MARCO MAZZARINO, D.Sc.
Facoltà di Economia
Università degli Studi di Trieste

Science in Traffic
Original Scientific paper
U.D.C. 656.073:168.2
Accepted: Oct. 19, 1998
Approved: Dec. 23, 1998

CRITICAL REVIEW OF FREIGHT TRANSPORT DEMAND MODELS

SUMMARY

The importance of freight transport for today's society need not be underlined. Yet, rather few attempts have been made in order to model freight transport demand in a quite "original" and appropriate way, especially if the literature on freight transport is compared with the huge amount of works done on passenger transport.

The aim of the paper is twofold: first, a selective survey is conducted on the issue of freight transport demand modelling, also from an "evolutionary" standpoint; second, such a body of literature is arranged in order to provide an original classification of available models.

1. INTRODUCTION

Even if the role of freight transport appears to be increasingly crucial in the present day society, it is quite surprising that very few studies have been made on freight transport demand modelling, particularly if one compares them to the relevant literature on passenger transport [1,2].Why is this so? There are relevant differences between freight transport modelling and passenger transport modelling:

- freight transport involves a great number of decision-makers: shippers and receivers, forwarders, shipping companies, multimodal transport operators, terminal operators, depots and warehousing, etc.. Each of them has a specific economic objective, which may differ from that of the others. On the other hand, decision makers are more homogenous in passenger transport,
- it is not always simple to collect good and reliable data on freight transport,
- location factors: as freight transport demand is a derived demand, the level of movements and their origins and destinations depend largely on raw material and market locations,
- physical factors: there are many ways to deliver products, based on their physical characteristics: bulk, in light vans, in refrigerated containers, etc. There is, therefore, a great variety of vehicles and cargo units,
- dynamic factors: seasonal changes in demand and in consumer tastes affect freight movement patterns,
- price factors: freight prices are not always published and are more flexible due to dealings and bargaining power,
- freight cargoes are not homogenous and their profitability depends on specific characteristics such as perishability and value/weight ratio. As a matter of fact, the product range is larger than the most disaggregated and fragmented passenger analysis by user type and journey purpose.

Conversely, other authors highlight some aspects which make freight transport demand more adequate to economic analysis than passenger transport:

- it is clear that the optimisation process for freight transport decision makers is strictly economical (profit-maximising or cost-minimising),
- travel behaviour variability is more pronounced in passenger transport,
- freight transport demand is a "derived" demand, in the sense that it can be understood by studying the nature of the production, consumption and marketing processes,
- goods can be stored, so flow observations reflect the "expected" flows rather than the actual flows.

The aim of this paper is to review the existing literature on freight transport demand modelling, with specific reference to intercity transport. There can be two general approaches in this field:

- macroeconomic (aggregate) models, based on the "four-stages" model,
- microeconomic (disaggregate) models, based on microeconomic analysis and discrete choice models.

2. MACROECONOMIC MODELS

Such an approach deals with the analysis of aggregate flows. It consists of models based on the four-

stages model used for passenger transport planning, with some adaptations to freight transport [3,4,5,6]. It deals with:
- the assessment of generation and attraction physical points for the movement of goods,
- the distribution of generated traffic flows in order to satisfy generation and attraction constraints,
- the movements assignment by mode and paths (mode choice models, assignment/equilibrium models).

2.1. Generation and attraction models

In multi-equation models, the freight transport demand derives from the land-use structure and the spatial location characteristics of the system. The location choice by a firm is a decisive factor of freight transport demand but it is a complex process. Location models relate mostly to long-term freight transport demand studies. In the long-term one has to analyze the integrated choices and relations between location and transport (integrated transport-land use models). In the short-term, location choices tend to be fixed and one can analyse how the economic activities react to short-term demand variables. In this paper, only the short-term approach is considered.

In order to serve the purpose of modelling freight transport demand, forecasts on land-use patterns must be converted into physical appraisals of generated and attracted shipments in the different geographic points. This exercise is traditionally made by means of the so-called "trip generation and attraction" models which usually take the following form [7]:

$$O_i = f_1(SE_i) \qquad \text{Trip generation model} \qquad (1)$$

$$O_j = f_2(SE_j, LU_j) \qquad \text{Trip attraction model} \qquad (2)$$

where:

O_i, O_j – number of movements originating from zone i; number of movements attracted to zone j;

SE_i, SE_j – socio-economic characteristics of zone i and zone j;

LU_j – land-use characteristics of zone j.

When forecasting freight, this means estimating how many tons of freight must be transported into a plant or area in order to manufacture certain commodities, and how much transport away from the area is required to remove the final goods from their production sites to their markets [20]. Trip generation assesses transport requirements in each geographic point being considered (origins and destinations), but does not assess the direction of such requirements. It is therefore a point-oriented analysis, in the sense that it affects the origin and destination but not the freight flows between them.

The following types of trip generation models exist:
- input-output analysis,
- multiple regression analysis and growth factor analysis.

2.1.1 Input-output models

In these models transport is considered an economic sector exchanging inputs and outputs with the rest of the economy. Such models are macroeconomic and they assess intersectorial aggregated freight flows between economic sectors. Each economic sector needs inputs from other economic sectors in order to produce its own output. By considering transport as one of the sectors, it is possible to convert such input/output requirements into freight traffic flows. The idea of arranging intersectorial flows in input-output matrices is due to Leontief [8]. The basic structure of such models consists in identifying the freight flows x_{ij} between the economic sectors, i being the producing sector and j the acquiring sector. The total output of a sector is given by:

$$x_i = \sum_{j \in E} x_{ij} \qquad (3)$$

where E indicates all the sectors in the economy. Sectors whose demands can be assumed as exogenous and unaffected by flows are then identified, grouped and referred to as "final demand". They include government, net exports, investments and also households. The above equation is then modified so that the total output of each sector is given by the sum of flows between it and every other production sector and consists of the amount needed to meet the final demand, in addition to the requirements of all the other production sectors:

$$x_i = \sum_{j \in P} x_{ij} + D_i \quad \forall i \in P \qquad (4)$$

where P is the set of all the production sectors of the economy, D_i is the final demand for output of sector i and x_{ij} is the flows between production sectors, as previously defined. Input-output analysis has several advantages due to its simplified assumptions:
- each sector provides homogenous products which can be considered unique in the analysis, and each product is provided by only one sector,
- each firm of a sector applies a technology which can be referred to as an average type,
- an equilibrium between total supply and total demand for outputs can be reached,
- technology is assumed to be a time-constant.

By using such sets of assumptions, the model defines for each sector a "technical coefficient" a_{ij} as the amount of flow between sector i and sector j for the total output unit of the sector j. In other words:

$$a_{ij} = \frac{x_{ij}}{x_j} \quad (5)$$

Such technical coefficients are also referred to as "direct requirements", as they describe the requirements for each sector j from each other production sector i. By combining (4) and (5), we get:

$$x_i = \sum_{i \in P} a_{ij} \, x_j + D_i \quad \forall j \in P \quad (6)$$

which can be better stated in matrix notation as:

$$X = A\,X + D \quad (7)$$

where D represents a final demand vector, X an output vector and A the $[a_{ij}]$ matrix.

Let us consider now an open system in which consumption (which buys the final products and provides labour) is seen as a sector separated from the production sectors and is independent from labour supply. Given a set of final products, what will be the gross output each sector will have to provide so as to produce these final products? In other words, what will be the output level exactly satisfying final demand? From (7) we obtain:

$$D = (I - A)\,X \quad (8)$$

By assuming $(I - A)$ to be a non singular matrix, it is possible to convert the above equation in a model that predicts the output on the basis of a given final demand:

$$(I - A)^{-1}\,D = X \quad (9)$$

where $(I - A)^{-1}$ is the inverse of $(I - A)$ matrix and is referred to as the "matrix of direct and indirect requirements". Output equilibrium levels can be found if the inverse matrix $(I - A)^{-1}$ is known.

In this manner, by treating transportation as a production sector, one can predict transportation requirements by observing the total output for transportation and the row and column values for that sector in the input-output table. By using (9), it is possible to forecast the transport requirements due to any change in the economy, such as an increase in the final demand of some product.

The simple input-output model has limited applications in transportation demand and policy analysis. Its single-region structure and simplified assumptions limit its real application at a national level. The enlargement of the single-region model to a multi-region structure for the input-output table was initially proposed by Isard. Since then, many developments have followed [9,10]. Multiregional input-output models are built by introducing the spatial dimension in intersectorial flow tables. The construction of such tables is not easy since origins and destinations of inputs and outputs must be identified by regions and industries. There are different approaches for doing this in such a way to simplify the data collection and statistical problems (see [11,12,13]).

In conclusion, multiregional input-output models are a powerful tool for analyzing the transport demand at a macroeconomic level. Despite their simplified assumptions, they remain the most feasible models for regional and national planning. Many research efforts still must be done in order to develop such an approach and ease many of the simplified assumptions.

2.1.2. Multiple linear regression and growth factor method.

These methods link the land-use to the transport demand. The aim is to determine the number of movements that finish or start in each traffic zone. By defining O_i as the number of movements produced or attracted in i, a basic equation of such models is the following one:

$$O_i^k = f_i^k\,(e_i^1, e_i^2, \dots, e_i^q) \quad (10)$$

where e_i^e is the e-th explanatory variable of zone i. The function f is chosen based on its good fit with the traffic data. A specific model which has been proposed is the following one [35]:

$$T_i = F_i\,t_i \quad (11)$$

whereT$_i$ and embed t_i are future and current movements respectively (possibly in disaggregated terms) in zone i, and F_i is the growth factor (to be explicited and estimated). Such a factor is usually linked to variables such as population (P), income (I) and car ownership (C). For example:

$$F_i = \frac{f\,(P_i^{t'}, I_i^{t'}, C_i^{t'})}{f\,(P_i^{t}, I_i^{t}, C_i^{t})} \quad (12)$$

where f can be a certain function of variables, t' is the future year being considered and t is the current year.

2.2. Distribution or zone interchange models.

In the aggregate approach the commodity flow demand between regions is directly derived from a certain measure of economic interaction between them. Given the number of movements originating in and destined to each area, such models provide a description or forecast of movements between areas. They usually take the following form [7]:

$$T_{ij} = f_3\,(SE_i, SE_j, D_{ij}) \quad (13)$$

where D_{ij} is a measure of impedance (deterrence function) between i and j. The type of objective function is fundamental in so far as the characteristics of the model depend on the behavioural assumption in the objective function selected. Two extreme cases can be distinguished with an in-between case. One extreme case is the "entropy-maximising" model which assumes randomness or complete lack of information and can generate gravity models. The second is the linear programming model which assumes a cost minimising process with perfect information by a single decision maker. The in-between case is the "intervening-opportunities" model whose aim consists in minimising total journey time, subject to the condition that each area has a given probability of being accepted.

2.2.1. Gravity models

The origin of gravity models is related to Newton's universal law of gravity: the number of movements T_{ij} from origin i to destination j directly depends on two quantities, A_i and B_j, measuring origin generation and destination attraction forces, and inversely on an impedance function on arc (i,j) :

$$T_{ij} = A_i \, B_j \, f(c_{ij}) \tag{14}$$

where c_{ij} is a generalised transport cost.

Transport researchers have concentrated their efforts on the explanation of the attraction and generation forces and on the impedance function, trying to give the model an autonomous content far from that of physics. The first problem is: which is the functional form of T_{ij} and c_{ij} relation? The following ones have been proposed:

hyperbolic : $T_{ij} = \gamma \, c_{ij}^{-\alpha}$

exponential: $T_{ij} = \gamma \, e^{-\beta c_{ij}}$

Gamma: $T_{ij} = \gamma \, c_{ij}^{-\alpha} \, e^{-\beta c_{ij}}$

The second problem concerns the quantification and specification of the attraction and generation forces. The first assumption was that the traffic flows between zones i and j depend on the attraction and generation parameters such as population and income, and on the impedance parameters such as the transport costs. Subsequently these parameters were substituted by tonnage originating in i and destined in j. The first attempts are due to Black [14]. The form becomes:

$$T_{ij} = \gamma \, O_i \, D_j \, f(c_{ij}) \tag{15}$$

subject to:

$$\sum_j T_{ij} = O_i \tag{16}$$

$$\sum_i T_{ij} = D_j \tag{17}$$

It is therefore necessary to introduce two balance factors A_i and B_j coming up with the general formulation:

$$T_{ij} = A_i \, O_i \, B_j \, D_j \, f(c_{ij}) \tag{18}$$

This is the general formulation of a doubly constrained gravity model. Also, "free" models and partially constrained models exist (when one has constrains only on destinations or on origins). In the model, calibration parameters are A_i and B_j and the coefficient β of the deterrence function (if an exponential form is used).

An interesting interpretation of the gravity model, trying to give it an economic and statistical content rather than a physical one, is the entropy model. It is due to Wilson et al. [15] and Wilson [16] . The objective function of an entropy model is the following:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} T_{ij} \log T_{ij} \tag{19}$$

It can be demonstrated [17] that a general gravity model formulation is derived by maximising an entropy function subject to the following constraints:

$$\sum_{j=1}^{n} T_{ij} = O_i \qquad i = 1,2,\ldots,n \tag{20}$$

$$\sum_{i}^{m} T_{ij} = D_j \qquad j = 1,2,\ldots,m \tag{21}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} T_{ij} \, c_{ij} = C \tag{22}$$

with:

$$T_{ij} > 0 \qquad \forall i, j \tag{23}$$

2.2.2. Optimisation models

Another approach to distribution modelling is that of optimisation models. The simplest form consists in Hitchcock's transport problem. One has a set of origins O_i and destinations D_j. Origins represent surplus supply points, and destinations represent surplus demand points. Unit transport costs are given between each origin and destination and they are assumed to be fixed and not dependent on traffic flows. Flows between origins and destinations are then derived so as to minimise total transport system costs, in the same way Wardrop's second principle is used for traffic assignment procedures. In order to derive such flows linear programming procedures are used, and they take the form of a minimisation problem:

$$\min C = \sum_{i=1}^{m} \sum_{j=1}^{n} T_{ij} \, c_{ij} \tag{24}$$

subject to:

$$\sum_{j=1}^{n} T_{ij} = O_i \qquad i = 1,\ldots,m \tag{25}$$

$$\sum_{i=1}^{m} T_{ij} = D_j \qquad j = 1,\ldots,n \tag{26}$$

$$T_{ij} > 0 \tag{27}$$

where:

T_{ij} movements between zones i and j

c_{ij} generalised transport cost per tons between zones i and j

O_i total present flows originating in i

D_j total present flows destined to j

m number of origins

n number of destinations.

In this problem, one has to find flows between zones i and j so as to minimize the total cost. Linear programming approach has often been used in commodity flow analysis [18]. Nevertheless, it has a certain number of drawbacks. One of them concerns the unit transport costs: in applying the simple linear programming model they have to be taken as constant. In reality, this is a strong assumption that can lead to distortions in flow estimations. Unit transport costs can vary for many reasons and they generate a concave cost function. More advanced formulations concretely consider non linear cost functions in addition to more elaborate types of constraints taking into account the time dimension and the minimum shipment size.

Another limitation concerns the simple form in which transport problem is presented. It appears clear that much importance is given to transport costs within objective function: the case is that of one industry trying to satisfy its customers at minimum cost. However if an industry has several plants with different capacities and costs, the objective function could be better represented by profit maximisation or total market cost minimisation. Nevertheless linear programming models can be better used with low value goods for which transport costs incidence is significant.

Because of such drawbacks, the gravity model is considered more flexible. By changing β value within the deterrence function one can vary the relative importance of transport costs. Formal relation between linear programming approach and the gravity model has been studied by Evans [19]. She showed that:

- up to the limit, when $\beta \to 0$, the gravity model provides a movement matrix in which transport costs play no role
- as the β value increases, a situation in which transport costs are dominant is generated
- when $\beta \to \infty$, a linear programming solution is generated.

In this sense, it is possible to use the gravity model so as to represent the entire range of behaviours in destination choices, from one almost indifferent to transport costs (electronics for example) to the case of low value and big volume goods (such as concrete, sands, etc.) for which transport costs are of great importance.

2.2.3. Intervening opportunity models

This approach has the following characteristics:

- it deals jointly with generation and distribution problems
- it has a probabilistic approach
- the impedance function is conceived in a qualitative way and it is not quantified.

It is based on the assumption that the total travel time from a certain point is minimised, subject to condition that each destination has a given probability to be accepted. The basic concept is that each movement is carried out to the nearest acceptable location [20]. Given a flow starting from origin i, one supposes that it has different opportunities to stop each with a given probability q. By treating the number of opportunities as a continuos variable and named $P(n)$ the probability that a flow comes to a destination when the number of opportunities is less or equal to n, the probability that a flow comes to a destination between n and $n+\Delta n$ is:

$$P(n+\Delta n)-P(n)=[1-P(n)]q\,\mathrm{d}n \tag{28}$$

From this one the following results:

$$\frac{\mathrm{d}\,P(n)}{\mathrm{d}\,n}=[1-P(n)]q \tag{29}$$

If $P(0)=0$, the solution is obtained:

$$P(x)=1-e^{-qn} \tag{30}$$

When applying such a result to the real world, one must first classify the possible destinations $j(i)$ which are reachable from a given origin i based on an increasing degree of impedances. Let us indicate $j_h(i)$ the destination occupying level in the rank. Problems arise in measuring the number of opportunities which are offered by each destination. They are usually solved by making this number equal to the number of movements $Dj_h(i)$ that, starting from any origin, end in $j_h(i)$. If one indicates with $J_{i,h-1}$ the total number of opportunities which exist before $j_h(i)$ in the rank, i.e.:

$$J_{i,h-1}=\sum_{l=1}^{h-1} Dj_l(i) \tag{31}$$

from (30) one assumes that the probability for a movement starting from i to end in one of the first h-1 destinations is:

$$P(J_{i,h-1})=1-e^{-qJ_{i,h-1}} \tag{32}$$

and consequently the probability of a movement on the arc (i,j) takes the form of:

$$P[j_h(i)]=P(J_{i,h})-P(J_{i,h-1})=e^{-qJ_{i,h-1}}-e^{-qJ_{i,h}} \tag{33}$$

If O_i is the number of movements originating in i, the total flow between i and j will be:

$$T_{ij_h}=O_i\,e^{-qJ_{i,h-1}}(1-e^{-qDj_h}) \tag{34}$$

with:

$$\sum_{h=0}^{\infty} T_{ij_h} = 0 \tag{35}$$

Major difficulties in actual applications of such a model arise because of the q parameter estimation. Methods for q calibration are shown in [21] and [22].

It has been suggested that this approach, in an imperfect competition, could better represent observed behaviours and be more suitable than gravity models and the linear programming approach.

2.3. Choice of mode

The choice of a specific mode of transport or path, as it is specified by trip generation and zone interchange models, requires examination of economic and service characteristics of modes and available paths. Choice models mostly refer to the supply of the transport market, i.e. they evaluate costs, performances, capacities of transport system. Transport operators are assumed to minimise the total transport cost or maximise their utility.

The number of movements between i and j can be seen as a market for which modes compete to have the major share. One has to determine the operator's preferences related to different alternatives and model market share. One can consider a rational decision maker (shipper) having a random utility function by which he evaluates alternative modes and tries to maximise utility. It is not possible to forecast the exact choice he will make but only the probability of a certain choice. Basic assumptions are:

- decision maker d considers all available options (i.e. modes of transport) that constitute his choice set
- he attributes to each option a a utility value U_a^d and chooses the option with the maximum utility.

The utility value of each option depends on some variable values (attributes) of the option itself [7,23]

$$U_a^d = U_a^d\,(K_a^d) \tag{36}$$

where K_a^d is the attribute vector of *a-th* option for *d-th* shipper. Such attributes can be divided into:

- level of service attributes, such as journey time, transport costs, etc.,
- activity system attributes, such as land-use attributes,
- shipper's attributes.

Utility value of *a-th* option for shipper d is a random variable in so far as it is certainly not known because:

- different shippers can evaluate the same attributes differently,
- some attributes cannot be included in the model,
- some of the included attributes can only be approximately evaluated.

These are the reasons why the utility value U_a^d can be considered as composed of a deterministic part $V_a\,(K_a^d)$, which is the same for all shippers with the same attribute vector K_a^d, and a random part (residuals) ε_a, which explains the non-observed variables:

$$U_a^d = V_a\,(K_a^d) + \varepsilon_a \qquad \forall a \tag{37}$$

The deterministic utility is assumed to be a linear function of attributes:

$$V_a\,(K_a^d) = \sum_s \mu_s\, K_{sa}^d \tag{38}$$

where K_{sa}^d is the *s-th* attribute value and μ_s are coefficients to be estimated.

Given that a random term exists, we are unable to predict the exact choice of the shipper but only the choice probability. The shipper will choose an option if it features a greater utility than the other(s):

$$P^d\,(a) = P\,(U_a^d > U_r^d) \qquad \forall r \neq a \tag{39}$$

By considering (37) one gets:

$$P^d\,(a) = P\,(K_a^d - K_r^d > \varepsilon_r - \varepsilon_a) \qquad \forall r \neq a \tag{40}$$

The expressions (39) and (40) generate different mode choice models, linking choice probability to explanatory variables (attributes) as one makes different assumptions about residuals ε_j distribution. The simplest and most common of such models is the Logit model. It assumes that residuals are independently and identically distributed as a Weibull/Gumble distribution (double exponential). The mean of this distribution is zero and the variance is constant. As the residual variance decreases, the prediction capacity of the models increases, in the sense that the option with the maximum deterministic utility will be chosen. Conversely, as residual variance increases, different options tend to have the same probability. If one considers R options, the functional form of the model can be analytically derived, as:

$$p(a=1) = \frac{\exp(\alpha K_1)}{\exp(\alpha K_1) + \sum_{r=2}^{R} \exp(\alpha K_r)} \tag{41}$$

where $p(a=1)$ is the probability of option 1 to be chosen and α is a Weibull/Gumble parameter. This is called multinomial Logit model. When $R=2$ one gets the binomial Logit model. Moreover, explanatory variables (attributes) can be expressed also in terms of difference between options. By having observed choice probabilities and explanatory variables values, one can calibrate the model using both the least square method and the maximum likelihood method.

The other type of model is the Probit model. It assumes that residuals are distributed as a multivariate normal distribution, with zero mean and constant variance. The functional form of the choice probability cannot be analytically computed given the difficulties of solving the integral form. Approximate methods for choice probability to be estimated are used, such as the Monte-Carlo simulation and the Clark approximation method.

2.4. Route assignment and equilibrium models

The choice of the "best" path for carrying commodities from an origin to a destination is essentially shipper's decision. One must first model the network, having a network for each modal alternative. This is done by means of a graph in which nodes represent origins, destinations and intermediate points (terminals, etc.), and a generalised cost is assigned to links. It is a process describing the characteristics of arcs and nodes in a quantitative framework which is suitable for analysis and manipulation. The number of movements from an origin to a destination is assigned to links of several paths by means of a minimum path algorithm which selects the path with the minimum generalised cost on an O-D pair (journey time, monetary cost, etc.). If the arc capacity were infinite, the assignment problem would be simple and practically irrelevant. This is the case when generalised cost is constant as flow changes, i.e. one is in the constant part of the cost curve (network segment operates under its capacity). If, on the other hand, one has a network in which congestion plays an important role, link costs are a function of the assigned link flows: in this case the network is called "capacity constrained". Generally speaking, when link cost is expressed as $c_{ij} = c_{ij}(x_{ij})$, one has a congested network, while if c_{ij} is a constant one has a non-congested network.

At the same time, demand (flow) could be assumed to be independent from cost ("fixed demand") or to be a decreasing function of generalised cost ("elastic demand"). As flows on the network change, network performance in terms of cost changes too. Forwarders' reactions to capacity-performance relations can consist of shifts in routes within a single mode, or can also bring about mode changes. This suggests that mode choice and route assignment can be a simultaneous decision to be made by the shipper [20]. Thus, one has a demand depending on costs and costs depending on flows: a "balancing" behavioural process between costs and flows takes place and it must be modelled. In a deterministic context, if the user knows the network and the link costs perfectly, Wardrop's first principle is applied: we get to an equilibrium when for each O-D pair no user can reduce his own cost by choosing a different path. In other words, the path which is chosen is the "shortest" one. In the case of a congested network, the assignment problem is a non linear mathematical programming problem, since congestion is a non linear phenomenon. If, instead, one considers a non congested network, the assignment problem becomes a linear programming problem. In this case, the user and system objective function is the same: the problem consists in several subproblems, one for each O-D pair and demand can be entirely assigned to links belonging to shortest path between i and j, while no flow is assigned to other links. In the case of congested networks with fixed demand, the assignment problem is solved by means of heuristic methods which are based on the all-or-nothing criterion. One of the most popular methods is the "incremental assignment": demand is assigned through subsequent increases and link costs are updated at each assignment. The Frank-Wolfe algorithm is usually used. In the case of elastic demand, either approaches modifying the previous algorithm or different algorithms are used. One of these is the so-called "diagonalization algorithm".

When considering a stochastic situation, i.e. when perceived costs by users are not "actual" costs, then the concept of choice probability of path k comes into consideration, and it is modelled generally by means of a discrete choice model such as Logit or Probit.

2.5. Final observations

The framework described so far is the traditional one, i.e. sequential. It does not take into account the simultaneous decision and feedback which occur in the real world [20]. System performance affects mode choice and routing and also the number of shipments made and destinations. Again, system performance affects location decisions and land-use, in the sense that households and firms react in one way or another [7]. This is particularly important for long-range freight transport planning, but is also very difficult to model. In view of the complexity of this issue, the use of traditional land-use models such as Lowry's does not seem appropriate. The main advantage of such a traditional approach is that it facilitates conceptualisation and computer programming. One can expect improvements by iterative repeating, either for improving internal consistency and accuracy of flow prediction or for better simulation of changes and reactions in the transport system [20]. As a matter of fact, the major drawback of such an approach is that these macroeconomic and location feed-backs or loops are rarely analysed in depth.

3. MICROECONOMIC APPROACH (DISAGGREGATE)

The aim of transport demand analysis consists in evaluating the effects of different transport policies.

Over the years, the interest of transport planners has shifted from long-term transport policies to short-term analyses (or rather to real time analyses), in order to run an integrated multi-modal transport system. Such a trend has had an impact on transport forecasting methods: more flexible tools have been developed incorporating behavioural forces of single decision makers. At present, the approach tends to be of a microeconomic type (i.e. it analyses a single decision maker - firm or consumer), and not only macroeconomic, that is, analysing aggregate quantities. Conventional methods (four-stages):

- use analogies with physical systems: transport flows are treated as in hydraulic flow models,
- the components of the models are quite independent (sequential),
- they are not very policy-sensitive, in the sense that they consider a few variables which are really under political control.

The idea of the microeconomic (disaggregate) approach is that the transport demand is generated by individual choice behaviours and, more specifically, by utility function maximisation processes. It tends to stress market segmentation up to the individual decision maker level: the transport aggregate demand is the summation of individual transport demands. Within the microeconomic approach one considers different choices made by shippers or firms related to their economic activities. For some choices, choice set is continuos, i.e. a firm faces a continuous (infinite) set of alternatives; for other choices, choice set is discrete, i.e. a firm faces a discrete (finite) set of alternatives. In the former case, one uses continuous approach models, in the latter, discrete choice models. The main difference between these two types of approaches is [24]:

- in continuous models, decision-makers are assumed to make marginal adjustments as a response to market conditions,
- in discrete models, production or consumption is an all-or-nothing decision.

Choices by a firm are related to Chiang [25]:

a) location (long-term choices)

b) level of output, pricing, inputs (intermediate-run choices)

c) logistics: origins, destinations, modes of transport, shipment size, order frequency, level of inventory (short-run choices).

For a) the same things said for aggregate models still hold: since the short-term transport demand is analysed, one does not refer to location choice models at a microeconomic level. Intermediate-run choices are mostly referred to as continuous approach (marginal analysis). Finally, short-run choices are partly referred to as discrete approach (discrete choice models).

3.1. Continuous models

3.1.1. Continuous models for intermediate-run choices

The basic assumption of such models is that the decision unit is a firm producing some commodity for which inputs from different zones must be obtained and that outputs must be delivered to other zones within a marketing process. The firm could be also a retailer or other type of firm. Within the present approach one has to consider first the firm demand for goods and then derive the transport demand from that. In order to do so, one needs to consider the characteristics of the production process of the firm determining the output level. Let us suppose a firm producing the outputs which can be described by an output vector Y. Such outputs are obtained using an inputs set X. The production process for the firm can be described by some functional relation P between Y and X, where P is a production function:

$$P(Y,X)=0 \tag{42}$$

The choice of a specific production process is usually assumed as the result of an optimisation process in which the firm tries to find the minimum cost process given input prices. Let the prices of input X be given by vector prices W, the selection of a production process will follow an optimisation in which total production cost $C(Z,W)$ is minimised subject to the constraint that input combination follows the production function. The result of this optimisation is a set of values representing quantities of each input which is used to produce every level of output Y and the associated cost of this production as given by the cost function $C(Y,W)$. If prices W_i of input are assumed to be fixed and independent from quantities X_i, then demand functions of the firm for input X_i can be obtained by using a relation referred to as Shephard's lemma:

$$\frac{\partial C(Y,W)}{\partial W_i}=X_i(Y,W) \tag{43}$$

This demand function is called "compensated" (Hicksian) and it gives the relation between quantity X_i and price W_i for a given level of output. Alternatively, it can be assumed that the firm maximises its profit given input prices and outputs. Optimum values for production and inputs constitute respectively the firm supply function and the input demand function [24]. Such a demand function is called "ordinary" (Marshallian).

The transport demand for a firm can be derived from the input demand. This can be done directly by considering the cost function and by re-specifying unit costs for each input in such a way as to include transport costs. By considering an input i which will be transported at a known and fixed cost, the unit cost for input becomes:

$$W_i^{'} = W_i + c_i \tag{44}$$

where c_i is the unit transport cost for input i. In this case the transport demand function for input (commodity) i can be expressed by:

$$T_i = \frac{\partial C(Y, W_i^{'})}{\partial c_i} \tag{45}$$

where c_i is the transport volume for commodity i which is required by the firm.

Two things must be underlined in this simplified continuous model:

- it is too abstract for a realistic disaggregate analysis to be implemented. In fact, the decision on transport (how many commodities must be moved) is a result of a sequence of choices, considering also modes, origins, logistics, etc.,
- transport demand functions as given by (45) are similar to trip generation models in the aggregate approach and have the same conceptual limits.

An attempt to introduce mode choice in continuous models is due to Friedlaender and Spady [26].

3.2. Discrete choice models

In the choice theory, the decision maker chooses the most desirable alternative among a set of alternatives. The desirability of a choice generally depends on its attribute and on the decision maker's attributes. Generally speaking, alternatives can be: location, shipment frequency, destinations, shipment schedule, modes of transport, route. Actually, analysts concentrate on specific ones (e.g. mode choice) by considering the others as given. The theoretical basis of discrete choice models can be found in 2.3.

In freight transportation most studies have concentrated on the mode choice [27,3], often using logit model (see Hashemian [28]). The flexibility of discrete choice models permits the construction of a very general utility function, incorporating:

- transport service characteristics, such as prices, journey time, reliability, losses and damages, minimum shipments, etc.,
- commodity attributes, such as type of commodity, volume/weight ratio, value/weight ratio, perishability, inventory system, ownership, etc.,
- market characteristics, such as relative prices, firm size, availability of loading/unloading facilities, etc.,
- shipment firm attributes, such as output level, sales prices, plants location, availability of facilities, stock policy, etc..

Chiang et al. [25] have used a logit model for describing the joint choice of mode and shipment size for different commodity classes. Most choice models for commodities are based on a choice function representing the generalised cost and logistics.

3.2.1. Deterministic mode choice

In a deterministic context, the mode choice directly derives from total transport cost minimisation. It has been developed mostly for abstract modes and commodities [29]. By ignoring inventory and safety stock costs, the choice between modes A and B can be done based on a simple cost function such as:

$$c_i = r_i + \theta t_i \tag{46}$$

where:

c_i cost per unit for mode i;

r_i freight rate per unit;

t_i travel time;

θ_i time cost, which embodies capital value.

No other cost factor is included in this model. It is also to be pointed out that only r_i and t_i are assumed to be mode-dependent, while θ_i is commodity-dependent. This model states that, given two modes with r_A, t_A and r_B, t_B, and given a commodity whose time value is θ, the deterministic process will choose A if $r_A + \theta r_A < r_B + \theta r_B$. For a different value of θ, a set of indifference curves between modes with given r and t can be developed. For example, given two modes, one being a less expensive but slower mode and the other being a fast but expensive mode, it could be possible to make a comparison by locating them on the map.

3.2.2. Stochastic mode choice

Even if the shipper is assumed to choose on the basis of the minimum cost principle, randomness in choice function is due to inaccuracy in perception and/or measurement of costs. Logit and probit models are used in this case. In microeconomic approach such models are not applied to aggregate commodity flows but are based on individual shipper data: for example, Chiang et al. [25] used the following variables: transport charges; capital carrying cost in storage; capital carrying cost in transit, or tied up in loss and damage claims for emergency shipments; same as previously, but with regular shipments; order costs; loss of value in transit or in storage; distance variable for private truck; variable for air shipments value; rate for rail shipments; other dummy variables.

A number of further applications of choice models can be found in the existing literature. For example, Kullman [30] has calibrated a binary logit model for rail and truck choice by using freight rate, mean and standard deviation of transit time, mileage and yearly tonnage as explanatory variables. Boyer [31] has introduced the shipment size as an additional explanatory variable in a binary choice model. Levin [32] has calibrated a multinomial logit model for three mode choices: truck, rail and piggyback. Winston [33] has applied a binary probit model to rail and truck choice analysis. Sasaki [34] has studied the allocation between rail and barge for coal movements in the Ohio River Basin.

4. CONCLUSIONS

Discrete choice models have had significant results in terms of policy sensitiveness, particularly when applied to specific sub-markets or commodities. For example, Ortuzar [35] has used stated preference data for examining the possibility of providing a new service (refrigerated container) for international maritime cargoes. Such an approach has also been used by Fowkes and Tweddle [36] .

It must be stressed that microeconomic approach, particularly discrete choice models, is an approach and not a specific model. Discrete choice models are not "models" or model sets, but are rather characterised by a behavioural approach. In particular, one can build a microeconomic model completely similar to conventional models as regards data and variables used. But then one can break out the model by better describing decision maker behaviour and by using other explanatory variables. Conventional and microeconomic models are different in the *degree* of analysis: microeconomic models highlight individual behaviour regularities, while macroeconomic models enhance physical regularities of aggregate flows. Once again, micro and macro models differ in the number of explanatory variables used. However, it must be stated that behind every macroeconomic model there is a microeconomic one and vice versa: the aim is that of finding regularities. Future efforts are needed in order to produce good results both from the theoretical and practical standpoint.

RASSEGNA CRITICA DEI MODELLI PER LA DOMANDA DI TRASPORTO MERCI

L'importanza del trasporto merci per la società odierna risulta talmente evidente da non dover essere sottolineato. Purtuttavia, gli sforzi prodotti dalla letteratura scientifica in termini di modellizzazione, originale ed appropriata, di questo sistema non risultano estremamente ampi, in particolare se gli stessi vengono comparati con la letteratura disponibile sul trasporto passeggeri.

Lo scopo del paper è duplice: innanzitutto, viene realizzata una rassegna critica dei principali modelli proposti per la domanda di trasporto merci, anche da un punto di vista evolutivo; successivamente, il materiale raccolto viene organizzato in maniera critica e viene proposta una specifica classificazione dei modelli disponibili.

LITERATURE

[1] **Barber, J.**: *Appraisal of existing freight traffic models*, UK Department of the Environment Mathematical Advisory Unit, Technical Note 204, London, 1971.

[2] **Wilson, A. G.**: *Some new forms of spatial interaction models: a review*, Transportation Research, 9(2), 1975.

[3] **Van Es, J. V.**: *Freight transport, an evaluation*, ECMT Round Table 58, Paris, 1982.

[4] **Kim** and **Hinkle**: *Model for statewide freight transportation planning*, Transportation Research Record, 889, 1982, 15-19.

[5] **Friesz, T. L., Tobin, R.** and **Harker, P.**: *Predictive intercity freight network models: the state of the art*, Transportation Research, 17A(6), 1983, 409-17.

[6] **Harker, P. T.**: *The state of the art in the predictive analysis of freight transport system*, Transport Reviews, 5(2), 1985, 143-64.

[7] **Domencich, T. A.** and **McFadden, D.**: *Urban travel demand: a behavioral analysis*, New York, 1975.

[8] **Leontief, W.**: *Studies in the structure of the American economy*, New York, Oxford University Press, 1953.

[9] **Chenery, H., Clark, P. G.** and **Cao-Pinna, V.**: *The structure and growth of the Italian economy*, U.S. Mutual Security Agency, Roma, 1953.

[10] **Leontief, W.**, and **Strout, A.**: *Multiregional input-output analysis*, Conferenza Internazionale sulle tecniche input-output, Ginevra, 1961.

[11] **Hirsch, W. Z.**: *Interindustry relations of a metropolitan area*, Review of Economics and Statistics, XLI, 1959.

[12] **Artle, R.**: *Studies in Structure of the Stockholm economy: toward a framework for projecting metropolitan community development*, Stockholm, 1959.

[13] Henderson, **J. M.**: *An economic analysis of the Upper Midwest Region*, four papers on methodology, in Upper Midwest Economic Study Technical Paper, n.1, 1961.

[14] **Black, W. R.**: *Interregional commodity flows: some experiments with the gravity model*, Journal of Regional Science, vol.12, n.1, 1972, 107-118.

[15] **Wilson, A. G.** and **Senior, M. L.**: *Some relationship between entropy maximising models, mathematical programming and their duals*, Journal of Regional Science, 14, 1974,. 207-215.

[16] **Wilson, G. W.**: *Economic analysis of intercity freight transportation*, Bloomington, Indiana University Press, 1980.

[17] **Pitfield, D.**: *The useful interpretation of freight transport models for regional investment policy*, in Transport, Location and Spatial Policy, K.J. Button and G. Gillingwater, 1983.

[18] **Roberts, P. O.** and **Kresge, D.**: *Simulation of transport policy alternatives for Columbia*, Amer. Econ. Rev., vol.58, 1968, 341-393.

[19] **Evans**: *A relationship between the gravity model for trip distribution and the transportation problem in linear programming*, Transportation Research, 7(1), 1973, 39-61.

[20] **Meyer, J. R.**: *Regional economics: a survey*, American Economic Review, vol.53, 1963, 19-54.

[21] CSST: *Modelli econometrici per la domanda di trasporto*, Roma, 1976.

[22] **Ruiter, E. R.**: *Toward a better understanding of the intervening opportunities model*, Transportation Research, 1,1, 1967.

[23] **Cascetta**, **E.**: *Metodi quantitativi per la pianificazione dei sistemi di trasporto*, Padova, 1990.

[24] **Oum**, **T. H.**: *Derived demand for freight transport and intermodal competition in Canada*, Journal of Transport Economics and Policy, 13(2), 1979, 149-68.

[25] **Chiang**, **Y. S.**, **Roberts**, **P. O.**, **Ben-Akiva M.**: *Development of a policy sensitive model for forecasting freight demand*, report CTS81-1, Center for Transportation Studies, Cambridge, Mass, 1980.

[26] **Friedlaender**, **A. F.** and **Spady**, **R. H.**: *Freight transportation regulation*, The MIT Press, Cambridge, Mass.,1981.

[27] **Gray**, **R.**: *Behavioural approaches in freight transport modal choice*, Transport Reviews, 2(2), 1982, 161-84.

[28] **Hashemian**, **H.**: *Mode choice between air and truck in short haul market*, Ph.D. dissertation, Institute of Transportation Studies, Berkeley, California,1981.

[29] **Baumol**, **W. J.** and **Vinod**, **H. D.**: *An inventory theoretics model of freight transportation*, Management Science, vol. 16, 1970, 413-421.

[30] **Kullman** , **B.**: *A model of rail truck competition in the intercity freight market*, report R74-35, vol.15, Chap.6., in Studies in railroad operations and economics, MIT Dept. of Civil Engineering, Cambridge, Mass. ,1973.

[31] **Boyer**, **K. O.**: *Minimum rate regulation, modal split sensitiveness and the railroad problem*, Journal of Political Economy, vol.85, 1977, 493-512.

[32] **Levin**, **R. C.**: *Allocation in surface freight transportation: does rate regulation matter?*, Bell Journal of Economics, vol.9, 1978, 188-45.

[33] **Winston**, **C.**: *Mode choice in freight transportation*, Berkeley Dept. of Economics, Berkeley, California, 1978.

[34] **Sasaki**, **B.**: *Predicting transporter's choice mode*, Transport. Research Board Annual Meeting, 1978.

[35] **Ortuzar**, **J.** de **D.**: *Determining the preferences for frozen cargo exports*, in WCTR, Transport Policy, Management and Technology Towards 2001, Western Periodicals Co, Ventura, Ca., 1989.

[36] **Fowkes, A. S.** and **Tweddle**, **G.**: *A computer guided stated preference experiment for freight mode choice*, Proceeding 16th PTRC Summer Annual Meeting, University of Bath, Sept. 1988, England.