

Mode choice analysis using discrete choice model from transport user (Case study: Jakarta LRT, Indonesia)

Dwi Novi Wulansari^{1,*}, and Milla Dwi Astari²

¹17 Agustus 1945 Jakarta University, Civil Engineering Department, Jakarta, Indonesia

²Yapis Papua University, Civil Engineering Department, Jayapura, Indonesia

Abstract. Jakarta Light Rail Transit (Jakarta LRT) has been planned to be built as one of mass rail-based public transportation system in DKI Jakarta. The objective of this paper is to obtain a mode choice models that can explain the probability of choosing Jakarta LRT, and to estimate the sensitivity of mode choice if the attribute changes. Analysis of the research conducted by using discrete choice models approach to the behavior of individuals. Choice modes were observed between 1) Jakarta LRT and TransJakarta Bus, 2) Jakarta LRT and KRL-Commuter Jabodetabek. Mode choice model used is the Binomial Logit Model. The research data obtained through Stated Preference (SP) techniques. The model using the attribute influences such as tariff, travel time, headway and walking time. The models obtained are reliable and validated. Based on the results of the analysis shows that the most sensitive attributes affect the mode choice model is the tariff.

1 Introduction

One of the drivers of economic activity is the availability of good transportation services. However, the current use of public transportation in Indonesia is still not able to facilitate the movement that occurs efficiently. Improper vehicles, accident risks, long waiting times, over capacity, crime and pollution are some of the most common issues facing most of the users of public transport, especially during peak hours. To overcome these problems, the active role of some parties involved in public transportation is required. In the Law of the Republic of Indonesia number 22 [1] regarding traffic and road transport, it is mentioned that the government is responsible for public transport in order to meet the needs of safe, safe, convenient and affordable transportation.

Through the Jakarta Governor's decree number 84 year 2004 regarding determination of macro transportation pattern in DKI Jakarta, stipulating the direction of development of transportation system, one of them is to add primary road network, bus priority, light rail transit and mass rapid transit (MRT). Currently, the government of DKI Jakarta is developing a transportation system, one of which is to build Jakarta Light Rail Transit (Jakarta LRT). Jakarta LRT as a transport system using a light rail is planned to connect

* Corresponding author: dwi.novi@uta45jakarta.ac.id

Jakarta with the surrounding cities such as Bekasi, Depok and Bogor. The development of Jakarta LRT will be integrated with other modes, such as TransJakarta Bus, KRL-Commuter Jabodetabek and MRT.

The objective of this study is to obtain a mode choice model that can explain the probability of choosing Jakarta LRT based on travel attributes, ie tariff, travel time, headway and walking time. In addition, this study also aims to determine the sensitivity of mode choice when changes to travel attributes are made. Changes in travel attributes will affect the probability of choosing between Jakarta LRT and Bus Transjakarta or KRL-Commuter Jabodetabek. Thus the attribute value of Jakarta LRT can be determined in such a way as to avoid overloading on both public transport.

2 Literature

2.1 Discrete choice models

According to Tamin [2], generally, discrete choice model expressed as probability of each individual to choose an option that is a function of socio-economic characteristics and the attractiveness of such options. To declare the appeal of an alternative, used the concept of utility (defined as something that is maximized by each individual). Domencich and McFadden, and Williams, as quoted from Tamin [2], suggests that each set of utility options U_{in} for each individual n . Modelers who is also an observer of the system does not have complete information about all of the elements considered by individuals who make choices. So, in making the model assumed that U_{in} can be expressed in two components, namely:

- V_{in} is measured as a function of the measured attributes.
- Random choice ϵ_{in} , reflecting the particulars of each individual, including mistakes made by modelers.

In general, these effects can be expressed:

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

where:

- U_{in} = utility of alternative i for decision makers n
- V_{in} = deterministic utility function of modes i for the individual n
- ϵ_{in} = random error (random error) or stochastic component and a particular distribution function

2.2 Binominal logit model

Decision-making on the binomial logit model is determined on a pair of discrete alternatives, which will be selected alternative is to have the largest utility, the utility in this case is seen as random utility. With two alternate modes of the equation can be written as follows.

$$P_2 = \frac{\exp^{U_2}}{\exp^{U_2} + \exp^{U_1}} \quad (2)$$

$$P_1 = 1 - P_2 \quad (3)$$

where:

P_1 is the probability for mode 1

P2 is the probability for mode 2

By using linear regression method, the probability that the individual chose mode 2 is a utility function of the difference between the two modes. Assuming a linear utility function, then the difference can be expressed in the form of utility difference in the number of n relevant attributes between the two modes, formulated as follows:

$$U_2 - U_1 = \beta_0 + \beta_1(X_{12} - X_{11}) + \beta_2(X_{22} - X_{21}) + \dots + \beta_n(X_{n2} - X_{n1}) \quad (4)$$

Where $U_2 - U_1$ is the individual response to the statement of choice, β_0 is a constant, β_1 , β_2 and β_n are the coefficients of each attribute are determined by multiple linear regression. Thus, the probability value which is reviewed both modes can be written in the form of the following equation.

$$P_2 = \frac{\exp^{U_2}}{\exp^{U_2} + \exp^{U_1}} = \frac{\exp^{(U_2 - U_1)}}{1 + \exp^{(U_2 - U_1)}} \quad (5)$$

$$P_1 = 1 - P_2 = \frac{1}{1 + \exp^{(U_2 - U_1)}} \quad (6)$$

Utility value as the individual response can also be expressed in terms of the probability of selecting a particular mode, known as transformation Berkson-Theil, the equation is as follows.

$$\ln \left[\frac{P_2}{1 - P_2} \right] = \beta_0 + \beta_1 \cdot (X_{12} - X_{11}) + \beta_2 \cdot (X_{22} - X_{21}) + \dots + \beta_n \cdot (X_{n2} - X_{n1}) \quad (7)$$

2.3 Survey techniques

According to Nobuhiro Sanko [3], Revealed Preference (RP) data is observation data / questions based on what people actually do. In these data, because the behavior of individuals who are actually known, which is usually assumed that the information is reliable and can be obtained from retrospective questionnaire. While Stated Preference (SP) data is data base on reaction given by the respondent is not the actual behavior, but only a statement of preference. SP technique is characterized by the use of experimental design to build an alternative hypothesis to the situation, which is then presented to the respondents. Several factors need to be considered in the design of the SP experiment is as follows:

1. Response Form (Ranking/ Rating/ Choice/ Degree of Preference)
2. Analytical Method
 Available analytical method is related to response form. Pearmain et al. [4] introduces four types of analytical methods, i.e.,
 - Naive or graphical methods
 - Non-metric scaling
 - Regression
 - Logit and probit models
3. Number of samples
 Data collection needs huge cost. After the analytical method has been determined, we need to decide the necessary number of samples.
4. Attributes (Measurement)
 How many levels should be treated and how to set attributes (absolute value, percentage and so on) should be considered.
5. Attribute Levels

This factor considers how many levels should be treated and how to set attributes (absolute value, percentage, etc.). Attribute levels in experimental design usually is 'orthogonal', namely to ensure that the attributes presented to respondents varied independently of one

another. The result is that the effect of each attribute on the response level more easily isolated. This is to avoid a 'multi-colinearity' between attributes, which is a common problem with the data RP.

3 Research methodology

The analytical approach in this research is using discrete choice analysis. The discrete choice function is solved by binomial logit model, where the research data is obtained by using the survey technique Stated Preference (SP) so that the respondent preference of the mode model attribute is obtained. Juan de Dios Oruzar and Luis G. Willumsen 2011, stated that in determining the amount of sample stated preference can be done 75-100 respondents for each segment can already give the right results. In this study used a sample of 277 respondents for both models of mode choice. Research questionnaires were distributed around the location of the Jakarta LRT station plan. Choice modes were observed between 1) Jakarta LRT and TransJakarta Bus, 2) Jakarta LRT and KRL-Commuter Jabodetabek. Assuming that utility functions are linear, the utility differences can be expressed in terms of differences in the number of n relevant attributes between the two modes, formulated as follows:

$$U_{LRT} - U_{Mode} = \beta_0 + \beta_1 \cdot (X_{1LRT} - X_{1Mode}) + \beta_2 \cdot (X_{2LRT} - X_{2Mode}) + \dots + \beta_n \cdot (X_{nLRT} - X_{nMode}) \quad (8)$$

Where $U_{LRT} - U_{Mode}$ is the individual response to the mode choice statement, β_0 is the constant, β_1 , β_2 and β_n are the coefficients of each attribute determined through multiple linear regression. Thus, the probability values of both modes are reviewed as follows.

$$P_{LRT} = \frac{\exp(U_{LRT} - U_{Mode})}{1 + \exp(U_{LRT} - U_{Mode})} \quad (9)$$

$$P_{Mode} = 1 - P_{LRT} = \frac{1}{1 + \exp(U_{LRT} - U_{Mode})} \quad (10)$$

Where P_{LRT} is the probability for Jakarta LRT and P_{Mode} is the probability for TransJakarta Bus or KRL-Commuter Jabodetabek. The utility value is an individual response expressed in terms of the probability of choice of the two modes offered. Choice responses are analyzed using point ratings presented on the probability scale. The probability scale is then transformed into a numerical scale, called the Berkson-Theil transform, the equation is as follows.

$$\ln \frac{P_{LRT}}{1 - P_{LRT}} = \beta_0 + \beta_1 \cdot (X_{1LRT} - X_{1Mode}) + \beta_2 \cdot (X_{2LRT} - X_{2Mode}) + \dots + \beta_n \cdot (X_{nLRT} - X_{nMode}) \quad (11)$$

The numerical scale value is the dependent variable on the regression analysis whereas the independent variable is the difference of the attribute value between the two modes (tariff, travel time, headway and walking time). After the transformation, then the process of estimating the equation of utility function is done using least square regression with the assumption that the equation of utility difference function is linear. Validation of the model is done to determine the level of confidence obtained model. Furthermore, model sensitivity analysis was conducted to find out the attributes that most influence the probability of choosing Jakarta LRT.

Table 1. Transformation value

Point Rating	Transformation value	
	Probability Scale	Numerical Scale
Definitely choose Jakarta LRT	0,9	2,197
Maybe choose Jakarta LRT	0,7	0,847
Point Rating	Transformation value	
	Probability Scale	Probability Scale
Do not know	0,5	0,000
Maybe choose existing mode	0,3	-0,847
Definitely choose existing mode	0,1	-2,197

4 Analysis

Based on 300 questionnaires that have been distributed, obtained 277 eligible answers and the remaining 23 answers are not eligible. Furthermore the data is summarized and displayed in the following table.

Table 2. Data characteristics of respondents

Number	Data		Percentage (%)	
			Bus	KRL
1	Gender			
	a.	Male	43,33	55,12
	b.	Famale	56,67	44,88
2	Age			
	a.	< 17 years	3,33	1,57
	b.	17 - 20 years	22,67	27,56
	c.	21 - 30 years	30,67	35,43
	d.	31 - 40 years	30,00	18,90
	e.	41 - 50 years	6,00	8,66
	f.	> 50 years	7,33	7,87
3.	Profession			
	a.	Student	31,33	41,73
	b.	Government Employees	9,33	7,09
	c.	Employees of BUMN / Private Sector	29,33	36,22
	d.	Entrepreneur	19,33	8,66
	e.	Indonesian State Army (TNI/POLRI)	1,33	0,00
	f.	Pensionary	0,67	1,57
	g.	Housewife	8,67	4,72
4.	Income Level			
	a.	≤ Rp 1.000.000	28,00	32,28
	b.	Rp 1.000.001 - 2.000.000	20,00	17,32
	c.	Rp 2.000.001 - 3.000.000	21,33	14,96

Number	Data		Percentage (%)	
			Bus	Bus
	d.	Rp 3.000.001 - 4.000.000	18,00	18,11
	e.	Rp 4.000.001 - 5.000.000	5,33	7,09
	f.	Rp 5.000.001 - 6.000.000	4,00	5,51
	g.	Rp 6.000.001 - 7.000.000	2,00	1,57
	h.	Rp 7.000.001 - 8.000.000	0,67	1,57
	i.	> Rp 8.000.000	0,67	1,57
5.	Transportation Cost per Month			
	a.	≤ Rp 50.000	8,00	7,09
	b.	Rp 50.001 - 100.000	10,00	6,30
	c.	Rp 100.001 - 150.000	28,00	15,75
	d.	Rp 150.001 - 200.000	14,67	12,60
	e.	Rp 200.001 - 250.000	10,67	11,81
	f.	Rp 250.001 - 300.000	8,00	15,75
	g.	Rp 300.001 - 350.000	6,00	9,45
	h.	Rp 350.001 - 400.000	4,67	5,51
	i.	> Rp 400.000	10,00	15,75

Table 3. Travel characteristics of respondents

Number	Data		Percentage (%)	
			Bus	Bus
1.	Income Level		54,15	45,85
2.	Travel frequency			
	a.	1 - 14 times per month	34,67	28,35
	b.	15 - 28 times per month	5,33	7,09
	c.	29 - 42 times per month	40,00	40,94
	d.	43 - 56 times per month	6,00	7,09
	e.	> 56 times per month	14,00	16,54
3.	Travel purpose			
	a.	Work / Trade	48,67	45,67
	b.	School / College	30,67	41,73
	c.	Daily shopping	4,67	1,57
	d.	Recreation / Treatment	16,00	11,02
4.	Travel time			
	a.	1 - 15 minute	8,67	7,09
	b.	16 - 30 minute	24,00	17,32
	c.	31 - 45 minute	11,33	9,45
	d.	46 - 60 minute	18,67	18,90
	e.	61 - 75 minute	2,00	0,79
	f.	76 - 90 minute	14,00	17,32
	g.	> 90 minute	21,33	29,13
5.	Origin of travel area			
	a.	Central Jakarta	10,67	10,24
	b.	South Jakarta	15,33	3,94
	c.	East Jakarta	8,00	3,15
	d.	West Jakarta	15,33	11,02
	e.	North Jakarta	7,33	5,51
Number	Data		Percentage (%)	
			Bus	Bus
	f.	Others	43,33	66,14

4.1 Logit-binominal model analysis

With linear regression-appraisal using the logit-binomial model, the results of the utility function are shown in the following table:

Table 4. Estimation of utility function (LRT vs Bus)

Number	Model Variables	LRT vs Bus			
		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-Value</i>
1	Constants	0,967	0,176	5,478	5,E-08
2	Tariff (X ₁)	-3,E-04	1,E-05	-22,999	6,E-99
3	Travel time (X ₂)	-0,141	0,040	-3,480	0,001
4	Headway (X ₃)	-0,009	0,007	-1,358	0,175
5	Walking time (X ₄)	-0,018	0,004	-4,348	1,E-05
	F	140,460			
	R ²	0,295			

From the estimation results, the following information is obtained.

1. Considering the F statistic value in the table, it can be seen that the F value is greater than the F-critical (F-critical value = 2,45, where the level of significance (α) = 0,05 = 5% and the observation number > 120), Can be concluded in the mode choice model, there is a relation between the attribute (X) and the choice response (Y).
2. The equation of the choice model has a value of R² of 0.295, which means that all attributes affect the utility change in the model is 29.5%.
3. In the table note that the coefficient of equation of utility function for tariff, travel time, headway and walking time is negative (-), it can be concluded that the bigger difference in attribute value of both modes will minimize the choice of Jakarta LRT.
4. Considering the value of P-value in the table, it can be seen that the attribute of tariff, travel time and walking time has a lower P-value value than the interest rate of 0.05 which means the attribute is acceptable in the utility equation. The headway attribute has a P-value greater than its importance level (> 0.05), so the headway attribute will be removed and the equation will be estimated in stepwise regression (backward elimination).

Table 5. Estimation of utility function (LRT vs KRL)

Number	Variabel	LRT vs KRL			
		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-Value</i>
1	Constants	1,1972	0,2203	5,4344	7,E-08
2	Tariff (X ₁)	-0,0003	1,E-05	-29,4022	8,E-142
3	Travel time (X ₂)	-0,2690	0,0828	-3,2499	0,0012
4	Headway (X ₃)	-0,0086	0,0069	-1,2422	0,2144
5	Walking time (X ₄)	-0,0184	0,0041	-4,4420	1,E-05
	F	224,081			
	R ²	0,441			

From the estimation results, the following information is obtained.

1. Considering the F statistic value in the table, it can be seen that the F value is greater than the F-critical (F-critical value = 2,45, where the level of significance (α) = 0,05 = 5% and the observation number > 120), Can be concluded in the mode choice model, there is a relation between the attribute (X) and the choice response (Y).
2. The equation of the choice model has a value of R² of 0.441, which means that all attributes affect the utility change in the model is 44,1%.

3. In the table note that the coefficient of equation of utility function for tariff, travel time, headway and walking time is negative (-), it can be concluded that the bigger difference in attribute value of both modes will minimize the choice of Jakarta LRT.
4. Considering the value of P-value in the table, it can be seen that the attribute of tariff, travel time and walking time has a lower P-value value than the interest rate of 0.05 which means the attribute is acceptable in the utility equation. The headway attribute has a P-value greater than its importance level (> 0.05), so the headway attribute will be removed and the equation will be estimated in stepwise regression (backward elimination).

Using binomial logit model, the probability of choice of Jakarta LRT is known by the following equation.

$$P_{LRT} = \frac{\exp(U_{LRT} - U_{Mode})}{1 + \exp(U_{LRT} - U_{Mode})} \quad (12)$$

And the probability of choice of existing mode is,

$$P_{Mode} = 1 - P_{LRT} \frac{1}{1 + \exp(U_{LRT} - U_{Mode})} \quad (13)$$

With the utility function being used as follows.

- For utility difference function between Jakarta LRT and Bus
 $U_{LRT} - U_{Bus} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$
 $U_{LRT} - U_{Bus} = 1,031 + (-0,0003 X_1) + (-0,141 X_2) + (-0,018 X_4)$
- For utility difference function between Jakarta LRT and KRL
 $U_{LRT} - U_{KRL} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$
 $U_{LRT} - U_{KRL} = 1,257 + (-0,0003 X_1) + (-0,269 X_2) + (-0,018 X_4)$

4.2 Model sensitivity analysis

To know the sensitivity of the model, it is necessary to illustrate the model sensitivity graph of the attribute change. From the graph, it can be seen the change of probability of choice of Jakarta LRT if gradual attribute value change is done.

Based on the sensitivity graph for both model mode model, it can be seen that the slope of the line shows the negative direction, stating that the greater difference between attribute value between Jakarta LRT and the existing mode will minimize the probability of choosing Jakarta LRT. The graph also has a steep slope, this means that if there is a change in the value of the attribute in the selection model it will cause a change in the probability of mode selection.

4.3 Model elasticity analysis

From model elasticity analysis result, it can be seen that tariff attribute has the highest elasticity coefficient value compared to other attributes, it shows that tariff is the most sensitive attribute in influencing mode choice.

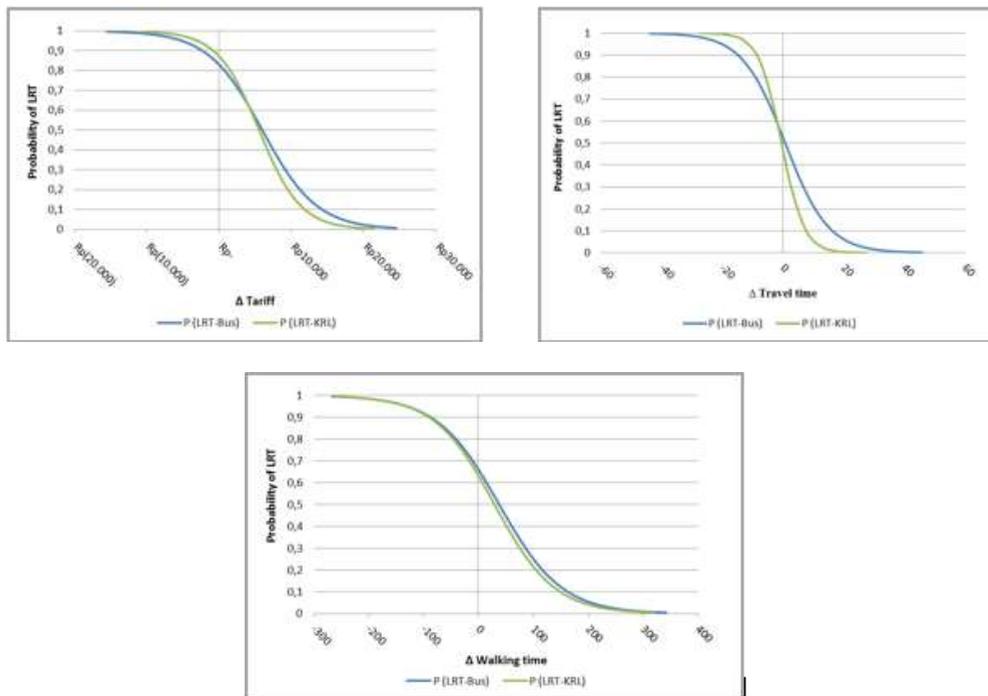


Figure 1. Attribute sensitivity graph

Table 6. Direct elasticity value

Number	Attribute	Coefficient of elasticity	
		Bus	KRL
1	Tariff (X_1)	-0,316	-0,512
2	Travel time (X_2)	0,191	0,248
3	Headway (X_3)	0	0
4	Walking time (X_4)	0	0

5 Conclusions

Based on the result of modeling study that influenced the Jakarta LRT mode choice in DKI Jakarta, the following conclusions are obtained:

1. Based on the individual response to the choice of travel scenarios, it can be seen that the choice attributes that are the consideration of respondents are tariff, travel time and walking time.
2. Based on the attribute sensitivity graph can be seen the slope of the line indicates the negative direction, stating that the greater value of attribute difference will reduce the probability of choosing Jakarta LRT.
3. Based on the results of elasticity analysis, it is known that the most sensitive attributes affect the probability of choosing Jakarta LRT is the tariff.
4. To obtain a probability value of 0.5 in the mode choice model, the difference in tariff attribute between Jakarta LRT and the existing mode (TransJakarta Bus or KRL-Commuter Jabodetabek) is Rp. 5,500 - 6,000.

Acknowledgment

Acknowledgments are addressed to the Directorate of Research and Community Service, Directorate General for Research and Development, Ministry of Research, Technology and Higher Education as a funding body.

References

1. *Laws Of The Republic Indonesia No. 22 Of On Road Traffic And Transportation.* (2009)
2. O.Z. Tamin, *Perencanaan, Pemodelan, & Rekayasa Transportasi: Teori, Contoh Soal, dan Aplikasi.* Penerbit ITB, page 388-433.(2009)
3. N. Sanko, *Guidelines for Stated Preference Experiment Design.* A dissertation submitted for the degree of Master of Business Administration, School of International Management. (2001)
4. D. Permain, and J. Swanson, *Stated Preference Techniques : A Guide to Practice, Steer Davies Gleave and Haque Consulting Group,* London, page 2-59. (1991)
5. D.N. Wulansari, *Analisis Pemilihan Moda Angkutan Penumpang Menuju Bandara (Studi Kasus : Bandar Udara Internasional Soekarno – Hatta),* Journal of Kajian Teknik Sipil, page. 90-100. E-ISSN 2502-8456.(2016)
6. S. Sihite, and M.S. Surbakti, *Kajian Pemilihan Moda Transportasi Antara Angkutan Kota Dengan Monorel Menggunakan Metode Stated Preference (Studi Kasus: Rencana Pembangunan Monorel Kota Medan).* In Proceedings of the 18th FSTPT International Symposium (Bandar Lampung, August 28th, 2015). (2015)