The application of random regret minimization on commuter's mode choice behaviour: Model-Fit comparisons with Rum-Modelling (case: comparison between Matsuyama and Yogyakarta)

Audinda Virsa Leinia^{1,*}, Imam Muthohar¹, and Ibnu Fauzi¹

¹Universitas Gadjah Mada, Civil and Environmental Engineering, Yogyakarta, Indonesia

Abstract. Modelling the mode choice behaviours of travellers is a key to design effective transport management policies, particularly in shifting travellers to public transport. Abundant studies have analysed the impact of level of services on mode choice preferences through its Random Utility Maximization (RUM), but the possibility of minimize the regret have been overlooked. This paper will discusses the possibility of using generalised Random Regret Minimization (G-RRM) model on choosing transportation modes. The study is performed in two cities for comparison: Jogjakarta in Indonesia and Matsuyama in Japan. A stated preference (SP) survey isconducted, in which respondents choose Bike or Bus under hypothetical situations. As the result of RUM revealed that travellers prefer the transportation mode with more ensuring level of service. While an empirical proof of concept, the G-RRM model is estimated on a stated mode choice dataset, and its outcomes are compared with RUM and RRM counterparts.

1 Introduction

Many modern cities are facing traffic problems due to rapid urbanisation. Increasing population induces increasing number of mobility demand, which results in problems in transportation. Particularly in developing Asian cities, the traffic problems have existed for many years. Growing economies have enabled the residents to own private vehicles, such as motorbikes and cars. This change makes the individuals more dependent on the private mode of travel and causes serious traffic jam on a daily basis. The problem of traffic congestion stems to disrupted traffic conditions, such as traffic accidents; and delay in mobility of people and goods, which cause huge economic loss every year.

A public transportation could give an outstanding solution to the above problems by attracting the travellers of private vehicles to mass transportation modes. However, the benefit of the public transportation systems is not easily attained. In general, private

* Corresponding author: audinda.virsa@gmail.com

vehicles are more attractive mode of transportation because of their comfort and speed. To devise an effective public transportation policy, it is essential to understand the relationship between travellers' mode choice decisions and the factors affecting it.

Abundant studies have analysed the impact of level of services, on mode choice preferences using Random Utility Maximization. However, the possibility of minimize the regret have been overlooked. To the best of authors' knowledge, Chorus, et al. [6] has been empirically compared between the application of Random Utility Maximization and Random Regret Minimization on consumer choice modelling. They identified that the RUM and RRM perform equally well on some aspects, but there are possibility of compromising effect on using RRM Although the study presents promising findings, it does not quantitatively evaluate the impact of using RRM while modelling the consumer choices.

2 Overview of Study Site

2.1 Yogyakarta, Indonesia

Yogyakarta is one of the most growing cities in Indonesia, and the mobility demand is dramatically increasing. In order to meet the demand, the government provided a public transportation system by operating Transjogja Bus in 2008. However, since its start, the number of passengers has been decreasing [2]. According to Department of Transportation of Yogyakarta, the designed time headway of Transjogja Bus is 16 minutes, and the targeted punctuality tolerance level is 5 minutes with average speed is 25 km/hour. In reality, a lot of schedules are missed from its tolerance level because of many obstacles in a road lane, which shuts away the passengers.

2.2 Matsuyama, Japan

Matsuyama has a well-developed transport network as the prefecture capital. It has a network of suburban railways run by the Iyo Railway. Iyo Railway also operates a system of trams and bus, which serve as the main modes of public transportation in the city. The public transportation network in Matsuyama is almost the same with Yogyakarta. Nevertheless, the majority of travellers tend to use private vehicles, such as bicycles, motorbikes or private cars.

3 Methodology

3.1 Questionnaire individual attribute

The first section is devoted for asking individual attributes of the respondents. The questions are designed to comprehend their socio-demographic characteristic, daily travel behaviours and past experiences in using public transportation. The list of questions includes the following:

1. Age
2. Gender
3. Occupation
4. Disposable income per month

The questions related to their travel behaviours seek for the daily commuting characteristics and the past experiences as listed below:

1. Daily mode of transportation for their commuting
2. Travel distance
3. Travel time
4. Travel cost
5. Private vehicle ownership (number of their own private vehicles, including car, motorbike and bicycle)
6. Motorbike availability
7. Experience of encountering crime such as thieves or gender harassment in using public transportation (never, once, twice, or more)
8. Experience of encountering or self-involvement in traffic accident in using private vehicles (never, once, twice, or more)

3.2 SP survey

The Stated Preference (SP) survey presents hypothetical travel scenarios where the respondents are required to choose either Bus (as a public transportation) or Motorbike (as a private vehicle) under various conditions. The factors considered in this study are listed below:

1. Travel Cost (TC)
2. Travel Time (TT)
3. Maximum Delay Time (Max. DT)
4. Average Waiting Time at the bus stop (Ave. WT)
5. Access Time to the bus stop (AT)
6. Frequency of traffic accidents while using the motorbike (Acc.)

3.3 Experimental design

Scenarios are created by assigning two levels of values to each of the six attributes. According to Rushing,H., Karl, A,et.al [8]Each of the six attributes has two cases, which creates 64 scenarios ($= 2^6$) in total. In order to ease the answering load for respondents, the scenarios are decreased based on 2^{k-2} fractional factorial design [3], where k represents the number of attributes. The 2^{k-2} fractional factorial design is constructed by listing down a basic design consisting of the scenarios consisting of a full factorial in $k - 2$ attributes. Then the remaining two attributes are added. Thus there will be 16 scenarios.

Table 1. Levels of attributes for Yogyakarta

Attribute	Bus	Motorbike
TC (Rp)	1000	2000
		3000
Acc. (accident/month)	0	1
		2
TT (minutes)	20	15
	25	
Max Delay (minutes)	4	5
	6	
Ave WT (minutes)	3	0
	5	
AT (minutes)	5	0
	8	

Table 2. Levels of attributes for Matsuyama

Attribute	Bus	Motorbike
TC (Yen)	50	100
		150
Acc. (accident/month)	0	1
		2
TT (minutes)	20	15
	25	
Max Delay (minutes)	4	5
	6	
Ave WT (minutes)	2	0
	4	
AT (minutes)	4	0
	6	

3.4 The-RUM model

According to Hensher,D.A [7], logistic regression is part of regression analysis that is used when dependent variable is dichotomous which represents whether an event appears or not. This regression forms a dependent variable as a linear combination of independent variables. The linear function, a utility function, is then transformed to probability using logit function. To predict if an alternative is chosen based on model, the value of utility function should be contrasted by alternative choice, transformed into a probability with score between 0 and 1.

Based on Issac K,et.al [3] cited from Ben-Akiva,et.al [4], the assumption of the model is that respondents would choose the travel mode for the greatest utility under certain condition, and the utility function of it is composed by fixed term and random term. The formulas are as follows:

$$U_{in} = V_{in} + \varepsilon_{in} \quad (1)$$

$$V_{in} = \beta_0 + \sum_{k=1}^K \beta_k X_{ink} \quad (2)$$

where U_{in} is utility function of the $i - th$ travel mode chosen by the $n - th$ respondent ($Bus (i = 1), Motorbike (i = 2)$); V_{in} is the fixed term of utility function (U_{in}); ε_{in} is the random term of utility function (U_{in}); K is the number of attributes (which is also called characteristic variables) affecting on the mode choice of respondents; β_k is the undetermined coefficients; X_{ink} is the $k - th$ factor of the $i - th$ travel mode chosen by the $n - th$ respondent. If the random term ε_{in} follows the Gumbel distribution and all variables are independent each other, the probability for the $n - th$ respondent choosing $i - th$ travel mode, P_{in} , is defined by the following equations:

$$P_{in} = \frac{\exp(V_{in})}{\sum_{i=1}^2 \exp(V_{in})} \quad (3)$$

3.5 The-RRM model

The RRM model has been designed to incorporate the notion of regret-based decision making in non-risky choice models. The RRM model hypothesizes that, when confronted with a choice set, the decision-maker chooses the alternative from the set that has minimum regret. The regret of alternative i is described by the sum of binary regrets where alternative i is compared to every other alternative in the choice task on each attribute. Overall regret is increasing with the number of attributes on which alternative i is outperformed as well as with the number of alternatives by which alternative i is outperformed (as denoted by the summation over $j \neq i$), and the importance of the attribute (as denoted by δm).

$$RRi = Ri + \varepsilon i = \sum_{j \neq i} \sum_{m=1}^M \ln(1 + \exp[\beta m . (X_{jm} - X_{im})]) + \varepsilon i \quad (4)$$

RRi : denotes the random (or: total) regret associated with a considered alternative i

Ri : denotes the ‘observed’ regret associated with i

εi : denotes the ‘unobserved’ regret associated with i

βm : denotes the estimable parameter associated with attribute Xm

Xim, Xjm : denote the values associated with attribute Xm for, respectively, the considered alternative i and another alternative j

3 Result

3.1 Attributes of respondents

The Socio-demographic attributes of the respondents are summarised in Table 3. In both cities, the majority of respondents are male aged 18-24 years old with an occupation as student. The disposable monthly income varies, but ranges in an average student level in both cities.

Table 3 Socio-demographic attributes

Attributes		Yogyakarta	Matsuyama
Gender			
	Female	36.00%	18.00%
	Male	64.00%	82.00%
Age			
	18-24	94.00%	88.00%
	25-49	6.00%	12.00%
Occupation			
	Student	90.00%	100%
	Full-time worker	4.00%	
	Self-employed	4.00%	
	Others	2.00%	
Disposable income			
	(Indonesian Rp)		
	<500,000	20.00%	-
	500,000 - 1,000,000	16.00%	-
	1,000,000 - 1,500,000	28.00%	-
	1,500,000 - 2,000,000	16.00%	-
	>2,000,000	20.00%	-
	(Japanese Yen)		
	<50,000	-	52.00%
	50,000-100,000	-	36.00%
	100,000-150,000	-	10.00%
	>150,000	-	2.00%
	N	400	400

Based on travel behaviour factors, the significance factor that could distinguished between Yogyakarta and Matsuyama respondents as Table 4.

3.2 RUM-Model

In order to create the valid input data, test of significance is conducted to figure out the attributes that affect the preference to the choices. The affecting factors are identified from the statistical test such as Wald-test for each variable. In the Wald-test, the null-hypothesis is $\beta_i = 0$ (Not significantly affecting), while the alternative hypothesis is $\beta_i \neq 0$ (significantly affecting). According to Ben-Akiva, et.al [5], the Wald-test calculation with α

= 0.05 and DF = 20 will has critical t-value as 1.96. The table 5 and 6 will show the significant affecting factors in both cities.

Table 4 Mode choice behaviours

Attributes		Yogyakarta	Matsuyama
Daily mode of travel			
	Car	6.00%	2.00%
	Bicycle	10.00%	88.00%
	Motorbike	84.00%	10.00%
Daily travel distance			
	<3km	46.00%	90.00%
	3-6km	24.00%	10.00%
	6-9km	14.00%	0%
	>9km	16.00%	0%
Daily travel time			
	<5 minutes	14.00%	44.00%
	5-10 minutes	42.00%	52.00%
	10-15 minutes	20.00%	4.00%
	>15 minutes	24.00%	0%
Daily travel cost			
	(Indonesian Rp)		
	0 - 1,000	4.00%	-
	1,000 – 3,000	32.00%	-
	3,000 – 5,000	28.00%	-
	5,000 – 7,000	22.00%	-
	>7,000	14.00%	-
	(Japanese Yen)		
	0 - 50	-	98.00%
	50 – 100	-	2.00%
	>100	-	0%
Private Vehicle Owned			
	0	6.00%	32.00%
	1	86.00%	58.00%
	2	4.00%	8.00%
	3 or more	4.00%	2.00%
Motorbike Availability			
	Yes	88.00%	24.00%
	No	12.00%	76.00%
Crime in public transport			
	Once	10.00%	2.00%
	Twice	2.00%	0%
	Three times or more	2.00%	0%
	Never	86.00%	98.00%
Traffic accident in Private			
	Once	34.00%	34.00%
	Twice	10.00%	0%
	Three times or more	36.00%	0%
	Never	20.00%	66.00%
	N	400	400

Table 5. RUM-Model estimation result
(Yogyakarta)

	Coefficient	Wald	P-value
TC	-0.004	1.961	0.041
TT	-0.272	19.964	0.000
Mode_1	-3.914	30.643	0.000
Dist_2	0.947	7.191	0.007
Dist_3	0.609	2.928	0.008
BK	-2.197	8.695	0.003
ExAcc_2	0.693	5.636	0.018
ExCrime	-0.260	0.559	0.015
Constant	5.228	24.093	0.000

Table 6. RUM-Model estimation result
(Matsuyama)

	Coefficient	Wald	P-value
TC	-0.011	9.328	0.002
TT	-0.061	2.732	0.038
Acc	-0.539	9.328	0.002
Dur_1	0.505	5.03	0.025
OwnVeh_1	-0.471	3.992	0.046
BK	-2.095	55.874	0.000
Constant	2.765	9.257	0.002

Based on tables above, the utility function equation in Jogjakarta using binary logistic regression model is

$$\ln\left(\frac{P_{bus}}{1 - P_{bus}}\right) = -0.004TC - 0.272TT - 3.914Mode_1 + 0.947Dist_2 + 0.609Dist_3 - 2.197BK + 0.693ExAcc_2 - 0.260Crime + 4.484$$

- Where :
- Pbus* : Probability that Bus is chosen
 - TC* : Travel Cost of Bus in a Scenario
 - TT* : Travel Time of Bus in Scenario
 - Mode_1*:Motorbike as respondent daily mode of transportation
 - Dist_2* : Daily travel distance about 3 – 6 km
 - Dist_3* : Travel distance more than >6 km
 - BK* : Bikeavailable(yes), The respondent has a motorbike and/or could use it
 - ExAcc_2*: The experience of traffic accidents
 - ExCrime*: The experience of encountering a crime on public transportation

Based on tables above, then the utility function equation in binary logistic regression model that suits in Matsuyama is:

$$\ln\left(\frac{P_{bus}}{1 - P_{bus}}\right) = -0.011TC - 0.061TT - 0.539Acc + 0.505Dur_1 - 0.471OwnVeh_1 - 2.095BK + 2.765$$

- Where :
- Pbus* : Probability that Bus is chosen
 - TC* :Travel Cost of Bus in a Scenario
 - TT* : Travel Time of Bus in Scenario
 - Acc* : Number of accident on private vehicle in scenario
 - Dur_1* : Travel duration that takes more than 5 minutes
 - OwnVeh_1*: One vehicle ownership
 - BK* : The respondent hasa motorbike and/or could use it

The goodness fit between Yogyakarta and Matsuyama also has different result due to the limitation number of respondent, but both model in Yogyakarta and Matsuyama has been fulfilled the criteria by the Omnibus test that state the significance of the model which is lower than 0.05 as α -value and also HosmerLemeshow value that is bigger than 0.05 which means that the model chosen is fit. Yet the variance in Indonesia can represent 54.7 % to the respondent variance, while in Matsuyama only represent 47.0%. In overall

percentage, both are more than the null hypothesis which is 50%, the overall percentage in Matsuyama is 64.8% and in Yogyakarta is 83.9%. The -2 Log Likelihood shown in Yogyakarta is 686.335 which if compared to Chi-critical ($\alpha = 0.05$ and $DF = 20$) which is 422.695 (Chi-square of Omnibus Test),then the value of -2LL is bigger and stated on rejection area, which means H_0 is rejected and proved that the unrestricted model is better than restricted model. While in Matsuyama the -2LL value which is 1000.201 is also higher than Chi-critical (108.835). Both cities stated the unrestricted model is better than restricted model. Restricted model is a model with only estimating the constant parameter, while the parameter inside scenario is not being used, while unrestricted model in involving socio-demograph factors, travel behaviour factors and all attributes in scenario.

3.3 RRM-Model

The respondents’ characteristics and the SP survey results are analysed for estimating the mode choice model. The RRM-Model will be differs in result with RUM since in RRM estimated attributes will not involving the socio-demograph and travel behaviour attributes, the proceeded attributes are the SP scenario only. Since the RRM model is not fit with the dummy variables in socio-demograph and travel behaviour attributes. The result in RRM-Model using Biogeme will consist of coefficient parameter, statistical t-test and statistical p-value as mentioned below:

Table 7. RRM-Model estimation result
(Yogyakarta)

Name	Value	Robust t-test	p- value
TC (X1)	-0.005	0.119	0.039
TT (X2)	-0.372	29.778	0
Mode_1 (X3)	-3.414	30.654	0
Dist_2 (X4)	0.931	7.281	0.008
Dist_3 (X5)	0.591	2.998	0.009
BK (X6)	-2.375	8.594	0.003
ExAcc_2 (X7)	0.658	5.364	0.019
ExCrime (X8)	-0.315	0.578	0.014
gamma	4.918	24.135	0
-Init log likelihood:		-277.259	
-Final log likelihood:		-678.457	
Rho-square		0.357	

Table 8. RRM-Model estimation result
(Matsuyama)

Name	Value	t-test	p- value
TC (X1)	-0.015	9.192	0.003
TT (X2)	-0.049	2.894	0.038
Acc (X3)	-0.521	9.772	0.002
Dur_1 (X4)	0.499	5.031	0.023
OwnVeh_1 (X5)	-0.456	3.679	0.041
BK (X6)	-2.101	55.473	0
gamma	2.923	9.599	0.001
-Init log likelihood:		-997.108	
-Final log likelihood:		-1001.121	
Rho-square		0.441	

Based on tables above, the utility function equation in Jogjakarta using binary logistic regression model is:

$$\text{Ln}(1 + \exp(\text{Pbus})) = -0.005\text{TC} - 0.372\text{TT}$$
$$- 3.414\text{Mode1} + 0.931\text{Dist2} + 0.591\text{Dist3} - 2.375\text{BK} + 0.658\text{ExCrime} -$$
$$0.315\text{ExAcc2} + 4.918$$

Where :

P_{bus} : Probability that Bus is chosen

TC : Travel Cost of Bus in a Scenario

TT : Travel Time of Bus in Scenario

Mode_1:Motorbike as respondent daily mode of transportation

Dist_2 : Daily travel distance about 3 – 6 km

Dist_3 : Travel distance more than >6 km

BK : Bikeavailable(yes), The respondent has a motorbike and/or could use it

ExAcc_2: The experience of traffic accidents

ExCrime: The experience of encountering a crime on public transportation

Based on tables above, then the utility function equation in binary logistic regression model that suits in Matsuyama is:

Ln (1+exp(P_{bus}))= -0.015TC -0.049TT
-0.521Acc+0.499Dur1
-0.456OwnVeh1-2.101BK +2.923

Where :

P_{bus} : Probability that Bus is chosen

TC :Travel Cost of Bus in a Scenario

TT : Travel Time of Bus in Scenario

Acc : Number of accident on private vehicle in scenario

Dur_1 : Travel duration that takes more than 5 minutes

OwnVeh_1: One vehicle ownership

BK: The respondent has a motorbike and/or could use it

3.4 Comparison RUM-Model and RRM model

There are slight differences between 0-LL in Yogyakarta and Matsuyama, but the value of Rho-square are quite different value in both cities. The significant factors both in Yogyakarta and Matsuyama are the same. In terms of model fit, RUM-Model generally fit the data significantly. In context of SP survey, the “opt out” or “none of these seems to performs better in RUM. The model also has an ability to provide correct choice probability predictions for no-choice option.

While the RRM-Model, Generally fit the data significantly. In context of SP survey, the “i am indifferent” seems to performs better in RRM models. The model has no ability to provide correct choice probability predictions for no-choice option, but it could capture compromising-effect, it recognizes alternatives with mediocre performance.

Table 9. Differences goodness of fit (Yogyakarta)

Parameters	RUM	RRM
Final LL	686.335 ^a	678.457
Rho-square	0.547	0.357

Table 10. Differences goodness of fit (Matsuyama)

Parameters	RUM	RRM
Final LL	1000.201 ^a	1001.121
Rho-square	0.470	0.441

Based on the model-fit test as the comparison of choices in actual site and choices in the observation, the proportion could be seen in Table 11 for Yogyakarta model and Table 12 for Matsuyama model.

Table 11. Site comparison in Yogyakarta

OBSERVATI ON		ACTUAL	
		Mode 1	Mode 2
	Mode 1	82.63%	17.37%
	Mode 2	27.12%	72.88%

Table 12. Site comparison in Matsuyama

OBSERVATI ON		ACTUAL	
		Mode 1	Mode 2
	Mode 1	65.87%	34.13%
	Mode 2	42.06%	57.94%

The difference in goodness of fit as in Table 11 and 12, as expected, is in favour of the RUM-model, which is a sign that RRM indeed has some difficulty with handling imputed choice sets. As in the model fit of site comparison in Yogyakarta, the approaching value for bus is 82.63% while in Matsuyama is 65.87%.

4 Conclusion

In transportation system, every aspects should be considered to affect the preference. In this model there are 12 variables, with 4 socio-demograph factors (age, gender, occupation, income per month) and 8 travel behaviour factors (daily mode of transportation, travel distance, travel duration, travel cost, vehicle ownership, availability of motorbike, self-experience on criminality and accident).

In terms of RUM-Model and RRM-Model that involves dummy variables using SPSS, factors that is significantly affecting the preference in Yogyakarta are are Travel Cost, Travel Time, Daily Mode of Transportation, Travel Distance, Availability of using Motorbike, Self-experience of Crimininality on public transportation and Self-experience of Accident on Private Vehicle. While in Matsuya the significant factors are Travel Cost, Travel Time, Safety while using the mode of transportation, Travel duration to campus, Private Vehicle Ownership, and Availability of using Motorbike.

The comparison analysis between RUM and RRM using Biogeme have slight difference, in terms of model-fit parameter and statistics test, and also the value of goodness of fit due to RUM ability to predict the probability of “no choice” option. The differences might getting large if the SP survey attributes level are vary and also the number of respondent is larger. The further statistical check like, t-test comparison and p-value is needed whether RRM and RUM model are fit to be used in terms of Mode Transportation Choices.

References

1. Adamowicz, W., Louviere, J., & Williams, M. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*, 26, 271-292. (1994)

2. Annual Report Jasa Marga. Access October 2016, http://www.jasamarga.com/download/AR_jm2016.pdf. (2011)
3. Baidoo, Issac K, & Nyarko, Eric. *Stated Preference Modeling for a Preferred Transportation Mode*. Ghana: iiste.org. (2015)
4. Ben-Akiva, M., and S. Lerman. *Discrete Choice Analysis*. London: MIT Press, (1985).
5. Ben-Akiva, M. & Morikawa, T. Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General*, 24(6), 485-495. (1990)
6. Cranenburgh, S., Guevara, C., & Chorus, C. New insights on random regret minimization models. *Transportation Research Part A: Policy and Practice*, 74, 91-109. doi: 10.1016/j.tra.2015.01.008. (2015)
7. Hensher, D.A., Rose, J.M., & Greene, W.H. *Applied Choice Analysis: A Primer* (Vol.1). Cambridge: Cambridge University Press. (2005)
8. Rushing, H., Karl, A., & Wisnowski, J. *Design and Analysis of Experiments by Douglas Montgomery: A Supplement for Using JMP*. North Carolina, USA: SAS Institute Inc., (2013)
9. Susilo, Y.O., Joewono, T.B. & Santosa, W. *An Exploration of Public Transport Users' Attitudes and Preferences towards Various Policies in Indonesia: Some Preliminary Results*. *Journal of the Eastern Asia Society for Transportation Studies*, 8. (2010)
10. Train, K. *Discrete choice methods with simulation*. United States: Cambridge university press. (2009)