

INTISARI

Pemodelan Pilihan Diskrit Menggunakan Model Probit Dan Model Mixed Logit Pada Respon Multivariat

Oleh

Jaka Nugraha
(05/1656/PS)

Pemodelan Pilihan diskrit (DCM : *Discrete Choice Model*) menggambarkan pilihan pembuat keputusan diantara sekumpulan alternatif yang didasarkan pada nilai utilitasnya. Pilihan atau respon berupa data nominal dan responden memilih salah satu yang mempunyai nilai utilitas tertinggi (maksimum). Permasalahan dalam DCM adalah mendapatkan model yang dapat menggambarkan perilaku pembuat keputusan (responden) dalam menentukan pilihan. Ada dua aspek yang menjadi perhatian, yaitu model yang digunakan dalam menggambarkan perilaku dan metode yang digunakan untuk mengestimasi parameter. Model yang telah banyak dibahas antara lain model Logit, model *Nested Logit*, model *Generalized Extreme Value* (GEV), model Probit, model *Mixed logit*. Metode estimasi yang biasa digunakan adalah Metode MLE (*Maximum Likelihood Estimation*), Momen, Bayesian.

Dalam penelitian bidang transportasi, riset pasar maupun politik, seringkali pengamatan melibatkan respon multivariat. Liang dan Zeger (1986) telah menyampaikan bahwa analisis Logistik maupun Probit pada data panel dengan menggunakan pendekatan univariat. Jika korelasi diabaikan maka penaksir menjadi tidak efisien. Demikian juga pada respon multivariat, jika antar respon terdapat korelasi dan dianalisis menggunakan pendekatan univariat maka akan mempengaruhi sifat-sifat penaksirnya. Biasanya pada kasus multivariat, model disusun menggunakan kombinasi semua alternatifnya. Pendekatan ini mengakibatkan sifat *Independence from Irrelevant Alternatives* (IIA) tidak terpenuhi.

Penelitian ini dimaksudkan untuk menyusun model dengan respon berupa data nominal multivariat, yaitu responden harus membuat beberapa keputusan sekaligus (lebih dari satu). Misalkan π_{ijt} adalah probabilitas responden i pada keputusan ke- t memilih alternatif j diantara J_t pilihan untuk $i=1, \dots, n$; $t=1, \dots, T$ dan $j=1, \dots, J_t$. Nilai π_{ijt} dipengaruhi oleh karakteristik responden i (X_i) dan karakteristik pilihan j (Z_{ijt}). Diasumsikan bahwa diantara keputusan saling berkorelasi. Model yang akan digunakan dalam penelitian ini adalah model Logit, model Probit dan model Mixed Logit. Metode estimasi yang digunakan adalah *Maximum Simulated Likelihood* (MSL). Metode MSL adalah sama dengan metode MLE, akan tetapi fungsi probabilitasnya dihitung menggunakan pendekatan simulasi. Sifat-sifat asimtotis metode MLE digunakan untuk menguji parameter dalam model. Metode simulasi yang digunakan untuk menghitung proporsi pilihan adalah metode GHK (Geweke-Hajivassiliou-Keane) dan Barisan Halton.

Kontribusi utama penelitian ini adalah mendapatkan metode untuk menyusun model DCM pada data respon nominal multivariat yang dapat mengakomodasi adanya korelasi antar respon. Pembahasan diawali pada kasus respon biner (dikotomi), yaitu masing-masing respon mempunyai dua alternatif. Pada respon biner, metode estimasi



parameter yang digunakan adalah MLE dan GEE. MLE pada model Logit disusun menggunakan *Bahadur's Representation*. Simulasi GHK digunakan untuk mendapatkan MLE pada model Probit dan simulasi Halton digunakan untuk mendapatkan penaksir MLE pada model Mixed Logit. Oleh karena *Bahadur's Representation* dan metode GEE hanya dapat diimplementasikan pada respon biner, penyusunan DCM pada respons nominal multivariat menggunakan model Probit dan model Mixed Logit. Parameter pada model Probit dan model Mixed Logit diestimasi menggunakan metode SMLE.

Telah dilakukan studi simulasi respon biner multivariat maupun nominal multivariat untuk mengetahui pengaruh besarnya korelasi terhadap sifat penaksirnya. Data dibangkitkan pada nilai parameter tertentu (diketahui) menggunakan software R.2.8.1. Berdasarkan data simulasi tersebut, penaksir yang diperoleh untuk masing-masing model dibandingkan dengan nilai parameternya. Pada respon biner multivariat, diperoleh beberapa kesimpulan. Pertama, penaksir GEE dan MLE pada model Logit adalah relatif sama. Penaksir tersebut tidak terpengaruh oleh besarnya korelasi antar respon. Penaksir parameter korelasi pada GEE yang didasarkan Liang-Zeger (1986) cenderung *underestimate* dan berbanding lurus dengan besarnya korelasi. Sedangkan metode GEE yang didasarkan Chaganti-Joe (2004) pada model Probit lebih akurat dalam mengestimasi parameter korelasi. Kedua, jika terdapat pelanggaran asumsi (yaitu besarnya variansi) maka akan mengakibatkan penaksir pada model Probit dan model Logit menjadi bias. Ketiga, model Mixed Logit dapat mereduksi bias yang terjadi pada model Probit dan model Logit. Model Mixed Logit juga dapat mengestimasi parameter korelasi secara akurat. Dari studi simulasi pada respon nominal multivariat juga diperoleh beberapa kesimpulan. Pertama, jika terdapat korelasi antar respon maka pendekatan univariat akan menghasilkan penaksir yang bias. Kedua, model Mixed Logit dapat mengestimasi secara akurat semua parameternya. Ketiga, penaksir yang diperoleh dalam model Probit yang dipengaruhi oleh penduga nilai awal mempunyai bias yang cukup besar.

ABSTRACT

Discrete Choice Model Using Probit Model And Mixed Logit Model On Multivariate Response

by

Jaka Nugraha
(05/1656/PS)

Discrete Choice Models (DCM) describe decision makers' choices among alternatives based on its utility value. Alternatives (responses) are nominal data and one of them having the highest utility value by the decision maker. The problem in DCM is how to find the proper model which can be used to describe a behavior of decision maker (respondent) in making decision. There are two aspects that need attention, those are model that will be used for describing behavior and method used to estimate the parameter. Models that have been discussed are Logit model, Nested Logit model, Generalized Extreme Value (GEV) model, Probit model and Mixed Logit model. The estimation methods utilized are MLE (Maximum Likelihood Estimation), Momen and Bayesian.

For research in the field transportation, marketing and politics, multivariate response analysis is often used. Liang and Zeger (1986) presented Logistic and Probit models on panel data by using univariate approximation. If correlations are ignored, estimators will be inefficient. As well as in multivariate response, if there is correlation between the response, analysis using univariate approximation can influence to the estimators properties. Usually in the case of multivariate, the models can be constructed from combination of all alternative. This approximation makes the Independence from Irrelevant Alternatives (IIA) properties not satisfied.

This research is aimed to construct model of the response of nominal multivariate data in which the respondent must make several decision in the same time. For example π_{ijt} is probability of i^{th} respondent at t^{th} decision choosing alternative j among J_i alternatives for $i=1, \dots, n$; $t=1, \dots, T$ and $j=1, \dots, J_i$. The value of π_{ijt} is influenced by the characteristic of respondent i (X_i) and alternative (Z_{ijt}). It is assumed that decisions are correlated. Models that will be used in this research are Logit, Probit and Mixed Logit. Maximum Simulated Likelihood (MSL) is utilized as an estimation method. MSL method is similar to MLE method but the function of its probability is calculated by simulation method. Asymptotic properties of MLE method is utilized to examine the parameters in the model. Simulation methods which used to calculate the proportion of alternatives are GHK and Halton's sequence.

Main contribution of this research is obtaining method to construct the DCM model on respon nominal multivariate data model, which accomodate the correlation between the responses. Discussion are started from the binary response (dichotomous), that each response has two alternatives. On binary response, methods of estimating parameters are MLE and GEE. MLE in Logit model is arranged by Bahadur's representation. GHK simulation is used to find the MLE in Probit model, and the Halton simulation is used to obtain MLE in Mixed Logit model. Because of Bahadur's



Representation and GEE method can be implemented only in binary response, DCM on nominal multivariate response used Probit and Mixed Logit models. Parameters in Probit and Mixed Logit models are estimated by SMLE.

Simulation study on multivariate binary responses and nominal responses to find correlation effects to estimator properties has been done. Data were generated on a certain parameter value (known) by R.2.8.1 software. Based on simulation data, the obtained estimator for each model was compared to the true parameter. In binary multivariate data, we have several conclusions. First, GEE and MLE in Logit model are relatively same. The estimators are not influenced by the correlation. The estimator of correlation in GEE proposed by Liang-Zeger (1986) tends to be underestimate. Meanwhile, estimator proposed by Chaganti-Joe (2004) in the Probit model is more accurate in estimating correlation parameter. For the second, if there is violation of assumption (specially in the value of variance), the estimator in Logit and Probit model will be biased. Third, Mixed Logit model can reduce the value of a bias estimator. Mixed Logit model can also be used to estimate the correlation parameter accurately. From the simulation in nominal multivariate, several conclusions are obtained. First, if there are correlations among responses, the estimator of univariate approximation will be biased. Second, Mixed Logit model capable to estimate all parameter more accurately. Third, the estimator in Probit model which is influenced by initial value has large bias.