MULTINOMIAL PROBIT WITH STRUCTURED COVARIANCE FOR ROUTE CHOICE BEHAVIOR

TETSUO YAI
Department of Civil Engineering, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo 152, Japan

SEIJI IWAKURA
Japan Transport Economics Research Center, 3-18-19 Toranomon, Minato-ku, Tokyo, Japan

and

SHIGERU MORICHI
Department of Civil Engineering, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 112, Japan

(Received 8 September 1994; in revised form 19 June 1996)

Abstract—We propose another version of the multinomial probit model with a structured covariance matrix to represent any overlapped relation between route alternatives. The fundamental ideas of the model were presented in Yai et al. (1993) and Yai and Iwakura (1994). The assumptions introduced in the model may be more realistic for route choice behaviors on a dense network than the strict assumption of the independent alternative property of the multinomial logit model. As the nested logit model assumes an identical dispersion parameter between two modeling levels for all trip makers, the model has difficulty in expressing individual choice-tree structures. To improve the applicability of the multinomial probit model to route choice behaviors, we introduce a function which represents an overlapped relation between pairs of alternatives and propose a multinomial probit model in which the structured covariance matrix uses the function in order to consider the individual choice-tree structures in the matrix and the estimatability of the new alternative's covariances. After examining the applicability of the multinomial probit model using empirical route choice data in a Tokyo metropolitan region, we also propose a method for evaluating consumer benefits on complicated networks based on the multinomial probit model.

1. INTRODUCTION

The applications of the multinomial probit model have not been adequately successful in spite of its advantages in flexibility of the model form. Certainly, the complexity of the computational process has deterred its use, compared to the wide applications of the multinomial logit models. Early advances in the estimation method of the multinomial probit model were achieved before the early 80s, by Daganzo (1977), Lerman and Manski (1981), Daganzo and Sheffi (1982) and Sheffi et al. (1982). Their work discussed alternative methods for estimating the covariance matrix simultaneously with utility function parameters. No accurate method was found during these earlier advances and thus the multinomial probit model was not widely applied (Horowitz et al., 1982; Horowitz, 1991). In the 1980s, most discrete choice models were calibrated by the multinomial logit model or expansion forms of the multinomial logit such as the nested logit model. Although most results were satisfactory in representing travel behaviors of modal choices, several behaviors which do not satisfy the assumptions of the multinomial logit model exist. Most probably, the cause of such behaviors is the interdependency of choice alternatives.

Recently, there have been advances in multinomial probit estimation (McFadden, 1989; Pakes and Pollard, 1989; Bunch, 1991; Bolduc and Ben-Akiva, 1991; Bolduc, 1992; Geweke et al., 1994). The method of simulated moments proposed by McFadden seems to encourage multinomial probit applications because of its computational efficiency in seeking model parameters. Bolduc focused on the estimation of the multinomial probit model with a large choice set using auto-regressive errors with distance related functions among alternatives for simplifying its covariance matrix. Bunch simplified the multinomial probit model's covariance matrix with his transformation method which lessens the estimation problem. Geweke et al. compared several
algorithms to estimate multinomial probit model parameters, such as the method of simulated moments or the simulated maximum likelihood estimators with the GM recursive probability simulator. Recent improvements of both the estimation algorithms and the covariance matrix simplification may be a motivation to apply multinomial probit models broadly in the field because they are fundamentally more flexible than the multinomial logit models.

Although the above advances make multinomial probit models applicable for a much wider range of travel behaviors, travel choices such as a choice of railway routes may still be restricted by the multinomial probit modeling. Let us consider the route choice of a rail commuter on a complicated network in a metropolitan area. There are several route alternatives available for rail commuters and sections of any two route alternatives often overlap. As this may lead to the invalidation of the independent alternative assumption of the multinomial logit model, the multinomial probit model appears to represent the behaviors more appropriately than the multinomial logit model. However, if the alternatives in a choice set include the combination usage of railway lines, the number of alternatives in the model may be unrealistically large. Selecting a small number of labeled alternatives to reduce the size of the choice set is not feasible because commuters' route alternatives do not correspond to a specific labeled alternative in the multinomial probit model. This is because the commuters have different origins and destinations and their transferring points may differ. The covariance matrix of the multinomial probit model requires specific labeled alternatives as its elements.

Consequently, this does not permit the application of the multinomial probit to model route choice behavior in situations where there are a large number of route alternatives. In addition, value assignment of a variance and covariances for a new alternative in the prediction stage using the multinomial probit model is, in general, problematic.

For the solution of the above problems, variable correlations among route alternatives may be expressed by a structured function. If the covariance between any two individual overlapped alternatives is explicitly described by the function, the multinomial probit model will be applicable to route choice behaviors even on a complicated network.

Hausman and Wise (1978) proposed the structured covariance matrix for the multinomial probit model to consider heterogeneity among individuals. They assumed parameter distributions of a utility function as normal and derived the covariance matrix in which explanatory variables of a utility function were incorporated. The covariance structure for error terms was described as follows

\[
\text{cov}(\varepsilon_r, \varepsilon_q) = \sum_k \sigma^2_{\beta_k} Z_{rk} Z_{qk}
\]

where \( r \) and \( q \) indicate routes, \( Z_{rk} \) is \( k \)th variable of route \( r \) and \( \sigma^2_{\beta_k} \) is an unknown variance of \( k \)th parameter \( \beta_k \). Individual differences in the covariance matrix are expressed by the equation.

Using this approach, let us consider an example in which the utility function of a route choice model contains a route distance variable which has heterogeneous parameters among individuals. The covariance between routes must be proportional to the product of two route distances from eqn (1). Assigning a larger covariance for the combination of longer distance routes seems to be reasonable as a covariance of errors, even though such a covariance structure may not represent the similarity among route alternatives.

Bolduc (1992) proposed the following error structure for the multinomial probit model.

\[
\varepsilon_i = \rho \sum_{j \neq i} w_{ij} \varepsilon_j + \xi_i
\]

In this equation, \( \varepsilon_i \) is a normal distributed error with a mean of zero and is correlated with the other errors \( \varepsilon_j \) (\( j \) is not equal to \( i \)). \( \xi \) is also distributed normally and \( \rho \) is a parameter. \( w_{ij} \) indicates a weight between alternatives \( i \) and \( j \). The model would be directly applicable to destination choice models with a large choice set and may be applicable to route choice behaviors by defining the distance related function to represent the relation among routes.

In this paper, we propose another version of the multinomial probit model with a structured covariance matrix to represent any overlapped relation between route alternatives. The fundamental ideas of the model were presented in Yai et al. (1993) and Yai and Iwakura (1994). The assumptions introduced in the model may be more realistic for route choice behaviors on a dense network than the strict assumption of the independent alternative property of the multinomial
logit model. As the nested logit model assumes an identical dispersion parameter between two modeling levels for all trip makers, the model has difficulty in expressing individual choice-tree structures. To improve the applicability of the multinomial probit model to route choice behaviors, we introduce a function which represents an overlapped relation between pairs of alternatives and propose a multinomial probit model in which the structured covariance matrix uses the function in order to consider the individual choice-tree structures in the matrix and the estimatability of the new alternative’s covariances. After examining the applicability of the multinomial probit model using empirical route choice data in a Tokyo metropolitan region, we also propose a method for evaluating consumer benefits on complicated networks based on the multinomial probit model.

2. MODELS

2.1. The multinomial probit model formulation

The multinomial probit model, expressed by the following equations, is widely known as the model that incorporates not only a strict utility $V_r$ for route $r$ but also a random error $\varepsilon_r$ into its final form. This is quite different from the simple multinomial logit form. A choice probability $P_r$ with a choice set size $R$ is calculated by multidimensional integration associated with $\varepsilon$ as

$$ P_r = \int_{\varepsilon_{r=-\infty}}^{\infty} \int_{\varepsilon_{r-\infty}}^{\varepsilon_{r-V_1}} \cdots \int_{\varepsilon_{r-V_R}}^{\infty} \Phi(\varepsilon) \, d\varepsilon_R \cdots d\varepsilon_1 $$

where the density function is described by

$$ \Phi(\varepsilon) = (2\pi)^{-\frac{R}{2}} \left| \sum \varepsilon \right|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \varepsilon \sum \varepsilon^T \right] $$

and the covariance matrix is

$$ \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1R} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1R} & \sigma_{2R} & \cdots & \sigma_R^2 \end{pmatrix} $$

We omit notation for an individual for simplicity but we can easily add the notation without any change in the model derivation. Elements of the above symmetric covariance matrix $\Sigma$, which are usually regarded as parameters, have been estimated simultaneously with parameters of the utility function $V_r$. A set of the estimated elements in the matrix is constant for the population. This model form is applicable to specified alternative situations such as a mode choice model whose alternatives are specified by car, rail and plane. If we utilize this model to represent a route choice behavior, we will be aware that an individual’s route pattern does not correspond to alternatives in the model and a constant value for the correlation between two alternatives cannot be assigned for each element of the covariance matrix.

For example, several alternative railway routes often exist in a large city and some alternatives have transfer stations on the route from the origin to the destination. This implies that some route alternatives overlap from or up to a transfer station. Such an overlapped situation between routes varies with the origin and destination location on the network. If the difference of such overlapped situations between routes cannot be considered explicitly as the similarity of alternatives in the choice structure, the multinomial probit models would be inapplicable to a complicated network. Let us consider a more detailed situation where we have three route alternatives. One of the three routes is independent from the other two, but the other routes overlap in part. A covariance of the overlapped two routes may be larger than that of an independent route and any remaining one. This is because the overlapped section is identical for the two routes and the similarity between them should be high. The following part of this paper concentrates on a new modeling idea using the multinomial probit model to represent the above mentioned similarity between alternatives.
2.2. Multinomial probit model with structured covariance

2.2.1. Derivation of the model. As discussed previously, we should clarify the following two possibilities in order to apply the multinomial probit model to route choice behaviors. One is the representability of individual differences in overlapped route alternatives, and another is the predictability of a new route probability by assigning its variance and covariances. The multinomial probit model with structured covariances derived from structured error terms may be an acceptable answer to the above problems.

The error term of the multinomial probit model, which has a normal distribution, is assumed to be divided into two parts,

\[ \varepsilon_r = \varepsilon_r^1 + \varepsilon_r^0, \quad r = 1, \ldots, R \]  \hspace{1cm} (6)

where \( \varepsilon^1 \) depends on the length of the route as will be explained later and \( \varepsilon^0 \) distributes independently by route. Both terms are defined to be normally distributed and to have a different variance matrix. We can easily write the covariance matrix of the total error term \( \sum \) by,

\[ \sum = \sum^1 + \sum^0 \]  \hspace{1cm} (7)

where \( \sum^1 \) and \( \sum^0 \) correspond to \( \varepsilon^1 \) and \( \varepsilon^0 \).

\( \varepsilon^1 \) may be expected to have a larger variance when a route’s distance or travel time is longer. We assume that \( \varepsilon^1 \) is a random disturbance distributed independently by unit length. This assumption makes it simple to understand the error depending on the route length and is consistent with the assumption of eqn (6). This assumption implies that the error possibly arises from any unit section in the route and the errors are independent from each other. The definition of ‘unit section’ is based on distance or time. We assume that the total distance or travel time determines the number of unit sections on the route between its origin station and destination one. Following such assumptions, the variance of route \( r \) is expressed by

\[ \text{var}(\varepsilon_r^1) = \sigma^2(x) + \cdots + \sigma^2(x) = n_r(x)\sigma^2(x) \]  \hspace{1cm} (8)

where \( \sigma^2(x) \) represents a unit variance, \( n_r(x) \) is the number of the unit sections, and \( x \) is the dimension of the unit, such as distance or travel time. For example, if \( x \) is defined by distance, using \( d_r \) which is the length of route \( r \) and \( \sigma^2 \) which is an unit variance by distance, the variance of route \( r \) can be rewritten by

\[ \text{var}(\varepsilon_r^1) = d_r\sigma^2 \]  \hspace{1cm} (9)

If defining \( x \) by time, the route variance should be explained by its travel time in a similar manner. In general, the assumption that the error of a route utility arises from every unit belonging to the route is clear for building the multinomial probit model though another error from a transferring terminal may arise independently. We can consider the later error to be included in the second term of eqn (6), as it characterizes the node and the whole route.

Now, two routes are correlated with each other under the overlapped situation. Let us consider the covariance between two overlapped routes in \( \sum^1 \). According to the assumptions mentioned previously, the covariance is easily expressed by the following equation,

\[ \text{cov}(\varepsilon_r^1, \varepsilon_q^1) = E(\varepsilon_r^1 \varepsilon_q^1) = E(\varepsilon_r^0 + \varepsilon_r^\text{ov})(\varepsilon_q^0 + \varepsilon_q^\text{ov}) = E(\varepsilon_r^0)^2 \]  \hspace{1cm} (10)

where \( r \) and \( q \) indicate routes and ov and nov indicate an overlapped and a non-overlapped section. As the unit error is independently distributed, a covariance between two non-overlapped sections should be zero. Finally, the covariance of two routes is identical to the variance of the overlapped section. If two routes do not overlap at any point, the covariance must be zero.

Because eqn (10) can be written by a similar notation to eqn (8), the general form of the element of \( \sum^1 \) is shown by

\[ \text{cov}(\varepsilon_r^1, \varepsilon_q^1) = n_{rq}(x)\sigma^2(x) \]  \hspace{1cm} (11)
where \( n_{rq}(x) \) is the number of units in the overlapped interval of route \( r \) and \( q \). If \( q \) is equal to \( r \), we can replace \( n_{rq}(x) \) by \( n_r(x) \) of eqn (8). It is important to know that \( n_{rq}(x) \) varies by individual but can be assigned before the estimation of the model parameters.

On the other hand, an element of the second error in eqn (6) may be a function of terminals or other specific characteristics for a route alternative. Since structuring the element appears to depend on the more specific conditions characterizing each network, we introduced a simple assumption into the second term \( \sum_{r} \) against its possible expansions.

\[
\text{cov}(\varepsilon_r^0, \varepsilon_q^0) = k_{rq} \sigma_k^2 + \delta_{rq} \sigma_0^2
\]

(12)

where \( k_{rr} \) is the number of transfer stations of route \( r \). \( k_{rq} \) is the number of the common transfer stations to routes \( r \) and \( q \). \( \sigma_k^2 \) is a unit variance for a transfer station and the variance characterizing the whole route: \( \sigma_0^2 \), is constant and identical to all routes. \( \delta_{rq} \) is defined by

\[
\delta_{rq} = \begin{cases} 
1, & q = r \\
0, & q \neq r 
\end{cases}
\]

(13)

Equation (12) requires the similar assumptions to eqn (11). We assume that the identical error arises from any transfer station on the route and the errors are independent from each other. The equation indicates that every route has the variance related to the number of its transfer stations and the covariance between two routes depends on the number of their common transfer stations.

Using the above derivation, the multinomial probit model with structured covariance is summarized as follows,

\[
P_r = \iint_{e_r=-\infty}^{e_r=+\infty} \iint_{e_q=-\infty}^{e_q=+\infty} \Phi(\varepsilon)e_{r-\varepsilon_1-\ldots-\varepsilon_k} \Phi(\varepsilon) \sum_{r=1}^{1} \exp[-\frac{1}{2} \varepsilon^T \sum_{r=1}^{1} \varepsilon^T] + \sigma_k^2 \begin{pmatrix} k_1 & k_{12} & \ldots & k_{1R} \\
k_{12} & k_2 & \ldots & k_{2R} \\
\vdots & \vdots & \ddots & \vdots \\
k_{1R} & k_{2R} & \ldots & k_R \end{pmatrix} + \sigma_0^2 I
\]

(14)

\[
\sum = \sigma_1^2 \begin{pmatrix} n_1(x) & n_{12}(x) & \ldots & n_{1R}(x) \\
n_{12}(x) & n_2(x) & \ldots & n_{2R}(x) \\
\vdots & \vdots & \ddots & \vdots \\
n_{1R}(x) & n_{2R}(x) & \ldots & n_R(x) \end{pmatrix} + \sigma_k^2 \begin{pmatrix} k_1 & k_{12} & \ldots & k_{1R} \\
k_{12} & k_2 & \ldots & k_{2R} \\
\vdots & \vdots & \ddots & \vdots \\
k_{1R} & k_{2R} & \ldots & k_R \end{pmatrix} + \sigma_0^2 I
\]

(15)

This model is different from the ordinary multinomial probit model because its covariance matrix is structured and parameterized only by three variances: \( \sigma_1^2, \sigma_k^2 \) and \( \sigma_0^2 \). In fact, if we define \( x \) by distance, \( n_r(x) \) is replaced by \( d_r \) and \( n_{rq}(x) \) is replaced by \( d_{rq} \) which represents the overlapped length of the two routes, \( k_{rr} \) and \( k_{rq} \) are the number of transfer stations. Therefore, we easily assign these values before the estimation of model parameters. The ratios of the following two variances can be estimated in addition to the utility function parameters.

\[
\eta = \sigma_1^2(x)/\sigma_0^2
\]

(17)

\[
\nu = \sigma_k^2/\sigma_0^2
\]

(18)

It is important to note that the number of these additional parameters never alter in accordance with the choice set size. Such a simplification of the covariance matrix structure enables us to utilize the following three advantages. First is the ability to simplify the parameter estimation by fixing the parameter size of the covariance matrix. Second is the ability to represent individual differences of the correlation among alternatives. Third is the ability to assign a variance and covariances, for a new route by the same equations.

2.2.2. Assumption of null terminal variance. We assume in eqn (16) that the error of the utility function arises not only from links but also from nodes on the route. If the variance at terminal station, \( \sigma_k^2 \), in eqn (12) can be ignored, the second term in eqn (7) becomes a simple form,

\[
\text{cov}(\varepsilon_r^0, \varepsilon_q^0) = \delta_{rq} \sigma_0^2
\]

(19)
The above equation indicates that every route has an identical variance and any covariance is zero as the second term in the equation. Using the above derivation and eqn (11), the structured covariance of the multinomial probit model is expressed as follows,

\[
\sum = \sigma^2 \left( \begin{array}{cccc}
n_1(x) & n_{12}(x) & \cdots & n_{1K}(x) \\
n_{12}(x) & n_2(x) & \cdots & n_{2K}(x) \\
\vdots & \vdots & \ddots & \vdots \\
n_{1K}(x) & n_{2K}(x) & \cdots & n_K(x) \\
\end{array} \right) + \sigma_0^2 I \tag{20}
\]

The covariance matrix is parameterized by two variances: \( \sigma^2(x) \) and \( \sigma_0^2 \). The ratio of the two variances, \( \eta \) in eqn (17), should be estimated in addition to the utility function parameters. Equation (20) is the simplest structured covariance of the multinomial probit model because an additional parameter should be estimated with the utility function parameters. The multinomial probit model with structured covariance in this paper can represent strong competition among similar alternatives like the nested logit model. However, the major advantage of the model is that the similarity of each pair of alternatives is not restricted by the nesting structure of the nested logit model. The structure, in fact, varies by the combination of origin and destination stations. The multinomial probit model with structured covariance can express the difference in structure even for the same railway line with different origin destination pairs.

2.2.3. Assumption of identical route length. The multinomial probit model with structured covariance can be further simplified to avoid using \( n_{rq}(x) \) directly as an element of the covariance matrix. We assume that the total variances of all routes are the same and the variance is identical to all passengers, as in the case of the multinomial logit model. This implies that \( n_r(x) \) is assumed to be identical across all individuals and routes so that, using \( n_r(x) (= n_r(x)) \), eqn (8) can be rewritten as

\[
\text{var}(\epsilon_r) = n(x)\sigma^2(x) \tag{21}
\]

and the overlapped ratio between route \( r \) and \( q \) can be defined as

\[
w_{rq} = \begin{cases} 1, & q = r \\ \frac{n_{rq}(x)}{n(x)} , & q \neq r \end{cases} \tag{22}
\]

where the overlapped ratio \( w_{rq} \) is regarded as the ratio of the overlapped size (time or distance) to the total size of the routes. As the number of units, \( n_r(x) \), is identical to all routes in eqn (21), from eqns (11) and (19), the covariance between two routes can be expressed by

\[
\text{cov}(\epsilon_r, \epsilon_q) = w_{rq} n(x)\sigma^2(x) + \delta_{rq}\sigma_0^2 \tag{23}
\]

Finally, the diagonal element of eqn (23) is constant and can be taken out of the covariance matrix to obtain

\[
\sum = (n(x)\sigma^2(x) + \sigma_0^2) \begin{pmatrix} 1 & \theta_{w_{12}} & \cdots & \theta_{w_{1K}} \\ \theta_{w_{12}} & 1 & \cdots & \theta_{w_{2K}} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{w_{1K}} & \theta_{w_{2K}} & \cdots & 1 \end{pmatrix} \tag{24}
\]

where \( \theta \), expressed by

\[
\theta = \frac{n(x)\sigma^2(x)}{n(x)\sigma^2(x) + \sigma_0^2} \tag{25}
\]

is regarded as a parameter. The covariance matrix in eqn (24) includes an unknown parameter \( \theta \) which should be estimated with the utility function's parameters. \( n(x)\sigma^2(x) + \sigma_0^2 \) of eqn (24) is also unknown but cannot be distinguished from parameters of the utility function and can be normalized to one. Therefore, estimation of an additional \( \theta \) is sufficient to calibrate the multinomial
probit model with structured covariance here. This model still considers the differences in covariance among routes and appears more flexible for application to the existing network than the ordinary multinomial probit or multinomial logit model. We already found that it is possible to employ the Section 2.2.2 model, if we prefer avoiding the computational complexity. Although the assumption of fixed route size (time or distance) is recognized as a severe simplification, the covariance form in eqn (24) has the advantage of requiring the minimum individual data related to the route structure: \( \sigma_{ij} \), when the difference of route distance or travel time as an element of covariance matrix may be too sensitive to express the individual choice behavior.

3. ESTIMATED RESULTS OF MULTINOMIAL PROBIT MODEL WITH STRUCTURED COVARIANCE

Mass Transit Passenger Survey data collected by the Ministry of Transport in 1990 was used to calibrate our models. The survey, started in 1960 and conducted every five years, collects data from nearly 400,000 commuters in three metropolitan regions: Tokyo, Osaka and Nagoya in Japan. The commuters were surveyed about their origin, destination, commuter mode, transfer stations, departure time, arrival time, personal profile, while they waited to purchase a monthly pass.

In Japan, commuters who purchase monthly passes exceed 90% of all passengers in peak travel hours. After compiling the survey results, we estimated congestion rates between stations by dividing estimated passenger volume by estimated capacity in order to utilize them for the explanatory variable of route choice models.

We selected more than a thousand individuals from the Tokyo samples to examine the characteristics of the multinomial probit model with structured covariance. The sample selection was conducted randomly but samples which had smaller than three route alternatives were omitted from the data set. The individuals who may have more than four alternatives were included in the data set but their route alternatives were restricted to three in accordance with the observed shares between their origin and destination. The overlapped length between alternatives was measured using the transportation map for every sample. In Tokyo, most passengers commute to the downtown area using nearly thirty radial railways from suburban areas. These lines cross the circumferential lines within the downtown and directly connect with subway lines. Alternative railway routes are quite different among selected samples. The model may require relatively complicated situations such as the existence of overlapped routes in order to obtain a significant parameter of \( \eta \) in eqn (17). For example, if most of the samples have independent alternatives or have nearly identical alternatives, the parameter significance is highly unlikely.

We employed the multinomial probit model with structured covariance with distance measured errors presented in eqn (20). The ratio of two variances, \( \eta \), is estimated with the parameters of the utility function. The utility function \( V_r \) with estimated parameters of the model using eqn (20) as its covariance matrix is presented below.

\[
V_r = -0.00583[\text{in-vehicle cost}] -0.127[\text{access time}] -0.151[\text{egress time}] -0.0695[\text{in-vehicle time}]
\]

\[
\begin{align*}
(5.22) & & (5.23) & & (5.45) & & (5.80) \\
-0.355[\text{time for going upstair}] & & -0.116[\text{time for going down stairs and walking horizontally}] \\
(2.44) & & (5.21) & & & & & \\
-0.118[\text{waiting time}] & & -0.382[\text{number of transfer stations}] & & -0.908[\text{congestion index}] \\
(3.90) & & (3.92) & & (0.82) & & & & & \\
\eta: 0.302 \quad (1.45) & & \text{[variables]} & & \text{t-statistics} & & & & & \\
\text{Likelihood ratio: 0.18} & & \text{Number of samples: 1074}
\end{align*}
\]

(26)

In the above equation, the cost variable is measured in terms of Japanese yen and all time variables are scaled by minute. The congestion index is defined by

\[
\text{congestion index} = \sum_h (\text{congestion rate}_h^2 \times \text{link travel time}_h)
\]
The square of the congestion rate at link \( h \) multiplied by travel time of the link is summed up for all links on the route. We utilized the program package GAUSS to estimate the parameters using the algorithm of numerical integration for a three alternative case. This is because the samples did not have more than four alternative routes and there is no computational complexity in the use of numerical integration. Although we estimated the parameters using the simulated maximum likelihood estimator with the GHK recursive probability simulator, the average difference of parameters was only 3.1% for 25 draws which are required to simulate the choice probabilities and that was 2.5% for 50 draws.

Since the quality of the service level is an important factor in transportation planning and commuters may perceive different satisfaction levels from different service variables, we introduced seven time variables separately into the model such as transferring time, waiting time, access time, in-vehicle time. The model including these statistically significant variables was considered to be successful. The parameter \( \eta \) is not so highly significant but the value is large enough to demonstrate its importance in the model. The differences in the parameters’ magnitudes may be consistent with the commuter’s perception. For example, the parameter of time for going up stairs at a station is distinctly larger than the parameter of time for going down and walking horizontally and those of access or egress time. This result is consistent with the difference of disutilities depending on fatigue or actual energy consumption. The trade-off ratios between the parameter of in-vehicle cost and those of time variables range from 11.9 yen/min for in-vehicle time to 57.5 yen/min for time for going up stairs. Other cost–time ratios are around 20 yen/min. The insignificance of the parameter of congestion index was unexpected but most of the samples have no choice but to use the existing highly congested lines. The congestion reduction policy is still essential in Tokyo. To obtain a significant parameter for the congestion index, the model may include individuals who can commute by less congested lines such as the circumferential railway lines. We found the incorporation of these numerous service variables into the model to be meaningful because we used only the actual chosen routes of commuters.

4. ESTIMATING CONSUMER BENEFIT FROM MULTINOMIAL PROBIT MODEL WITH STRUCTURED COVARIANCE

4.1. Consumer benefit estimation using multinomial logit function

The institutional process often requires an evaluation of the beneficial effects of proposed transportation projects. Measuring direct effects such as reduction of travel cost or an increase of utility is fundamental for evaluation of the project. The multinomial logit model has been regarded as the institutional model of Japanese governmental agencies and provision of the consumer benefit using the multinomial logit models has recently been applied in a practical manner in Japan (Yai, 1989; Yai et al., 1993).

Estimating consumer benefits based on random utility theory has been widely applied since the model was developed. As Williams (1977) showed, log-sum variables defined by a general cost function before and after the project were utilized to estimate the consumer benefit \( \Delta S \), even for route choice models.

\[
\Delta S = \frac{1}{\lambda} \sum_{ij} T_{ij} \left[ \ln \sum_r e^{-\lambda C_r} - \ln \sum_r e^{-\lambda C_r} \right]
\]  

(28)

where \( \lambda \) is a dispersion parameter, \( T_{ij} \) is a travel volume between origin \( i \) and destination \( j \), \( C_r \) is a composite cost for route \( r \), and \( a \) and \( b \) on \( C \) denote after and before the project. As the disutility \( C_r \) has the dimension of cost, a dispersion parameter is used to transform the cost to the utility.

This equation represents the total benefit for all origin destination pairs. A major advantage of eqn (28) is that it is not necessary to estimate travel demand changes in route transition after the improvement of transportation services in order to calculate the consumer benefit. Equation (28) is similar to the difference of the utility level weighted by choice probabilities and is considered to capture the effect of route transition. Therefore, it is enough to know the future travel volume between origins and destinations.
However, the limitation of representing several branches on a tree structure using nested logit models, and the problem of assuming a single dispersion parameter in a complicated tree structure, were pointed out in Williams (1977). We also mention here that route choice behaviors characterized by a few similar routes cannot be expressed by multinomial logit or nested logit models. We know that a method for estimating consumer benefits based on random utility theory, such as Williams' demonstration, is quite important to evaluate projects. Consequently, for the route choice situations, examining an evaluation method of consumer benefits consistent with the multinomial probit model with structured covariance would be necessary for network improvement projects. Such a method may be utilized in the evaluation of the priorities in constructing additional railway lines on the existing network.

As the expected value of maximum utilities indicates the utility level of a consumer, we obtain the consumer benefit index after transforming the utility level into the monetary term. However, a route choice is generally explained by many kinds of service variables, as mentioned in the previous section. A travel cost is nothing more than a variable and a composite cost is seldom available before the model estimation, as in Williams' case. We now focus on a travel cost variable in the utility function. Using the variable, we are able to calculate the cost reduction which makes the utility level before the project equivalent to that after the project. As the cost reduction is provided to users to increase their utility level and the idea is similar to an equivalent variation, we may say it is an equivalent gain. The equivalent gain principally depends on the parameter and the level of the maximum utilities. The gain, $\Delta C$ is describes in the following equation

$$E(\max, U^r_{\max}) = E(\max, (U^r_{\max} + \beta \Delta C)), \beta > 0$$

(29)

where $U^r_{\max}$ indicates the random utility of route $r$ after the project and $U^r_{\max}$ is that before the project. $\beta$ is the reversed parameter of cost in the utility function though the original sign of $\beta$ is always negative. Therefore, $+ \beta \Delta C$ implies the utility's increase. In addition, the equivalent gain $\Delta C$ is added to all alternatives before the project. This is because a user can choose any route and the choice is made only once. Consequently, the average of the gain weighted by choice probabilities is always equal to $\Delta C$ and a user finally attains $\Delta C$ as a fare reduction.

Let us explain the equivalent gain for the multinomial logit model case as a simple example.

If a random utility is expressed by the following composite cost function

$$U_r = -\beta C_r + \varepsilon_r$$

(30)

and the error term distributes as Gumbel, eqn (29) is replaced by

$$\ln \sum_r e^{-\beta C^r} = \ln \sum_r e^{-\beta C^r + \beta \Delta C}$$

(31)

then $\Delta C$ is easily derived as

$$\Delta C = \frac{1}{\beta} \left[ \ln \sum_r e^{-\beta C^r} - \ln \sum_r e^{-\beta C^r} \right]$$

(32)

This expression is identical to eqn (28) by Williams.

Now, if the utility function contains several variables, we can write it by

$$U_r = \alpha X_r - \beta C_r + \varepsilon_r$$

(33)

where $C_r$ is not a composite cost function but rather an out-of-pocket cost. $X_r$ is a vector of several variables including travel time and transfer time. The assumption of the Gumbel distribution leads to the form of $\Delta C$ as follows

$$\Delta C = \frac{1}{\beta} \left[ \ln \sum_r e^{\alpha X^r - \beta C^r} - \ln \sum_r e^{\alpha X^r - \beta C^r} \right]$$

(34)
The meaning of this equation is similar to eqn (32). The monetary transformation of the difference between the expected value of maximum utilities before and after the project is achieved by the equation.

4.2. Consumer benefit using multinomial probit model with structured covariance

The idea introduced above can be extended to the multinomial probit model with structured covariance. However, unlike the multinomial logit model, the multinomial probit model with structured covariance has the error term distributed normally, and thus the expected value of maximum utilities is impossible to express by the closed form. Our approach considers that the distribution function of maximum utilities is described by the probability of \( U^* \) in which any alternative's utility does not exceed \( U^* \). Now, let \( F(U^*) \) be the distribution function as follows

\[
F(U^*) = \text{Prob}(\max, U_r \leq U^*) = \text{Prob}(U_1 \leq U^*, ..., U_R \leq U^*) \tag{35}
\]

If the error term of the utility \( U_i \) distributes as normal, eqn (35) can be written by

\[
F(U^*) = \int_{t_1 = -\infty}^{U^* - V_1} \cdots \int_{t_R = -\infty}^{U^* - V_R} \Phi(\epsilon) d\epsilon_R \ldots d\epsilon_1 \tag{36}
\]

The density function \( \Phi(\epsilon) \) is obtained from the multinomial probit model like eqn (4) or eqn (15). The expected value of \( U^* \) is the same as the most frequent value of the density function when the distribution is symmetric. As we did not sufficiently examine the characteristics of the above function, numerical calculations associated with the first derivative of the function appeared almost symmetric for three alternatives cases.

\( U^* \), in which the second derivative of \( F(U^*) \) with respect to \( U^* \) equals zero, provides the most frequent value of maximum utilities for the multinomial probit model.

\[
\frac{\partial^2 F(U^*)}{\partial U^{*2}} = 0 \tag{37}
\]

In order to find the equivalent gain for the multinomial probit model, first we should calculate the most frequent value of maximum utilities after the project: \( \hat{U}^*(V^a) \), using the function

\[
F(U^*) = \int_{t_1 = -\infty}^{U^* - V_1} \cdots \int_{t_R = -\infty}^{U^* - V_R} \Phi(\epsilon) d\epsilon_R \ldots d\epsilon_1 \tag{38}
\]

Secondly, we can find the gain \( \Delta C \) which makes the most frequent value of maximum utilities derived from the following function equal to that from eqn (38).

\[
F(U^*) = \int_{t_1 = -\infty}^{U^* - (V^a + \beta \Delta C)} \cdots \int_{t_R = -\infty}^{U^* - (V^a + \beta \Delta C)} \Phi(\epsilon) d\epsilon_R \ldots d\epsilon_1 \tag{39}
\]

This function implies that the distribution of maximum utilities when \( \beta \Delta C \) is added to each utility. Following the above process, we can obtain the equivalent gain for the multinomial probit model, which satisfies the utility level equivalence associated with most frequent value of maximum utility as in eqn (29). A simple algorithm for finding the gain is a numerical integration method using eqns (37) and (39). The process for the multinomial probit model with structured covariance is identical to the one above.

However, \( U^* \) is the satisfaction function (see Daganzo, 1979) and if the covariance matrix is identical before and after the project development, we can utilize the same property of \( U^* \) with eqn (31),

\[
\bar{U}^*(V + a) = \bar{U}^*(V) + a \tag{40}
\]
\( \Delta C \) for the multinomial probit model is also obtained by

\[
\Delta C = \frac{1}{\beta} (\bar{U}^*(V^a) - \bar{U}^*(V^b))
\]  

(41)

\( U^*(V^b) \) can be calculated by eqn (37) using

\[
F(U^*) = \int_{\epsilon_1=-\infty}^{U^*+\epsilon_1} \int_{\epsilon_2=-\infty}^{U^*+\epsilon_2} \Phi(\epsilon) \, d\epsilon_1 \, d\epsilon_2
\]  

(42)

Whenever we find the values of \( U^* \) by the second derivative of eqns (38) and (42), we easily calculate \( \Delta C \) using eqn (41) which is the same equation as the multinomial logit case.

After finding equivalent gains for individuals, it is necessary to sum them up for the whole population to obtain the estimated value of the consumer benefit after the project. The consumer benefit (CB) is simply defined by,

\[
CB = \sum_{ij} t_{ij} \Delta C_{ij}
\]  

(43)

where \( i \) and \( j \) indicate any origin destination pair on the network. \( t_{ij} \) is travel demand between \( i \) and \( j \) and \( \Delta C_{ij} \) is an equivalent gain between \( i \) and \( j \). From the above derivations, the consumer benefit by eqn (43) enables us to evaluate service improvement of railway projects after estimating the model parameters. In addition, it is important to note that the estimated benefit from IID (independently and identically distributed) models may have a distinct bias, which may lead to incorrect evaluation for railway projects. This is because the models cannot reflect the similarities between alternatives. The amount of bias depends on several conditions associated with network structures and projects. Evaluating such bias may be required to understand the characteristics of most frequent value of maximum utility derived from the multinomial probit model with structured covariance.

As two simple examples of the most frequent value of maximum utilities for the multinomial probit with structured covariance and IID probit models, we demonstrate here differences of equivalent gains per person for a simple network which has a pair of origin and destination and three alternative routes. However, the result is easily extended to all passenger volumes for the existing networks. Let us consider three routes labeled A, B and C, whose lengths are identical. Routes A and B overlap for half of their length but C is not overlapped with A or B.

First we consider the improvement of the overlapped section between A and B, and 6 min were subtracted from total travel time in routes A and B. Generally speaking, different model assumptions may lead to the different consumer benefit values. In this example, the equivalent gain from the multinomial probit model with structured covariance with \( \theta = 1 \) (in the model of the Section 2.2.3) was 49 yen/person but the IID probit model with \( \theta = 0 \) provided 52 yen/person. We may say this difference is the bias from the assumption of IID. The equivalent gain must be overestimated by IID model in this case because the model regards two overlapped routes as independent and assumes that two routes are improved independently.

The second example is the case that the travel time on the non-overlapped section of route A was improved in 6 min. In this case, the estimated equivalent gain from the IID probit model is 30 yen/person but the gain from the multinomial probit model with structured covariance is nearly 33 yen/person. This is because one of the independent three alternatives is improved in the IID model but two of the three are, in fact, dependent on each other. Consequently, as the utility level estimated by the IID model is superior to the actual level, the improvement of a route has relatively less effect on the IID model.

These fundamental results imply that the bias of estimated equivalent gains depending on model assumptions may not be negligible in evaluating projects on existing complicated networks. The IID models probably behave inadequately in complicated networks which have several overlapped routes. In the above examples, two alternative projects were introduced and both models
indicated the superiority of the improvement of the overlapped section on routes A and B over that of the independent section on route A. The difference of two projects were estimated as 22 yen (52–30) by the IID model and 16 yen (49–33) by the multinomial probit model with structured covariance. The amount of the difference will become huge for aggregated passenger volume in multiple networks. Both over- and under-estimation may arise in several origin–destination pairs if we employ the IID model and thus may lead to inappropriate conclusions about railway sections to be improved. Consequently, utilizing the multinomial probit model with structured covariance is preferable to the IID models in the estimation of user benefits on complicated networks.

5. CONCLUSION

Our research in this paper provides a new utilization of the multinomial probit model for rail networks. The multinomial probit model with structured covariance had three major advantages of representing route choice behaviors. First was the ability to simplify the parameter estimation by fixing the parameter size of the covariance matrix. Second was the representability of individual differences in relationship among alternatives. Third was the ability to assign a variance and covariances for a new route by the same equations. These advantages enable us to apply the model to complicated railway route choice situations. In addition, the consumer benefit index based on the model appears to be more efficient than the IID models for networks with overlapped route alternatives. Some applications indicated that several service variables required to evaluate railway planning options had been successfully included in the utility function.

In addition to the choice behavior mentioned in this paper, the multinomial probit model with structured covariance by distance related measures is applicable to other behaviors, such as a combination choice of travel modes or destination choice. Even for route choices, we may extend the model for modeling more complicated situations with local and express rail services or with locational similarity among railway lines. Estimating full parameters of the multinomial probit model with structured covariance for more than four alternatives and evaluating the consumer benefit index in the existing networks are expected in further studies.

Acknowledgements—The authors thank Moshe Ben-Akiva, Frank Koppelman, Denis Bolduc and two anonymous referees for useful suggestions concerning an earlier draft of this paper. The authors also wish to thank the Japanese Ministry of Transport for providing the passenger survey data and acknowledge the contribution to this paper provided by Takahiro Nakagawa and Michelle Franciose.

REFERENCES


