Nonlinearity in Utility Function in Travel Mode Choice Model

Student Number: 03M16119 Name: Hiroki Sango Supervisor: Prof. Tetsuo Yai

交通手段選択モデルにおける効用関数の非線形性の検証

山後 宏樹

非集計行動モデルは交通需要予測やプロジェクト評価を行う際に様々な意味で重要な役割を担っており，その統計的推測は，実証分析上極めて重要な問題である．多くの理論・応用研究では，所要時間や費用などの説明変数に対して線形や対数線形の特定化を行った，比較的単純で取り扱いが簡単な効用関数が用いられてきた．しかし，効用関数の特定化が適切ではない場合（特に，変数とその部分効用との間に非線形関係が認められる場合），誤ったインプリケーションが導かれる可能性もある．そこで本研究では，関数の形状にアプリオリな仮定を与えないノンパラメトリック回帰分析の手法に基づいて，非集計行動モデルにおける効用関数の非線形性の検証を行った．具体的には，一般化加法モデルの枠組みのもとで，ノンパラメトリックの効用関数を持つ三肢（航空，鉄道，自動車）の交通手段選択モデルを構築し，幹線旅客純流動調査から得られた個票データを用いてモデルの統計的推測を行い，所要時間，費用，航空運行頻度等の説明変数と効用値の関係について実証的な考察を行い，これらの変数の非線形性を統計的に確認した．

1 Introduction

Behavioral models in transportation have extensively been used for quantitative assessment of travel behavior analysis. After 1970s, disaggregate behavioral models based on the principle of random utility maximization took the central stage in modeling travel decisions. Since the operational capability of these choice models were quite high, they are being frequently used as very effective and practical tools. As the time passes by, many researchers have pointed out different issues and proposed different modeling concepts that can easily be incorporated in the original models.

For example, Yai et al (1996) proposed a structured variance-covariance matrix of error parameters in traditional Probit model to analyze the route choice behavior of railway commuters. Train & McFadden (2000) developed mixed Logit models to estimate the random parameters of explanatory variables. Swait & Ben-Akiva (1987) and Morikawa et al (1991) developed choice set generation models.

As described above, most of the existing researches on disaggregate travel choice models are based on the random utility assumption. Generally, explanatory variables of utility function have linear form, log-linear form, and CES (Constant Elasticity of Substitution) form. The reasons for such simplified assumptions are computational tractability and easiness in application. These assumptions are a priori ones. There are few literatures in marketing research that have used the non-parametric approach to specify the functional form of variables and to estimate the parameters. To the author's knowledge, there are no literatures in transportation that have analyzed non-linear functional form of random utility and non-parametric approach to estimate the parameters. In addition, we assumed some uncertain explanatory variables such as travel frequency of airlines or congestion rate of railways.

In these modeling premises, the main objective of this study is to investigate non-linear functional form of explanatory variables and to estimate the parameters of these variables using nonparametric approach, which is frequently used in econometrics or medical statistics.

2 Theoretical background

2.1 Outline of Nonparametric Methods

The adopted theory in this research originates from Hastie & Tibshirani (1990). Nonparametric method is one type of smoothing method, which depicts smooth curve based on the observed data. The difference between nonparametric and parametric approaches is that in parametric approach we assume the distribution function and a small set of parameters, while in non-parametric method; we let the observed data distribution freely without reducing the number of parameters.

Nonparametric methods have two main advantages. First, it graphically depicts scatter plot between *Y* and *X*, which is very easy to see the trend or the relationships between *Y* and *X*. The second advantage is that it not only estimates the dependence of the mean *Y* on the predictors, but also serves as a building block for the estimation of additive models.

In particular, nonparametric method shows the piecewise polynomial function:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{l=1}^{L} \eta_l (x - t_l)_+^3 \quad (1)$$

where the notation $(\ )_+$ denotes the positive part of ( ).

It is called "smoothing cubic spline". This function has the following properties:

- f is cubic polynomial in any sub-interval $[t_l, t_{l+1}]$;
- f is twice differentiable; and
- f has a third derivative that is a step function with jumps at $t_1, \ldots, t_L$.

2.2 Outline of GAM

This theory is proposed by Hastie et al (1990). GAM (Generalized Additive Models) are the extensions of generalized linear models (GLIM), and GLIM are themselves a generalization of linear regression models.

Specifically, the predictor effects are assumed to be linear in the parameters, but the distribution of the responses, as well as the link between the predictors and this distribution, can be quite general. Many useful models fall into this class, including the linear logistic model for binary data. Here, the response Y is assumed to have a Bernoulli distribution with $\mu = pr(Y = 1 | x_1, \ldots x_p)$, and μ is linked to the predictors via $\log\{(\mu/1-\mu)\} = \alpha + \sum_j x_j \beta_j$. Other familiar models in this comprehensive class are log-linear models for categorical data and gamma regression models for responses with constant coefficient of variation. The family of generalized linear models provides an appealing framework for studying the common structure of such models, and as we will see, there is a convenient, unified method for their estimation.

GAM is an extension of the class of generalized linear model. They extend the linear models by replacing the linear form $\alpha + \sum_j x_j \beta_j$ with the additive form of $\alpha + \sum_j f_j(x_j)$. The logistic additive model is one of simple examples. When it is applied to binary response data, the model takes the form of $\log\{\mu/(1-\mu)\} = \alpha + \sum_j f_j(x_j)$.

3 Model

3.1 Nonparametric Multinomial Logit Model

The simple multinomial Logit model is formulated as follows:

$$\begin{aligned} U_{ki} &= \beta_1 x_{1ki} + \cdots + \beta_p x_{pki} + \varepsilon_n \\ &= \sum_{j=1}^{p} \beta_j x_{jki} + \varepsilon_n \\ &= V_{ki} + \varepsilon_n \end{aligned} \tag{3}$$

where, U_{ki} is the utility of choice k for an individual i, V_{ki} is the deterministic utility of choice k for an individual i, x_{pi} is the pth explanatory variable of individual i, β_p is the p th parameter, and ε_n is the error term.

Then, the choice possibility of alternative choice k for an individual i is as follows:

$$P_i(k) = \frac{\exp(V_{ki})}{\sum_{k=1}^{m} \exp(V_{ki})}, \quad \forall k \text{ and } i \tag{4}$$

In this paper, by applying nonparametric method and GAM to multinomial logit model, we can derive the forms of explanatory variables of utility function. Then, equation (3) is changed as follows:

$$V_{ki} = \sum_{j=1}^{p} f_j(x_{jki}) \tag{5}$$

where, $f_i(x_{jki})$ is x_{jki}'s nonparametric function it. This nonparametric function is assumed to be the smoothing cubic spline as equation (1).

Then, equation (4) can be rewritten as:

$$P_i(k) = \frac{\exp(\sum_{j=1}^{p} f(x_{jki}))}{\sum_{k=1}^{m} \exp(\sum_{j=1}^{p} f(x_{jki}))}, \quad \forall k \text{ and } i \tag{6}$$

3.2 Model Estimation

Using (5) and (6), we estimate the forms of explanatory variables. Normally, standard logit model can estimate with using maximum likelihood estimation (MLE). The actual estimation procedure maximizes the following log-likelihood function:

$$\ln L = \sum_{i=1}^{n} \sum_{k=1}^{m} y_{ki} \ln P_i(k) \tag{7}$$

where, y_{ki} is the indicator whether the alternative k is chosen by individual i or not (y_{ki} =1 if the alternative k is chosen by individual i and y_{ki} =0 for otherwise).

When we use the nonparametric method, the log-likelihood function can be conditional, and it is called penalized log-likelihood function.

The penalized log-likelihood method is initially proposed by Good & Gaskins (1971). Anderson & Blair (1982) apply this estimation method to logistic regression that examines the effect of individual characteristics on binary choices. In this paper, we examine a nonparametric multinomial logit model. Therefore, we use the penalized log-likelihood function. The penalized log-likelihood function is:

$$j(f_1, \ldots, f_p) = \sum_{i=1}^{n} \sum_{k=1}^{m} y_{ki} \ln P_{ki} - \sum_{j=1}^{p} \lambda_j \int (f_j''(x))^2 dx \tag{8}$$

where λ_j is the jth smoothing parameter to determine the degree of smoothness.

Since we are interested to obtain the functional form of explanatory variables, we partially differentiated the penalized log-likelihood with respect to V_{ik}. The estimates of V_{ik}s (in the vector form $\hat{\mathbf{V}}$) must satisfy the following equation.

$$\frac{\partial j}{\partial \mathbf{V}} = 0 \tag{9}$$

In practice, equation (9) can not be analytically solved. Therefore, we use the Newton-Raphson algorithm for estimation. This series of algorithm is called the local scoring procedure.

3.3 Selection of Smoothing Parameter

The first term of equation (8) represents the closeness to the data, while the second term of (8) penalizes curvature in the function.

The smoothing parameter λ_j's control the roughness of the fitted curve. Hence, the parameter λ_j plays the bias-variance trade-off. Large values of λ_j produce the smoother curves while smaller values produce more wiggly curves. At the one extreme, as $\lambda_j \to \infty$, the penalty term dominates, by forcing $f''(x) = 0$ for every domain, and the solution is the normal log-likelihood function. At the other extreme, as $\lambda_j \to 0$, the penalty term becomes unimportant and the solution tends to an interpolating twice-differentiable function.

Various methods to determine the optimal value of λ_j have been proposed in the literature. For its simplicity and popularity, we chose to use the generalized cross-validation (GCV) as the criterion to determine the optimal value of λ_j. The GCV function for selecting λ_j can be formulated as:

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{y_i - \hat{f}_\lambda(x_i)}{1 - tr(\mathbf{S}_\lambda)/n}\right\}^2 \tag{10}$$

where $\mathbf{S}_\lambda$ is smoother matrix. $\mathbf{S}_\lambda$ must fulfill the following relationship for any constants a and b:

$$S\left(a\mathbf{y}_1 + b\mathbf{y}_2 \left|\mathbf{x}\right.\right) = aS\left(\mathbf{y}_1 \left|\mathbf{x}\right.\right) + bS\left(\mathbf{y}_2 \left|\mathbf{x}\right.\right) \tag{11}$$

If we focus on the fit at the observed points $x_1, \ldots, x_n$, a linear smoother can be written as:

$$\hat{\mathbf{f}} = \mathbf{S}\mathbf{y} \tag{12}$$

In principal the optimal λ_j is determined by minimizing the GCV function with respect to λ_j.

4 Empirical Analysis of the Nonlinearity in Utility Functions of Travel Moce Choice Model

4.1 Data

We apply this model for travel mode choice model . Because the main objective of this study is estimate the form of nonparametric function, it is necessary to apply large amount of disaggregate data. We collected these data from "2000 Inter-regional Person Trip Survey" conducted by Land, Infrastructure and Transportation Ministry (MLIT). We collected LOS (level of service) data by NAVINET data and combined them to each disaggregate data.

In the model specification, the mode choice model has three alternatives (airline, railway, car). We used the business and private trip for three alternative travel models. We construct the model "trip time (min)", "trip cost (yen)" and "flight frequency (flight/day)". And we randomly chose approximate 3000-4000 individuals who chose air, rail and private cars from the overall sample. Here, we compare the following three model specifications: the first one has linear utility function, the second model has log-linear model, and the last model has nonparametric function. These specifications are as follows:

1. Linear model:

$$\begin{aligned} V = {} & \beta_{time} \times time + \beta_{cost} \times cost + \\ & \beta_{air\_freq} \times \left(air\_freq\right) + ASC's \end{aligned} \tag{13}$$

2. Log-linear model:

$$\begin{aligned} V = {} & \gamma_{time} \ln\left(time\right) + \gamma_{cost} \ln\left(cost\right) + \\ & \gamma_{air\_freq} \ln\left(air\_freq\right) + ASC's \end{aligned} \tag{14}$$

3. Nonparametric model:

$$\begin{aligned} V = {} & f_{time}(time) + f_{cost}(cost) + \\ & f_{air\_freq}(air\_freq) + ASC's \end{aligned} \tag{15}$$

where, *ASC*'s denote alternative specific constant.

4.2 Estimation Results

(1) The functional form of variable "Travel Time"

First, we show the form of travel time variable. We classify the trip into business trip and private trip.

Figure 1 and Figure 2 show both business trip and private trip variable might have similar function. In these graphs, full lines are nonparametric functions and dot-lines denotes 95% confidence intervals of the estimates of nonparametric functions. They are linear function before 600 minutes approximately for both business trip and private trip. Exceeding 600 minutes they are step-down. These are similar to log-linear functions. These functional forms support the law of

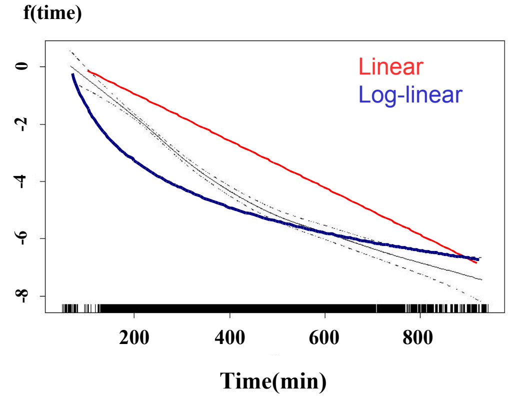

Fig.1 Variable “time” form (business) (3011 samples)

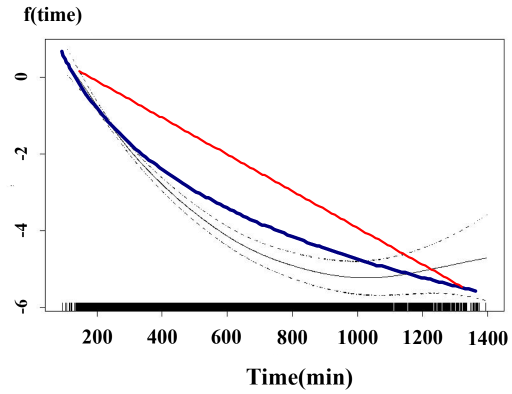

Fig.2 Variable “time” form (private) (3079 samples)

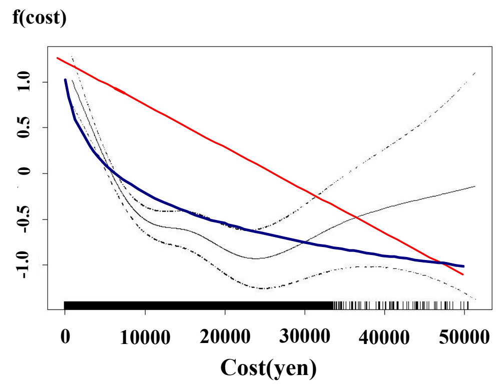

Fig.3 Variable “cost” form (business) (3011 samples)

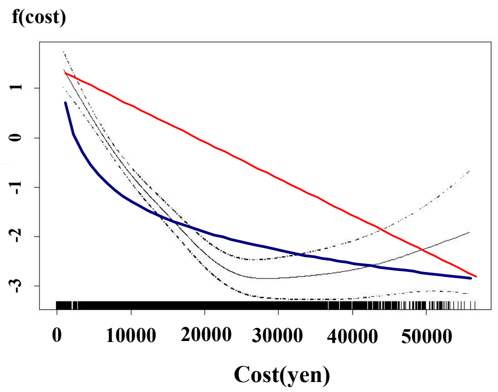

Fig.4 Variable “cost” form (private) (3079 samples)

decreasing marginal utility that is a basic concept of economics.

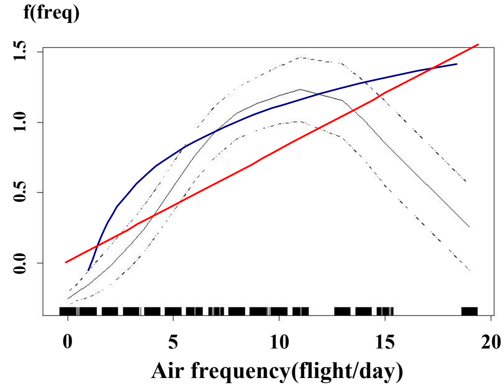

Fig.5 Variable “Flight frequency” form (business) (3011samples)

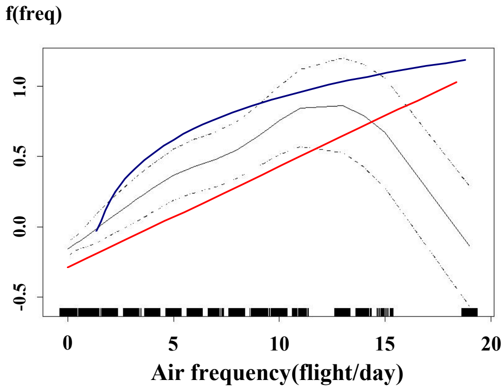

Fig.6 Variable “Flight frequency” form (private) (3079 samples)

(2) The functional form of variable “Travel cost”

Figure 3 and Figure 4 shows “cost” variables are step-down like “time” variable.

Note that, in Figure 3, “cost” variable of business has three inflection points. First inflection point is about 8000 yen, second is about 15000 yen, and third point is 25000 yen. It shows the bound from 8000 yen to 15000 yen, utility doesn’t decrease very well. The advantage that can be gained from the nonparametric method is to get these inflection points.

(3) The functional form of variable “Flight frequency”

Figure 5 seems more smoothing than Figure 6. This difference is considered because of the different trip purpose.

Business trip makers are more sensitive to the flight frequency than those of private trip makers. Therefore figure 5 is more important than Figure 6.

The first half of Figure 5 is similar to logistic curve. This curve is similar to our conceptual understandings. That is, if a value is in the low range, utility doesn’t increase well. Similarly, when the value is high, the change is utility is low. However, if a value is in the middle range, utility increase significantly.

From 12 to 20 in Figure 5 and 6, utilities are decreasing. Reason of this is these are the major air-route. When people use the airline at this route, people don't think about frequency because flight frequency is very high. Therefore air frequency is decreasing at this range.

(4) Model comparison dependent on trip length

We divided Japan into 8 blocks area (Hokkaido, Tohoku, Hokuriku, Shinetsu, Kanto, Tokai, Kinki, Chugoku, Shikoku, Kyusyu and Okinawa) and compared the variable's form by difference of trip length. That is, one group's trip in the same block (ex. From Kanto to Kanto trip) the other group's trip over two block (ex. from Kanto to Hokkaido or Tohoku trip). Here, we verify only private trip.

Intra trip (same block) has 2937 samples, and inter trip (different block) has 2426 samples. As shown in Figure 7 and Figure 8, these forms have approximately same shapes like log-linear form.

Intra trip diminishes gradually from 700, however inter trip diminishes from 500. There is difference between trip length. As shown in Figures 9 and Figures 10, these function s are almost same. However, Inter trip doesn't change over 25000, long trip people doesn't feel expensive.

Figure 11 and Figure12 shows that the estimated function for short trip is concave whereas that for long trip is convex. The conventional assumption of log-linear transformation of air-frequency may not be appropriate for short trip.

Therefore, we show that it isn't correct parameters of variables are assumed regardless of trip of length.

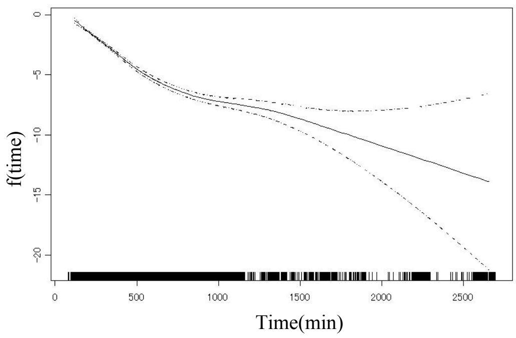

Fig.7 Variable "time" form of intra regional trip

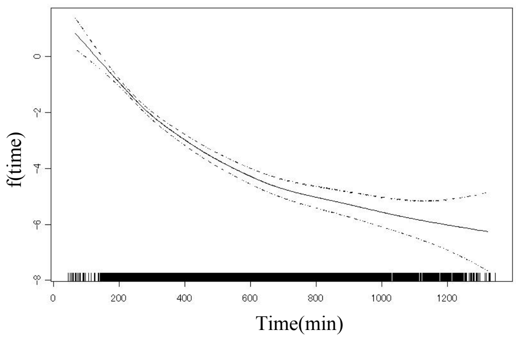

Fig.8 Variable "time" form of inter regional trip

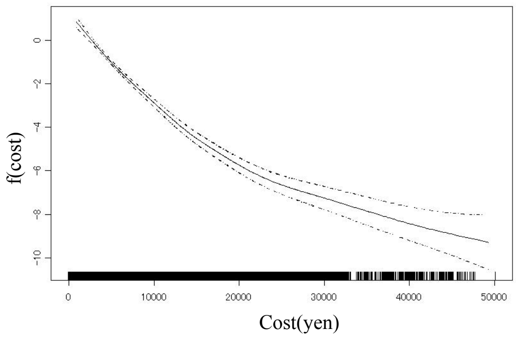

Fig.9 Variable "cost" form of intra regional trip

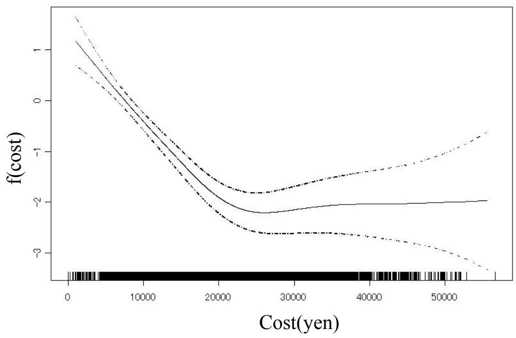

Fig.10 Variable "cost" form of inter regional trip

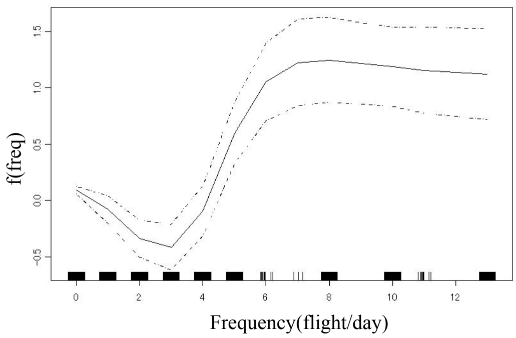

Fig11 Variable "cost" form of intra regional trip

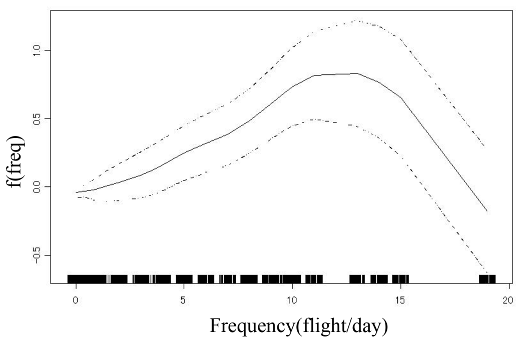

Fig.12 Variable "cost" form of inter regional trip

(5) Comparison of goodness fit

We conduct the comparison of goodness fit between the linear model and the nonparametric model.

This table shows that the likelihood ratio of nonparametric model is higher than the linear model's one. Therefore the nonparametric model is better model than the linear model.

5. Sensitivity analysis

Finally, we conduct the sensitivity analysis using nonparametric model. We choice the representative inter-regional trip (Tokyo-Osaka), and change the time and the cost variables of car in small steps (10 minutes and 500 yen). And we analyze whether the possibility of car changes (Figure 13 and 14).

These figure show that car choice Prob. is more sensitive in non-parametric model than in the existing linear model.

Because of the complexity of the functional form of travel cost variable, the sensitivity of car choice probability calculated from non-parametric model is highly sensitive at lower values of travel cost but gradually less sensitive at higher values of travel costs. Therefore nonparametric model shows more sensitive reaction than linear model.

Table.1 Goodness fit the model (business trip)

	Linear model	*Nonpara model*
Maximum likelihood	-2312.739	-2173.812
Likelihood ratio	0.283	0.326

Table.2 LOS of Tokyo-Osaka

	Time (min)	Cost (yen)
Air	222	19900
Rail	218	15140
Car	515	12170

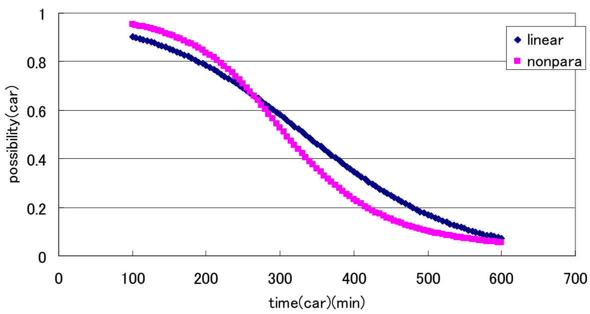

Fig.13 Variable "cost" form (business) (3011 samples)

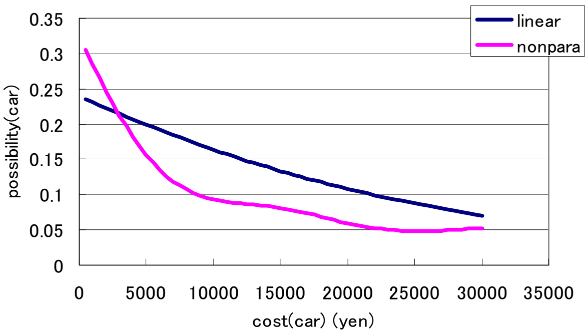

Fig.14 Variable "cost" form (business) (3011 samples)

6 Conclusions and Future Works

In this paper, nonparametric multinomial Logit model is used for analyzing the functional form of explanatory variables. The following conclusions can be drawn from the analysis:

1. This study proposes a non-parametric approach to estimate utility functions in disaggregate choice models.
2. We apply this approach to mode choice model from the data collected from inter-regional person trip survey.
3. We analyze functional forms of variables to compare with existing simple linear and log-linear models.
4. "Time" and "Cost" variables are similar to log-linear model.
5. The proposed non-parametric approach can be used as a tool to make diagnosis of parametric specifications.

We adapt nonparametric method only for Logit model. However, this method can be extended to other discrete choice models also. Non-parametric approach reveals more information to discrete choice models and hence, it is better to use this model to other situations. For example, estimating railway route choice congestion rate is an example. However congestion rate function isn't so explicit. This model will give effective implication.

References

[1] Yai, T. and Nakagawa, T.: Applicability of multinomial probit models with structured covariance, *Infrastructure Planning Review*, Vol.13, pp.563-570, 1996.

[2] McFadden, D. and Train, K.: Mixed MNL models for discrete response, *Journal of Applied Econometrics*, Vol.15, pp.447-470, 2000.

[3] Swait, J. and Ben-Akiva, M.: Incorporating random constraints in discrete models of choice set generation, *Transportation Research B*, Vol.21, pp.91-102, 1987.

[4] Morikawa, T., Takeuchi, H. and Kako, Y.: Destination choice analysis of vacation trips considering attractiveness of regions and probabilistic choice sets, *Infrastructure Planning Review*, No.9, pp.117-124, 1991.

[5] Good, I. and Gaskins, R.: Nonparametric roughness penalties for probability densities, *Biometrika*, Vol.58, pp.255-277,1971.

[6] Anderson, J. and Blair, V.: Penalized maximum likelihood estimation in logistic regression and discrimination, *Biometrika*, Vol.69, pp.123-126, 1982.

[7] Hastie, T. and Tibshirani, R.: *Generalized Additive Models*, Chapman & Hall, London, 1991.