Model selection of online motorcycle taxi and motorcycle modes on work trips

Risdiyanto1,4 *, A Munawar2, M Z Irawan3 and A A Nugraha4
1 Department of Civil Engineering, Universitas Islam Indonesia, Indonesia
2,3 Civil and Environmental Engineering, Universitas Gadjah Mada, Indonesia
4 Department of Civil Engineering, Universitas Janabadra, Indonesia

* risdiyanto@janabadra.ac.id

Abstract. Advances in information and communication technology are driving the rapid development of online transportation. The growth rate of online transportation has an impact on intermodal competition, especially in urban areas. One of the most-used online transportation applications in Indonesia is the online motorcycle taxi. This type of raid-hailing transportation offers flexibility, as it can filter through traffic, and affordable costs, making it a popular mode of transportation for the wider community. It is interesting to find out variables affecting people’s choice of online motorcycle taxi as compared to private vehicles of the same type, namely motorcycle. This study aims to build a mode selection model to compare between motorcycle - as the most widely used mode - and online motorcycle taxi. The data were collected through a home-interview survey of 150 workers in Yogyakarta Indonesia who normally travel to work using motorcycle or online motorcycle taxis on their daily basis. Data analysis were conducted using binary logistic regression in SPSS version 24 and included sociodemographic variables and travel characteristics. The analysis showed that gender, age, number of vehicle ownership, and travel time affected people’s choice on online motorcycle taxi or motorcycle.

1. Introduction

There has been a rapid development of online transportation in many countries worldwide. The growth of Uber online transportation, for example, was recorded to make 1,907, up 20% number of trips in 2019 compared to that in 2018 with 1,493 trips [1]. In Indonesia, the ride hailing giant Gojek recorded twice the revenue growth in 2019 compared to the previous year [2]. The presence of an online motorcycle taxi offers a satisfactory ride-hailing service with the range of 3.85 out of 5 scale [3]. The proliferation of online transportation in Indonesia with motorbike types, or commonly known as online motorcycle taxis (ojek), is inseparable from government policies that allows the use of online motorcycle taxis. In fact, the Ministry of Transportation has issued Regulation of The Minister of Transportation of The Republic of Indonesia Number PM 12 of 2019 concerning The Safety Protection of Riders of Motorcycle in Public Use [4].

Despite the currently proliferating use of online motorcycle taxi, there are only a few number of studies to address the factors to contribute to online motorcycle taxi operations in few countries. Some studies on online motorcycle taxis focused on the competition between conventional motorcycle taxis, online motorcycle taxis, and trans Jakarta buses, which highlighted that online motorcycle taxis were not competitors but were complement to the Trans Jakarta buses [5]. However, other opinions articulated that online transportation was strongly suspected to have a negative impact on other form of transportation since it reduced the conventional transportation market [6].

Although online motorcycle taxis are growing rapidly, the use of private motorbikes by the public in many cities in Indonesia continues to dominate. In Yogyakarta for example, the number of motorcycles in 2017 reached 222,195 units, while four-wheeled motor vehicles amounted to 12,746 [7]. With the development of online motorcycle taxis, comparing the mode choice transportation between motorbike and online motorcycle taxi is increasingly important as a way to gain a foothold in policy making in the field of transportation. The mode choice model links travel agents with the characteristics of available modes (such as time and cost of travel to each city, destination, and level of services), characteristics of the trip, and characteristics of transportation users (such as age, sex, and...
This study will develop a special model for work travel by focusing on the work trip for two reasons. First, workers normally earn regular income allowing them to freely choose their preferred modes of transportation available. Second, work travel is generally done on a regular basis. These two reasons, thus, signify the importance of online mode of transportation for the selected research population.

2. Literature review
The mode selection model that is currently being developed is in the form of a discrete choice model. This model analyzes the choice of consumers (travel agents) from a set of alternative modes of choice that compete with one another and cannot be chosen (used) together by more than one mode (mutually exclusive). The critical choice model was widely chosen because this model experienced success especially for the measurement and testing of hypotheses regarding demand, elasticity, and non-market valuations. More clearly it can be said that the discrete selection model is used to test "which" choices, whereas to test "how much" a regression model is used. Among the various types of discrete choice models, one of which is often used is the binary or binomial logit model. The binomial logit model is used when there are only two alternative choices.

In linear regression, the term of least squares was used to estimate model parameters, while the maximum likelihood (ML) estimation principle was used for logistic regression. The ML principle insinuates that population parameters are estimated by maximizing the likelihood of observation data. The estimator obtained from this method is known as the Maximum Likelihood Estimator (MLE). Likelihood is a function of data and model parameters. In the presence of binary data, the form of likelihood is 

\[ \eta_i = \ln \left( \frac{P_i}{1 - P_i} \right) = \eta_i \]  

if \( \eta_i \) is written in the form of a regression equation it becomes

\[ \eta_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \]  

where \( X_1, X_2, \ldots, X_k \) is a sociodemographic variable and travel characteristics.

From equations (1) and (2), the probability value is

\[ P_i = \frac{e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}}} \]  

3. Research methodology
This study examined the sociodemographic variables, travel time, and distance traveled by users of online mode of transportation in Padukuhan, Sendanga Adi Mlati, Sleman, Yogyakarta, Indonesia, with a number of population amounting to 1,078 people. The survey was conducted by way of direct interview of 150 respondents in March 2020 (before the spread of covid 19 in Yogyakarta) using a randomized home interview survey method. More than one respondents could be selected from each house provided that the respondents conduct work travel on their daily basis. Respondents were required to answer some questions related to sociodemographic conditions, travel characteristics and daily modes of transportation for their work travel. Routine indicators in choosing modes were welcome to the respondents’ perceptions without mentioning the frequency of use. After the survey,
the next step is to compile the data. Next, using SPPS software version 24, an analysis is performed to obtain the mode selection model.

4. Results and discussion

4.1. Sociodemographic conditions and travel characteristics

The survey reveals that 105 (70%) respondents used motorbikes and 45 (30%) used online motorcycle taxis. The sociodemographic conditions of the respondents and the characteristics of the trip are presented in the following table.

Table 1. Sociodemographic characteristics of respondents

| Variable | Sociodemography and travel characteristics | Code | Number of respondents (n) | % |
|----------|-------------------------------------------|------|---------------------------|---|
| X1       | Gender                                    |      |                           |   |
|          | Male                                      | 0    | 89                        | 59|
|          | Female                                    | 1    | 61                        | 41|
| X2       | Age (years)                               |      |                           |   |
|          | <= 20                                     | 0    | 5                         | 3 |
|          | 21 - 30                                   | 1    | 76                        | 51|
|          | 31 – 40                                   | 2    | 51                        | 34|
|          | 41 – 50                                   | 3    | 15                        | 10|
|          | > 50                                      | 4    | 3                         | 2 |
| X3       | Education                                 |      |                           |   |
|          | SD                                        | 0    | 3                         | 2 |
|          | SMP                                       | 1    | 9                         | 6 |
|          | SMA                                       | 2    | 63                        | 42|
|          | SMK                                       | 3    | 15                        | 10|
|          | D3                                        | 4    | 8                         | 5 |
|          | S1                                        | 5    | 51                        | 34|
|          | S2                                        | 6    | 1                         | 1 |
| X4       | Profession                                |      |                           |   |
|          | Government employees                      | 0    | 18                        | 12|
|          | Private sector worker                     | 1    | 60                        | 40|
|          | Entrepreneur                              | 2    | 32                        | 21|
|          | Freelancer                                | 3    | 24                        | 16|
|          | National army and police                  | 4    | 1                         | 1 |
|          | Etc                                       | 5    | 15                        | 10|
| X5       | Revenue (in rupiah / month)               |      |                           |   |
|          | < 1.000.000                               | 0    | 9                         | 6 |
|          | 1.000.000 – 2.000.000                     | 1    | 64                        | 43|
|          | > 2.000.000 – 4.000.000                   | 2    | 64                        | 43|
|          | > 4.000.000 – 6.000.000                   | 3    | 11                        | 7 |
|          | > 6.000.000 – 10.000.000                  | 4    | 2                         | 1 |
|          | > 10.000.000                              | 5    | 0                         |   |
| X6       | Based on residence                        |      |                           |   |
|          | With parents                              | 0    | 75                        | 50|
|          | Rent a house                              | 1    | 30                        | 20|
|          | Living in an owned house                   | 2    | 45                        | 30|
Table 1 indicates that 59% of respondents were male, 85% of whom were aged 21-40 years old. Most of whom completed high school/vocational education, and some others completed undergraduate level. Around 61% worked in the private sector. 91% had a maximum income of 4 million rupiah each month. Those living in an owned-house or with parents amounted to 80% of respondents. The largest number of family members was between 1-2 people in each household. The respondents mostly have lived in Yogyakarta for more than 10 years, 85% had a SIM C (driver license for motorcycle), 52% of respondents traveled on a scheduled basis (based on vehicle schedule) for 20 minutes or less. 66% of respondents traveled a distance of no more than 10 km.

4.2. Logit Regression Analysis
The first analysis, which was done by entering all X values using SPSS 24, revealed the following results:

| Variabel | B    | Significance | Information |
|----------|------|--------------|-------------|
| X1 = Gender | -2.114 | 0.000 | received |
| X2 = Age     | 0.864  | 0.049 | received |
| X3 = Education | -0.354 | 0.037 | received |
From the SPSS running program, it appears that there are only 6 variables that have a significant influence on the mode choice (sig <0.05), namely $X_1$, $X_2$, $X_3$, $X_5$, $X_8$, and $X_{11}$. In the same way, the program runs the process repeatedly and eliminates insignificant variables. In the end, it resulted in a variable that has a significance level of <0.05 namely the variables $X_1$ (gender), $X_2$ (age), $X_8$ (number of motorized vehicles), and $X_{11}$ (travel time on vehicles). Value of $\chi^2$ program results 49,480$>\chi^2$ table in df 4 (number of independent variables 2) is 9,488 or with a significance of 0.000 (<0.05) so the null hypothesis is rejected, which indicates that the independent variable has a significant impact on the model. Nagelkerke R Square value of 0.398 and Cox & Snell R Square 0.281 indicate that the ability of the independent variable ($X_1$, $X_2$, $X_8$, and $X_{11}$) in explaining the dependent variable is 39.8% and there are 60.2% other factors outside the model that explain the dependent variable. Meanwhile, from the results of running the SPSS program, the predicted values are the same as those observed by 23 people for online motorcycle taxis and 98 for motorbikes from 150 respondent. In the interpretation of logistic regression indicates that the overall percentage value of (23 + 98) / 150 = 80.7%. It implies that the accuracy of model is 80.7%.

Table 3. Classification

| Observed   | Predicted               | Percentage Correct |
|------------|-------------------------|--------------------|
|            | online motorcycle taxi | Motorcycle         |
| Step 1 Y  |                          |                    |
| online motorcycle taxi | 23          | 22              | 51.1       |
| Motorcycle | 7                  | 98              | 93.3       |
| Overall Percentage |            |                  | 80.7       |

The variable in the equation table 4 show that all independent variables have a significant $p$-value of <0.05, meaning that each variable has a significant partial effect on $Y$ in the model. The magnitude of the effect is indicated by the value of exp (B) or also called Odds Ratio (OR). For example, the travel time variable is ($X_{11}$) of 2.459, meaning that the increase in travel time is 1, and that there will be an increase in the probability of choosing a motorcycle by 2.459 times.

Table 4. Variables in Equations

|            | B     | S.E. | Wald | Df   | Sig.  | Exp (B) |
|------------|-------|------|------|------|-------|---------|
| Step 1     |       |      |      |      |       |         |
| $X_1$      | -1.838| 0.440| 17.460| 1    | 0.000 | 0.159   |
| $X_2$      | 0.952 | 0.390| 5.979 | 1    | 0.014 | 2.592   |
| $X_8$      | 0.957 | 0.431| 4.946 | 1    | 0.026 | 2.605   |
| $X_{11}$   | 0.900 | 0.290| 9.637 | 1    | 0.002 | 2.459   |
| Constant   | -1.854| 0.818| 5.145 | 1    | 0.023 | 0.157   |
The value of B in the table forms the following equation model:

$$\ln \left( \frac{P}{1-P} \right) = -1.854 - 1.838 \text{Gender} + 0.952 \text{Age} + 0.957 \text{Number of motorized vehicles} + 0.9 \text{Travel time}$$

We can also use the formula derived from the above equation as mentioned below:

$$P = \frac{\exp (-1.854 - 1.838 \text{Gender} + 0.952 \text{Age} + 0.957 \text{Number of motorized vehicles} + 0.9 \text{Travel time})}{1 + \exp (-1.854 - 1.838 \text{Gender} + 0.952 \text{Age} + 0.957 \text{Number of motorized vehicles} + 0.9 \text{Travel time})}$$

where P is the probability of choosing a motorcycle

Based on the model, of the four independent variables, gender is the most influential variable because it has the greatest absolute value of the X coefficient. The regression coefficient of gender was -1.838, showing that, in the travel mode choice of work trips, female were more inclined to choose online motorcycle taxi. The regression coefficient of age, number of motorized vehicles, and travel time was positive, indicating that the older the respondents’ age, the more their number of motorized vehicles, the longer their travel time, the greater their probability of using a motorcycle. For example, if there is a male, aged 25 years, has 1 motorized vehicle, and the travel time on the vehicle is 15 minutes, then the probability of using a motorcycle is 72%. If the travel time goes up to 25 minutes - with other variables fixed - then the probability of using a motorcycle becomes 87%. The regression coefficient of age, number of motorized vehicles, and travel time is around 0.9, which indicates that the impact of the three variables on the choice of mode is almost the same.

5. Conclusion
This study concludes that the sociodemographic/socioeconomic variable and the characteristics of travel in the form of travel time significantly affect people in determining their mode choice on work trips. Further studies are needed, especially those addressing the variable cost of travel and the level of service from each mode in the model to build.

6. References
[1] Uber Investor, “Uber Announces Results for Fourth Quarter and Full Year 2019,” Uber Investor, 2020. [Online]. Available: https://investor.uber.com/news-events/news/press-release-details/2020/Uber-Announces-Results-for-Fourth-Quarter-and-Full-Year-2019/. [Accessed: 31-May-2020].
[2] R. K. Hastuti, “Pendapatan Gojek Tumbuh Double di 2019,” CNBC Indonesia, 2020. [Online]. Available: https://www.cnbcindonesia.com/tech/20200212125118-37-137286/pendapatan-gojek-tumbuh-double-di-2019. [Accessed: 31-May-2020].
[3] J. Alfred and Risdiyanto, “Kajian Kinerja Pelayanan Transportasi Ojek Online (Gojek) Wilayah Kota Yogyakarta,” Universitas Janabradra Yogyakarta, 2018.
[4] Kementerian Perhubungan, “Peraturan Menteri Perhubungan Republik Indonesia No. 12 Tahun 2019 Tentang Perlindungan Keselamatan Pengguna Sepeda Motor Yang Digunakan untuk Kepentingan Masyarakat,” 2019.
[5] M. Z. Irawan, P. F. Belgiawan, T. B. Joewono, and N. I. M. Simanjuntak, “Do Motorcycle-Based Ride-Hailing Apps Threaten Bus Ridership? A hybrid Choice Modeling Approach with Latent Variables,” Public Transp., vol. 12, no. 1, pp. 207–231, 2020.
[6] Fer and Yan, “Peminat Naik Angkot Turun Drastis,” Jabar Ekspres, 2020. [Online]. Available: https://jabarekspres.com/2020/peminat-naik-angkot-turun-drastis/. [Accessed: 01-Jun-2020].
[7] C. M. Wardhani, “Wow, Jumlah Kendaraan di Kota Jogja Naik Hingga 200%,” TribunJojga.com, 2018. [Online]. Available: https://jogja.tribunnews.com/2018/02/26/wow-
jumlah-kendaraan-di-kota-yogya-naik-hingga-200. [Accessed: 18-Jun-2020].

[8] M. A. Abdulsalam, B. Miskeen, A. M. Alhodairi, R. Atiq, A. Bin, and O. K. Rahmat, “Modeling of Intercity Transport Mode Choice Behavior in Libya: a Binary Logit Model for Business Trips by Private Car and Intercity Bus,” vol. 7, no. 1, pp. 302–311, 2013.

[9] J. Setyono, “Model Pemilihan Moda Antara Angkutan Umum dan Sepeda Motor Untuk Maksud Kerja,” pp. 66–72, 2008.

[10] F. Miro, Perencanaan Transportasi untuk Mahasiswa, Perencana, dan Praktisi. Jakarta: Penerbit Erlangga, 2004.

[11] C. Rungie, “A Review of Latent Variable Models for Discrete Choice,” Int. Choice Model. Conf., no. Dcm, 2013.

[12] J. Nugraha, S. Guritno, and S. H. Kartiko, “Model Discrete Choice dan Regresi Logistik,” 2006.

[13] I. Putu Widiarta and I. Wardana, “Analisis Pemilihan Moda Dengan Regresi Logistik Pada Rencana Koridor Trayek Trans Sarbagita,” J. Ilm. Tek. Sipil, vol. 15, no. 2, 2011.