign and magnitude of coefficient estimates) as well as more formal tests (robust t-test) have been performed. A parsimonious model specification has been sought.

Summary goodness of fit statistics indicate that the mixed logit model provides a superior fit (a lower final log-likelihood, a higher ρ^2 and corrected ρ^2, only at the expense of a single additional estimated parameter, i.e. the random coefficient). The random coefficient is also very significant, which suggests that including it in the model specification was appropriate. The magnitude of the coefficients (which is higher for the mixed logit model) also suggests that this model more accurately captured the drivers' behaviour.

Most of the coefficients are significant at the p=5% level. The coefficients for the income variables and the low education variable are significant at the p=10% level. These coefficients have been retained in the model since the informal specification tests (sign and relative magnitude) indicate that these coefficients have been estimated according to prior expectations and would therefore provide intuitive results for the sensitivity analysis.

The inclusion of the variables in the final model implies that the behaviour of the surveyed drivers in relation to an intensification of enforcement that would result in a travel time increase and risk reduction (for complying drivers), or a monetary fine (for non-complying drivers) depends on the following variables:

- Age group
- Income level

- Education level
- Trip duration (prior to intensification of enforcement)
- Trip time increase (due to intensification of enforcement)
- Risk change (due to intensification of enforcement)

Table 3: Estimation results.

	Binary logit		*Mixed logit - random effects*[a]	
Utility parameter name	Value	t-test	Value	t-test
$\beta_{Compliance}$	2.5589	8.2984	3.3262	6.3203
$\beta_{Age\ 35-44}$	1.1843	4.7182	1.6885	3.5026
$\beta_{Age\ 45-54}$	0.7346	3.4713	0.9725	2.6388
$\beta_{Age\ 55+}$	1.8336	5.1579	2.3494	3.4951
$\beta_{Low\ income}$	0.4946	2.2330	0.8390	1.8858
$\beta_{High\ income}$	-0.4783	-2.6156	-0.5975	-1.8013
$\beta_{Low\ education}$	-0.5316	-2.6463	-0.7255	-1.7882
$\beta_{Trip\ duration}$	0.0047	3.0714	0.0070	2.4780
$\beta_{Trip\ duration\ increase}$	-0.0075	-2.2456	-0.0108	-2.5328
$\beta_{Risk\ change}$	0.1136	6.6275	0.1466	6.7072
σ_{panel}	---[b]	---[b]	1.6487	7.7446

	Binary logit	Mixed logit - random effects
Number of Halton draws:	---[b]	200
Number of estimated parameters:	10	11
Sample size:	1184	250[c]
Null log-likelihood $LL(0)$:	-820.69	-820.69
Final log-likelihood $LL(\beta)$:	-490.23	-454.18
Likelihood ratio test:	660.92	733.01
ρ^2 (Rho-square):	0.4027	0.4466
Adjusted ρ^2 (rho-square):	0.3905	0.4332

[a] 200 Halton draws have been used (results had stabilized well below 100 Halton draws)

[b] --- denotes not applicable

[c] Sample size for the mixed logit refers to individual respondents (each providing up to 5 responses). Number of observations is again 1184.

A discussion of the sign and magnitude of the estimated coefficients (also called informal specification tests) is presented next. It is important that signs and relative magnitudes of estimated coefficients agree with a priori expectations (Ben-Akiva and Lerman, 1985, pp. 157-160). The positive alternative specific constant suggests that there is some a priori tendency of the drivers to choose the conservative option of compliance to the increased police enforcement and not risk getting fined to decrease travel time. This finding is intuitive and consistent with expectations.

Age has been incorporated into the model as three coefficients (using age group 18-34 as the base). All estimated parameters are positive, indicating that the younger base group is less likely to comply with police enforcement and thus implicitly more risk-prone, which is again consistent with expectations. With the exception of the age group 45-54, compliance increases with age. Drivers in the age group 45-54 show the second highest tendency towards non-compliance. It is useful at this point to revisit the choice of individual dummy variables for each age group, instead of using a single ordinal variable (with values e.g. 1 through 4, where 1 would be the age group 18-34 and 4 would be 55+). Such a model might give a significant coefficient (and indeed it does, capturing the overall trend of increasing risk aversion with age) but would miss the fact that the drivers in the age group 45-54 actually appear to exhibit more risk-prone behaviour than drivers in the age group 35-44. This misspecification would have significantly altered the sensitivity analysis results presented in the following sections (and hence the conclusions drawn from this research).

Income has been modeled using two dummy variables (one for low and one for high income) with medium income serving as the base. The coefficient for the low-income dummy variable has a positive sign, indicating that drivers with low income are more likely to comply and not risk paying a fine. High-income drivers, on the other hand, are more likely to risk non-compliance. This is an intuitive finding, since the cost of the fine (~□120 or ~US$150 according to the survey setup) is less of a disincentive for higher-income drivers. Similarly to the Age variable, modeling income levels as an ordinal variable with three levels would have resulted in erroneous results: even though the order of the levels is retained, the magnitude of the coefficients varies.

Education has also been modeled through two dummy variables (low and medium), while high education is used as the base. Note that low education in this level is the two first levels of education (elementary and junior/high-school) combined. However, only the "low education" variable has been retained in the model, as the coefficients for the others were not statistically significant. The negative estimated coefficient indicates that drivers with low education are less likely to comply with the increased police enforcement, thus risking to get fined and/or involved in an accident.

The duration of the trip is coded as a continuous variable (even though it only takes two values, i.e. 120min and 300min). A positive coefficient suggests that drivers are getting more compliant as the duration of the trip increases. This may be due to the fact that drivers associate longer trips with a higher probability of actually getting caught (or getting involved in an accident). This issue deserves deeper investigation; ideally a follow-up study would include a more detailed mapping of the variation of the trip duration.

The coefficient associated with the increase in the duration of the trip is negative, implying that the drivers' tendency to comply with the intensified enforcement (and avoid getting fined and/or hurt) is inversely proportional to the additional travel time in which it would result. Furthermore, the coefficient associated with the risk change is positive, as expected.

Finally, the random error term capturing the intra-personal correlation in the responses is significant, thus confirming that the mixed logit model is effective in capturing the unobserved heterogeneity among respondents. The sign of this parameter is not relevant.

5. Sensitivity analysis

The model application in the previous section provided some insight into the behaviour of the driver population in response to compliance with police enforcement intensification. A sensitivity analysis with respect to some of the main variables is presented in this section. This analysis is indicative of the breadth of similar analyses that can be performed using such models to illustrate the modeled behavior of the sampled population. Such analyses can be used to develop policies and strategies, estimating the potential impact of alternative scenarios.

In all cases, the dependent variable is the proportion of non-compliant drivers. Figure 1 presents the sensitivity with respect to age group for trips with a duration of 2 hours, a risk reduction of 10%, medium income level and high education. Both the ranking of the curves, and their concavity, are consistent with expectations. Furthermore, it is important to note that the proportion of drivers choosing to not alter their driving patterns in response to the hypothetical enforcement intensification and thus risking to be fined does not increase strictly with age group, as the age group 45-54 is less likely to comply to increased enforcement than the age group 35-44.

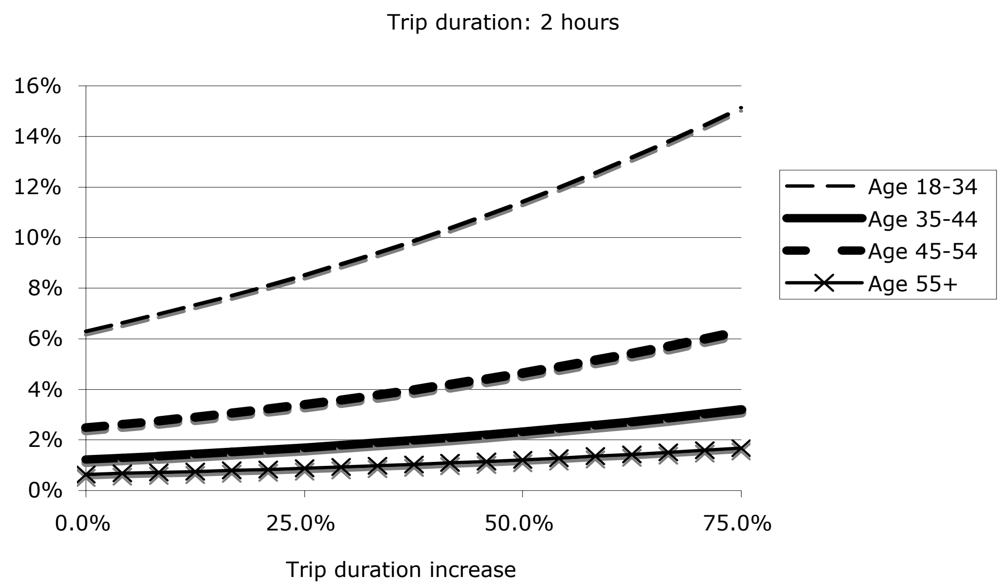

Figure 1: Sensitivity analysis w.r.t. age group.

Figure 2 presents the sensitivity with respect to income level for trips with duration of 2 hours, a risk reduction of 10%, and highly educated drivers in the 35-44 age group. Not surprisingly, drivers with high income are less likely to comply as the financial cost of getting fined affects them less.

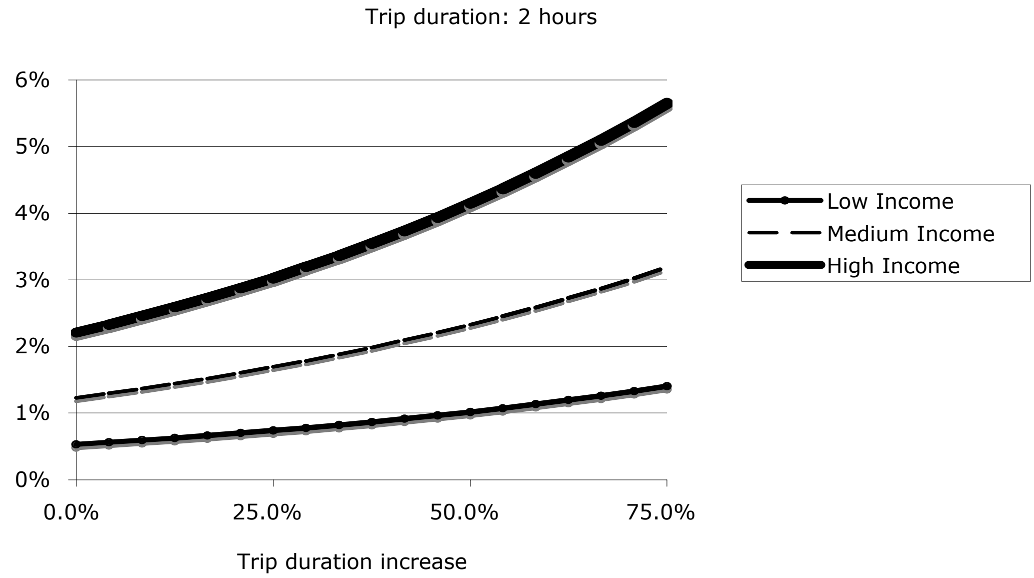

Figure 2: Sensitivity analysis w.r.t. income level.

Figure 3 presents the sensitivity with respect to increase in trip duration for a risk reduction of 10% and young drivers (i.e. the 18-34 age group) with medium education and income level. For travel time increase up to almost 60% of the original trip duration, young drivers in shorter trips (2 hours) would be less likely to comply than for longer trips (5 hours). It should be stressed, however, that the x-axis in Figure 3 corresponds to percent increase in travel time with respect to the original trip duration. This means that a 30% increase for a 2-hour trip is 36 minutes, while a 30% increase for a 5-hour trip is 90 minutes. The difference in the trip duration may also be the primary reason for the lower curvature of the shorter trip curve, while the curve for the longer trip is clearly concave.

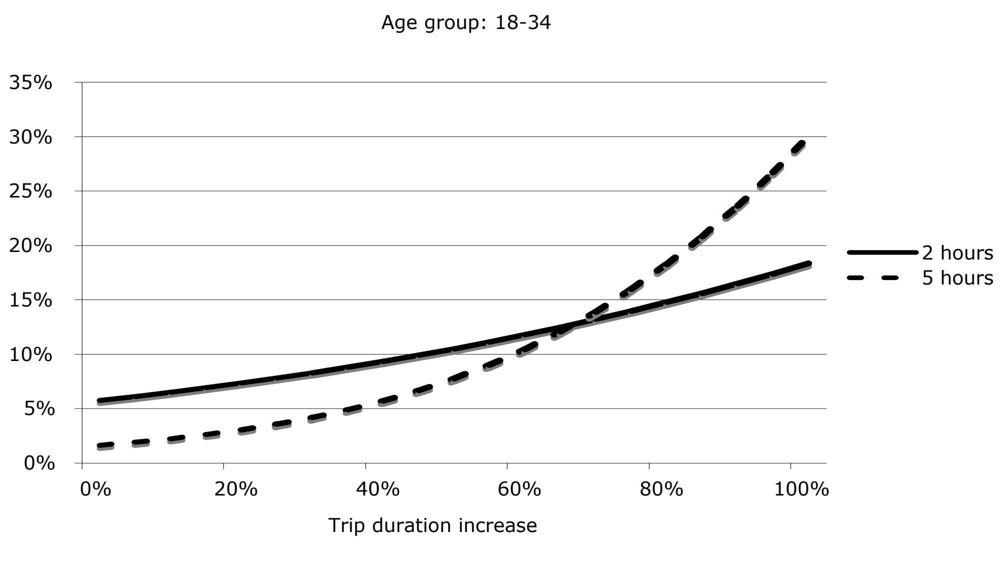

Figure 3: Sensitivity analysis w.r.t. trip duration.

6. Conclusion

A parsimonious mixed logit model that captures the unobserved heterogeneity between individual respondents thus modeling the correlation among several responses from the same individual (panel data) has been developed. Model parameters were estimated using results from a questionnaire-based survey among highway drivers in Greece. Based on the presented analysis, the behaviour of the surveyed drivers towards an intensification of police enforcement that would result in a travel time increase and accident risk reduction (for complying drivers, but also arguably indirectly for non-complying drivers) or a monetary fine (for non-complying drivers) depends on the following variables:

- Age group: younger drivers exhibit in general less compliant and consequently more risky behaviour (with the exception of the age group 45-54 showing more aggressive behaviour that the 35-44 group).
- Income level: wealthier drivers are less likely to comply, possibly due to their decreasing marginal utility of money, which makes them more indifferent to paying a fine.
- Education level: drivers with higher education show a higher tendency to comply with the traffic laws.
- Trip time increase (due to intensification of enforcement): as in all travel related models, travel time is typically considered as "impedance". Therefore, an increase in travel time due to a change in behaviour (in this case strict compliance to traffic enforcement) is translated to a trade-off: the drivers weigh the increase in travel time against the probability of getting fined and/or getting involved in a traffic accident.

The analysis also provides evidence that the existing trip duration (before the intensification of police enforcement) affects the driver's preferences. Ceteris paribus, drivers who intend to make a 5-hour trip are found to be more "patient" or compliant than those who intend to make a 2-hour trip, as long as the increase in time duration is below 60% of the original trip duration.

Based on these findings, it can be argued that while the "typical" Greek driver may be compliant, there are some segments of the population that show a higher tendency to violate traffic laws (even if they know that they can get fined and/or involved in an accident with injury). This is a useful and practical finding for road safety policy makers, which could use it to develop targeted police enforcement campaigns (or targeted media campaigns, special education initiatives), aimed at the demographic segments with a higher tendency for the violation of traffic laws. Especially, this sensitivity analysis allows policy makers to better define the effective range of the enforcement intensification for the various driver categories. The findings of this research can also be applicable to other cases, as long as the particularities of these cases are met through adequate adaptations of the proposed methodology.

An additional contribution of this research is that it demonstrates how a state-of-the-art modelling technique (mixed logit) can be used by road safety practitioners and policy makers in general, in their quest to identify critical characteristics of driver behaviour and develop the related road safety measures and programmes. New computing and software advances make the use of the more flexible mixed logit models more accessible, instead of the simpler binary logit models often preferred by practitioners.

This research has provided valuable insight into the behaviour of the Greek drivers in a situation where an intensification of police enforcement forces them to choose between complying to an intensified police enforcement scenario or being subject to a fine or a trip time increase. Similar initiatives in other countries of the EU (and the world) could show which of the underlying behavioural patterns are shared across populations and which are specific to the Greek population.

Acknowledgment

Special acknowledgment belongs to Mr. Stamatis Halapas, NTUA civil engineer for the coordination of the survey execution and for the data processing.

References

Abdel-Aty, M., Kitamura, R. and P. Jovanis (1997) "Using stated preference data for studying the effect of advanced traffic information on drivers' route choice". *Transportation Research (C)*, Vol. 5, No. 1, pp. 39-50.

Ben-Akiva, M. and Lerman, S. (1985) *Discrete Choice Analysis*. The MIT Press.

Bierlaire, M. (2003) *BIOGEME: a free package for the estimation of discrete choice models*, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Bierlaire, M. (2005) *An introduction to BIOGEME* (Version 1.4). December 16, 2005. URL: roso.epfl.ch/mbi/biogeme/doc/tutorial.pdf

Bierlaire, M., Thémans, M. and Axhausen, K. (2006) "Analysis of driver's response to real-time information in Switzerland", *European Transport*, Issue n° 34, December 2006.

Bjornskau, T. and Elvik, R. (1992) "Can road traffic law enforcement permanently reduce the number of accidents?" *Accident Analysis and Prevention,* Vol. 24, pp. 507-520.

Chamberlain, G. (1980) "Analysis of Covariance With Qualitative Data". *Review of Economic Studies*, 47, pp. 225-238.

European Transport Safety Council (ETSC 1999) *Police enforcement strategies to reduce traffic casualties in Europe.*

European Commission (EC 2001) *European transport policy for 2010: Time to decide* (COM(2001) 370 final, 12 September 2001).

Golias, J., Matsoukis, E. and Yannis, G. (1997) "An analysis of factors affecting road safety: The Greek Experience". *Journal of the Institute of Transportation Engineers*, Vol. 67, No. II, pp. 26-31.

B. Han, Algers, S. and Engelson, L. (2001) *Accommodating Drivers' Taste Variation and Repeated Choice Correlation in Route Choice Modeling by Using the Mixed Logit Model*. Proceedings of the 80th Annual Meeting of the Transportation Research Board, January 7-11, 2001, Washington, D.C.

Heckman, J. (1981) "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process", In: *Structural Analysis of Discrete Data with Econometric Applications*. Manski, C. and McFadden, D. editors. MIT Press, Cambridge, MA.

Hess, S., Bierlaire, M. and Polak, J. (2005) "Estimation of value of travel-time savings using Mixed Logit models", *Transportation Research Part A,* 39(3), pp. 221-236.

Holland, C. and Conner, M. (1995) "Exceeding the speed limit: An evaluation of the effectiveness of a police intervention". *Accident Analysis and Prevention,* Vol. 28, pp. 587-597.

Hoorn, A. van der Kroes, E., and Meijer, E. (1984) "Mode choice and mode capacity interlocal commuting: Two Dutch studies on modal split in congested transportation corridors". *Transport policy decision-making,* 2, pp. 373-388.

Hsiao, C. (1986) *Analysis of panel data*. Cambridge University Press, Cambridge, U.K.

Kanellaidis, G., Yannis, G. and Harvatis, M. (1999) "Attitude of Greek drivers towards road safety". *Transportation Quarterly*, Vol. 53, No.1, 1999, pp. 109-121.

Khattak, A., Polydoropoulou, A. and Ben-Akiva, M. (1996) "Modeling Revealed and Stated Pre-trip Travel Response to Advanced Traveler Information Systems". *Transportation Research Record*, 1537, 46.

Lin, L., Pioche, A. and Stander, P. (1986) "Estimating sales volume potential for innovative products with case histories". *Proceeding of the 1986 ESOMAR congress*, Monte Carlo.

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

Luce, R. D. and Tukey, J. W. (1964) "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement", *Journal of Mathematical Psychology* 1, pp. 1-27.

Matsoukis, E. C., Golias, J.C. and Yannis, G.D. (1996) "An Overview of Road Safety in Greece". *Transport Reviews*, vol. 16, no. 4, pp. 357-367.

McFadden, D. and Train K. (2000) "Mixed MNL models of discrete response", *Journal of Applied Econometrics* 15, pp. 447–470.

Newstead, S., Cameron, M., Mark, L. and Leggett, W. (2001) "The crash reduction effectiveness of a network-wide traffic police deployment system". *Accident Analysis and Prevention*, Vol. 33, pp. 393-406.

Polydoropoulou, A., Ben-Akiva, M., Khattak, A. and Lauprete, G. (1996) "Modeling Revealed and Stated En-Route Travel Response to Advanced Traveler Information Systems", *Transportation Research Record*, 1537.

Rothengatter, J. A. and Harper, J. (1991) "The scope and design of automatic policing information system with limited artificial intelligence". *Advanced Telematics in Road Transport*, pp. 1499-1515. Amsterdam: Elsevier.

Rizzi, L. and Ortuzar, J. de D. (2003) "Stated preference in the valuation of interurban road safety". *Accident Analysis and Prevention*, 35 (1), pp. 9-22.

Sándor, Z. and Train, K. (2004) "Quasi-random simulation of discrete choice models". *Transportation Research* Part B: Methodological, Volume 38, Issue 4, May, pp. 313-327.

SARTRE (1998) *Social attitudes to road traffic risk in Europe* - Contextual analysis. *SWOV*, Leidschendam, The Netherlands.

SARTRE (2004) *Making our drivers and roads safer.* Selected results from a European survey. INRETS, October 2004.

Sivakumar, A., Bhat, C. R. and Okten, G. (2005) "Simulation Estimation of Mixed Discrete Choice Models with the Use of Randomized Quasi-Monte Carlo Sequences". *Transportation Research Record*: Journal of the Transportation Research Board 1921, pp. 112-122.

Tay, R. (2005) "General and Specific Deterrent Effects of Traffic Enforcement. Do we have to Catch Offenders to Reduce Crashes?" *Journal of Transport Economics and Policy*, Volume 39, Part 2, pp. 209–223.

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

Washington, S., Karlaftis, M., and Mannering, F. (2003) *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall / CRC.

Yannis, G., Kanellopoulou, A., Aggeloussi, K. and Tsamboulas, D. (2005) "Modelling driver choices towards accident risk reduction", *Safety Science*, Vol.43/3, 2005, pp. 173-186.

Zaal, D., (1984) *Traffic law enforcement: a review of the literature*. Report Number 53. Monash University accident Research Center, Melbourne, Australia.

European Transport \ Trasporti Europei n. 37 (2007): 78-92

Choice of departure station by railway users

G. Debrezion [1*], E. Pels [1], P. Rietveld [1]

[1] *VU University Amstedam, Department of Spatial Economics*
De Boelelaan 1105, 1081 HV Amsterdam
The Netherlands

Abstract

This paper applies a multinomial logit model to the choice of a departure railway station by Dutch railway passengers. This is a relevant theme since about 50% of Dutch railway passengers do not travel via the nearest railway station. The passengers' choices for departure stations are aggregated at the four digit postal code area level. We applied three functional forms for the underlying systematic utility of a station, namely a linear effect of attributes, cross effect of distance and frequency of service, and a translog formulation on distance and frequency of train services. With 3,498 post code areas and 360 railway stations our analysis found consistent effect sizes for distance, frequency of service, intercity status of the station and the presence of park-and-ride facility on the choice of departure station. The effect of distance on the choice of a departure station declines smoothly. The effect of frequency of service is relatively small compared to the effect of distance. A frequency of service increase by a hundred trains per day is equivalent to being 600 m closer to the station. The Intercity status of the station plays the biggest role in the choice of departure station. It has an equivalent effect of a change in 2 km distance or about a frequency of service of 300 trains per day. In addition, the presence of park-and-ride facility in the station poses a sizable effect in the departure station choice. In most cases its effect reaches about 35% of the intercity status effect.

Keywords: Railway station choice; Logit model.

1. Introduction

Railway transport constitutes a sizable share of the daily travel made by the Dutch travellers. The figures from the central bureau of statistics (CBS) in 2002 reveal that railway transportation accounts for about 8% of the over all passenger kilometres. This figure is among the highest shares of railway transport in Europe and the world. In the US the overall public transit share (which includes railway and bus services) is about 2% (bureau of transportation statistics 2005). The modal split of passenger kilometres shares for the fifteen members of the European Union are given also in Table 1. Following Austria and France, the Netherlands has the third highest market share for

[*] Corresponding author: Ghebreegziabiher Debrezion (gdebrezion @feweb.vu.nl)

rail transport. Railway transport is, therefore, expected to be an important travel alternative for the Dutch households.

Table 1: Modal split by country for passenger transport: EU-15 (5 modes) year 2002.

	Passenger kilometer (percentage)				
	CAR	BUS	RAILWAY	TRAM & METRO	AIR
BELGIUM	79.8	9.9	6.0	0.7	3.6
DENMARK	74.3	11.1	6.8	0.0	7.8
GERMANY	78.8	8.6	7.8	0.9	3.9
GREECE	65.9	17	1.4	1.0	14.6
SPAIN	71.2	10.6	4.5	1.2	12.5
FRANCE	83.1	4.5	8.2	1.2	3.0
IRLAND	72.8	12.4	3.2	0.0	11.5
ITALY	80.2	11	5.3	0.6	3.0
LUXEMBOURG	74.7	12.8	5.1	0.0	7.4
NETHERLANDS	81.5	4.1	8.1	0.8	5.5
AUSTRIA	70.7	13.6	8.4	2.8	4.5
PORTUGAL	79.7	8.3	3.1	0.5	8.3
FINLAND	77.7	aw10.3	4.4	0.7	7.0
SWEDEN	74	8.0	7.2	1.8	9.0
UNITED KINGDOM	80.9	5.9	5.1	1.1	7.1

Adapted from EU energy and transport in figures: statistical pocket book 2004.

Once one has decided to travel by train, the logical next question that follows is which station to use as a departure point. The decision on this issue is expected to be based on the assessment of relevant access and station features. Easily accessible railway stations are expected to be chosen more often as departure stations than similar stations that are less accessible. Moreover, the choice of a departure station also depends on the quality of the station itself. The quality of railway station is generally explained by the quality of rail and supplementary services provided by or at the station. The frequency of train services, network connectivity and coverage are some examples of the rail service. The presence of other supplementary facilities such as availability of parking spaces, the park-and-ride possibility, the availability of bike stand and safes also boosts the attractiveness of a station as a departure station.

The revealed choice data for departure stations shows that in the Netherlands about 47% of the cases passengers choose a departure station other than the nearest station to their place of residence. This reveals that there is a need to understand the decision process for a departure station. Understanding the valuation and decision mechanism leading to the choice of a particular railway station as an access (departure) station has several practical implications. In the first place, it enables us to define the catchment area (or market area) for stations. This enhances the predictions of travel demand at the station level. Based on the sensitivity of travellers towards access and station features, it gives a basis to the station operator to increase travellers' turnover. In addition, it can be used as a basis for site selection for new line development or extension planning for existing lines, as well as parking, park and ride facility and feeder public transport operation planning.

In this paper we analyze the choice for departure railway station made by Dutch railway travellers. This will in turn be used in the calculation for a general railway accessibility index for zones where people live. In most real estate price studies railway station accessibility is just given by the distance to the nearest of the railway station from the property in question. However, railway station accessibility encompasses all aspects that are involved in the choice process for a departure station. The general accessibility index for areas based on the choice analysis is expected to perform better than the simple distance proxy for the accessibility to the railway station.

This paper is organized as follows. Section 2 briefly reviews the literature in the area. Section 3 gives the specification of the multinomial logit model that we apply in our choice analysis. This is followed by the description of the data used in our analysis. Section 5 presents the methodology for our analysis. Various specifications for the utility function are considered. Section 6 gives the estimation results followed, by the discussion of these results. Section 7 concludes the analysis.

2. Literature review

The literature in this area is generally scarce. One of the early rail transit station choice models was developed by Kastrenakes (Kastrenakes 1988) in an effort to prepare a basis for forecasting rail ridership in New Jersey area. With origin-destination pair data, Kastrenakes analyzed the choice for a departure station based on access time required to reach the station, frequency of service at the boarding station, whether the boarding station is located in the locality of the residence of the passenger, and generalized cost of the train trip between the departure station and destination station (Kastrenakes 1988). The study found, as expected, positive effects for frequency of service and location of the station in the locality of the passenger's residential area on the choice of departure station choice. Similarly, negative effects were found for access time and the generalized cost of the rail trip as expected. In another study, Wardman and Whelan (1999) studied railway station choice for the inter-urban trips to London. This study was done in relation to parking attractiveness for station choice. It is indicated that availability of parking area in a station and other station facilities are important features for station choice (Wardman and Whelan 1999).

Some studies on this theme have also incorporated access mode choice in a nested structure (Fan et al. 1993; Wardman and Whelan 1999; Davidson and Yang 1999). Generally speaking, the preferred model structure has the access mode choice in the upper level. Fan el al. (1993) included several variables for the lower level transit station choice part. Travel time including access and in vehicle time, fare, peak hour frequency of trains, and the number of parking places were among the included variables. Expected positive signed coefficients for frequency of service and parking and negative signed coefficients for travel time and fare were found. Wardman and Whelan (1999) on the other hand compared the access mode-station choice for business and leisure travels. They found the value of time is highest for business trips and lower for leisure trips. Other variables included were journey time, journey headway, facilities at station and parking availability. They all show expected signs and significant effects on the choice of the departure station.

Choice analysis of this form has been popular in the literature on the choice for a departure airport (Ashford and Bencheman 1987; Hess and Polak 2004; Pels et al. 2001;

Pels et al. 2003; Basar and Bhat 2004). Fares (airport tax), access time, frequency of service, and other facilities are important features used in airport choice. Some studies also include time series historic data in the choice feature of those commuters who tend to use an airport that they have previously used. The analyses on departure airport choice have some relevance to the railway station choice. Most of the time we do not see fare difference between railway stations, so the fare does not play a relevant role in the choice among stations. However, access features like access time and access cost obviously are relevant for the railway station analysis. The frequency of service, as indicated by the number of trains leaving the station per given time and/or the number of destinations served directly from the station, plays an important role in the station choice analysis. The same holds for the nature of the station and facilities at the station. Obviously, international and intercity stations are expected to enjoy higher choice probabilities compared to express or stop train stations[1]. Stations with better public and passenger related facilities are also expected to be attractive compared to stations with lesser or no facilities. As the access time increases, the attractiveness of the station for departure declines.

3. Multinomial Logit model for a station choice

The choice made on which departure station to use forms an important choice decision for the household. In the short run the nature of a station is not expected to vary. Thus, generally speaking, the households' choice remains unchanged. However, in the long run a railway station can undergo major changes that can cause the travellers to look for different departure stations. At the same time households can have multiple destinations and multiple trip purposes. Thus, it is natural to observe different railway stations to be chosen as departure stations for different trip purposes and destination combinations. However, our data does not include information on trip purpose and destination. The analysis in this paper, therefore, will be limited to the aggregate choice of departure stations made by households in a post code area without looking at the purpose of the trip and destination.

We assume that any household's choice for a departure station is based on the assessment of an underlying utility function. That means that depending on the nature of the origin of the trip, each railway station provides certain utility level. These utilities may differ not only on the relative location from the origin, but also because stations differ from each other in their nature. As rational choice makers, travellers select a departure station among a set of feasible alternatives so that the utility is maximized. Different features can enter the utility function. The access distance, i. e. from the origin to the railway station, is an important feature in the departure station choice model. In addition, the service levels provided at the station determined by the frequency of trains, the number of destinations that can be reached from the station and other station facilities like parking and bicycle stands are potentially relevant station characteristics

[1] In the Netherlands there are four types of railway services namely: the all station rail services called stop train; semi fast also called express rail services which call at main and medium cities, inter-city service rail services that only call at main cities, and international trains that only stop at a very limited number of stations.

in the station choice utility function. The utility function can assume several functional forms. Later in our model specification we will give three functional forms for the utility, namely a simple linear additive, linear additive utility with cross distance frequency of service product and transcendental logarithmic function. For the purpose of outlining the multinomial logit model, in this section, we specify the two general components of the total utility function as follows:

$$U_j = V_j + \varepsilon_j \tag{1}$$

Where U_j is the total utility level of alternative j; V_j represents the systematic component of the utility for alternative j and ε_j is the stochastic component of the utility for alternative j.

The probability that alternative j is chosen is given by the probability that the utility corresponding to alternative j exceeds the utility levels of other alternatives. If we assume that ε_j is independently and identically distributed and follows the Gumbel extreme value distribution (see Ben-Akiva and Lerman 1985 for a detailed description of Gumbel extreme value distribution) the probability of choosing station j is:

$$P(j) = P(U_j > U_k) \text{ For all } k \neq j \tag{2}$$

McFadden (1973) has shown that if the distribution of the stochastic component of the utility (ε) follows an extreme value distribution function, the choice situation results in multinomial logit model. The multinomial logit model is a family of discrete choice model, which allows a choice situation with multiple alternatives. The probability of choosing alternative j is thus given by the multinomial logit model given by (3):

$$P(j) = \frac{\exp(V_j)}{\sum_{k \in J} \exp(V_k)} \tag{3}$$

An important property of the multinomial logit model is the independence from irrelevant alternatives (IIA). That means that the ratio of the station choice probabilities of two stations is not affected by the systematic utilities of any other alternatives. This property holds true for the departure station choice situation analyses in this paper.

4. Data

The data used in our analysis were acquired from the Dutch national Railway Company (NS). We employed two data sets: post code area related data and railway station related data. The choice outcome for the departure station choice is aggregated at the post code area level. The Netherlands is composed of 4,004 post code areas. However, because of the data incompleteness our final analysis is based on 3,498 post-code areas. This accounts for 87% of the country. For each of the post code areas a set of three most frequently chosen stations is identified. The overall number of stations covered in the analysis is 367. Due to the aggregation process, the choices for the departure station are explained by the share of each of the three stations receives at the post code area.

GIS information on the location of the centroid for post code areas and the railway stations was used to calculate the distance between the centroid of the post-code centre and the railway stations in the choice set. The distance assumes a Euclidean measure. The data set combines the shares of the choice each of the three stations receive in the postcode area, and the railway station features including the access distance.

Descriptive of Station and Accessibility Characteristics

Railway station accessibility is generally explained by two factors namely, the ease of reaching the stations and the service levels provided at the stations. The ease of reaching the stations is linked to the distance between the departure point (the centroid of the post code area in this case) and the railway station. The level of services provided at the stations is related to the frequency of trains leaving the station per time period and network connectivity as determined by the number of destinations that can be reached without a transfer. In addition, it includes facilities that supplement railway transport. Table 2 below, gives the descriptive statistics of the railway station characteristics and the accessibility indicators for the post code areas. For the purpose of showing the variation, in Table 2 below, we only give the distance to the first most frequently chosen station from the post code area. In addition, Table 2 gives the station features. Included are the indicators of railway service, type of station, and facilities in the station.

Table 2: descriptive statistics for the railway station characteristics (2001/2002).

	Number of stations/ post code areas	*Min*	*Max*	*Mean*	*Std. Deviation*
Rail service					
Frequency of trains per day		18	788	113	103
Destinations reached without a transfer		1	114	16	14
Station type					
Inter-city stations	64			0.18	
Station Facilities (dummies)					
Train taxi	109			0.30	
Bicycle stand	96			0.27	
Bicycle safe	264			0.74	
Bicycle rent	114			0.31	
Park-and-ride	49			0.14	
Parking	326			0.91	
Taxi	163			0.45	
Car rent	1			0.00	
Luggage deposit	64			0.18	
International connection	22			0.06	
Accessibility from Post code areas					
Distance to the most frequently chosen station (m)	3498	95	34250	6756	6129

5. Methodology

Our analysis identifies several factors which have an impact on the choice for a departure station. One of these factors is the distance from the postcode area to the railway station. We apply categories of distance classes in our analysis. To see the smoothness of the effect we use a 500 meters range categories, except in the two inner circle categories of the station, which are 250 meters each. Thus we have 21 categories of distances up to 10,000 meters, where the category above 10,000 meters is taken as reference group. Each of these categories is represented by dummy variable. We identify a number of station facilities that have utility bearing nature in the general sense. However, these facilities can be departure station related or destination station related in nature from the passenger's point of view. For instance a car rental facility is a typical destination station feature. A parking lot on the other hand is a departure station feature. As our main focus here is the choice of a departure railway station, we stick to the departure station related features of the station.

Departure related features of a station that can be used in our analysis include the frequency of train, the number of direct destinations, the availability of a bicycle stand, the presence of a park-and-ride facility, the availability of parking and the type of railway station, and inter city status. Frequency of service and number of direct destinations are expected to be highly correlated; we, therefore, only include frequency in our analysis. Bicycle stands are not expected to affect the choice of departure station, since in general passengers will find a place for their bicycle anyhow. In addition, we include the presence of park-and-ride facility, and the intercity status of the station. All the features are expected to influence the choice of a departure station positively. However the effect of intercity status of a station can differ regionally. Thus, we use it as specific to the provinces of the country. The Netherlands is made up of 12 provinces. The province of Utrecht, North Holland and South Holland constitute the most urbanized area of the country. This region is mostly called as the Randstad (rim city), since it constitutes an extended chain of cities. The effect of the inter city status of a station on the choice of a departure station is expected to be less in these regions as compared to the effect of inter city status of a station for the choice of a departure station in the more peripheral provinces.

Specification of the systematic utility function

For comparison purposes we give three functional specifications for the systematic utility of a station: the linear additive, linear additive utility with cross distance frequency of service product and transcendental logarithmic formulation. The first is aimed at capturing the separate effects of distance and other station features. The second is aimed at determining the effect of train service frequency on the choice of a departure station at different distance segments. The transcendental logarithmic station utility function is aimed at determining the general smoothed utility function for the station choice. They are given respectively in equations 4, 5, and 6.

$$V_j = \sum_{c=1}^{21} \beta_c * Dcateg_{jc} + B_f * freq_j + B_{P\_R} * P\&R_j + \sum_{p=1}^{12} B_{IC-p} * IC_j * Prov_{jp} \quad (4)$$

$$V_j = \sum_{c=1}^{21} \beta_{cfreq} * Dcateg_{jc} \times freq_j + B_{P\_R} * P \& R_j + \sum_{p=1}^{12} B_{IC-p} * IC_j * Prov_{jp} \quad (5)$$

$$\begin{aligned} V_j = {} & \beta_{dist} \ln(dist_j) + \beta_{distsq} (\ln dist_j)^2 + \beta_{freq} \ln(freqT) + \beta_{freqsq} (\ln freqT_j)^2 \\ & + \beta_{distfreq} (\ln dist \times \ln freq_j) + \beta_{P\&R} P \& R_j + \sum_{p=1}^{12} B_{IC-p} * IC_j * Prov_{jp} \end{aligned} \quad (6)$$

Where $Dcateg_{jc} = 1$ if station j is located in the distance category c from the centre of the post code area and zero otherwise; dist is the distance between the centroid of the post code area and the railway station in continuous measure; $freq_j$ is the frequency of trains at station j; $P \& R_j$ is an indicator for the presence of park-and-ride facility at station j; IC_j is an indicator for an intercity status of station j; $Prov_{jp}$ is a dummy variable for the province in which station j is located. It takes the value 1 if station j is located in province p, and 0 otherwise.

6. Estimation results

1. Linear additive utility function: Piecewise distance measure

Here we formulate the utility function for the railway stations in a linear additive way. Distance is measured in a piecewise fashion. This gives a detailed effect of distance compared to other continuous treatment of distance, because the area over which averaging is made is quite limited. In addition we include the frequency of trains at the stations, the availability of park-and-ride and the IC status of the station into the utility function. The systematic utility specification is given by equation 4 above. The estimation output for the multinomial logit model for the station choice is given below in Table 3. All coefficients are significant, with a dominant role for the distance effect. The effect of distance as expected has a positive sign. The values of the coefficients for the distance categories are relative to the zero reference value for areas beyond 10 km.

The value of the pseudo R-square (given as $1 - \log L_{\text{estimated model}} / \log L_{\text{base model}}$), used as a measure of goodness of fit, shows the model has a good prediction power. According to the empirical relationship drawn (Domencich and McFadden 1975) against the R-square of the linear models, the R-square of the above logit model is equivalent to an R-square close to 0.8, in the linear regression models.

From the value of the coefficients we observe a smooth decline in the effect size with distance. That means that the closer the postcode area is to the station, the higher the probability of choosing that station as a departure station. The value found for the frequency of service effect is relatively small compared to the effect of distance. A frequency of service increase by a hundred trains per day is equivalent to being about 600 meters closer to the station.

Table 3: Multinomial logit model estimation for station choice.

Variable	*Coefficient*	*t- value*	*p-value*
DIST250	7.379	7.360	0.000
DIST250_500	6.852	16.190	0.000
DIST500_1000	6.328	27.154	0.000
DIST1000_1500	5.734	27.195	0.000
DIST1500_2000	5.147	26.058	0.000
DIST2000_2500	4.797	25.488	0.000
DIST2500_3000	4.150	22.916	0.000
DIST3000_3500	3.723	21.676	0.000
DIST3500_4000	3.411	19.758	0.000
DIST4000_4500	2.940	17.630	0.000
DIST4500_5000	2.765	16.837	0.000
DIST5000_5500	2.471	15.274	0.000
DIST5500_6000	2.282	13.505	0.000
DIST6000_6500	1.953	11.354	0.000
DIST6500_7000	1.902	11.776	0.000
DIST7000_7500	1.748	9.901	0.000
DIST7500_8000	1.626	9.273	0.000
DIST8000_8500	1.487	8.635	0.000
DIST8500_9000	1.143	6.347	0.000
DIST9000_9500	1.237	6.682	0.000
DIST9500_10000	0.955	4.873	0.000
Frequency	0.004	12.633	0.000
Park & Ride	0.419	6.226	0.000
IC-Groningen	1.947	7.875	0.000
IC-Friesland	1.785	8.318	0.000
IC-Drenthe	1.510	3.841	0.000
IC-Overijssel	1.101	5.038	0.000
IC-Gelderland	1.052	5.751	0.000
IC-Utrecht	-0.272	-0.970	0.332
IC-North Holland	0.724	3.427	0.001
IC-Zuid-Holland	0.534	3.041	0.002
IC-Zeeland	3.235	5.436	0.000
IC-Noord-Brabant	0.971	6.340	0.000
IC-Limburg	1.769	9.173	0.000
IC-Flevoland	-3.054	-3.446	0.001
Number of observations 3396			
Log likelihood function -2312.613			
R-sqrd 0.38014 RsqAdj 0.37693			

The province specific coefficients for the intercity status of a station have generally positive effect with the exception of the province of Flevoland. A positive coefficient shows that intercity status has a positive effect on the choice of a departure railway station compared to a non intercity station. As expected the coefficients for the provinces making up the Randstad area, have lover value. For example, the coefficient for the province of Utrecht is found to be insignificant. This indicates that an intercity status of a station does not make a significant difference in the choice for a departure station in the area as compared to the non intercity stations. This finding makes much sense because this province is in the heart of the Netherlands and the most accessible one. The provinces with the highest value for the intercity status include Zeeland, Groningen and Friesland. These are the peripheral provinces of the country. Generally the intercity status of a railway station has a big effect on the choice for a departure station. On average, the effect is equivalent to the frequency effect of about 300 trains per day. Further, the availability of a park-and-ride facility has a sizable impact on the choice of a departure station. It has an equivalent effect of about 105 trains per day for the frequency of service level.

2. Linear additive utility function: piecewise distance and frequency of trains cross product

This model is a slightly adapted form of the above formulation. The formulation is aimed at assessing the effect of frequency of trains on the choice of a departure station at the different distance categories. Thus, distance categories are cross multiplied with frequency. This approach gives some flexibility in allowing the effect of frequency on the choice of a station to differentiate across distance categories. The systematic utility specification is given by equation 6 above. The estimation results for this specification are given in Table 4 below.

Table 4: Cross product of distance categories and frequency of trains.

Variable	*Coefficient*	*t-value*	*p-value*
FRQ250	0.0717	6.183	0.000
FRQ250_500	0.0549	13.171	0.000
FRQ500_1000	0.0441	17.907	0.000
FRQ1000_1500	0.0326	17.739	0.000
FRQ1500_2000	0.0249	16.651	0.000
FRQ2000_2500	0.0196	16.781	0.000
FRQ2500_3000	0.0167	16.195	0.000
FRQ3000_3500	0.0146	16.102	0.000
FRQ3500_4000	0.0103	12.564	0.000
FRQ4000_4500	0.0087	12.227	0.000
FRQ4500_5000	0.0069	9.411	0.000
FRQ5000_5500	0.0065	9.053	0.000
FRQ5500_6000	0.0053	7.599	0.000
FRQ6000_6500	0.0046	5.798	0.000
FRQ6500_7000	0.0043	6.602	0.000
FRQ7000_7500	0.0040	5.031	0.000
FRQ7500_8000	0.0036	5.691	0.000
FRQ8000_8500	0.0032	5.069	0.000
FRQ8500_9000	0.0026	3.765	0.000
FRQ9000_9500	0.0026	3.550	0.000
FRQ9500_10000	0.0016	1.825	0.068
Park & ride	0.2509	4.091	0.000
IC-Groningen	0.8691	4.304	0.000
IC-Friesland	1.0502	5.921	0.000
IC-Drenthe	1.4540	4.616	0.000
IC-Overijssel	0.3014	1.695	0.090
IC-Gelderland	0.3964	2.540	0.011
IC-Utrecht	-0.2983	-1.118	0.264
IC-North Holland	0.4322	2.138	0.033
IC-Zuid-Holland	0.4792	3.263	0.001
IC-Zeeland	2.7405	5.295	0.000
IC-Noord-Brabant	0.6605	5.071	0.000
IC-Limburg	0.7897	4.931	0.000
IC-Flevoland	-1.9936	-2.284	0.022
Number of observations	**3396**		
Log likelihood function	**-2693.256**		
R-sqrd .27812 RsqAdj .27449			

From the table we see that the frequency of trains at a station has the expected effect sign and pattern on choice of a departure station across the distance categories. The effect size at each segment is given in comparison to the base category of frequency of service at distances of more than 10 kilometres. Frequency is given in trains per day. Similarly, the regional effect of intercity status of a station on the choice of a departure station is higher on the peripheral provinces as compared to its effect on the Randstad area.

In Figure 1 below we show the smoothness of distance and frequency effects on the choice of a departure station given in tables 3 and 4. On the Y-axis we have the size of the coefficients and on the X-axis we have the distance categories.

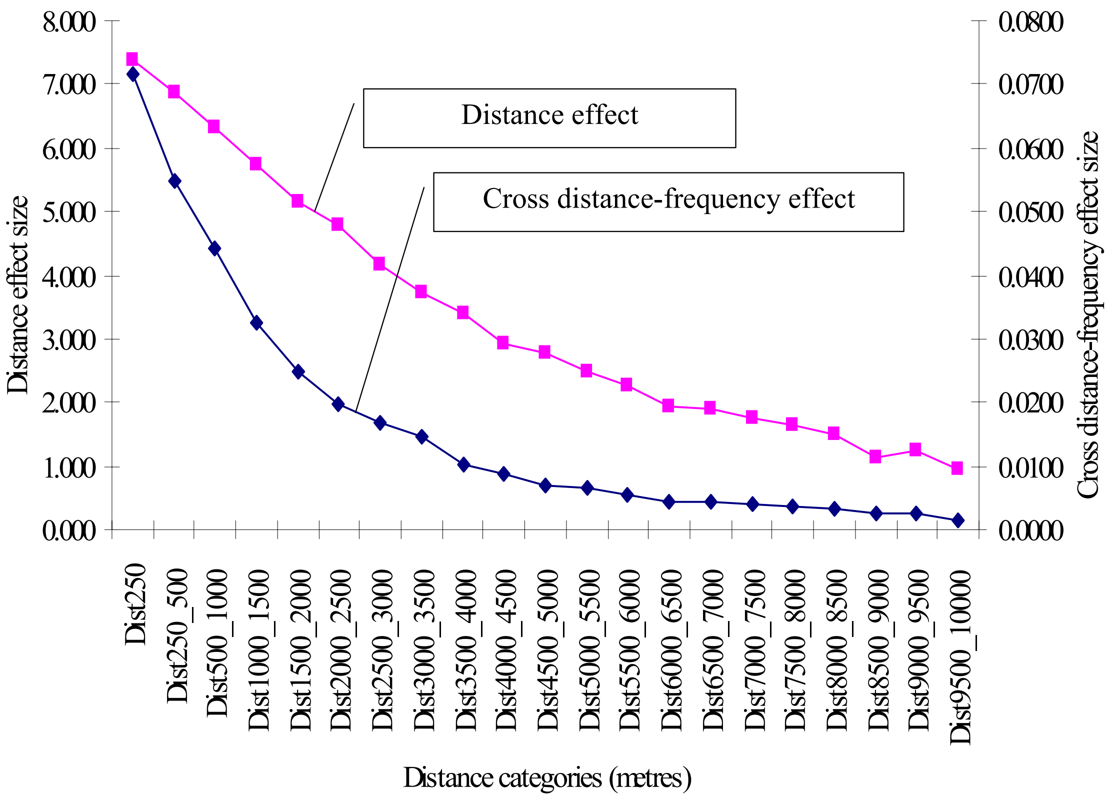

Figure 1: Distance and frequency effect on the departure station utility.

Comparing the two models we conclude that a tendency can be discerned that the effect of frequency improvements on the utility of a station is larger for residents that live nearby a station compared with living further away. However, it should be noted that the fit of the model without the cross effect is substantially higher than that of the model with the cross-effect (.29 versus .38). In order to shed more light on the issue we explore a third specification via the translog function which allows a more refined analysis of interaction effects.

3. Transcendental logarithmic formulation

In addition to the linear additive formulation of the utility function we used the transcendental logarithmic formulation. Here, distance, frequency of trains and park-and-ride variables are included. Distance is treated in the continuous form. The use of the translog function does not give a detailed treatment of the effect of distance as the stepwise treatment of distance does. The translog model is better in dealing with the effect of frequency, in particular the extent to which frequency effects are different for a traveller close to stations and a traveller further away. The systematic utility for this formulation is given by equation 7 above. The multinomial logit estimation output based on this specification is given below.

Table 5: Transcendental logarithmic utility function.

Variable	*Coefficient*	*t-value*	*p-value*
Log (distance)	1.341	2.305	0.021
Log (square distance)	-0.826	-1.224	0.221
Log (frequency)	-0.124	-1.954	0.051
Log (square frequency)	-0.192	-4.771	0.000
Log (distance) * log (frequency)	0.327	5.867	0.000
Park & ride	0.369	5.149	0.000
IC-Groningen	2.067	7.990	0.000
IC-Friesland	2.111	9.360	0.000
IC-Drenthe	1.743	4.027	0.000
IC-Overijssel	0.483	2.121	0.034
IC-Gelderland	0.584	2.968	0.003
IC-Utrecht	-0.507	-1.786	0.074
IC-North Holland	0.446	2.034	0.042
IC-Zuid-Holland	0.306	1.660	0.097
IC-Zeeland	3.260	5.305	0.000
IC-Noord-Brabant	0.611	3.814	0.000
IC-Limburg	1.362	6.627	0.000
IC-Flevoland	-2.623	-2.951	0.003
Number of observations	**3396**		
Log likelihood function	**-2195.868**		
R-sqrd .41144 RsqAdj	**.40987**		

Mapping the above output enables us to see the utility level for the different levels of frequency at different distances from the station. The graphical illustration is given in Figure 2 below. This confirms the relevance of the cross-product specification used above. Note also that the fit of model 3 is clearly better than that of models 1 and 2, even though the number of parameters is smaller. From the figure we see that the utility level for the stations smoothly declines with distance for all frequency levels. In addition, the curve corresponding to higher frequency of trains assumes a flatter pattern, indicating that the catchment area for the station expands with an increase in frequency of service at the station. The curves represent the combined utility contribution of the distance and frequency of service from the translog station utility formulation.

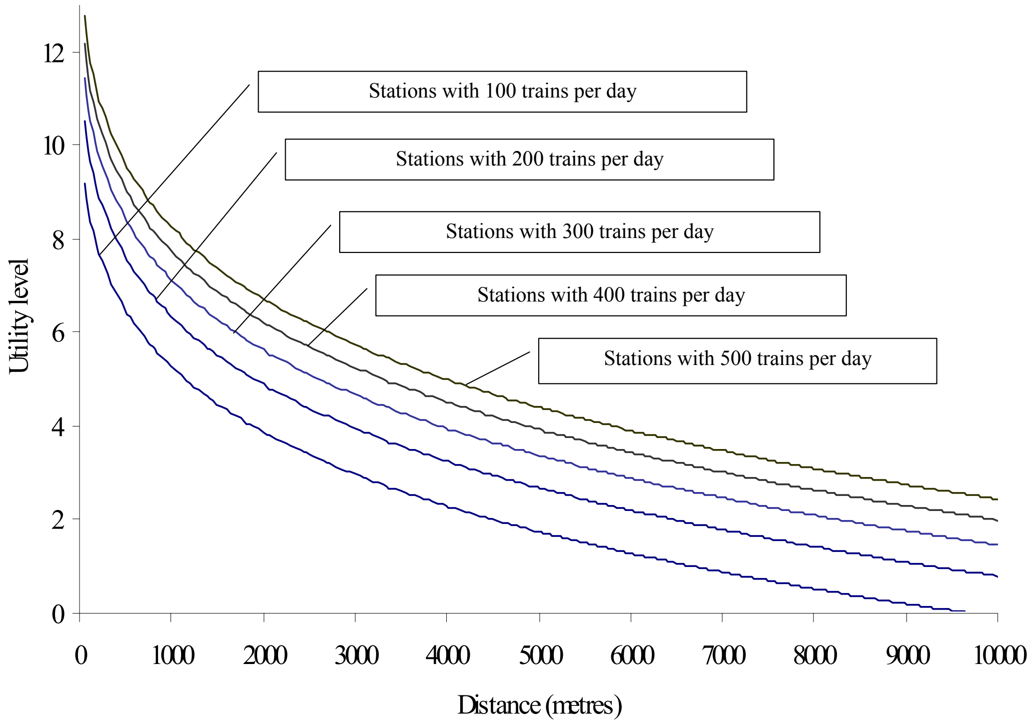

Figure 2: Railway station utility for different utility levels.

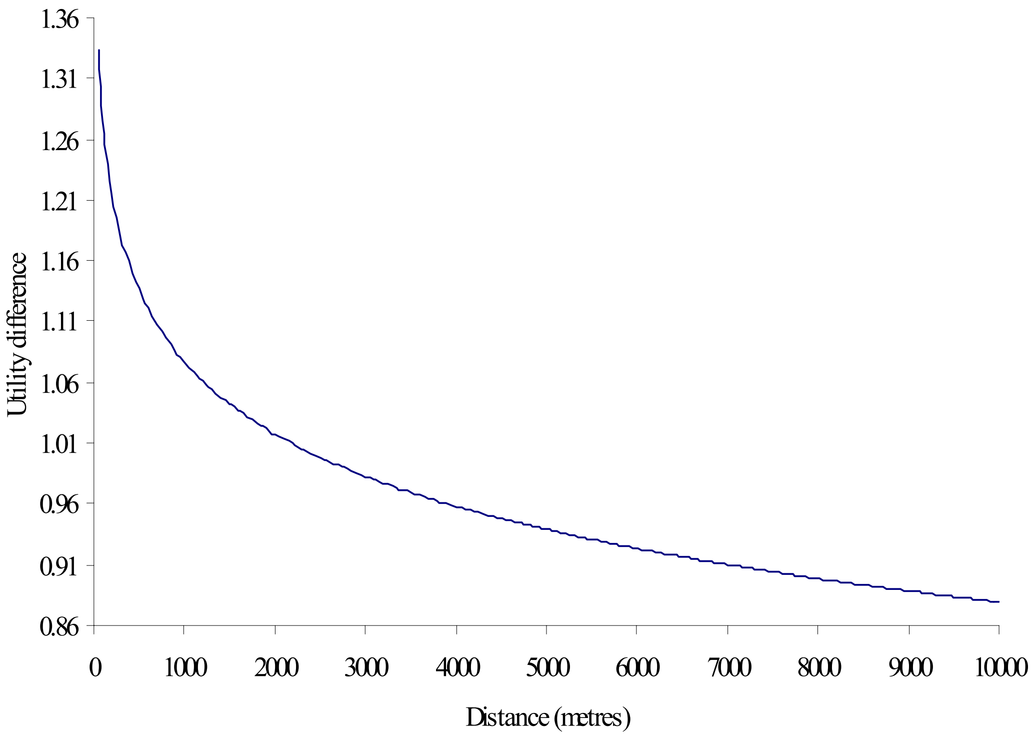

Figure 3: The effect of an increase in frequency from 100 to 200 trains per day on utility for a range of distances.

The lower curve corresponds to a frequency level of 100 trains per day, whereas the outer curve corresponds to a frequency of 500 trains per day. The frequency interval between the curves is fixed to 100 trains per day to facilitate comparison on the effect of an additional train. As one moves upward, the curves get closer. Thus, the graph reveals there is a diminishing effect of increasing frequency of trains on the utility level.

We can also show that the effect of frequency on utility declines with distance. Taking the difference in utility between frequency level 100 and 200 and mapping it with distance gives the curve in Figure 3 below.

7. Conclusion

This paper discusses the choice of a railway station as a departure station. Aggregated choices of households at the post code level were analyzed. For each post code area a set of three alternative stations was determined. We applied a multinomial logit model to determine the choice process. A number of access and station features were included in the station utility function. Distance between the centroid of the postcode area and the railway station was taken to include all access features. The station features considered in the utility function include frequency of service at the station, availability of park-and-ride facility and the intercity status of the railway station. Applying three model specifications for the utility function (linear additive, linear additive with the cross product of distance and frequency of service and the transcendental logarithmic), we found that all access and station features have significant expected effect sign. The probability of choosing a particular station declines with distance: a nearby station has a higher probability of being chosen. Moreover, the higher the frequency of service at a station, the higher is the probability that the station is chosen for departure. On the other hand, we found that the effect of frequency of a change in service declines with distance. The effect of frequency is higher for closer post code areas than for post code areas farther away. The intercity status of the station plays the biggest role in explaining the choice of a departure station. The intercity status of a station has on average an equivalent effect of a decrease of 2 km in distance or an increase in frequency of 300 trains per day. In addition the presence of a park-and-ride facility in the station poses a sizable effect in the departure station choice. In most cases its effect reaches about 35% of the intercity status' effect.

Our model may however suffer from endogeneity problem. The choice of the station stems on the characteristics of the station and other facilities in the station. Most of the time the facilities at the station can be explained by the demand for travel the station generates, which in return corresponds with the choice variable. For example decisions for parking lots around stations are made on the basis of the number of commuters accessing the station. Endogeneity issues will be explored in a more advanced model.

These results may be further developed in various directions. One is to correct for possible endogeneity in the explanatory variables such as the park-and-ride variable. A second and more drastic development would be the use of more detailed data on access mode, which would allow for the estimation of a joint access mode and railway station choice model.

Acknowledgement:

The authors thank Mark van Hagen and his colleagues at the Dutch railway company (NS) for providing us with the necessary data and comments on an earlier draft. This paper was written in the context of three research programs: NWO-VEV, Transumo and Habiforum.

References:

Ashford, N. and Benchemam, M. (1987) "Passengers' choice of airport: an application of the multinomial logit model". *Transportation Research Record,* 1147, pp. 1-5.

Basar, G. and Bhat, C. (2004) "A Parameterized Consideration Set Model for Airport Choice". *Transportation Research Part B,* 38, pp. 889-904.

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

Davidson, B. and Yang, L. (1999) "Modeling commuter rail station choice and access mode combinations". Presented at the *Transportation research Board* 1999 annual meeting Washington D.C.

Domencich, T. and McFadden, D. (1975) *Urban Travel Demand a behavioural Analysis.* Elsevier, New York and Oxford.

European Commission (2004) *EU Energy and Transport in Figure: statistical pocketbook 2004*

Fan, K., Miller, E. and Badoe, D. (1993) "Modeling Rail Access Mode and Station Choice". *Transportation Research Record,* 1413, pp. 49-59.

Hess, S. and Polak, J. W. (2005) "Mixed Logit Modelling of Airport choice in Multi-Airport Regions". Presented at the 45th annual congress of the European regional Science Association.

Kastrenakes, C. R. (1988) "Development of a Rail Station Choice Model for NJ Transit" *Transportation Research Record,* 1162, pp. 16-21.

Lythgoe, W. F., Wardman, M. and Toner, J. P. (2005) *Enhancing Rail Passenger Demand models to Examine Station Choice and Access to the Rail Network.* (draft).

McFadden, D. (1973) "Conditional Logit Analysis of Qualitative Choice Behavior". In: Zarembka, P. (ed.), *Frontiers in Econometrics*. Academic Press, New York.

Pels, E., Nijkamp, P. and Rietveld, P. (2003) "Access to and competition between airports: a case study for the San Francisco Bay area". *Transportation Research Part* A 37, pp. 71-83.

Wardman, M. and Whelan, G. A. (1999) "Using Geographical information systems to improve rail demand models". Final report to Engineering and Physical science Research council.

European Transport \ Trasporti Europei n. 37 (2007): 93-104

Bus speed estimation by neural networks to improve the automatic fleet management

G. Salvo [1], G. Amato [1], P. Zito [1*]

[1] *Università degli Studi di Palermo, Dipartimento di Ingegneria dei Trasporti*

Abstract

In the urban areas, public transport service interacts with the private mobility. Moreover, on each link of the urban public transport network, the bus speed is affected by a high variability over time. It depends on the congestion level and the presence of bus way or no. The scheduling reliability of the public transport service is crucial to increase attractiveness against private car use. A comparison between a Radial Basis Function network (RBF) and Multi layer Perceptron (MLP) was carried out to estimate the average speed, analysing the dynamic bus location data achieved by an AVMS (Automatic Vehicle Monitoring System). Collected data concern bus location, geometrical parameters and traffic conditions. Public Transport Company of Palermo provided these data.

Keywords: Radial Basis Neural Network; Public Transport Performances; AVM system.

1. Introduction

Liberalization and privatisation of the public transport impose the adoption of appropriate strategies to improve the efficiency and the competition. The rationalization of the resources and the competition among different sectors (costs reduction, waste elimination, profits increase, improvements of the perceived service quality) have required the use of telematics and control systems for the automatic fleet management. Furthermore, an improved service quality of the public transport can attract new users. Under this aspect: by the company's view point, it produces an increase of revenue; whereas by the citizens' view point, it implies a reduction of the travel time and delay; whose effects can be led up to an increase the life quality, de-congestioning urban areas, and reduction of noise and air pollution levels.

The recent advances in information and communication technologies, which support phases of collection, storage, processing and dissemination of information, allow to have spatially referenced data. Intelligent Transport Systems (ITS) defined as: *the*

[*] Corresponding author: Pietro Zito (zito@ditra.unipa.it)

application of communications and information technologies in the sector of transport, allows:

- to improve mobility of vehicles fleet (by i.e. Automated Vehicle Monitoring, bus priority system), increasing transport safety, maximizing the performance and the comfort, and reducing conflict points with private cars;
- to keep travellers/drivers more informed (with pre and on trip information by i.e. internet kiosks, variable message sign at the bus stop, mobile phone, on board speaker and so on);
- to improve the accessibility transport infrastructures increasing their capacity, optimizing the traffic flows and keeping the existing transport/traffic infrastructures;
- to improve traffic management and increase the average travel speeds;

The main outcome of ITS applications, for the public transport has been a new approach for the management of fleet as such as AVMS (Automatic Vehicle Monitoring System). An AVMS is generally composed: by GPS (Global Position System) that updates the fleet vehicles' positions frequently, by remote control system (with an equipment of sensors on-board) to check the safety and security of vehicles and by a communication system that broadcasts information to the Control Centre (CC), whose task is essentially to coordinate the vehicles fleet. The CC can also receive real-time updates on the state of the road network (i.e. an incident on a link) and uses together with position data to determine a new scheduling of the affected bus. The AVMS allows to communicate in real time to the Control Centre the exact vehicle's position and various other parameters of each vehicle as load condition on board, vehicle diagnostic conditions (check of the engine and closure of doors).

The main task of Control Centre is to check the positions on the road network of the transit fleet, respect to scheduled ones at the bus stops. Whenever delays due to traffic congestion or difficulties in loading and unloading passengers occur on public transport network, the CC can provide alternative solutions. Another specialised application of AVMS is the emergency service, to improve security of the driver and passengers (in case of crimes on-board and/or an incident). Whether an incident occurs involving the bus, the control system will broadcast automatically its position directly to security services (police, ambulance, fire fighters, etc.). Moreover, this service is particularly useful in rural areas or in case of night incident. Thus, the AVMS allows of having a whole control of bus fleet in real-time, improving quality service under the following aspects:

- to check any non recurrent events (incidents, dangerous situations, etc.), taking the right decision to keep a high service level by on-line fleet management data;
- to plan a demand-oriented service (routes, frequencies, etc) by off-line fleet management data;
- to up-date information provided for customers, as such as forecasted arrival/departure time at the bus stop, by using various communication devices (mobile phone, web site, variable message sign at bus stops, etc).

In the time scheduling of the public transport service, the bus speed is assumed constant on every road link, nevertheless the speed varies over time. It depends essentially on traffic and road conditions that in particular cases of traffic congestion can cause service reliability issues and delays on routes.

In literature, the automatic vehicle location (AVL) systems appeared since 1990, they used data coming from AVL for traveller information (Balogh 1993, and Daley 1994).

Furthermore, Daley (2003) forecasted transit arrival/departure time using Kalman Filter and automatic vehicle location data. Data coming from the AVMS allows the estimation of the average speed of buses on any road link of public transport network, planning scheduled times on each route in a more reliable way. This work aims to point out a novel approach to estimate the average speed of buses using data coming from an AVMS and elaborated by a neural network. The knowledge of the bus speed allows to plan and to keep the maximum levels of service improving its quality (efficiency, frequency, and regularity of the courses). In literature, many works were carried out to estimate the arc cost (running time) for private cars by the knowledge of the road characteristics and traffic conditions (slope, width of lane, tortuosity, number of intersection per km, flow and capacity (for further details see High Capacity Manual, 1990, and Cascetta, 1998). Nevertheless, for the public transport and in particular for bus service, similar relationships have been used by a car equivalency factor. Thus, these relationships can be improved by using tools (as neural models) which better fit to the real traffic conditions. The main task of the analysis is to model the bus average speed in terms of the road characteristics and traffic conditions, by using neural network.

The neural networks for the transport sector were mostly applied to forecast short-term traffic flows Mussone (1996), Dougherty (1996, and 1997), Chen (2001) and Yin et al. (2002), using a feed forward neural network and the back propagation learning algorithm or hybrid approaches (neurofuzzy and Kohonen maps). Neural networks were also implemented to estimate the average journey time (Tafti, 2001 and La Franca et al. 2004). Amin et al. (1998) used a radial basis function network to predict the flow of traffic. In 1997, Costa and Markelos appraised the public transport efficiency by neural networks. More recently Celikoglu and Cigizoglu (2007) developed a radial basis function to forecast daily trip flows of public transport.

The neural networks can approximate highly non-linear functions and do not require a prior knowledge on the nature of these relationships. In this work, a comparison was carried out between different neural paradigms namely: Multi Layer Perceptron (MLP) and Radial Basis Function network (RBF). Two neural networks were used to estimate the average speed analysing the dynamic bus location data achieved by AVMS. The Radial Basis Function network (RBF) is an effective method to solve complex mapping problems between input and output variables. Furthermore, few works based on RBF networks are present in literature among which it is notable Celikoglu's one.

In this paper, the collected data concern (used as input variables in our models): bus locations, geometrical parameters and traffic conditions on each link of public transport network (number and width lanes, rate flow/capacity, reserved lane, number of intersections with or without traffic lights per kilometre, number of bus stops per kilometre, legal or illegal parking on the way, etc.). Public Transport Company of Palermo provided them. Moreover, the calibrated neural networks provide the estimated speed for two traffic conditions (rush hour or off peak conditions).

The paper has the following structure: next section describes the RBF network and its advantages; description of input data is included in Section 3, Section 4 presents the methodology, whereas Section 5 shows results of the model proposed and finally Section 6 discusses conclusions and further research direction.

2. Multi Layer Perceptron and Radial Basis Function Networks

The multi layer perceptron (MLP) or feed forward neural network consists of a set of simply interconnected neurons. The output of each neuron (except those in the input layer) is scaled by the connecting weights, modified by a transfer or activation function which can be either linear or non-linear, and fed forward to be an input to the neurons in the next layer of the network.

The multi-layer feed-forward neural networks (MLP) have the well-known ability to learn through training. The back propagation algorithm (BP) is usually used to train MLP networks. The learning is based on some nonlinear optimization techniques, which avoid during the learning period of falling into local minima. During training, the neural network is repeatedly presented with the training data and the weights in the network are adjusted until the output vector, produced by the network, does not match the target vector within a certain error. The training process uses this error to adjust the weights of the network according to the gradient descent learning algorithm to minimise the error.

An RBF network is regarded as a feed-forward neural network with a single layer of hidden units, whose responses are the outputs of radial basis function. Broomhead and Lowel (1988) developed the concept of Radial Basis Function (RBF) networks and applied it in the neural network modelling. The radial basis neural networks are another class of neural networks mainly used to approximate functions and for cluster classification. An RBF network not only has good performance of generalization, but it avoids the over-elaborated, lengthy computing like Back Propagation (BP) algorithm. Its learning speed is 10^3– 10^4 time faster than the BP algorithm, and this makes the RBF network applied widely.

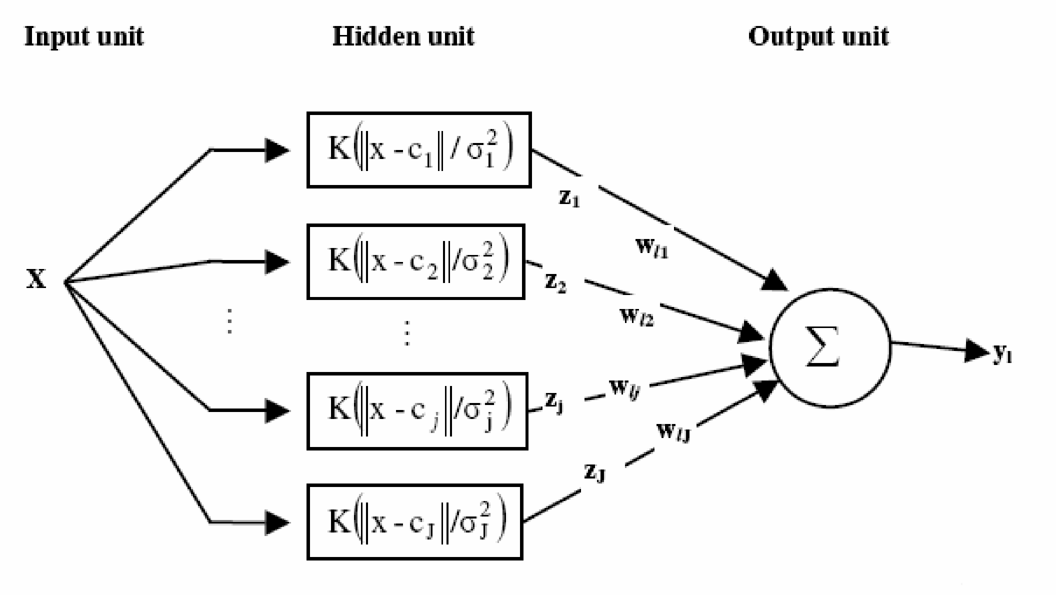

Figure 1: The structure of an RBFNN.

The adjustable parameters of such networks are the centres (the location of basis functions), the width of the receptive fields (the spread), the shape of the receptive field and the linear output weights. The Figure 1 shows the structure of an RBFNN. Radial basis networks have their origin in techniques for performing exact interpolation of a data set. The input of each radial basis function of a RBF neural network is the distance between the input vector (activation) and its centre (location). The radial basis function approach introduces a set of N basis functions, which take the form $\phi(\| x - x_n \|)$, where ϕ is a non-linear function. Thus the n^{th} function depends on the Euclidean distance $\| x - x_n \|$ between x and x_n. A basis function commonly used is the Gaussian:

$$\phi(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{1}$$

where σ is a parameter whose value controls the smoothness of the interpolating function.

The radial basis function networks are similar to the feed forward neural networks and assume the following form:

$$y_k(\mathbf{x}) = \sum_{j=1}^{M} w_{k,j}\phi_j(\mathbf{x}) + w_{k0} \tag{2}$$

$$\phi_j(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\mu}_j\|^2}{2\sigma_j^2}\right) \tag{3}$$

where $\mathbf{x}$ is the n-dimensional input vector with elements x_i, and μ_j is the vector determining the centre of basis function ϕ_j and has elements $\mu_{j,i}$. By matrix notation, we can write:

$$\mathbf{y}(\mathbf{x}) = \mathbf{W}\Phi \tag{4}$$

where $\mathbf{W} = (w_{k,j})$ and $\mathbf{\Phi} = (\phi_j)$. The weights can be achieved by minimization of a sum of square error function given by:

$$E = \frac{1}{2}\sum_n \sum_k \left\{y_k(\mathbf{x}^n) - t_k^n\right\}^2 \tag{5}$$

where: t_k^n is the target value for output unit k when the network is presented with input vector $\mathbf{x}^n$. An important aspect of radial basis network is the distinction between the roles of the first and second layer of weights. The first layer weights contain the parameters governing the basis function (σ_j and μ_j), which can be determined through unsupervised training. The second phase consists to find the weights $w_{k,j}$ of second layer, keeping fixed the parameters of radial basis functions, by a supervised training. Therefore, an unsupervised procedure allows the optimisation of the basis function parameters, which depend only on the input data from training set, and which ignores any target information. The basis function centres μ_j can be considered as prototypes of the input vectors. The determination of the weights $w_{k,j}$ of second layer, knowing the target vector, is achieved by a supervised procedure, resolving a linear matrix equation (Bishop, 1995):

$$\mathbf{W}^{\mathbf{T}} = \mathbf{\Phi}^{\dagger}\mathbf{T} \tag{6}$$

where: $\mathbf{T} = (t_k^n)$, $\mathbf{W}^{\mathbf{T}}$ is the transpose matrix of weights, and $\mathbf{\Phi}^{\dagger}$ is the pseudo inverse of $\mathbf{\Phi}$. Girosi and Poggio (1990) showed that radial basis function networks possess the property of best approximation. The optimum number of nodes required in the hidden layer is related to the complexity of the input and output mapping, the amount of noise in the data and the amount of training data available. During training process, the RBF network will automatically add nodes into the hidden layer when needed until the actual error is below the given one.

3. Data collection and analysis

The main data source was the Public Transport Company of Palermo (AMAT). Dynamic bus location data along some bus routes were achieved by AVMS implemented by AMAT for fleet management. Collected data concern bus positions, geometrical parameters and traffic conditions.

Let **F/C** be the rate flow/capacity; **RL** the reserved lane (dummy variable 1 if present 0 otherwise); **NJ** the number of intersections with or without traffic lights per kilometre; **NB** the number of bus stops per kilometre; **IP** the illegal parking on the way (dummy variable); **FP** the free parking (dummy variable); **NI** the number of inflows per kilometre; **NO** the outflows per kilometre; **PC** the number of pedestrian crossings; **CA** with or without commercial activities (dummy variable). The neural networks provide the estimated speed for various traffic conditions (rush hour or off peak conditions) on each link of public transport network.

A cleaned dataset of 112 data points, each representing 40-seconds location of the line number 102 along its route, was used for the analysis (see figure 2).

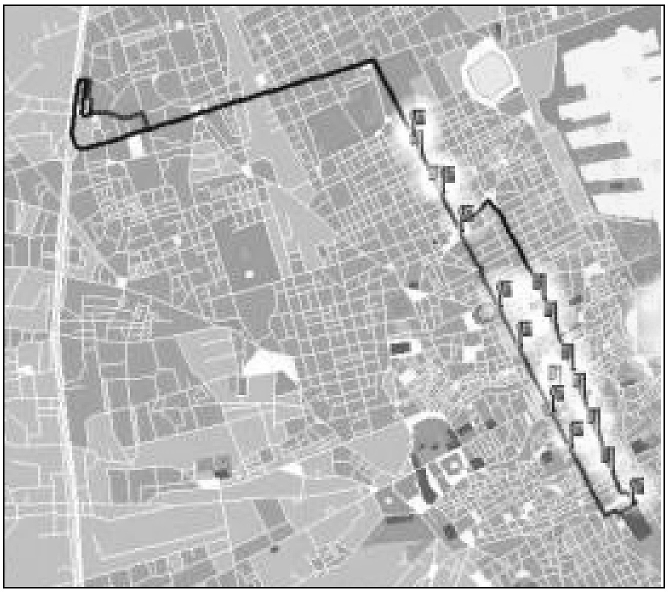

Figura 2: The route of the bus line number 102.

After a statistical analysis of position data was carried out in order to delete outlier points. For each time interval between two positions of the bus in succession by GPS, the travel time was estimated. Thus, the average speed of the bus was computed for every time interval. The average bus speed is the target vector used during training process. Training data set took into account both rush hour and off peak traffic conditions.

The statistical analysis was carried out to determine whether any obvious correlations were present. The linear regression analysis carried out by SPSS, highlighted low coefficients of determination equal to 0.28 (see table 1).

Table 1: Results of the statistical model.

Multiple R		0.537					
R^2		0.288					
Adjusted error		0.268					
Standard error		0.2121					
ANOVA		**DF**		**Sum of square**		**Mean Square**	
Regression		3		1.967		0.565	
Residual		108		4.859		0.045	
F = 14.572		**Sign. F** = 0.0000					
Variable	**B**	**SE B**	**Beta**	**Toler.**	**VIF**	**T**	**Sign. T**
Const	.456	.044				10.242	.000
F/C	-.389	.070	-.465	.949	1.053	-5.585	.000
RL	.117	.044	.218	.977	1.023	2.651	.009
IP	.0958	.041	.194	.961	1.041	2.337	.021

4. Methodology

All data collected was normalised in a range (0,1) ensuring that values of different input variables are in the same range, in order to avoid overflows due to very large or very small weights.

Let $\hat{z}$ be the normalised values; $z_{\min}$ and $z_{\max}$ the minimum and maximum values of z respectively.

$$\hat{z} = \frac{z_i - z_{\min}}{z_{\max} - z_{\min}} \tag{7}$$

The performance index used during the training process was the mean sum of squared errors between the estimated t_i and the actual values a_i, N is the number of observations, according the following equation:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (t_i - a_i)^2 \tag{8}$$

After the initial training of the MLP and RBF network, some statistics indexes (the Root Mean Squared Error RMSE, the Mean Absolute Percentage Error MAPE, the correlation and determination coefficient R and R^2) were determined in order to compare different models and to appraise the goodness of fit.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_i - a_i)^2} \tag{9}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|t_i - a_i|}{t_i} \times 100 \tag{10}$$

The structure of MLP was found by a pruning approach whereas the RBF during training process, the learning algorithm adds automatically nodes in the hidden layer until when the error function is below a given threshold.

Finally, a sensitivity analysis was carried out for input variables to see which of them is the most important to the estimation of the bus speed along its route. The sensitivity analysis investigated the behaviour of the RBF network to various increments of input variables. The conclusion from the sensitivity analysis was that if an increase in an input variable causes significant change (either positively or negatively) to the output variable, this input variable is regarded as an important input and should be retained in the model. A task of this work is to study a generic approach, which allows the neural networks developed for the selected route to be transferable to other bus routes. The goodness of results allows to extend the use of the calibrated neural network on new links and routes.

5. Results

The outcomes achieved after the training process of the RBF and MLP network are showed in table 2. Such values highlight the goodness of fit between real and estimated bus speed. Also, for sake of note RBF outperforms MLP in any studied case.

Table 2: Training and testing performance of RBF and MLP.

	RBF Network		***MLP Network***	
	On Training Set Bus line 102	**On Test Set Unseen data**	**On Training Set Bus line 102**	**On Test Set Unseen data**
RMSE	1.675	1.985	3.438	8.701
MAPE	6.010	8.507	22.429	33.789
R	0.950	0.932	0.786	0.405
R^2	0.903	0.865	0.618	0.164

Figure 3 and 4 show the comparison between real and estimated bus speed in rush hour and off peak traffic conditions respectively for RBF. It should be noted that the estimated bus speed is very close to real one. Nevertheless, in the range between 18 and 28 (numbers of bus location along route corresponding to Maqueda Street) the RBF network provides underestimated values of speed in off peak traffic conditions whereas it provides overestimated values of speed in rush hour traffic conditions. Probably, since this street is characterised by traffic conditions and commercial activities much heavy that cause low performances in speed estimations.

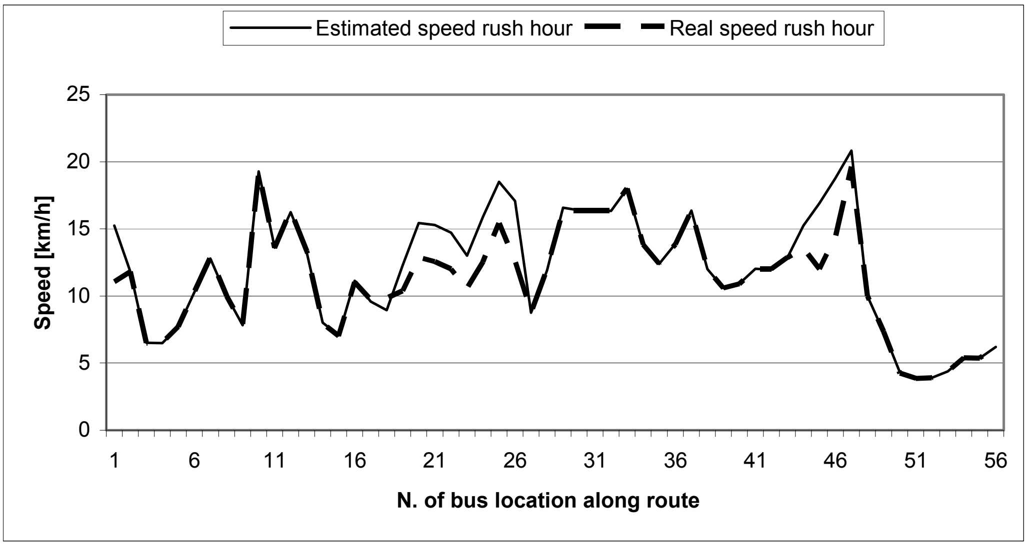

Figure 3: The comparison between real and estimated bus speed in rush hour traffic condition and RBFNN.

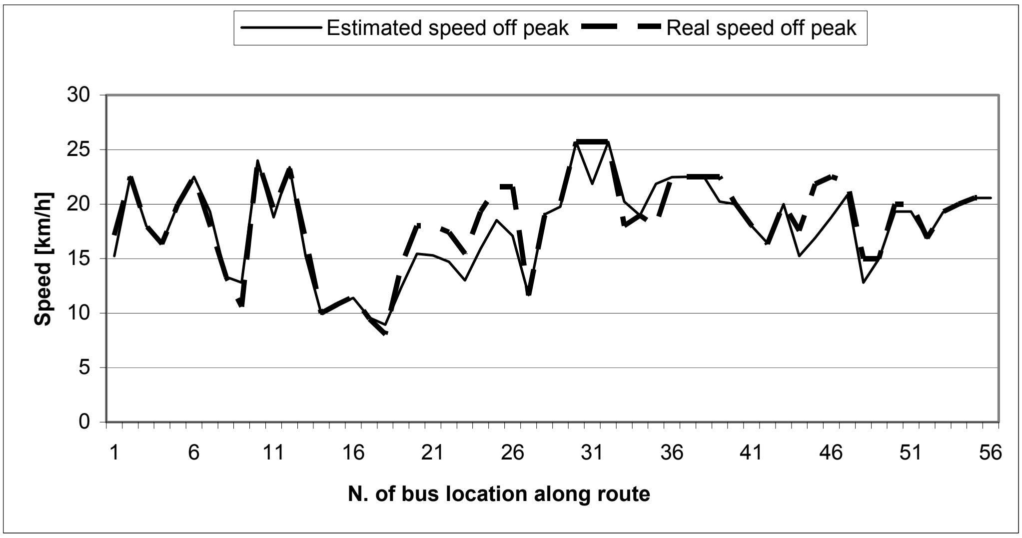

Figure 4: The comparison between real and estimated bus speed in off peak traffic condition and RBFNN.

Finally, the figure 5 shows sensitivity analysis discussed before. It should be noted that major important input variables are: the rate flow/capacity F/C, the reserved lane RL, commercial activities CA and illegal parking IP.

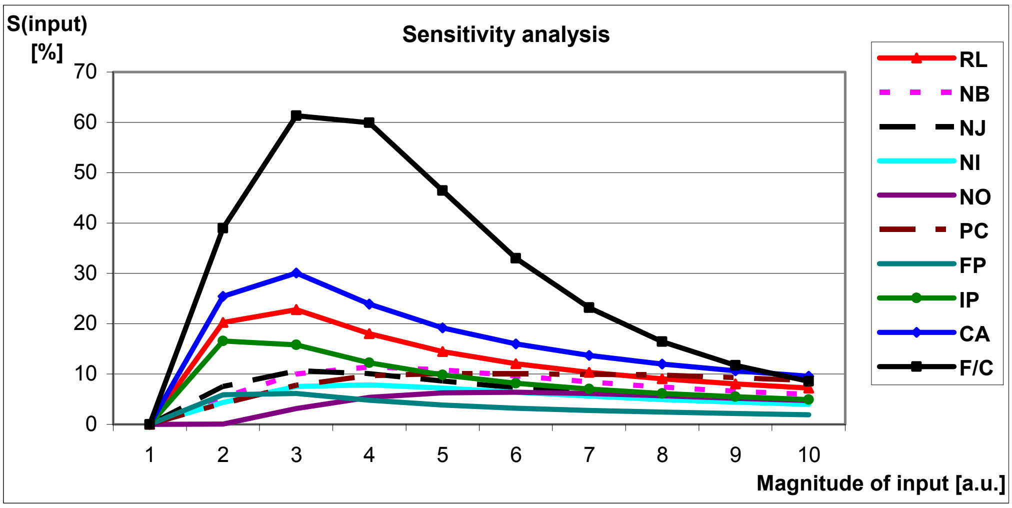

Figure 5: The sensitivity analysis.

The maximum error between real and estimated bus speed is equal to 4.5 km/h in Maqueda Street. After trained, the neural network was tested with unseen data on a new bus route along Strasburgo Road in Palermo. The test carried out gave a correlation coefficient between real and predicted values of 0.93, showing that the approach used still produces good performance on transferability (see table 2).

6. Conclusions

The effect on the average speed of bus produced by various parameters was investigated by using data coming from an AVMS and a neural network. The AVMS allows the communication in real time of the exact vehicle location and various other parameters of each vehicle as load condition on board, vehicle diagnostic conditions. Collected data concern bus location, geometrical parameters and traffic rules.

A comparison was carried out between Radial Basis Function network (RBF) and Multi Layer Perceptron (MLP). These neural netoworks were used to estimate the average speed analysing the dynamic bus location data achieved by AVMS. Input parameters are divided into geometrical parameters and traffic conditions (number and width lane, rate flow/capacity, reserved lane, number of intersections with or without traffic lights per kilometre, number of bus stops per kilometre, legal or illegal parking on the way). It was showed that RBF outperformed MLP in any studied case. The neural networks provide the estimated speed for various traffic conditions (rush hour or off peak conditions) on some link of public transport network. The RBF neural network (as well as MLP) was tested on different links of the public transport network with different geometric and traffic characteristics from ones used for training, producing good results compared with the real measured average speed and highlighting a good generalization capabilities without over fitting issues. The calibrated neural model, on some links of public transport network, is able to estimate the bus average speed producing good results compared with the real measured average speed by knowledge of geometric and traffic characteristics.

The use of trained neural network to estimate the running average speed along all arcs of road public transport network, for various traffic condition (rush hour and off peak conditions) allows to assess the round-trip time achieved by sum of estimated running times for every arc along the bus route and to compare it with scheduled one.

The estimation of running time in terms of traffic conditions implies advantages both the Public Transport Company (optimising crew and means scheduling) and the costumer (reliability of the trip time and the scheduled timetable of transit at the bus stop).

The results presented in the paper are a specific case but encouraging for further applications of the proposed methodology on an enhanced database. The goodness of results allows to extend the use of the calibrated neural network on new links and any routes.

References

AA.VV. (1985) "Highway Capacity Manual" Transportation Research Board, National Research Board.

Amin, S. M., Rodin, E. Y. and Garcia-Ortiz, A. (1998) "Traffic Prediction and Management via RBF Neural Nets and Semantic Control", *Computer Aided Civil and Infrastructure Engineering*, 13(5), pp. 315-327.

Balogh, S. and Smith, R. (1993) "London transport's countdown system – a leader in the bus transport revolution", *IEE Colloquium* on "Public Transport Information and Management Systems. IEE London UK.

Bishop, C. M. (1995) *Neural Network for Pattern Recognition*, Oxford University Press, Oxford.

Broomhead, D. and Lowe, D. (1988) "Multi variable functional interpolation and adaptive networks", *Complex Systems* 2, pp. 321–355.

Cascetta, E. (1998) *Teoria e metodi dell'ingegneria dei sistemi di trasporto*, Utet.

Celikoglu, H. B. and Cigizoglu, H. K. (2007) "Modelling public transport trips by radial basis function neural networks", *Mathematical and computer modelling*, pp. 480-489.

Chen, H. and Grant Muller, S. (2001) "Use of sequential learning for short term traffic flow forecasting" *Transportation Research Part* C, 9, pp. 319-336.

Costa, A. and Markelos, R. N. (1997) "Evaluating public transport efficiency with neural network models", *Transportation Research Part C*, 5, pp. 301-312

Crainic, T. G. and Laporte, G. (1998) *Fleet Management and Logistics*, Kluwer Academic Publishers.

Dailey, D. J. and Cathey, F. W. (2003) "A prescription for transit arrival/departure prediction using automatic vehicle location data" *Transportation Research Part C*, 11, pp. 241-264.

Daley, D. and Haselkorn, M. (1994) "Demonstration of an advanced public transportation system in the context of an ivhs regional architectur". *First World Congress on Application of Transport Telematics and Intelligent Highway System*, pp 3024-3031.

Dougherty, M S (1998) "Exploring traffic systems by elasticity analysis of neural networks". In: Himanen, V., Nijkamp, P. and Reggiani, A. (eds), *Neural networks in transport applications*. (Ashgate)

Dougherty, M. S. and Cobbett, M. R. (1997) "Short term inter urban traffic forecasts using neural networks" *International Journal of Forecasting* 13, pp.21-31.

Dougherty, M., Van Der Voort, M. and Watson, S. (1996) "Combining Kohonen maps with ARIMA time series models to forecast traffic flow" *Transportation Research Part C*, 4, pp. 307-318.

Greenfeld, O. and Koncz, N. (1995) "The development of a gis-based transit advanced traveller information system". ARISA Proceedings. *Papers fron the Annual Conference of the urban and*

Regional Information Systems Association. Urban & Regional Inf. Syst. Assoc., Washington, DC, USA, pp. 695-709.

Hornik, K., Stinchcombe, M. and White, H. (1989) "Multilayer feed forward networks are universal approximators", *Neural Networks*, 2(5), pp. 359-366.

La Franca, L., Migliore, M., Salvo, G. and Carollo, F. (2004) "The automatic vehicle monitoring to improve the urban public transport management", *XI convegno CODATU Towards a more attractive Urban Transportation,* Bucharest, 22-24 aprile 2004.

Mussone, L. and Florio, L. (1996) "Neural network models for classification and forecasting of freeway traffic flow stability" *Control Engineering Practice*, 4, pp. 153-164.

Poggio, T. and Girosi, F. (1990) "Network for approximation and learning". *IEEE Proceeding* 78, pp. 1481-1497.

Tafti, M. F. (2001) "The application of artificial neural networks to anticipate the average journey time of traffic in the vicinity of merges" *Knowledge-Based Systems*, 14, pp. 203-211.

Yin, H., Wong, S. C., Xu, J. and Wong, C. K. (2002) "Urban traffic flow prediction using a fuzzy – neural approach" *Transportation Research Part C*, 10, pp. 85-98.

Instructions to Authors

Papers should be written in English and submitted electronically to trasportieuropei@istiee.org and, by regular mail, in printed form in duplicate to:

prof. Romeo Danielis
Dipartimento di Scienze Economiche e Statistiche, Facoltà di Economia
Università degli Studi di Trieste. P.le Europa, 1, 34100 Trieste, Italy
Phone: +39-0405587076 - Fax: +39-040567543

Submission of a manuscript is considered to be a representation that it has been neither copyrighted (or if copyrighted is clearly marked so that the appropriate permission can be obtained) nor published, that it is not being submitted for publication elsewhere, and that, if the work results from a military contract, it has been released for open publication. As a condition of final acceptance of a paper for publication in European Transport\Trasporti Europei, the author(s) must indicate if their paper is posted on a working paper website, other than their own. They are responsible for assuring that, if any part of the paper has been copyrighted for prepublication as a working paper, the copyright can and will be transferred to ISTIEE when the paper has been accepted. This includes both print and electronic forms of the paper. On acceptance, the text, or any link to full text, must be removed from the working paper websites, other than the author's own website. Other material such as book reviews and announcements should also be sent to the Editor.

Manuscripts should contain only endnotes. Figures are required in a form suitable for photographic reproduction. Any one of a number of forms will be acceptable, e.g., laser printer drawing, original black ink drawings, or high-quality glossy prints. Lettering should be uniform in size and style and sufficiently large to be legible after reduction. Figures should be designated by arabic numbers and referred to in the text by number. Figure legends should be collectively provided on a separate sheet rather than placed on the figures themselves. Tables may be typed on sheets separated from the text. Each table should have a caption that makes the table entries clearly independent of the text; complicated column headings should be avoided. All tables should be numbered and referred to in the text by number.

In mathematical expressions, authors are requested in general to minimize unusual or expensive typographical requirements; for example: authors are requested to use the solidus wherever possible in preference to built-up fractions, to write complicated exponentials in the form exp() and to avoid subscripts and superscripts on subscripts or superscripts. Subscripts and superscripts should be shown large and clear, Greek letters and unusual symbols should be labeled on first occurrence, as should subscript "zero", to distinguish it from the letter "oh". Whether each letter is capital or lower case should be unambiguous. Equation numbers must be at the right.

Each paper must be accompanied by an abstract of about 100 words. The abstract should be adequate as an index and should summarize the principal results and conclusions. The first section of the article should not be numbered. References to related previous work should be reasonably complete, and grouped at the end of the paper. References in the text should be cited by the author's surname and the year of publication, e.g.: (Jansson, 1980), (Marguier and Ceder, 1984).The following format should be used for references:

Article in a journal:
Swait, J. (2001) "A non-compensatory choice model incorporating attribute cut-offs" *Transportation Research, Part B: Methodological*, 35, 10, pp. 903-28.

Chapter in a book:
Regan, A. and Garrido, R. (2001) "Modelling Freight Demand and Shipper Behaviour: State of the Art, Future Directions", In: Hensher, D. (eds) *Travel Behaviour Research: The Leading Edge*, Pergamon, Amsterdam.

Working paper:
Gavish, B. and Graves, S.C. (1981) "Scheduling and Routing in Transportation and Distribution Systems: Formulations and New Relaxations", *Working Paper 8202*, Graduate School of Management, University of Rochester, Rocherster, NY.

Book:
Urban, D. (1993) *Logit – Analyse. Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen*, Gustav Fischer, Stuttgart.

Dissertation:
Jaillet, P. (1985) Probabilistic Traveling Salesman Problems, *Ph.D. thesis*, Massachusetts Institute of Technology, Cambridge, MA

Presentation at a conference:
Maggi, R. and Bolis, S. (1999) "Adaptive Stated Preference Analysis of Shippers' Transport and Logistics Choice", *World Transport Research* -Proceedings from the 8th World Conference on Transport Research, (H. Meersman, E. Van de Voorde, W. Winkelmans eds.), Pergamon, Amsterdam.

Authors are responsible for revising their proofs, and should limit alterations to the strict minimum. The editorial management of ISTIEE reserves the right to accept only those changes that affect the accuracy of the text.