#### **A MODEL FOR ANALYZING THE MODAL SPLIT IN FREIGHT TRANSPORTATION**

**Lionello NEGRI Responsible for the "Innovation Valorization and Management Division" National Research Council Rome - Italy** 

**Livio FLORIO Researcher University "La Sapienza" Rome - Italy** 

### **1. PRELIMINARY REMARKS**

In Italy, the freight modal split is characterized by a marked lack of balance, since road transportation is strongly prevailing.

This research work aims at both quantifying the flows of different modes which satisfy the domestic freight transportation demand and defining appropriate analytical tools for a correct interpretation of the collected data.

The map of the goods flows across the Italian regions was drawn on the basis of data provided by official sources and in accordance with marketable goods categories fixed by TSN (Traffic Statistic Nomenclature) for the road, rail, and coasting transportation modes. Air transportation was not taken into account, because the carried goods represent a negligible percentage of the whole freight traffic. Instead, intermodal transport, although not much developed as yet, was considered and included, together with other goods, in category No. 9, where it represents the most relevant part.

#### **2. THE STRUCTURE OF THE GENERALIZED COST OF TRANSPORT**

At national level, and international as well, goods traffic arises mostly from business or business related needs. Consequently, it seemed appropriate to develop the present methodological analysis from the industrial enterprises viewpoint. For the company, the transportation cost depends on three main factors: the technique adopted, the logistics of supplies, and the characteristics of the shipment.

These factors are interactive. In fact, the use (and consequently the cost) of a specific technique depends, among others, on the size and frequency of supplies as well as on the typology of goods to be transported.Therefore, the generalized cost of transport can be decomposed into five components [61:

- **i) out-of-pocket cost of transport,** i. e., the amount paid by the customer, which depends on both the shipment size (weight or volume) and the journey length, since the unit tariff lowers proportionally to the increase of the journey length. This cost - which also includes insurance, storage, freight handling (in terminals and transhipment facilities, packaging, and any other due (custom tariffs, taxes, etc.) - is strongly affected by the transportation mode;
- **ii) cost of time,** which includes: a) losses of goods value due to perishability (e. g., food-stuffs) or obsolescence (e. g., newspapers); b) tied-up capitals referred to travelling goods, which can be quite considerable for products with a high unit value or for large shipments. Actually, users give a subjective economic value to the time factor. In fact, such a value can exceed even markedly the above-defined costs, since users believe empirically that delays in goods delivery can negatively affect the sales. For this reason, it is very difficult to determine the cost of time, which is also closely related to the reliability cost (see point v) below);
- **iii) storage cost,** which usually depends on the type of goods as well as on the average stocks level, the latter being a function of size and frequency of shipments. The storage cost includes also the cost of the tied-up capital for stocks. It is known that users sending/receiving goods characterized by a high unit value consider it advantageous to keep the stocks to a minimum and prefer to run the risk of requiring emergency supplies in order to avoid large tying-up of capitals;
- **iv) ordering cost,** which depends on both mode and nature of transportation. In fact, costs related to frequent orders to the same operator are lower than those arising from occasional orders to different operators and, at the same time, the costs of emergency orders tend to be higher,
- **v) reliability cost,** which results from increases in the stocks size and from necessities of emergency orders at higher costs. In general, the reliability cost is influenced by the transportation mode.

The impact of the above-analyzed components on the generalized cost varies mainly according to the type of goods transported. Generally speaking, the main components are the out-of-pocket cost and the cost of time, followed by the reliability cost, which is particularly high in the case of basic necessities and strategic goods. Although storage and ordering costs come under the cost function, usually customers do not take them into consideration, since storage is often imposed by the specific requirements of production and ordering follows empirical rules. Therefore, the storage cost and the ordering one represent real "shadow costs". In most cases, it would be reasonable to state that they do not affect the choice of the transportation mode.

The accessibility **cost** has to be taken into account too, in order to choose the most appropriate transportation mode. As far as the accessibility costs are concerned, road transport, able to reach any place, prevails on rail and coasting ones, which are closely related to the availability and dissemination of fixed facilities, thus involving "load disruption" and load transhipments from one transportation mode to another. The quantification of this cost item must necessarily be conventional, i. e., related to each specific transportation case.

## **3 . THE MODAL SPLIT MODEL**

## **3.1 Basic hypotheses of the model**

The generalized cost model described in this paper is strongly restricted because of the lack of reliable data about the Italian freight transportation sector. Therefore, it was necessary to work on some approximations and, subsequently, to formulate simplified and empirical hypotheses for the quantification of the cost items which are unknown .

In accordance with the formulation of the model structure, the following hypotheses have been adopted as regards the items of the generalized cost components.

As far as the **out-of-pocket cost** is concerned, by first approximation it was assumed that the item "packaging" has no influence on the choice of the transportation mode, since it does not change meaningfully from one mode to another. The tariff to be charged was calculated with reference to standard shipment of standard goods representative of each marketable goods category (see Table 1). Moreover, the cost at the terminals was taken into account only in the case of carriage by sea. This is due to the fact that the cost at the terminals is already included in the tariff quoted for rail transport and it is not difficult to be determined for road transport, toll interports and carports being not very widespread in Italy.

The **cost of time** should be evaluated in a deterministic way, by considering the tied-up capital through the all transportation time. The loss of value due to perishability should be calculated only in the case of very particular goods, which represent the lower share of freight global trade. On the basis of the information provided by recent studies on this matter, it was deemed expedient for the present research work to estimate the generalized cost on a probabilistic way. Therefore, it was assigned to time the value arising from the calibration of the modal split model.

The **storage cost** was not intentionally included in the generalized cost model for two reasons: the difficulty in quantifying the average level of stocks per product or per marketable goods category and the fact that the choice of the transportation mode is not directly affected by such a cost.



Standard goods representative of the various marketable goods categories



The **ordering cost** too was not taken into consideration, since the models developed up to now allow to analyze and to interpret the yearly shipments of goods, in which constant supplies in different economic sectors are prevailing, due to their running regularly and continuously.

As regards the **reliability cost** it has to be pointed out that the transportation reliability of given goods on a given link and by a given mode is a variable highly difficult to be determined and to be quantified. Thus, this cost is often related to the users' subjective evaluations and, consequently, it is hard to be analytically interpreted. In this study, the reliability of transport was evaluated on the basis of two variables: a) the variance of transport time expressed for each mode; b) the modal specific dummy, a characteristic parameter of each mode, for the interpretation of the "subjective" users' behaviours.

The evaluation of the accessibility cost too is necessarily subjective, since no data are available on the tariff surcharges applied in case of transhipments. From the theoretical point of view, accessibility is one of the fundamental factors for the choice of the transport mode. Therefore, the accessibility cost was included in the model by relating it to two regional indexes for rail and coasting transportations  $1$ . These indexes are equal to the inverse value of the "relative density" of the facilities existing in the territory:

$$
I_R = \frac{S_R}{N_{RK}}
$$

where  $S_R$  is the area of region R and  $N_{RK}$  is the number of available facilities (port

# wharves for coasting transport and railways stations) with reference to mode k.

### **3.2 Formulation of the model**

The proposed modal split model is based on the logit multinomial formulation  $2[1;$ 2; 3; 7; 9] and its general expression is:

(2) 
$$
P\left(k:A_t\right) = \frac{\exp\left[-\beta\left(f_k + a_t t_k + a_v v_k + a_a A_k + d_k\right)\right]}{\sum_{m \in A_t} \exp\left[-\beta\left(f_i + a_t t_i + a_v v_i + a_a A_i + d_i\right)\right]}
$$

The symbols have the following meanings:

- $P (k:A<sub>t</sub>) =$  probability of choosing mode k within a finite set of alternatives  $(A<sub>t</sub>)$ ;<br>  $B =$  calibration parameter, which provides a measure of the users' average.
- $=$  calibration parameter, which provides a measure of the users' average sensitivity to the generalized cost of transportation;



 $a_{\rm a}$  = average value assigned to the accessibility of transportation mode k.

Most of the data necessary to calibrate the model were taken from analyses carried out in previous years [4]. As regards the variances, a different procedure was adopted for each mode. Variances concerning road transportation times were derived from studies on long-distance freight traffic and, for a few traffic links, they were estimated by adopting a regression model capable to relate the actual variances to the journey length. Variances referring to rail transportation times were determined by expression (3):

$$
\sigma_{\rm r} = \frac{T_{\rm d} - T_0}{2}
$$

where:

 $\sigma_r$  = variance of rail transportation time;<br>  $T_0$  = rail transport time measured accord

 $T_0$  = rail transport time measured according to the Italian Railways official timetable;<br> $T_d$  = assured delivering time, calculated according to the instructions contained in [5]. assured delivering time, calculated according to the instructions contained in [5].

To determine the variance of coasting transportation time, reference was made to the average delays recorded in the Italian sea-harbours 3. For two marketable goods categories (0 and 6), it was impossible to consider some traffic links, since coasting tariffs are not registered because of lack of shipments.

### 3.3 **Calibration of the model**

The flows related to Sardinia were excluded from the calibration procedure, since

freight transport from/to this region presents a particular tariff structure, which differs completely from that of the other interregional links. The freight carriage by sea interesting Piedmont and Lombardy were not considered too, because these regions do not benefit from sea-harbours.

Moreover, the "maximum likelihood" calibration criterion was chosen as an alternative to the "least squares" one. In fact, from a theoretical standpoint, such a criterion complies better with the hypothesis of random distribution of components, i. e., each single shipments, which is the basis for the formulation of the model itself.

The probability of choosing transport mode k, defined by equation (2), for shipping a unit of goods (e. g., one ton) on traffic link j, can be simplified to expression  $(4)$ :

(4) 
$$
P (k: A_{i}) = \frac{\exp (\Sigma_{i} \alpha_{i} x_{ijk})}{\Sigma_{i} \exp (\Sigma_{i} \alpha_{i} x_{ijk})}
$$

where coefficient B is implicitly included in coefficients  $\alpha_i$  and  $x_{ijk}$  is the value assumed by the exogenous variable i (tariff, transportation time, etc.) with respect to link j and to mode k. Once the total goods traded by transportation mode k across the Italian regions considered two by two  $(n_{jk})$  is known, the probability that the whole system can assume the configuration previously observed is given by the "likelihood function" (5):

(5) 
$$
\lambda = \pi_j \pi_k \left[ \frac{\exp \left( \sum_i \alpha_i x_{ijk} \right)}{\sum_i \exp \left( \sum_i \alpha_i x_{ijk} \right)} \right]^{n_j k}
$$

The "maximum likelihood" calibration method lies in the determination of the vector of coefficients  $\alpha$ , which maximizes equation (5). In other words, the "likely" hypothesis according to which the observed configuration is the most probable one among those that can be assumed by the examined system is formulated. From a mathematical viewpoint, the problem can be solved by finding the roots of the simultaneous equations (6):

$$
\frac{\partial \lambda}{\partial \alpha_i} = 0
$$

However, equation (5) can be replaced by its natural logarithm  $(\lambda^*)$ , a continuous positive function being in the real field. In fact, this manipulation does not modify the solution and, at the same time, makes much easier the development of equation (6), which can be written as follows:

$$
\text{(7)} \hspace{1cm} \lambda^* = \ln \ \lambda = \Sigma_j \ \Sigma_k \ n_{jk} \ \big\{ \big( \ \Sigma_i \ \alpha_i \ x_{ijk} \ \big) \cdot \ln \big[ \ \Sigma_k \ \exp \big( \ \Sigma_i \ \alpha_i \ x_{ijk} \ \big) \big] \big\}
$$

Therefore, simultaneous equations (6) become:<br> $2a^*$ 

Which can be written as follows:

\n
$$
\lambda^* = \ln \lambda = \sum_{j} \sum_{k} n_{jk} \left\{ \left( \sum_{i} \alpha_i x_{ijk} \right) - \ln \left[ \sum_{k} \exp \left( \sum_{i} \alpha_i x_{ijk} \right) \right] \right\}
$$
\nTherefore, simultaneous equations (6) become:

\n
$$
\frac{\partial \lambda^*}{\partial \alpha_i} = -\sum_{i} \sum_{k} n_{jk} x_{ijk} - \frac{\sum_{k} x_{ijk} \exp \left( \sum_{i} \alpha_i x_{ijk} \right)}{\sum_{k} \exp \left( \sum_{i} \alpha_i x_{ijk} \right)} = 0
$$

Simultaneous equations (8) are obviously non-linear ones and can be solved by means of the several available numerical methods. The best-known is the method of Newton-Raphson, which, in this case, is expressed by equation (9):

(9) 
$$
\left\{\alpha_{m}\right\} = \left\{\alpha_{m-1}\right\} - \left[\left(\partial \lambda \partial \alpha\right)_{\alpha = \alpha_{m-1}}\right]^{-1} \left\{\left(\lambda^{*}\right)_{\alpha = \alpha_{m-1}}\right\}
$$

where:

 $(\alpha_m)$  = vector of the roots at iteration m;  $\{\alpha_{m-1}\}$  = vector of the roots at iteration m-1;  $\left[ \left( \partial \lambda \right) \partial \alpha \right)_{\alpha = \alpha_{m-1}} \right]^{-1}$  inverse matrix of the partial derivatives of  $\lambda^*$  functions calculated at points  $\alpha_{m-1}$ ;

 $({(\lambda^*)}_{\alpha = \alpha_{m,1}}$  = vector of the  $\lambda^*$  functions calculated at points  $\alpha = \alpha_{m,1}$ .

Taking, for notational convenience:

$$
G_{jk} = \exp\left(\sum_{i} \alpha_i x_{ijk}\right)
$$

it then follows that simultaneous equations (8) change in:

(11) 
$$
\frac{\partial \lambda_i^*}{\partial \alpha_r} = -\Sigma_j \sum_k n_{jk} \frac{\left(\sum_k G_{jk}\right) \left(\sum_k x_{rjk} x_{ijk} G_{jk}\right) - \left(\sum_k x_{ijk} G_{jk}\right) \left(\sum_k x_{rjk} G_{jk}\right)}{\left(\sum_k G_{jk}\right)^2} = 0
$$

Assuming as initial vector the "banal" solution  $(\alpha_0 = 0)$  corresponding to a situation of neutral choice among the different alternative modes, the unknown coefficients were calculated by establishing the error condition  $|(\alpha_m) - (\alpha_{m-1})| < 10^{-10}$ . Despite the minimum error range imposed, the solution was reached by no more than 10 iterations for all marketable goods categories reported in Table **1.** 

# **3.4 Checking and statistical testing of the model**

Model capability and its fitness were evaluated by means of the so-called "likelihood index" [1; 8]:

(12) 
$$
\rho^2 = 1 - \frac{\lambda^*(\alpha)}{\lambda^*(0)}
$$

where  $\lambda^*(\alpha)$  is the natural logarithm of the likelihood function calculated according to parameters values corresponding to those defined during the calibration phase and  $\lambda^*(0)$  is the value assumed by the same function when coefficients  $\{\alpha\}$  of the utility function are considered null. It can be demonstrated that the likelihood function reaches the lowest value when the probability of choosing the different alternatives is the same ( $\alpha$ ) = 0),

i. e., when the neutral condition represents the least probable status for the system taken into consideration. In practice,  $\rho^2$  coefficient gives a measure of the gap existing between the analyzed system (characterized by the set of coefficients  $\alpha_i$ ) and the least probable status. In theory, the "likelihood index" can assume values ranging from  $0$  to 1: the closer the index is to 1 the more reliable the model is. Actually  $\rho^2$  has a lower limit which is 0 if, and only if, the expression of the utility function does not contain any dummy variable specifically referred to the set of alternatives of transportation. In this case, relation (13) is applied:

(13) 
$$
\rho_{\min}^2 = 1 - \frac{\sum_{i} \frac{N_i}{T} \left( \ln \frac{N_i}{T} + \ln \frac{1}{k_i} \right)}{\ln \frac{1}{\sum_{i} k_i}}
$$

where:



Therefore, the presence of specific dummy variables makes the value of the  $p<sup>2</sup>$ index less significant mainly whenever one alternative is greatly preferred with respect to others (the latter circumstance actually happening in the transportation system of the present investigation). In these conditions, the analytic expression (12) of the "likelihood index" is modified as follows:

$$
\rho_{\mathbf{d}}^2 = 1 - \frac{\lambda^*(\alpha)}{\lambda_{\mathbf{d}}^*}
$$

where  $\lambda^*_{d}$  is the natural logarithm of the likelihood function containing only the dummy variables specific of the set of alternatives. In this way, the variability of  $\rho^2$  between 0 and 1 is restored and, hence, the above-mentioned drawbacks are overcome. Besides a lower limit, the  $\rho^2$  coefficient affords an upper limit too, as clearly shown by its expression: in fact, value 1 corresponds to  $\lambda(0) = \infty$ , such a circumstance occurring only if the utility function is made of an infinite number of terms. However, the upper limit cannot be determined analytically, as was the case for  $p^2_{\text{min}}$ . Moreover, as for the correlation index, the concept of  $p^2_{\text{adjusted}}$  (15) can be introduced to allow the number of degree of freedom of the system to be considered:

(15) 
$$
\rho_{\text{adjusted}}^2 = 1 - \frac{\frac{\lambda^*(\alpha)}{\Sigma_t(I_t - 1)} - K}{\frac{\lambda^*(0)}{\Sigma_t(I_t - 1)}}
$$

A further statistical indicator can be used for this class of models, i. e., coefficient  $\chi^2$  which can be expressed in the following way:

(16) 
$$
\chi^2 = 2 [\lambda^*(\alpha) - \lambda^*(0)] \quad \text{if there are no specific dummy variables}
$$

(17) 
$$
\chi^2 = 2 [\lambda^*(\alpha) - \lambda^*_{d}]
$$
 if there are specific dummy variables

On the other hand, the conventional statistical indicators should be considered with caution, since they presuppose some basic hypotheses (in particular, with regard to the distribution law of the random components) which make them suitable only for the models calibrated by the "least squares" method.

### 4. ELASTICITY ANALYSIS OF THE DEMAND

The elasticity of the demand for a given transportation mode is a fundamental element to evaluate users' sensitivity to each single component of the relevant generalized cost of transportation. In the probability type models, elasticity allows to evaluate how the probability of choosing a given alternative varies as a function of a unitary change of a variable related to the same alternative ("direct" elasticity) or to another one ("crossed" elasticity). For the logit type models and for a specific traffic link, the following expressions apply:





The so-called "aggregate elasticity", a quantity representative of the users' average sensitivity with respect to average unitary changes in the values of the generalized cost of transportation, can be calculated by expression  $(20)$ :

(20) 
$$
E_{x_{i\mathbf{i}\mathbf{k}}} \left[ P \right] = \left\{ \Sigma_{t} P \left( i: A_{t} \right) E_{x_{i\mathbf{i}\mathbf{k}}} \left[ P \left( i: A_{t} \right) \right] x_{jx} / x_{j\mathbf{i}\mathbf{k}} \right\} / \sum_{t} P \left( i: A_{t} \right)
$$

where  $x_{ik}$  is the average value assumed by vector  $\{x_{ik}\}\$ . The elasticity values calculated as described above are still "punctual" values, since they refer to probability changes related to the present conditions.

#### 5. ANALYSIS OF THE RESULTS

The results of some elaborations are reported in Tables 2 and 3 just as an example

of the analytical method described in this paper. Calibration was carried out for each marketable goods category, with/without the dummy variables and by/without considering the accessibility variables previously defined. Since road transport prevails on the other transportation modes, high values were foreseen (and confirmed later) for the specific rail and coasting dummies. Moreover, various calibration attempts were made with the aim of stressing the importance of "accessibility" variables in such particular conditions as those existing in Italy.

Generally speaking, it can be stated that the "accessibility" variable has proved to be able to replace the dummy variables. Such a circumstance implies a marked influence of the transhipment costs on the choice of the mode. Furthermore, time variances are seldom included in the model explicative variables owing to a correlation between these variables and transportation times, although the degree of such a correlation is decidedly lower than it could be expected *a priori.* 

Tables 4 and 5, made in the matrix form, give evidence of the results of the elasticity analysis of the demand. Each element  $E_{1i}$  of the matrices provides a measure of the percent change in the probability of choosing mode i, caused by a unitary percent change in the variable considered for the mode j. The elements of the main diagonal, obviously negative, refer to "direct" elasticity, whereas the remaining elements apply to "crossed" elasticity.





### 6. CONCLUSION

From the strictly mathematical point of view, the presented model is rather approximate owing to the high level of aggregation of the marketable goods categories, to the small number of transportation modes considered (only three), and to the poor reliability of the available data. At the present stage of development, the model gives already the possibility to evaluate quantitatively the influence on transportation mode choice of such typologies of intervention (such as those which determine increases/reductions in travelling time, tariff changes, and structural modification in supply), all of them considered in terms of "accessibility" indexes. In fact, some applications to the Italian scenario have permitted to assess their consequences, by adopting some basic hypotheses and using the

well-known "what-if' technique. Therefore, the model represents a valid supporting tool for Italian policy-makers, who intend to make large investments in the transportation sector, especially in road and rail infrastructures, as well as in rail services quality.

We aim at improving the model by introducing a greater number of transportation modes and by getting more representative measures for both transportation and accessibility variables.



Table 3



### 7. ENDNOTES

- 1) Accessibility is virtually infinite in the case of road transportation.<br>2) It is worth reminding that this formulation derives from the theor
- It is worth reminding that this formulation derives from the theory of the so-called "random utilities" and requires two basic hypotheses to be satisfied:
	- i) the probability of choosing a given alternative has to be proportional to the probability that the associated utility was not less than the utilities of the other possible alternatives;
	- ü) the utility can be considered composed of: a) a deterministic part, which depends on the visible characteristics of each alternative; b) a stochastic (or random) part, which takes into account other elements that are not quantifiable.
- 3) Informal communications by the authorities of a certain number of sea-harbours.

## BIBLIOGRAPHY

[1] Ben Akiva , M., and Lerman, S. R. Discrete choice analysis. Cambridge: MIT Press, 1985.

[2] Chiang, Y. S., Roberts, P. 0., and Ben Akiva, M. Development of a policy sensitive model for forecasting freight demand. Cambridge: Center for Transportation Studies, 1980.

[3] Chiang, Y. S., Roberts, P. O., and Ben Akiva, M. A short-run freight demand model: the joint choice of mode and shipment size. Cambridge: Center for Transportation Studies, 1980.

[4] CRF - Centro Ricerche FIAT. Metodologie e modelli per l'analisi del traffico merci interreeionale. Rome: National Research Council, Project on Transportation Research, Research Reports RF 21070-21071, 1985.

[5] Ente Ferrovie dello Stato (Italian Railways). Condizioni e tariffe per i trasporti delle cose sulle Ferrovie dello Stato. Rome: Report, 1979.

[6] Kanafani, A. Transportation Demand Analysis. New York: McGraw-Hill, 1983.

[7] Richards, M., and Ben Akiva, M. A disaggregate travel demand model. Saxon House: Lexington Books D. C. Heath and Company, 1975.

Table 4 Elasticity analysis of the demand - Marketable goods category No. 5 (metallurgical products)

Variable No. 1 - Transportation time			
	Coasting	Rail	Road
Coasting Rail	$-0.2585$ 0.2109	0.0712 -1.0417	0.0715 0.0976
Road	0.1928	0.1544	$-0.0740$
Variable No. 2 - Transportation tariff			
	Coasting	Rail	Road
Coasting	$-0.2394$	0.0428	0.7961
Rail	0.2076	$-0.5243$	1.0206
Road	0.1948	0.0742	$-0.7994$
Variable No. 3 - Variance of transportation time			
	Coasting	Rail	Road
Coasting	$-0.2882$	0.0786	0.0595
Rail	0.2949	- 1.1968	0.0899
Road	0.2767	0.1637	$-0.0675$
Variable No. 4 - Accessibility of coasting transportation			
	Coasting	Rail	Roed
Coasting	$-0.0224$		
Rail	0.0543		
Road	0.1015		
Variable No. 5 - Accessibility of rail transportation			
	Coasting	Rail	Road
Coasting		0.0126	
Rail		$-0.2760$	
Road		0.0536	

[8] Timothy Tardiff, J. A note on goodness of fit statistics for probit and logit models. Transportation No. 5. Amsterdam: Elsevier, 1976. Pages 377-388.

[9] Wilson, A. G. A statistical theory of spatial distribution models. Transportation Research No. 1. Oxford: Pergamon Press, 1967. Pages 253-269.

# ACKNOWLEDGEMENTS

The authors wish to thank Marco Ghersi for his valid contribution in improving and setting up the modeL Thanks go also to Natalia Allegretti for reviewing the English text.

#### Table 5

Elasticity analysis of the demand

Marketable goods category No. 9 (machines, vehicles, manufactured articles, and special transactions)

