Effect of Air Pollution on Travel Mode Choice

Vaishali Yadav[1], Shiv babu Yadav[2], Arvind Gupta[3]

1, 2 M. Tech Scholar, Department of Energy and Environmental Engineering, J.C. Bose University of Science & Technology, YMCA

3, Professor, Centre for Energy Studies, J.C. Bose University of Science & Technology, YMCA

1. Abstract

With the progressions in the transportation sector, new modes of conveyance are being discovered for the mobility of people. India is one of the fastest developing nations in the world and the makeup of the vehicle fleets, air pollution management system, road designs and driving behaviours in Indian cities are different in comparison to all other regions of the world, hence the results from studies done outside cannot be directly implemented in Indian cities. People can choose from a variety of transportation modes for the same distance and route. These modes can be broadly classified as public modes, private motorized modes, para-transportation modes and active modes. The choice of any mode depends on many factors like meteorological factors including weather and seasons, socioeconomic factors like household income, number and type of vehicles owned, gender, and the nature of the occupation etc., and finally trip profile characteristics like the number of trips, trip length, purpose and cost. The pollution and poor air quality caused by the pollutants generated by these motor sources have a significant detrimental effect on the health of people of all ages in the community. Breathing in poor quality air can cause a variety of symptoms from short-term to long-term complications. Some of the chemicals that contribute absolutely to bad air quality include NO_x, SO_x, VOCs and particulate matters like PM_{10} and $PM_{2.5}$. Each pollutant has its adverse effects but due to having a high deposition rate the particulate matter gets easily deposited in the respiratory tract with smaller particles getting accumulated in the inner part of the lungs, while other chemicals may get out of the body exhaling. For this reason, only particulate matter is considered in this study. The multinomial regression method can be used to predict the probability of different mode choices based on multiple independent variables (meteorological factors, socioeconomic factors, trip characteristics) which can be either categorical or continuous. For this study, a questionnaire-based survey was performed to find out the stated preferences for the mode choice. People were also inquired about their preferred mode choice at higher pollution levels. The findings of this study can be used by numerous organisations at the local and national levels to implement policies that can increase the share of public transportation, decrease the number of private vehicles on the road, and help reduce air pollution even more significantly. This will enable the area to develop more sustainably.

Keywords: Air pollution, Air quality, NO_x, SO_x, VOCs, Particulate matters, PM_{10} and $PM_{2.5}$.

2. Introduction

Air pollution is the presence of non-natural or harmful substances which alters the proportion of various gases in the atmosphere. These pollutants can be in any form like gases (e.g. CO, NH_3, SO_2, NO_x, CH_4, CFC's), The diameter of pollutants allows for the division of biological molecules or particulate materials [1]. These might contain organic or inorganic materials. In India, Central Pollution Control Board (CPCB) gave a value to an industrial areas, residential areas, rural and other areas and ecologically sensitive areas which are notified in National Ambient Air Quality Standards providing upper limit for 12 air pollutants [2]. All these pollutants cause harmful effects on a human being to such an extent that they canlead to death, which causes many diseases and allergies [3]. These symptoms may be classified as long-term or short-term symptoms. Old people and children are highly susceptible to the harmful effect of these pollutants [4]. Epidemiological and toxicological studies have shown that the air pollution primarily affects functioning of heart and lungs leading to adverse complications by entering through the respiratory tract and accumulating in lung cells. Prolonged exposure to these pollutants results in lung cancer and chronic obstructive pulmonary disease (COPD) [5]. A decline in cognitive ability is also seen along with neuro-degenerative diseases like dementia. While for short term effects problems related to eyes, nose and throat are most common [3]. These may compile if exposed for long duration leading to bronchitis, pneumonia, asthma, and allergic reactions. A number of air pollutants, including smog, particulate matter, ozone, nitrogen dioxide, and sulphur dioxide, may irritate the ears, nose, and/or throat[6]. Many immediate consequences of air pollution on human skin may aggravate conditions including atopic dermatitis, skin cancer, psoriasis, and acne[7]. Not only living beings, it can cause damage to non-living agents like materials of structures or monuments and it also greatly affects the environment [6].

In new developing world various steps are made towards faster as well as greener transportability. The world is seeing a shift from motorised to non-motorized mode, but bad air quality can be a difficulty in this direction [7]. Developing countries like India and China are also now focusing on improving their air quality as most of the cities with maximum air pollution lie in these South-East Asian and East Asian regions [8]. People these days can choose from a variety of transportation modes for the same distance and route. These modes can be broadly classified as public modes, private motorised modes, para-transportation modes and active modes [9]. The choice of any model depends on many factors like meteorological factors including weather and seasons, and socioeconomic factors like household income [10]. Higher income group people generally prefer a more private mode of travel as it is faster, while lower-income groups are captive to public transit and active transit because of their low cost of operations. The number and type of vehicles owned were also affected in the same manner. Gender and occupation affect the type of vehicle owned by the person [11]. Finally, and importantly, trip profile characteristics like the number of trips, trip length, purpose and cost affect the mode choice directly or indirectly [12]. Among these modes of transportation, generally public-owned mass transit systems produce less pollution as compared to private vehicles in per-capita terms [13]. Pollutants emitted from these vehicular sources combine with fog and dew particles especially at dawn times to form smog and lead to poor air quality. Even the good health impacts of the active mode of transportation can be overshadowed by the negative effects of bad air quality being breathed by the commuter on the polluted road segment [14]. The emitted particulate matter gets easily deposited in the respiratory tract with smaller particles getting deposited in the inner part of the lungs, while other chemicals may get out of the body while exhaling. For this reason, only particulate matter is considered in this study. $PM_{2.5}$ is a part of PM_{10} [15]. Thus, being smaller in size $PM_{2.5}$ has the highest ill effect on the health of people. These values of $PM_{2.5}$ are measured by various fixed monitoring stations located at various areas in the study area. The pollution concentration which is obtained from here can be labelled as good, satisfactory, poor, very poor and severe [16].

This mode choice analysis is done to predict the mode which a person will choose and the choices available being discrete choices, mode choice can be classified as categorical variables [1]. Thus, a multinomial logit model fits the best for discrete choice modelling. For this study, a questionnaire-based survey was performed to find out the stated preferences for the mode choice for different levels of perceived pollution and to find to which mode a person will switch if the pollution level is increased [17]. The questionnaire survey was collected outside various interchange metro stations in the city of Delhi in India. The number of samples was based on the relative population of that area [10]. A survey was divided broadly into three parts, first targeted to understand people's understanding of air quality and problems caused by air pollution [18]. The second targets the socio-economic characteristics of the person while the third to get information based on the trip profile of the person. A descriptive analysis is done separately on the three parts followed by the development of a logit model with walking being considered as a base choice i.e. probability of any mode choice is found in comparison to walking [12]. From the descriptive analysis, it can be found out which variables affect the mode choice and what is the correlation of various variables.

3. Methodology

Study region

Air Pollution has already become a major concern in our country, affecting health, habits, ecology and tourism. The topography of Delhi significantly influences how pollution affects the city. Delhi is separated into five segments for this research. (east Delhi, west Delhi, north Delhi, south Delhi, centre Delhi). The life expectancy of people living in Delhi has dropped as AQI have exceeded severe levels. Although having the 12th highest population, Delhi is one of the world's top 21 most populated urban agglomerations and the 10th most populous metropolitan region.

According to some studies on Delhi's pollution, it has been proved that Delhiites have 8% higher risk of further chest pain while they are exposed to hazardous level of PM2.5 last year. The pollution level was more than 4 times trend that set off Public Health Emergency by UN's World Health Organization (WHO).

In Delhi, there has been rapid population growth of residents and vehicles, leading to compounded levels of air pollution over the last 30 years. The main idea behind this study is to investigate how the people will react to change in air quality in terms of their mode of choice. As the private motorised modes of transportation produce more vehicular emissions per capita, it becomes necessary to understand the increase or decrease in number of private vehicles if the air quality deteriorates. As discussed in the literature review logistic regression model can be used to study discrete variables choices like mode choice. For this study a questionnaire was made based on the possible variables as suggested by literature review. For this firstly information about air pollution perception was collected followed by trip characteristics and their socio- economic information. A road-side interview survey was performed and obtained data was analysed using SPSS.

Designing Survey Sample

The sample size is an important consideration in any empirical research that aims to infer information about a population from a sample. Even though a bigger sample size would require more time and effort without improving accuracy, larger samples are frequently more accurate when predicting unknown parameters. Because of this, the ideal sample size for the given study's objectives sits somewhere in the middle. The following statistical formula may be found in the majority of statistical textbooks:

$$n = \frac{(Z_{\frac{\alpha}{2}})^2 * \pi * (1-\pi)}{E^2 + \frac{(Z_{\frac{\alpha}{2}})^2 * \pi * (1-\pi)}{N}}$$

n: is the sample size

Z: the statistics tables' normal distribution factor

α: is the 1-confidence level

π: response distribution factor
E: acceptable margin of error
N: size of population

Questionnaire

Questionnaires are among the main sources for the primary data collection. Questionnaires can be used to collect both qualitative data as well as quantitative data. Based on the literature review all the possible data required for the survey was reviewed and a questionnaire was constructed consisting of all the data required for the study. Most important factor about this questionnaire is the division of the questionnaire into three parts: first is air pollution perception information, second is socio-economic information and third is trip profile. In these three-section interviewers ask questions which may be directly or indirectly affecting the factors which may help a person in deciding the mode of choice for their commute.

Model Development

The multinomial logit model presupposes that each independent variable has a single value for each occurrence and that data are case-specific. The multinomial logit model also assumes that the dependent variable cannot always be accurately predicted from the independent variables. The appeal that a person associates with a certain travel determines the utility of a particular method of transportation. The person will pick the technique that appeals to them the most given several considerations including the cost and duration of the journey, their gender, their age, etc. Utility maximisation is the term used to describe this idea. A mathematical equation may be used to explain a mode m's utility function.

$$U_{ni} = þ_{ni1}x_{ni1} + þ_2x_{ni2} + \cdots + þ_kx_{nik}$$

Where:

U_{ni} : is the mode m for person i's net utility function.

$x_{ni1}, \ldots x_{nik}$: are k different mode m qualities for person i.

$þ_1, \ldots þ_k$: are k numbers of coefficients that must be inferred from the survey data, or weights that are assigned to each characteristic.

The random utility model, which considers utility as a random variable with two clearly separable components: a measured conditioning component and an error component, may be used to explain the decision behaviour. Therefore,

$$U_{ni} = V_{ni} + E_{ni}$$

Where:

Vni : is the systematically observed part of the mode m's utility for person I.

Eni : is the unobserved error component of utility of mode m for individual I as a result of the influence of unobserved variables.

The likelihood of a person I choosing a mode n out of the total M possible modes is provided in Equation, which briefly summarises the framework.

$$P_{in} = \frac{e^{V_{in}}}{\sum_{m=1}^{M} e^{V_{im}}}$$

Where:

Vin : is the mode n's useful function for person i.

Vim: The utility function of any mode m among the available options for an individual I is known as vim.

PIN: is the likelihood that person I will choose mode n.

M : is the total number of trip options for person I in the choice set.

Although these equations are often approximated using the greatest likelihood technique, the multinomial logit model makes use of mathematical frameworks.

The values of the parameters for which the observed sample is most likely to have happened are estimated using the maximum likelihood technique. The collection of parameters that are "most likely" to have led to the decisions shown in the data are determined using the maximum likelihood technique.

Data Analysis and Calibration of The Model

After submitting the questionnaire, all of the completed data were entered into a computer programme (SPSS) and used in a multinomial logistic method that estimates probabilities using a logistic function, in this case the cumulative logistic distribution, to determine the relationship between the categorical dependent variable and one or more independent variables.

In order to create the mode choice model, data were analysed and calibrated using logistic regression rather than conventional multiple regression, as the assumptions needed for conventional regression are not satisfied, taking into account each of the chosen transportation modes as well as all potential factors that could have an impact on this mathematical model.

Data Analysis

The data analysis is done in 2 parts. First, the descriptive statistics is done to understand the nature of the variables followed by correlation analysis as Multiple logistic regression requires the variables to be non collinear. Hence if two variables show collinearity only of them is used in the model development. Using these two methods data is filtered and then a model is developed based on the filtered observation data.

Descriptive statistics

According to the descriptive data, more than 80% of people are aware that one way to gauge the quality of the air we breathe is by using the AQI. As most individuals learn about AQI information via mobile apps, it is also clear from the response data that mobile applications were crucial in raising awareness about air quality. Although they don't check it on a regular basis and mostly people check it once or twice a week. As the pilot survey was conducted in the month of December most of the responses for people's opinion on currentAQI came out to be in the poor category while the main survey conducted in late winters and early springs showed people's response shifting towards moderate category. Although a bias was observed as 87% people were not married with most of them being in the age category of 19-25 years being students. Most of the people use public transportation with 60% people. While this study reveals that 10% of the population using active transportation takes on the highest number of ill-effects of air pollution. It is seen people will switch to private modes as the pollution increases. Most of the trips by these people are from home to work/office.

4. Results

Due to the fact that eliminating the influence of any variable does not increase the degrees of freedom, the final reduced model is comparable to the full model. Travel cost was found to be most statistically significant for the final model. Also the McFadden pseudo-R^2 value was obtained to be 0.38. Also the Nagelkerke pseudo-R^2 value was nearing 0.7 which shows the model to be good. From the model fitting statistics we can conclude the significance of likelihood ratio test is very less than 0.05, thus we can conclude that the model showed a good result for detecting preferred modes of travelling for people based on the environmental conditions. Although socio-economic parameters also become important while using this model, one should keep in mind the distance of travel and cost for the same plays major roles along with environmental factors.

Table 2 : Model fitting

Model	Model fitting criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	sig.
Intercept Only Final	996.687 616.260	380.427	256	.000

Model fitting information

We must first evaluate if the model improved our ability for outcome prediction before we examine the implications of each explanatory variable in the model. To do this, we contrast a model with no explanatory factors (the baseline or "Intercept Only" model) with a model that includes all explanatory variables (the "Final" model), which at the time of writing only includes gender. To determine whether the final model considerably improved the data fit, when comparing it to the baseline model.

The deviation has limited usefulness on an intuitive level since it depends on the sample size, the quantity of parameters in the model, the degree of fit, and other variables. If the model represents the data better than the baseline model, the deviation (-2LL), which can be tested against the chi-square distribution to generate a p value, should be much lower.

R^2 equivalents for logistic regression

Another way to judge the effectiveness of a regression model is to determine how strong the relationship is between the explanatory variable or variables and the outcome variable. In a linear regression study, the R2 statistic served as a representation of this. Logistic regression also allows for the calculation of R2, or rather a variant of it. There are several versions, which might be confusing. This is due to the fact that the various versions employ pseudo-R2 statistics to estimate rather than accurately measure the amount of variance explained. Bear in mind that possibilities are at play here. Despite this, it can occasionally be helpful to evaluate them in order to determine the actual worth of your model.

The two versions that are most frequently used are R2 by Hosmer & Lemeshow and R2 by Nagelkerke. These words relate to the proportion of the variance in the result that the model can explain. These values vary from "0" to "1," similar to R2 in multiple regression, with a value of "1" signifying that the model fully accounts for the variance in the outcome and "0" signifying that it does not. Be careful since they are decided differently and may provide contradictory results! We'll go over our examples in the pages that follow and explain how to comprehend these data, which are available with SPSs.

Table 3 : Pseudo - R squares

Cox and Snell	.652
Nagelkerke	.696
McFadden	.382

5. Conclusion

In the developed model a high goodness of fit was seen. Although all the variables from pilot survey were not used in the final model development as few variables were dropped from the questionnaire in the final survey based on obtained results. High range of pseudo R^2 suggests few redundancies are present in the data thus a more extensive filter needs to be applied in the results obtained from the final survey. Only 27% of those surveyed are familiar with and comprehend the AQI. The categories of "moderate" and "poor" air quality were seen as troublesome by more than 60% of the respondents. Accordingly, 68% of commuters stay away from areas with extremely bad or severe air quality. For primary and secondary travels, the percentage of the most exposed user group—walking, bicycling, using motorised two-wheelers, and using auto-rickshaws—is around 38% and 35%, respectively. Additionally, it was discovered that when air quality is bad or really bad, more people are prone to utilise closed modes of transportation. The study shows that commuters are unable to make choices that would lessen their exposure while travelling since there is no information on air pollution exposure available. This research suggests that government decision-makers develop or put into place a real-time information system that can help commuters change their minds.

Limitations

This paper presents a mode choice analysis using surveys. The major issue with that is people are not ready to share their point of view. also, only the broad categorisation of the mode of transportation is stated in the questionnaire, which can be further specified like bus metro motorbikes and cars. The data was collected on random days whereas it can be collected differently for weekdays and weekends during some of the most polluted days of the year.

Multinomial logistic regression is a powerful tool for modelling categorical data. However, there are some limitations to its use. Particularly susceptible to minute changes in the data is the multinomial logit model. This could produce unreliable outcomes. Moreover, huge data sets do not lend themselves well to the multinomial logit model.

Another limitation is that the model cannot be used to predict the probability of a particular event occurring; also the multinomial logit model cannot identify the order of the explanatory variable.

Future work

According to our research, the quality of the transportation system has both direct and indirect effects on how many cars people buy and use (through outcome expectations and attitudes). Therefore, actions that enhance the transit system's effectiveness in terms of speed and dependability, accessibility, and trip quality might potentially make it more alluring than driving. The crucial caveat is that, regardless of how effective and handy the transportation system is, once a car is owned, the owner's driving habit is not changed by it. There is no doubt that both short- and long-term human behaviour are strongly impacted by air pollution. Travelers are, naturally, the most exposed demographic. This may interfere with transportation regulations intended to make the system more cost-effective and ecologically friendly. The items listed below can be looked into further.

1. The study used data and took into consideration how perceptions alter with the conditions. A perception survey should be conducted throughout the year to gather information.

2. Although stated preferences are taken into account in this study, a preference survey should also be conducted to determine the preferred method of transportation when air pollution levels are high.

References

[1] R. K. Gupta, P. K. Singh, V. K. Mishra and P. P. Chauhan, “A Comparative Study for Assessing the Air Quality Status for Industrial Areas [Gida, Gorakhpur and Talkatora, Lucknow],” *Int J Sci Technol Eng,* vol. 2, p. 949–953, 2016.

[2] S. Tiwari, A. Thomas, P. Rao, D. M. Chate, V. K. Soni, S. Singh, S. D. Ghude, D. Singh and P. K. Hopke, “Pollution concentrations in Delhi India during winter 2015–16: A case study of an odd-even vehicle strategy,” *Atmospheric Pollution Research,* vol. 9, p. 1137–1145, 2018.

[3] D. Absalon and B. Ślesak, “The effects of changes in cadmium and lead air pollution on cancer incidence in children,” *Science of the Total Environment,* vol. 408, p. 4420–4428, 2010.

[4] R. Beelen, G. Hoek, D. Vienneau, M. Eeftens, K. Dimakopoulou, X. Pedeli, M.-Y. Tsai, N. Künzli, T. Schikowski, A. Marcon and others, “Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe–The ESCAPE project,” *Atmospheric Environment,* vol. 72, p. 10–23, 2013.

[5] D. M. N. Abdulhaq, “Transportation Mode Choice Model for Palestinian Universities Students: A Case Study on An-Najah New Campus,” 2016.

[6] A. Bayaga, “Multinomial Logistic Regression: Usage and Application in Risk Analysis.,” *Journal of applied quantitative methods,* vol. 5, 2010.

[7] D. J. Briggs, C. de Hoogh, J. Gulliver, J. Wills, P. Elliott, S. Kingham and K. Smallbone, “A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments,” *Science of the Total Environment,* vol. 253, p. 151–167, 2000.

[8] A. Ernst and J. D. Zibrak, “Carbon monoxide poisoning,” *New England journal of medicine,* vol. 339, p. 1603–1608, 1998.

[9] N. Armaroli and V. Balzani, “Towards an electricity-powered world,” *Energy & Environmental Science,* vol. 4, p. 3193–3222, 2011.

[10] R. T. Xu, S. F. Pan, J. Chen, G. S. Chen, J. Yang, S. R. S. Dangal, J. P. Shepard and H. Q. Tian, “Half-century ammonia emissions from agricultural systems in Southern Asia: Magnitude, spatiotemporal patterns, and implications for human health,” *GeoHealth,* vol. 2, p. 40–53, 2018.

[11] J. Antony, “Six Sigma in the UK service organisations: results from a pilot survey,” *Managerial auditing journal,* 2004.

[12] M. E. Ben-Akiva, S. R. Lerman, S. R. Lerman and others, Discrete choice analysis: theory and application to travel demand, vol. 9, MIT press, 1985.

[13] S.-J. Cao, X.-R. Kong, L. Li, W. Zhang, Z.-P. Ye and Y. Deng, “An investigation of the PM 2.5 and NO 2 concentrations and their human health impacts in the metro subway system of Suzhou, China,” *Environmental Science: Processes & Impacts,* vol. 19, p. 666–675, 2017.

[14] J.-P. Gabriel, F. Saucy and L.-F. Bersier, “Paradoxes in the logistic equation?,” *Ecological Modelling,* vol. 185, p. 147–151, 2005.

[15] B. A. Greene and J. D. Hovis, “TQM Process Action Team reduction of receiving cycle time from 3 weeks to 3 days at Kennedy Space Center,” in *Space Logistics Symposium, 5 th, Huntsville, AL,* 1993.

[16] K. Guo, Y. Du, Y. Wu, X. Mi, X. Li and S. Chen, “Morphology and FT-IR analysis of anti-pollution flashover coatings with adding nano SiO2 particles,” in *IOP Conference Series: Materials Science and Engineering,* 2017.

[17] H. Harmens, D. A. Norris, G. R. Koerber, A. Buse, E. Steinnes and Å. Rühling, “Temporal trends in the concentration of arsenic, chromium, copper, iron, nickel, vanadium and zinc in mosses across Europe between 1990 and 2000,” *Atmospheric Environment,* vol. 41, p. 6673–6687, 2007.

[18] P. Doyle, Echo and reverb: Fabricating space in popular music recording, 1900-1960, Wesleyan University Press, 2005.

[19] M. Eeftens, R. Beelen, K. De Hoogh, T. Bellander, G. Cesaroni, M. Cirach, C. Declercq, A. Dedele, E. Dons, A. De Nazelle and others, “Development of land use regression models for PM2. 5, PM2. 5 absorbance, PM10 and PMcoarse in 20 European study areas; results of the ESCAPE project,” *Environmental science & technology,* vol. 46, p. 11195–11205, 2012.

[20] A. Dutta and W. Jinsart, “Assessing short-term effects of ambient air pollution on respiratory diseases in Guwahati, India with the application of the generalized additive model,” *Human and Ecological Risk Assessment: An International Journal,* vol. 27, p. 1786–1807, 2021.

[21] M. Jerrett, R. T. Burnett, J. Brook, P. Kanaroglou, C. Giovis, N. Finkelstein and B. Hutchison, “Do socioeconomic characteristics modify the short term association between air pollution and mortality? Evidence from a zonal time series in Hamilton, Canada,” *Journal of Epidemiology & Community Health,* vol. 58, p. 31–40, 2004.

[22] S. Velasco, A. J. Kedaigle, S. K. Simmons, A. Nash, M. Rocha, G. Quadrato, B. Paulsen, L. Nguyen, X. Adiconis, A. Regev and others, “Individual brain organoids reproducibly form cell diversity of the human cerebral cortex,” *Nature,* vol. 570, p. 523–527, 2019.

[23] Y. O. Khaniabadi, R. Polosa, R. Z. Chuturkova, M. Daryanoosh, G. Goudarzi, A. Borgini, A. Tittarelli, H. Basiri, H. Armin, H. Nourmoradi and others, “Human health risk assessment due to ambient PM10 and SO2 by an air quality modeling technique,” *Process safety and environmental protection,* vol. 111, p. 346–354, 2017.

[24] C. A. Klöckner and T. Friedrichsmeier, “A multi-level approach to travel mode choice–How person characteristics and situation specific aspects determine car use in a student sample,” *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 14, p. 261–277, 2011.

[25] S. Menard, Applied logistic regression analysis, Sage, 2002.

[26]

IJREET
International Journal for Research
in Engineering and Emerging Trends