韓國港灣學會誌 第7卷 第1號 pp.25~42, 1993. 6.

論　　文

選擇基盤 貨物데이타를 利用한 個別로짓模型의 適用에 관한 研究

南　寄　燦*

A Study on the Application of Disaggregate Logit Models from Choice-Based Freight Data

K. C. Nam

Key Word : Freight Mode Choice Model(화물수송수단 선택모형), Random Sample(확률표본), Choice-based Sample(선택기반 표본), Disaggregate Behavioural Model (개별행태모형), Maximum Likelihood Method(最尤추정방법), Elasticity(탄력성)

요　약

지난 20여년간 화물수송 분야에는 큰 변화가 있었다. 수송 공급 측면에서는 보다 다양하고 기술적으로 앞선 수송수단들이 등장했으며, 수송 수요 측면에서는 로지스틱스 개념의 도입으로 화주들의 보다 높은 수송 서비스가 요구 되었다. 수송수단의 수송 분담에 있어서도 특히 철도에서 공로로의 두드러진 화물이동 현상이 나타났다. 이러한 변화는 수송현안 해결에 대한 관심을 높이고 화물수송수요 예측기법의 이론적, 개념적인 발달을 가져왔다. 그 중 두드러진 발달은 화주의 행태를 반영하는 행태모형의 개발과 새로운 자료수집 방법 및 자료형태이다. 전통적으로 화물수송 및 교통 연구에 널리 사용된 행태모형은 확률표본을 사용하여 왔으나, 80년대 부터 비확률 표본 사용에 관심이 높아졌다. 그 대표적인 것으로 기반근거 데이터를 들 수 있다. 이 데이터는 제한된 정보를 제공한다는 자료자체의 한계를 지니고 있으나, 자료수집이 용이하고 비용이 저렴하다는 장점을 가지고 있다. 화물수송 분야에서 선택기반 데이터를 이용한 연구는 현재까지 두 편이 발표 되어 있다.

따라서 본 연구는 선택기반 데이터를 이용한 개별선택모형의 잠재력을 검증하는 것을 그 목적으로 하고, 네 종류의 제조품 그룹을 대상으로 기반근거 데이터를 수집하여 로짓모형을 추정하였으며, 추정결과를 이전 연구들의 결과와 비교하여 그 타당성을 검토 하였다.

추정된 결과는 통계적으로 유의하며 직관적으로 타당한 것으로 나타났다. 또한 그 결과는 문헌의 결과와도 일치하였다. 수송계획에 있어서 자료수집비용 절감의 필요성을 생각할 때 이것은 중요한 의미를 지닌다.

* 정회원, 한국해양대학교 항만운송공학과 교수

I. Introduction

Over the last two decades there has been a significant shift of freight traffic between transport modes, particularly from railway to road, and this has been a world wide phenomenon. This has raised various freight transportation issues, resulting in motivating conceptual improvements in the field of freight demand analysis. Particularly, a significant progress has been made in such areas as the specification of models relating to behavioural hypothesis (behavioural specification) and new methods, both in sample and preference analysis (Williams, 1987).

Generally, for the estimation of logit models and their probabilistic discrete-choice models, either a simple random sample of the population concerned or a stratified random sample is used. For the former, individual observations of the population have an equal chance of being sampled, and similarly, for the latter, each observation of a stratum has an equal probability of being sampled; the proportion of the total sample drawn from each stratum being determined by the analyst. These samples that are unstratified or stratified on population characteristics other than choice set are termed "exogenous" in which a sequence of decision makers are drawn and their choice behaviours observed.

In the 1980s the use of alternative sampling designs has been adopted for both reasons of efficiency and ease of data acquisition. In particular, a choice based sampling, which is termed "endogenous", has been given increasing attention. In this sampling method, a sequence of chosen alternatives is drawn and the characteristics of the decision makers selecting these alternatives are observed. Such sampling, which is less widely used, is said to be an easier and cheaper sampling method than exogenous sampling for a model of transport mode choice, particularly where some alternatives are so infrequently used that exogenous sampling is unlikely to obtain an adequate number of observations for that mode (Cosslett, 1981). Another situation in which choice-based sampling is more appropriate is when consideration is confined to a particular freight transport segment. In that case, it appears to be difficult to create an appropriate sampling frame for random sampling, as generally shippers are concerned with several different freight transport segments. Moreover, random sampling, if undertaken, may comprise a number of observations which are not relevant to the selected transport market. For instance, Benabi (1984) with a well designed sampling frame, obtained consignments ranging from 10 Kgs to over 10 tonnes within a single industrial sector. Thus, to obtain an appropriate number of observations, a substantial sample size may be necessary. In that case, choice-based sampling can obtain consignments directly relevant to the selected market more easily. Consequently, it would appear that a properly designed choice-based sample can often provide more precise estimates than can a random sample of the same total size resulting in a reduction of the size and cost of the sample.

In the freight context, the large majority have used a random sample method and to our knowledge there have been two application of the choice based approach (Winston, 1981; McFadden et al. 1985). The present study contributes to an improved understanding of this area. The overall aim of this study is to examine the potential for application of disaggregate models for commodity transport estimated from a choice based sample.

Following introduction, section II describes the general aspects relevant to the method of analysis employed in the study; these include the specification of models and parameter estimation. Particular emphasis is placed on description of the use of choice based data for modal choice models. In section III the sources of data used for the study and the ways in which the variables introduced in the models are constructed are described. This section is also concerned with two aspects of model development: the evaluation and interpretation of the results of the logit models and evaluation of the elasticities. Concluding comments are presented in section IV.

II. Estimating Logit Models with Choice Based Data

1. Specification of the Model

The logit model formulation is developed in a utility context, in which it is assumed that a shipper selects a mode which gives maximum utility after comparing the utilities of a feasible set of alternatives. The probability of any alternative i being selected by shipper n from choice set Cn is given as follows:

$$Pn(i \mid Cn) = Prob(Uin > Ujn) \quad \cdots\cdots (1)$$

where, in the binary choice context Cn contains two alternatives, i and j, and the utility can be expressed as:

$$U(Mi, Co, Sn) = V(Mi, Co, Sn) + e \quad \cdots\cdots (2)$$

where, Mi = a set of characteristics of mode i
Co = a set of characteristics of consignment
Sn = a set of characteristics of shipper
U = utility function
V = systematic components of the utility which is non-random
e = disturbance which is random utility

If e, the disturbance in the utility function, is distributed weibull(Gumbel Type I) then the general form of the logit model in a binary choice situation of two alternatives, i and j, can be written as follows:

$$Pn(i) = \frac{\exp[V(Mi, Co, Sn)]}{\exp[V(Mi, Co, Sn)] + \exp[V(Mj, Co, Sn)]} \quad \cdots\cdots (3)$$

where, $Pn(i)$ = the probability that consignor n chooses mode i.

For computational convenience, it is generally assumed that the functional form, V, is linear in its parameters (Ben-Akiva and Lerman 1989). Thus, by that assumption the equation 3 can be rewritten as follows:

$$Pn(i) = \exp^{\beta' xin} / [\exp^{\beta' xin} + \exp^{\beta' xjn}]$$
$$= 1 / [1 \times \exp^{\beta'(xjn - xin)}] \quad \cdots\cdots (4)$$

where, β' = the vector of unknown parameters
x = a vector of attributes, i.e. xin = h(Mi, Co, Sn)

As to the form of variables, traditionally the difference form has been adopted in both passenger and freight studies, and will be applied here.

2. Estimation Methods from Choice-Based Sample

The most general and widely used procedure for the estimation of the parameters from a sample of observations is the maximum likelihood procedure, which is based on the notion that "although a sample could originate from several populations, a particular sample has a higher probability of having been drawn from a certain population than from others" (Ortuzar and Willumsen, 1990, p. 207). This procedure is thus to derive the "value of the parameters for which the observed sample is most likely to have occurred" (Ben-Akiva and Lerman, 1989, p. 20).

Considering a random sample of N observations, the likelihood of the entire sample occurring is the product of the likelihoods of the individual observations, written as :

$$L=(\beta_1, \beta_2, \cdots \beta_k)= \prod_{n=1}^{N} Pn(i)^{yin}\ Pn(j)^{yjn} \cdots\cdots (5)$$

where, Pn(i) and Pn(j) = the probabilities that consignment *n* will go by mode *i* and *j* respectively

Yin = 1 if consignment *n* goes by mode *i*, otherwise 0

Yjn = 1 if consignment *n* goes by mode *j*, otherwise 0

$\beta_1, \beta_2, \cdots \beta_k$ = parameter estimates

It is more convenient to maximise the log of the likelihood function presented as :

$$Ln(\beta)= \sum_{n=1}^{N} [Yin\ \log\ Pn(i)+Yjn\ \log\ Pn(j)] \cdots\cdots (6)$$

To solve for the maximum of *Ln, Ln* is differentiated with respect to each of the β's and the partial derivatives are set equal to zero as :

$$\frac{\partial Ln}{\partial \beta_k} = \sum_{n=1}^{N} [Yin \frac{\partial Pn(i) / \partial \beta'_k}{Pn(i)} + Yjn \frac{\partial Pn(j) / \partial \beta'_k}{Pn(j)}] = 0,$$

$$\text{for } k=1, \cdots ,K \cdots\cdots (7)$$

When the first partial derivatives of the equation approach zero and the second partial derivatives are negative, the maximum likelihood estimates are obtained. Equation 7 constitutes a set of non-linear coupled equations in the unknown parameters β_k, $k=1, \cdots K$. Detailed descriptions of the process of solving for estimators of parameters can be found in Hensher and Johnson (1981).

In general the application of such classical maximum likelihood estimation to choice-based samples is impractical, except in certain circumstances, due to computational intractability (Ortuzar and Willumsen, 1990). However, the standard method may be amended to produce appropriate estimations. To this end, a set of weights W(i) for each commodity group concerned are introduced. The function W(i), ($i \in C$), is defined as W(i) = A(i) / H(i), where A (i) denotes the aggregate market share of alternative i and H(i) denotes the sampling fraction which is determined by the design of the sample. It may be shown (Ben-Akiva and Lerman, 1989 ; Cosslett, 1981) that valid estimations can be obtained by maximising the weighted exogenous-sample log likelihood with respect to the parameter β.

$$Log\ Lw(\beta) = \sum_{n=1}^{N} W(in)\ \log\ P(in \mid xn, \beta) \cdots\cdots (8)$$

This estimator therefore simply weights the observations and then treats the weighted sample as if it were an exogenous sample. The resulting estimator of β is termed the "weighed exogenous-sample maximum likelihood (WESML) estimator". Manski and Lerman (1977) proved that the WESML estimator is consistent and asymptotically normally distributed with a covariance matrix that can be estimated from the data. The weighting procedure can be understood by a simple intuitive argument. Suppose a choice-based sample has proportionately twice as many rail users in it as actually occur in the population. The weight factor for such observations would be 1/2, and its effect in the WESML objective function would be to weight each rail-using observation by 1/2 ; this results in reducing the effect of each one on the value of β_k which maximises the WESML objective function. Conversely, for under-sampled alternatives the observations would have the weight factors greater than one, increasing their relative importance.

3. Definition of Elasticities for Logit Models

One useful property of mode choice models is the concept of an elasticity which is a measure of the sensitivity of the predicted responses of shippers to changes in the model's explanatory variables. Direct elasticity is the percentage change in the probability of choosing a particular mode with respect to a given percentage change in one of the attributes describing the utility of that mode. Cross elasticity however, is the percentage change in the probability of choosing a particular mode with respect to a given percentage change in one of the attributes in the utility function of a competing alternative. The arc elasticity of y with respect to a variable x is defined in the usual way as :

$$E = \frac{\triangle Y}{\triangle X} \cdot \frac{X}{Y} \quad \cdots\cdots (9)$$

where, $\triangle X$ and $\triangle Y$ = the change in X and Y respectively.

The limit of $(\triangle Y / \triangle X)$ as $\triangle X$ approaches zero, is the derivative of the function $\partial Y / \partial X$. Therefore, the direct point elasticity for the logit model can be written as :

$$E^{pn(i)}_{xikn} = \frac{\partial Pn(i)}{\partial Xikn} \cdot \frac{Xikn}{Pn(i)} \quad \cdots\cdots (10)$$

where, $E^{pn(i)}_{xikn}$ = the elasticity of probability of choosing mode i for observation n with respect to a change in the ***kth*** variable describing the utility of ith mode for observation n.

pn(i) = Probability of choosing mode i for observation n.

xikn = kth attribute describing mode i for observation n.

The direct point elasticity in equation 10 may be written as :

$$E^{pn(i)}_{xikn} = \frac{\partial in\ Pn(i)}{\partial in\ Xink} = \beta_k\ Xink(1-Pn(i)) \quad \cdots\cdots (11)$$

The cross point elasticity can be evaluated similarly as :

$$E^{pn(i)}_{xjkn} = \frac{\partial in\ Pn(i)}{\partial in\ Xjnk} = -\beta_k\ Xjnk(1-Pn(j)) \quad \cdots\cdots (12)$$

The elasticities described so far are concerned with individual observations. Most real-life decisions are largely based on the forecast of some aggregate demand such as the amount of freight shipped between different city pairs, and thus some procedures to aggregate individual observations to determine market demand elasticities are needed. One simple approach is by use of so-called mean elasticities which are to evaluate the equations, 11 and 12 at the sample average Xik and average estimated P(i). However, using this mean elasticity for the logit model which is nonlinear will cause over-estimation of the aggregate elasticity of the independent variable, because the calibrated model curve would not necessarily pass through the point elasticity evaluated at the population mean (Hensher and Johnson, 1981). Therefore, a preferable approach is to evaluate equations 11 and 12 for each individual observation and then aggregate, weighting each individual elasticity by the individual's estimated probability of choice (Hensher and Johnson, 1981). Referring to this technique, termed the sample enumeration method, the aggregate direct and cross elasticities of market demand can be written respectively as follows from the equations 11 and 12 :

$$E^{p'i}_{xik} = \sum_{n=1}^{N} (Pn(i) \cdot E^{pn(i)}_{xikn}) \Big/ \sum_{n=1}^{N} Pn(i) \qquad \cdots\cdots (13)$$

$$E^{p'i}_{xjk} = \sum_{n=1}^{N} (Pn(i) \cdot E^{pn(i)}_{xjkn}) \Big/ \sum_{n=1}^{N} Pn(i) \qquad \cdots\cdots (14)$$

where, Pn(i) = an estimated choice probability of mode i for observation *n*.

p'i = aggregate probability of choice of mode *i*.

III. Model Estimation

Literature on modal choice analysis of freight transport has revealed that the freight transport market is highly diversified, resulting in great variability in modal choice behaviour. Thus, it seems desirable that the total freight market be divided into appropriate and relatively homogeneous groups in order to obtain a better understanding, both of the factors which influence mode choice, and a quantitative assessment of response to changes in these factors which will represent particular policies. Accordingly, the freight market segment concerned in this study is confined to the small consignment market, whose segmentation is based on the proposals of Rimmer and Hicks (1979, p. 536).

As to the choice set Cn which is defined as route trucking and the railway, there appears to be good reasons for defining a binary choice situation. First, the two modes constitute feasible alternatives for users concerned with the movement of small consignments. Second, preliminary survey found the two modes to be used almost exclusively in the concerned consignment market.

1. Data

The data used for the analysis of mode choice was drawn from different sources for the different modes of rail and route trucking. It consists of a large number of shipments which are confined to the small consignment category, covering several manufactured commodity groups and origin-destination pairs. The sample represents shipping that took place during the 1988－early 1989 period.

The data for route trucking is from a study on route trucking carried out by the Korean Transport Institute (KOTI, 1990a). The Institute undertook a nation-wide survey on route trucking companies and records of actual freight shipment were gathered from the carriers concerned in the survey. 326 business depots of route trucking companies out of the 1785 across the country were involved in the survey which accounted for 18.3 per cent. The information on the shipments which took place on 15th of each month during 1988 were gathered from these depots. A detailed description of the data and the sampling methods used is contained in KOTI (1990a). This survey was compiled by KOTI, based on the 19 commodity groups at two digit level (Table A3) and the 264 administration zones. The variables included are commodity type, shipper type, consignment weight, freight charge, and origin and destination pair. From this data, a large number of consignments were drawn randomly by the author to be used in this study ; thus the sample is choice-based, which means that choice rather than decision makers is sampled.

The sample pertaining to rail was drawn from the waybill record of the Korean National Railway (KNR), which contains raw information about shipments that belong to the small consignment category. The waybill record contains raw information about origin, destination, date shipped, consignee, detail of commodity, weight, date arrived, and train used. In addition, in some cases, information on the date delivered and freight rate charged is also available. From this waybill record, the sample was drawn randomly ; thus this sample is also choice-based. The drawn raw information was coded by the author in the same way as the route truck data for analyses. For instance, the name of the station is coded, based on the 264 administration zones, and the commodity detail is based on the 19 commodity groups. Here, the use of the spatial unit, zone, is merely for the convenience of coding, and is different from that used by the aggregate approach for which zone is used as the unit of analysis. The origin and destination in the data for both modes correspond to the carriers' terminals in the cities concerned, and thus the zone represents the terminals.

For the construction of the proportion weights required for the calibration of logit models with choice-based data (see section II), the national share of transport modes by 18 commodity groups compiled by KOTI (1986) is chiefly used. The Statistical Yearbook of Transportation (KMOT, 1990) and the Statistical Yearbook of Railroad (KNR, 1991) are also referred. The aggregate *population* shares of the commodity groups concerned are presented in Table A5.

2. Construction of the Variables from Choice-Based Data

The two samples, based on several criteria, were then screened in order to make them appropriate for this study. As the study is concerned solely with inter-city movement of small consignments, the sample included only city origin-destination pairs which are provided with both transport services concerned. The commodity groups were also confined to those found appropriate and the size of consignment confined to those found appropriate for the small consignment category.

Theoretically, the disaggregate approach uses actual information on consignments, transport users and modes for individual observations. In reality, however, the limitation on available data causes researchers to make certain compromises by drawing information on explanatory variables from other sources. This is more likely to be the case for studies using waybill record data than for studies based on shipper survey data ; the shipper survey data providing much richer information relevant to individual consignments.

It is common for disaggregate studies on both freight and passenger mode choice analysis to use estimated measures for service variables for both chosen and alternative modes. For instance, Roberts (1977) suggested three separate classes of level of service models which are based on a regression technique for the construction of loss and damage, freight rate, and transit time and reliability variables. Hartwig and Linton (1974), Daughety and Inaba (1981), Winston (1981), and Prins and Schultheis (1987) also used a similar method for the construction of level of service variables. DeDonnea (1971) and Grayson (1981) followed this method for passenger transport. Accordingly, the variables are often average measures over spatial units rather than actual information pertaining to individual cases.

In the present study, as both the data for the route truck and the railway have limited information pertaining to the chosen mode only, the necessary information on the chosen and alternative modes has been obtained from various sources. Thus, some variables take estimated values in average form and, as there is no door-to-door information available, it is necessary to use a proxy variables to represent the influence of access times and costs on mode choice decision. This is not uncommon in inter city disaggregate models (DeDonnea, 1971). With reference to the experience of previous research which can provide implication on the importance of mode choice attributes, several variables are selected within the limitations of the data. The variables included for the analyses are shipment weight, freight charge, origin-destination pair, commodity type, distance, mean transit time, mean accessibility, and daily service frequency. The sources from which these variables were obtained, as well as the description of these variables, are described in detail.

The actual distances for both rail and route truck were measured between each origin-destination pair on the sample from the "Chart of Zone Demarcation (Freight)" (KNR, 1988) and the "Route Network Charts of Route Trucking Companies" published by KOTI (1990b) respectively. In the case of the route truck which has more complicated route networks, much effort was spent in identifying the actual route the shipment would take. When there were several different routes between a particular origin-destination pair, the most dense was assumed to be used.

The rates for rail and route truck shipments were determined from the published Korean Ministry of Transport (KMOT) rate list and Korean National Railway (KNR) rate list respectively as little information on the actual rate charged for shipments was available for rail from the waybill records. The rate system for the two modes is very similar, and is based on the weight of consignment, distance of haul, and commodity type. Thus, the rate for each consignment in the sample can be identified easily by commodity type, individual weight and distance of haul.

Since transit time information is not available in the data set for both modes, other sources were employed. For the rail transit time estimation, some train operation times relevant to trains offering services for small consignments were obtained from the train operation time tables (KNR, 1989). There are two types of rail services offered for small consignments : passenger trains and the small consignment train. As their operation speeds differ, a weighted regression equation is estimated, based on the proportion of consignment moved by each train in the sample. The following regression equation estimated from the timetable was used to generate rail mean transit time for each consignment.

$$RTTi = 0.725 + 0.0168\ DISi \quad R\ square = 0.70$$
$$sample\ size = 31\ trains$$

where, RTT = rail transit time in times of shipment i

DIS = railway distance of shipment i in Km

For truck transit time estimation, the transit times of 26 truck trips on various routes have been obtained by the author by interviewing staff in several terminals representing four large route-trucking companies. Based on the data obtained, a linear regression equation was estimated for generating truck mean transit time for each consignment as follows :

$$TTTi = 0.312 + 0.0201\ DISi \quad R\ square = 0.89$$
sample size = 26 trucks

where, TTT = truck transit time of shipment i in times
DIS = *truck distance of shipment i in* Km

Mean accessibility, which was introduced in order to incorporate the important aspects of access suitability, and access time and costs, takes the form of average over city zones as the actual proximity to the carriers' terminal pertaining to individual consignments in the sample is not available. To formulate the variable, for each city zone included in the data set, the number of truck terminals and rail stations in each city zone was divided by the area of the zone as follows (see Table A2 for accessibiligy measure) :

$$RACCESS\ i = \frac{\text{Number of Station in } i}{\text{Area } i}$$

$$TACCESS\ i = \frac{\text{Number of Terminal in } i}{\text{Area } i}$$

where, RACCESSi = *the value of accessibility* for rail for city i
TACCESSi = the value of accessibility for truck for city i

The daily service frequency for rail for each shipment was easily determined from the train operation time table of KNR (1989), while truck frequency for each shipment was determined more laboriously from the Route Condition of Route Trucking (KOTI, 1990b), For the latter, *the highest* service frequency for each origin-destination pair is chosen as the frequency of route trucking.

In addition to these variables, there are several ones perceived to be important, but omitted here due to the difficulty of defining and measuring them; we, therefore, emphasise that although they are not included in the measured set, their influence will exist in the random component of the logit utility function.

The statistical summary for the choice-based data for both rail and route truck is presented in Table A4. Given the data, the weights for this study are constructed on the basis of national share of a commodity group for each mode relative to the share in the sample. For estimation of the parameters, a GLIM package (see Healy,1988 and Aitkin et al., 1989 for details) was used.

3. Estimation Results

Before attempting to include the variables described above in models, it may be useful firstly to examine the multicollinearity which is likely to exist between variables. Examination of correlation matrices (Table A1) reveals that high interdependencies exist between distance of haul variable and transit time ; their correlation coefficients are over 0.889 for all commodity groups. This is attributable to the fact that transit *time is taken as a linear function of distance in* this study. The preliminary estimation of the model including distance variable instead of transit time provided almost identical results with the model including transit time. As transit time is an important policy variable, the transit time variable is included, and accordingly, the distance variable is not taken into account. The parameter estimates for the four commodity groups are displayed in Table 1 together with the corresponding t values, the likelihood ratio and the rho-squared statistics. In all cases, the five variables - accessibility, rate, transit time, daily service frequency and size of shipment- are included. As the rail mode was coded 1 and the road mode as 0, the parameters displayed in the table are those relevant to the probability of choosing the rail mode.

All the mode variables have the correct signs conforming to what is expected on intuitive grounds. For instance, the negative para meter of the variable, transit time, indicates that an increase of a particular mode's transit time or a decrease in the alternative mode's transit time, results in the decreasing probability of that particular mode being chosen, other things being equal. Conversely, the positive sign of accessibility implies that an increase of the rail mode's accessibility leads to an increase in the probability of choosing the rail mode.

Table 1. Logit Model Estimations for Each Commodity Group
Coefficient. (S. E), [t value]

	Textiles	Paper	Chemical	Basic Metal
Constant	1.822	3.066	3.878	4.001
	(0.504)	(0.822)	(0.731)	(0.783)
	[3.62]	[3.73]	[5.31]	[5.11]
Accessi-bility	7.848	11.29	11.68	12.36
	(1.755)	(2.507)	(2.18)	(2.563)
	[4.47]	[4.50]	[5.36]	[4.82]
Transit Time	−0.854	−1.139	−0.661	−1.831
	(0.246)	(0.352)	(0.267)	(0.349)
	[−3.47]	[−3.24]	[−2.48]	[−5.25]
Frequency	0.080	0.387	0.250	0.123
	(0.035)	(0.125)	(0.088)	(0.050)
	[2.29]	[3.10]	[2.84]	[2.46]
Rate	−0.0001	−0.0024	−0.0007	−0.0014
	(0.0003(	(0.001)	(0.0005)	(0.0006)
	[−0.33]	[−2.47]	[−1.40]	[−2.33]
Size of Consignment	−0.028	−0.054	−0.053	−0.049
	(0.0056)	(0.015)	(0.012)	(0.010)
	[−5.00]	[−3.67]	[−4.42]	[−4.90]
Rho-squared	0.24	0.45	0.48	0.49
Likelihood Ratio	288.71	130.10	152.28	178.78

The variable consignment size has negative sign, which means that the rail share decreases linearly with consignment size. This is contrary to the generally considered advantages of the railway in large consignment size, which are intuitively justified in terms of the relative cost structures of the two modes : e.g. as the size of consignment increases, the rail mode becomes increasingly advantageous relative to the truck mode. The negative sign of size of consignment can be explained by the fact that the rail mode handles only parcel-size consignments, even though there is no weight limit, while the route truck serves a much wider range of consignment in terms of weight.

Regarding the contribution of particular variables to the overall fit, the four variables—accessibility, transit time, daily service frequency and size of shipment—turn out to be statistically significant determinants of mode choice at the 95 per cent level. However, the variable rate is found to be insignificant in the model for chemical, implying that service variables have more explanatory power for shippers' mode choice decision than freight rates. The likelihood ratio statistic and the rho-squared statistic for the model imply a good fit of this model. (a rho-squared with values between 0.2 and 0.4 are considered good fits 「Hensher and Johnson, 1981」).

In summary, the three variables transit time, accessibility and service frequency are always statistically significant across all commodity groups concerned, while rate variable is commodity specific.

4. Evaluation of the Elasticities

By utilising the sample enumeration method, the aggregate direct point elasticities are estimated across the six commodity groups (see equation 13). The procedure is to estimate elasticities for each individual observation and then aggregate, weighting each individual elasticities by the individuals' estimated probability of choice. The results of the estimation of aggregate direct point elasticities with respect to quantitative policy variables are presented in Table 2.

All the estimated direct elasticities of demand for the selected commodity groups have correct signs.

Table 2. Direct Elasticities

		Rate	Frequency	Transit Time	Accessibility
Textiles	Rail	−0.004	0.166	−2.179	0.108
	Road	−0.002	0.050	−0.952	0.438
Paper	Rail	−0.759	0.641	−2.148	0.115
	Road	−0.253	0.129	−0.842	0.435
Chemicals	Rail	−0.264	0.314	−1.036	0.105
	Road	−0.107	0.107	−0.566	0.547
Basic Metal	Rail	−0.540	0.157	−2.553	0.102
	Road	−0.212	0.053	−1.355	0.506

The elasticity of rate variable is relatively low for both modes across all commodity groups ; for the commodity groups, paper and basic metal, rates however have relatively high elasticity for the rail mode, implying that the rail users of this commodity group are quite sensitive to rate. The values for the rail mode are higher than those for the road mode across commodity groups, implying that shippers choosing the rail mode are more sensitive to unit change in rate than are truck mode users. The elasticity for the frequency variable is inelastic for both modes across the commodity groups, but for the commodity groups, paper, this variable has considerable influence on rail users. Transit time is highly elastic for the rail mode across all commodity groups ; for the road mode, it has a little lower elasticity than for the rail mode. Demand for the rail mode is inelastic with respect to the accessibility variable across the commodity groups, but the variable has moderate influences on truck users; the values are higher for the road mode than for the rail, which means that truck mode users are more sensitive to changes in accessibility than are rail users.

Consequently, the results of direct elasticities indicate that transit time exerts the greatest influence on the shippers' mode choice response for all commodity groups for both modes; rate and accessibility have some influences for rail and truck users respectively, while service frequency has a less of an influence for both modes. Users of the truck mode seem less sensitive to any change in the attributes except for accessibility, than do the users of the rail mode.

5. Comparison with Other Studies

The comparison of the coefficients with those from other studies is very difficult as studies are concerned with different commodity markets with different units of observation and different variables. Table 3 provides a limited comparison of the importance of explanatory variables.

In general, it is said that shippers of high−value commodities responded more positively to changes in speed than price, while those of low−value commodities−mainly bulk− respond positively to changes in freight rates. The table

Table 3. Demand Elasticities of Road and Rail Freight

Studies	Commodities Concerned	Elasticities			
		Truck		Rail	
		Time	Rate	Time	Rate
Benabi(1982)	Plastics	−0.793	−0.420		
Winston(1981)	Chemical(Bulk)	−0.46	−1.87	−0.58	−2.250
Levin(1978)	42 Manufactured Commodities			−0.7	−0.350
Current Study	Paper	−0.842	−0.253	−2.148	−0.759

verifies this : for manufactured commodities Benabi(1982) and Levin(1978) derived higher direct point elasticities for transit time than rate, while for bulk commodities (chemicals), Winston (1981) obtained higher elasticities for rate than transit time. This study, which is concerned with manufactured commodities, displays an identical results with other studies. And this seems to be significant for the verification of the use of choice−based data.

Ⅳ. Conclusions

The modelling focus of the present study relates to the use of choice based samples for micro-model calibration. It has been argued that the choice-based sample has some advantages compared with exogenous sampling in certain mode choice circumstances. These include both ease and cost efficiency in the construction of a sample, and for a particular market segment (here the small consignment market), the ease of creation of the sampling frame.

The choice-based sample was successfully applied to the disaggregate models of commodity transport in the Korean context. The logit estimation results of choice based data for four commodity groups have showed significant results with respect to the explanatory power of the models and statistical significance of explanatory variables. The rho-squared statistic ranges from 0.24 to 0.53 across commodity groups. The three variables, transit time, service frequency and accessibility, are always statistically significant across all commodity groups concerned, while rate variable is commodity specific.

The evaluation of transport users' responsiveness to modal choice with respect to changes in the explanatory variables has disclosed that transit time exerts the greatest influence on the shippers' mode choice for all commodity groups for both modes; accessibility has some influence on rail and truck users, while service frequency and rate have less of an influence for both modes. These results are consistent with previous studies which found that shippers of manufactured commodities responded more positively to changes in speed than price, while those of low-value commodities−mainly bulk−responded positively to changes in freight rates.

The above findings imply sound application of a choice based sample to commodity mode choice analysis. Although there have been a few applications of the choice based approach in the United States, to my knowledge, this is the first application in the context of a developing country. The implementation of such an approach will be very useful for transport planners for policy analysis of commodity modal choice in developing countries where there is particular emphasis on low cost methods.

The objectives and results of this research suggest a number of directions for further research. It has been argued that the advantages of the choice based approach relate to the ease with which commodity choice data may be obtained. However, this must be paid for in two respects, firstly the limited range of variables (often used as proxies), which can be incorporated, particularly in relation to access times, and secondly it is necessary to use the averages of certain level of service variables (although this is usually the case in practical application of disaggregate models). To some extent the problems can be overcome with the random sample approach in which relevant decision makers are surveyed. This implies that the relative advantages and disadvantages of each approach should be further considered, and further research needs to be conducted on the comparison of choice based vs random sample methods; and hybrid approaches in which different sampling methods are combined; that is random samples are enriched by a choice based sample, as in the case of some passenger demand applications.

Another direction is the extension of the scope of analysis. This study has been concerned with a specific freight category, termed the small consignment market. An extension to a wider variety of commodity markets, which reflect different transport activities, is desirable, as different freight transport segments will manifest different mode choice behaviour and thus respond differently to any policy change. In this context it is felt that there is significant scope for further research through the use of disaggregate methods to examine the whole commodity market in the context of long term freight transport planning.

References

1) Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989) "Statistical Modelling in GLIM", Clarendon Press, Oxford.

2) Benabi, B. (1984) "Elasticity Parameters of Disaggregate Freight Modal Choice". in Pitfield, D.E. (ed), Discrete Choice Models in Regional Science, London Papers in Regional Science 14, Pion Limited, London, pp.122–140.

3) Ben-Akiva, M. and Lerman, S.R. (1989) "Discrete Choice Analysis : Theory and Application to Travel Demand". 3rd Printing, the MIT Press, Cambridge, Mass.

4) Boyer, K.D. (1977) "Minimum Rate Regulation, Modal Split Sensitivities, and the Railroad Problem", Journal of Political Economy, Vol.85, pp. 493–512.

5) Cosslett, S.R. (1981) "Efficient Estimation of Discrete-Choice Models", in Manski, C.F. and McFadden, D. Structural Analysis of Discrete Data with Econometric Applications, The MIT Press, Cambridge mass., pp. 51–111.

6) Daughety, A.F. and Inaba, F.S. (1981) "An Analysis of Regulatory Change in the Transportation Industry", The Review of Economics and Statistics, Vol.63, pp. 246–255.

7) DeDonnea, F.X. (1971) "The Determinants of Transport Mode Choice in Dutch Cities", Rotterdam University Press, Rotterdam.

8) Grayson (1981) "Disaggregate Model of Mode Choice in Intercity Travel", Transportation Research Record 835, pp. 36–42.

9) Hartwig, J.C. and Linton, W.E. (1974) "Disaggregate Mode Choice Models of Intercity Freight Movement", unpublished MSc. dissertation, Northwestern University.

10) Healy, M.J.R. (1988) "Glim : An Introduction", Clarendon Press, Oxford. Heggie, I.G. (1972) "Transport Engineering Economics", McGraw-Hill Book Co., England.

11) Hensher, D.A. and Stopher, P.R. (1979) "Behavioural Travel Modelling", Croom

Helm, London.
12) Hensher, D.A. and Johnson, L.W. (1981) "Applied Discrete-Choice Modelling", Croom Helm London.
13) KMOT (1989) "Statistical Yearbook of Transportation", Korean Ministry of Transportation, Seoul.
14) KMOT (1990) 'Statistical Yearbook of Transportation", Korean Ministry of Transportation, Seoul.
15) KNR (1988) "Statistical Yearbook of the Railroad', Korean National Railway, Seoul.
16) KNR (1989) "Train Operation Time Table", Korean National Railway, Seoul.
17) KNR (1991) "Statistical Yearbook of the Railroad", Korean National Railway, Seoul.
18) KOTI (1986) "A Study on Improvement in Freight Transport System", Korean Transport Institute, Seoul.
19) KOTI (1989) "A Study on the Basic Direction of Long Term Policy for Physical Distribution", Korean Transport Institute, Seoul.
20) KOTI (1990a) "A Study on Activating Route Trucking Operation", Korean Transport Institute, Seoul.
21) KOTI (1990b) "Route Network Charts of Route Trucking Company", Korean Transport Institute, Seoul.
22) Kullman, B.C. (1973) "A Model of Rail/Truck Competition in the Intercity Freight Market", unpublished Ph.D. dissertation, Department of Civil Engeering, Massachusetts Institute of Technology.
23) Lerman, S.R., and Manski, C.F. (1979) "Sample Design for Discrete Choice Analysis of Travel Behaviour: the state of the art", Transportation Research, 13A(1), pp. 29－44.
24) Levin, R.C. (1978) "Allocation in Surface Freight Transportation : Does Rate Regulation Matter?", Bell Journal of Economics, Vol.9, No.1, pp. 18－45.
25) Manski, C. and Lerman, S. (1977) "The Estimation of Choice Probabilities from Choice-based Samples", Econometrica, Vol.45, pp. 1977－1988.
26) McFadden, C., Winston, C. and Boersch-Supan, A. (1985) "Joint Estimation of Freight Transportation Decisions under Nonrandom Sampling",in A.F. Daughety (ed) Analytical studies in Transport Economics, Cambridge University Press pp. 137－157.
27) Ortuzar, J. de D. and Willumsen, L.G. (1990) "Modelling Transport", John Wiky and Sons, England.
28) Oum, T.H. (1989) "Alternative Demand Models and their Elasticity Estimates", Journal of Transport Economics and Policy, XXIII (2), May, pp. 163－187.
29) Prins, V. and Schultheis, M. (1987) "Measuring the Impacts of Freight Transport Regulatory Policies", Transportation Research Record 1154, pp. 15－22.
30) Rimer, P.J. and Hicks, S.K. (1979) "Urban Goods Movement : Process, Planning Approach and Policy", in Hensher and Stopher (1979), pp. 525－552.
31) Roberts, P.O. (1977) "Forecasting Freight Demand", Proceedings of the 3rd World Conference on Transport Research, Martinus Nijhoff, The Hague, 247－264.
32) Soliman, A.H., Gadi, A.M., Wyatt, D.A. and Easa, S.M. (1991) "Regulatory Reform and Freight Mode Choice", Transportation, Vol. 18, pp. 261－284.
33) Williams, H.C.W.L. (1987) "Some Reflections on Transportation Models", in Nijkamp, P. and Reichman, S. (eds), Transportation Planning in a Changing World, Gower, pp. 254－265.

34) Winston, C. (1981) "A Disaggregate Model of the Demand for Intercity Freight Transportation", Econometrica, Vol.49, No.4, pp. 981－1006.
35) Winston, C. (1983) "The Demand for Freight Transport: Models and Application", Transport Research, Vol.17A, No.6 pp. 419－427.
36) Winston, C. (1985) "Conceptual Developments in the Economics of Transportation, An Interpretive Survey", Journal of Economic Literature Vol.23, Spring, pp. 57－94.
37) Zlatoper, T.J. and Austrian, Z. (1989) "Freight Transportation Demand : A Survey of Recent Econometric Studies", Transportation, 16, pp. 27－46.

Table A1. Correlation Tables

For commodity Group 06

	WEIGHT	RATE	ACCESS	DIST	FREQU	TIME	WEIG.DU	DIST.D
WE	1.0000							
RA	-.0660	1.0000						
AC	-.0700	.4238	1.0000					
DI	.0931	.2552	.3583	1.0000				
FR	.0175	.2221	.4448	.0930	1.0000			
TI	.0277	.5336	.5208	.9132	.2044	1.0000		
WD	-.6927	.0647	.1174	-.0577	.0689	.0163	1.0000	
DD	.1301	-.5075	.1965	.2231	-.3414	-.0937	-.1413	1.0000

For commodity Group 08

	WEIGHT	RATE	ACCESS	DIST	FREQU	TIME	WEIG.DU	DIST.D
WE	1.0000							
RA	-.1944	1.0000						
AC	-.2397	.2858	1.0000					
DI	.0906	.1278	.1524	1.0000				
FR	-.0821	.3206	.4126	.1567	1.0000			
TI	.0354	.3304	.3298	.9327	.3240	1.0000		
WD	-.6376	.0979	.3074	-.1228	.1066	-.0815	1.0000	
DD	.0933	-.4447	-.2997	.1525	-.5259	-.1127	-.1546	1.0000

For commodity Group 09

	WEIGHT	RATE	ACCESS	DIST	FREQU	TIME	WEIG.DU	DIST.D
WE	1.0000							
RA	-.2450	1.0000						
AC	-.2362	.3832	1.0000					
DI	.0929	.0101	.2051	1.0000				
FR	-.1498	.3914	.4888	.1377	1.0000			
TI	.0270	.2876	.4528	.9050	.3403	1.0000		
WD	-.6287	.1287	.2427	-.1061	.2338	-.0311	1.0000	
DD	.1708	-.5519	-.4252	.2497	-.5463	-.0835	-.2888	1.0000

For commodity Group 12

	WEIGHT	RATE	ACCESS	DIST	FREQU	TIME	WEIG.DU	DIST.D
WE	1.0000							
RA	.1084	1.0000						
AC	-.1389	.3375	1.0000					
DI	.2148	.0940	.1343	1.0000				
FR	-.0435	.2293	.3553	.0589	1.0000			
TI	.1614	.3916	.3800	.8894	.1793	1.0000		
WD	-.6205	-.1817	.1695	-.2244	.0919	-.1967	1.0000	
DD	.1667	-.5028	-.3216	.2182	-.5273	-.0966	-.1697	1.0000

Table A2.

Assessibility Measure

Zone	City	Population	Area	Route Truck Depot(1)	(1)/Area	Railway Station(2)	(2)/Area
1	Seoul	10,287	605.4	193	0.319	10	0.017
25	Anyang	433	58.5	7	0.120	1	0.017
26	Kwangmyong	253	38.8	3	0.077	1	0.026
29	Uijongbu	200	81.8	4	0.049	1	0.012
30	Guri	102	30.1	4	0.133	1	0.033
36	Tongduchon	70	95.2	3	0.032	1	0.011
38	Suwon	544	105.6	8	0.076	1	0.009
40	Osan	52	42.0	5	0.119	1	0.024
41	Songtan	52	41.4	6	0.145	1	0.024
42	Pyongtaek	75	43.0	6	0.140	1	0.023
61	Inchon	1,616	208.3	22	0.106	2	0.010
67	Chunchon	175	53.3	5	0.094	2	0.038
69	Kangnung	151	72.4	5	0.069	1	0.014
70	Tonghae	94	180.1	10	0.056	1	0.005
71	Samchok	52	56.7	4	0.071	1	0.018
72	Taebaek	115	259.0	3	0.012	1	0.003
73	Wonju	162	84.2	5	0.059	1	0.012
89	Chongju	417	119.0	7	0.059	1	0.008
90	Chungju	121	97.8	6	0.061	1	0.010
91	Chechon	100	89.3	7	0.078	1	0.011
102	Taejon	1,021	543.8	25	0.046	3	0.006
108	Onyang	61	44.9	4	0.089	1	0.022
109	Chonan	188	83.5	8	0.096	1	0.012
110	Taechon	56	46.2	4	0.087	1	0.022
127	Chonju	491	155.3	9	0.058	1	0.006
128	Iri	206	83.2	7	0.084	1	0.012
129	Kunsan	202	54.2	7	0.129	1	0.018
131	Chongju	82	109.4	7	0.064	1	0.009
132	Namwon	61	52.1	3	0.058	1	0.019
146	kwangju	1,116	500.7	25	0.050	4	0.008
150	Naju	29	60.6	7	0.116	1	0.017
151	Sunchon	151	88.5	6	0.068	1	0.011
152	Kwangyang	61	59.1	1	0.017	1	0.017
153	Yochon	59	106.3	1	0.009	1	0.009
154	Yosu	177	45.2	7	0.155	1	0.022
155	Mokpo	250	45.6	8	0.175	1	0.022
177	Taegu	2,239	455.6	58	0.127	1	0.002
184	Kumi	190	126.7	9	0.071	1	0.008
185	Yongchon	55	72.4	5	0.069	1	0.014
186	Kyongsan	144	410.3	5	0.012	1	0.002
187	Kimchon	79	60.6	5	0.083	1	0.017
188	Sangju	62	109.8	4	0.036	1	0.009
189	Chomchon	53	44.5	4	0.090	1	0.022
190	Andong	177	83.2	4	0.090	1	0.012
191	Pohang	302	74.3	7	0.094	1	0.013
192	Kyongju	139	218.9	8	0.037	1	0.005
193	Yongju	88	60.5	4	0.066	1	0.017
218	Pusan	3,751	435.8	75	0.172	10	0.023
230	Chinhae	123	110.8	1	0.009	1	0.009
232	Changwon	253	124.4	4	0.032	1	0.008
233	Masan	484	73.0	10	0.137	1	0.014
237	Chinju	243	69.6	8	0.115	1	0.014
238	Miryang	145	796.2	9	0.011	3	0.004
239	Ulsan	618	180.8	9	0.050	2	0.011

Population : one thousand people
Area : Square kilometre
Soure : KOTI(1990) p. 185 and KNR(1989)

Table A3. Lists of Commodities

Code No.	Commodities
01	Coal Mineral
02	Lime Mineral
03	Crude Petroleum and Natural Gas Product
04	Metal Minerals, Other Mineral
05	Foods, Beverages and Tobacco
06	Textile, Wearing Apparel and Leather
07	Wood and Wood Products including Furniture
08	Paper and Paper Products, Printing and Publishing
09	Chemicals and Petroleum, Coal, Rubber and Plastic Products
10	Cement and Cement Products
11	Non-Metallic Products, Except Products of Petroleum and Coal
12	Basic Metal and Mineral Products and Fabricated Metal Products
13	Electrical Houseware and Machinery for Consumption
14	The Other Products
15	Agricultural Products
16	Domestic Animal and Sericulture
17	Forest Products
18	Marine Products
19	The Others

Table A5.
Aggregate Share of Commodity Groups

Aggregate Share			Sample		Weight	
	Truck	Train	Truck	Train	Truck	Train
G6	76	24	57	42	1.31	0.58
G8	96	4	82	18	1.17	0.22
G9	92	8	83	17	1.11	0.47
G12	90	10	86	14	1.05	0.71
Pooled	73	27	63	37	1.16	0.73

Note : G6 : Textile, wearing apparel and leather
G8 : Paper, paper products, printing and publishing
G9 : Chemicals, rubber and plastic products
G12 : Metal and fabricated metal products

Table A4.
Summary of the Data

(S. D,)

	Textiles	Paper	Chemical	Metal
Length of Haul				
Rail	229	255	231	238
	(79)	(59)	(88)	(53)
Truck	263	264	270	210
	(125)	(111)	(130)	(116)
Consignment Size (kg)				
Rail	32	25	25	28
	(19)	(18)	(14)	(16)
Truck	70	75	92	96
	(67)	(85)	(105)	(98)
Rate per Kg				
Rail	52	32	43	41
	(30)	(21)	(24)	(22)
Truck	32	28	30	24
	(31)	(36)	(34)	(27)
Transit Time				
Rail	4.61	5.00	4.61	4.73
	(1.29)	(1.00)	(1.49)	(0.89)
Truck	5.59	5.63	5.75	4.53
	(2.51)	(2.23)	(2.62)	(2.34)
Frequency				
Rail	3.1	3.5	3.1	3.4
	(1.3)	(0.9)	(1.2)	(1.0)
Truck	7.1	9.5	11.4	6.9
	(9.5)	(10.2)	(11.8)	(7.7)
Accessibility				
Rail	0.026	0.027	0.026	0.027
	(0.007)	(0.006)	(0.007)	(0.007)
Truck	0.309	0.368	0.362	0.316
	(0.123)	(0.103)	(0.118)	(0.116)
Sample Size				
Rail	145	71	137	118
Truck	196	165	179	204