Modelling and Exploring the Impacts of E-Grocery Shopping on Trip Generation in India

Suprava Jena[1], Momi Deb[2] and Manish Dutta[3]

[1]*Assistant Professor Department of Civil Engineering, National Institute of Technology Silchar, Assam-788010, India*

[2] *PhD Student, Department of Civil Engineering, National Institute of Technology Silchar, Assam-788010, India*

[3] *Assistant Professor, Department of Civil Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat- 382481, India*

{Corresponding author's email: momidebsrv18@gmail.com}

Abstract - The limited space for the road widening and an increasing number of vehicles leads to never ending congestion in the peak hours. However, the trip generation can be reduced or manipulated by targeting any essential activity such as grocery shopping (which has been carried out even in the strictest lockdown due to the covid-19 pandemic). In this study, a questionnaire survey is done and data of 434 respondents are analysed to see how the grocery shopping trip behaviour of the people residing in tier I cities of India is influenced by the chosen demographic variables and online grocery shopping. A linear model is formulated by multiple linear regression to forecast the number of grocery shopping trips, which is validated with the observed number of trips. The model reveals that purchasing groceries from online shopping do reduce the number of shopping trips generated and the same can be used by urban planners while sketching out the plans for transportation infrastructure

Keywords – Online grocery shopping; Shopping trips; Trip generation; Multiple linear regression; India

INTRODUCTION

The expansion of road widths on limited and expensive urban lands would not be feasible after a stage in a populous and dense country like India. The intent of reducing congestion in urban areas by constructing flyovers or limited- access high capacity roads has gone in vain with the passage of time as from practical experience and researchers have demonstrated that when highway capacity increases, vehicle travel also increases as it attracts potential demand as well as reduces public and non-motorised transportation use due to compromised access for these modes of transport. Ref. [1] reveals India is the fastest growing E-commerce market and is expected to grow at approximately 1,200% by 2026 with grocery and fashion/apparels likely to be the key drivers. Also, grocery is the largest consumer segment and Indians spend more than 50% of their monthly income on groceries and is conducted frequently with high repeat rate, so much so that it was carried out even during the lockdown. Studying the activity and travel pattern of grocery shopping and considering the same while modelling to forecast the grocery shopping trips, then not only the congestion in the markets can be reduced but also suggestions can be put forward for planning an infrastructure for trip reduction which will counteract the booming number of private vehicles on urban roads.

Ref. [2] demonstrated that ICT could influence leisure activities in four ways: substitution, complementary, modification and neutrality. The positive relationship between ICT usage and travel were found by a few researchers [3], [4], [5], [6]. Ref. [7] found that men between the ages of 25 and 40 who were highly educated, had a high income, living in a less urbanised region are mostly the online buyers. Ref. [8] concluded that online shopping might facilitate changing travel behaviour by employing eco-friendly means of transportation in Sweden. Individuals in developing countries prefer in-store shopping practises, according to evidence from Bandung by Ref. [9]. Few researchers showed that frequency of online purchases was positively related to the frequency of shopping trips [10], [11].

The aforesaid previous studies were done in the European countries, USA, and China. Few studies were done on the transportation impacts of ICT use [5], [6]. While few others explored the impacts of online shopping on travel behaviour [8], [7].The relationship between online searching, online buying and in-store shopping were also investigated by some researchers [10], [11], [12], [13]. However, none of them have focussed particularly on online grocery shopping, which is an indispensable activity, that too keeping the current pandemic in view.

In India, transportation impacts of teleworking have been studied till now [14], [15], [16], but online shopping impacts on travel behaviour are yet to be explored. Impacts of online grocery shopping in particular on travel behaviour is also not explored anywhere across the globe The economically and culturally diverse population and the largest middle-class section of Indian society might respond different to the transforming grocery shopping (physical to virtual) and thus can give valuable inputs to the planners.

The objective of this study is to investigate the impacts of e-grocery shopping on grocery shopping trip generation by developing a model which will predict the number of grocery shopping trips generated. The paper is structured in the following ways. The next section consists of methodology which includes selection of study areas and

input variables, data collection and modelling approach. Results and analysis are presented in the final section which is succeeded by concluding remarks.

METHODOLOGY

A. *Selection of Study Areas & Input Variables*

The study areas are the tier I cities of India i.e., Delhi, Mumbai, Bangalore, Kolkata, Pune, Ahmedabad and Hyderabad. They are chosen based on the availability of online grocery shopping services like Big Basket, Jiomart etc.

The input variables i.e., dependent variable chosen is number of trips generated for grocery shopping even after they have purchased online and the independent variables include socio demographic factors and grocery shopping information.
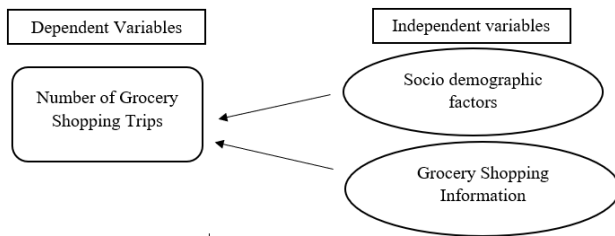

Fig.1 Input Variables

B. *Data collection*

The data was collected from the above shown cities though questionnaire survey by sharing the google forms via social media sites, messenger apps and mails.

The Questionnaire survey consists of three sections. The first section contains demographic information of the grocery shopper of the family like age, gender, educational background, type of work and the city they inhabit. The second section comprises of household information of the grocery shopper like monthly household income, no. of kids (<5 years) and no. of senior citizens (>60 years) in the family, no. of family members with driving license ownership, and no. of two wheelers and four wheelers. The last section inquires grocery shopping information whether they make the purchase online or offline, how many times a month they opted for store shopping, how and when they make the purchase in a day, how many times they went on their foot or in a vehicle.

C. *Modelling Approach*

Since the dependent variable (number of trips) was continuous and there were six independent variables, multiple linear regression is used for modelling expressed in the following equation:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon \qquad (1)$$

Where, β_0 is the intercept; $\beta_1, \beta_2 ..... \beta_n$ are the corresponding coefficients to be calculated; $X_1, X_2 ..... X_n$ are the independent variables; ε is the standard error.

RESULTS AND ANALYSIS

A. *Analysis from the survey*

543 responses were received and got updated in the google sheet simultaneously, from which, after screening,

434(80%) have been selected for analysis. Once the data were received, the independent or predictor variables were finalised. Gender, type of work (fixed/flexible/ work from home/ student/homemaker), number of kids below five years of age, number of senior citizens, number of two wheelers and online grocery shopping time (not applicable, before/after work, while working) were selected for further analysis.

It is observed from the data that out of 434 respondents, the higher income groups (50k-100k and >100k) indulge more in e-grocery shopping (57% and 55% respectively). The middle- and lower-income groups do purchase groceries from online sites but the percentage is lower than those who do not at all (less than 45%). It must be due to the financial capability and hectic working hours of the respondents that make them opt for e-grocery shopping. Another tendency was seen in respondents who go for the in-store grocery shopping mostly in the evening (69%), followed by morning (17%) and afternoon (14%). This might be because people finish off their work mostly by evening (in Indian scenario) and goes for purchasing groceries post-work. Also, the heat in the daytime might discourage shoppers to go to the stores in the afternoon and thus purchasing starts as the sun sets. This finding can be used to suggest for encouraging masses to use public transportation by improvising the infrastructure, solely for the purpose of shopping in the evening. This would decrease congestion to quite a significant extent in the peak hours of evening. Upon asking the reason for not opting the E-grocery shopping, 33% ogave the reason of intangibility which is missing in the virtual platforms. 15% claim that groceries are cheaper in stores than in online sites which contradicts our assumption of discounts and cashbacks being the magnetic reason for increase in online grocery shopping. High delivery charge prevents 10% to make online grocery purchase.

B. *Outcomes of Multiple Linear Regression*

Table I shows if the overall regression model is significant or not which is determined by checking if the p value is less than 0.05 or not. The ANOVA test results shows that the model is significant since $p \leq 0.05$. In Table II the R Square value is interpreted first. R square value indicates that 77% percentage of variance is accounted for in number of trips generated by all the independent variables or predictors included in the model. The Standard error of estimate shows the measure of accuracy of the prediction is 69.1%.

Table III describes how the predictor variables impact the dependent variables. The intercept of the model is found to be 3.991. All the independent variables are significant since p value is less than 0.05. The coefficient of gender is 1.614 which suggests that if the grocery shopper is male, then the trips generated will be 1.614 times more than that of female. This justifies that woman has multiple roles to play and going multiple number of times from their tight schedule for grocery shopping purpose is least expected from them.

TABLE I
(ANOVA TEST RESULTS)

Model	Sum of Squares	df	Mean Square	Sig.
1				.000ᵇ
Regression	688.623	6	114.770	
Residual	203.764	427	.477	
Total	892.387	433		

TABLE II
(SUMMARY OF THE MODEL)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.878	.772	.768	.691

The type_of_work variable has the coefficient of -0.211 which means that the one with the most liberal working hours (homemaker=4, student=3) will generate lesser trips than that of fixed timing workers (fixed=0, flexible=1, WFH=2) i.e., if a homemaker is the grocery shopper of the home, the shopping trips generation decreases by 84.4%. This indicates that the respondents who have fixed working schedule tend to generate more grocery shopping trips. The coefficient of no_of_kids is 0.812 and keeping the values assigned into account (0 kid=3, 1 kid=2, 2 kids =1 and >2 kids =0), it can be interpreted that the number of shopping trips increases by 2.44 if a household has no kid below 5 years of age. The presence of kids below 5 years of age requires constant attention which does not let the adults in families (mostly nuclear in structures) make too many frequent trips to grocery stores Whereas no_of_senior_citizens has a coefficient of -0.881 which shows as the number of senior citizens in a household increases, the shopping trip generation increases by 88.1. This positive co-relation between dependent and independent variables in spite of the negative sign associated with the coefficient is because of ordinal nature of independent variable (no_of_senior_citizens: - 0=3,1=2,2=1,>2=0). This depicts that those households which have a greater number of senior citizens are more inclined towards in-store shopping and thus more trips will be generated. The coefficient of no_of_two_wheelers indicates that with the increasing number of two wheelers, the trip generation will decrease by 88.7%. This is quite contradictory since the greater number of two wheelers should have made the grocery shopping trips more convenient and easier but here the decrease in trip generation is justified after analysing data (from our questionnaire survey) that a household with a greater number of two wheelers indicates a higher monthly household income and thus they are more inclined towards online grocery. The last predictor variable Online_shopping_time has a negative association with the dependent variable with a coefficient of -0.383. This coefficient clearly indicates the negative relationship between online grocery shopping & in-store grocery shopping trips generated.
The t-test associated with a b-value is significant (sig.<0.05) then the predictor is making a significant contribution to the model. All are significant independent variables for this model. The larger the value of t, the greater the contribution of that predictor

TABLE III
(MODEL PARAMETERS)

Model	Unstandardized Coefficients B	Std. Error	Standardized Coefficients B	t	Sig.
(Constant)	3.991	.156		25.565	
Gender	1.614	.072	.548	22.330	0.000
Type_of_Work	-.211	.025	-.205	-8.418	0.000
No_of_Kids	.812	.041	.474	19.608	0.000
No_of_Senior_Citizens	-.881	.041	-.517	-21.517	0.000
No_of_Two_Wheelers	-.887	.035	-.629	-25.353	0.000
Online_Shopping_Time	-.383	.046	-.206	-8.341	0.000

Since B-values indicate the individual contribution of each predictor to the model, the model obtained by replacing B-values in (1) is as follows:
$$NOT = 3.991 + 1.614G - 0.211TOW + 0.812NOK - 0.881NOS - 0.887NOTW - 0.383OST + error \qquad (2)$$
Where, NOT= Number of grocery shopping trips generated,
G = Gender,
TOW= Type of work,
NOK= Number of Kids below 5 years of age,
NOS= Number of senior citizens,
NOTW= Number of Kids Two wheelers,
OST= Online grocery shopping.

C. Validation of the Regression Model
The model obtained after applying multiple linear regression on the input variables is validated by plotting a scattered graph between observed number of shopping trips and predicted number of shopping trips, after choosing 20% data randomly from the response sheet. Fig. 2 shows that there is 80.13% similarity between the predicted number of shopping trips, which is obtained after putting the assigned values of the independent value in (2) and observed number of trips obtained from the questionnaire data.
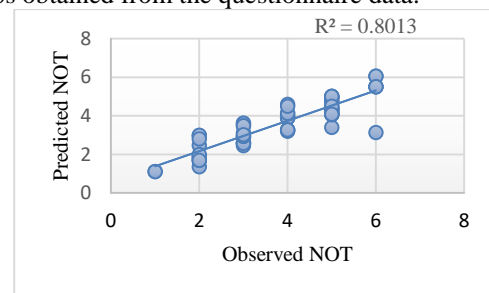

Fig 2 Graph between observed and predicted trips.

CONCLUSION

This study explores the grocery shopping trip behaviour of people of tier I cities of India and how the shopping trip generation gets impacted by the different independent variables. After the literature review and gap identification, the determination of the objective followed. The study area was chosen, and data collection was done using a questionnaire survey. The modelling approach employed on the collected data was the multiple linear regression. The results from this project work and their applications are summarised below:

- E-grocery shopping is done more by the middle- and higher-income groups which justifies their financial capability as obtained from the data.
- The in-store grocery shopping is done mostly in the evening (72%). So, the probable congestion due to this activity can be solved by providing public transportation facilities by the urban planners meant specifically for shopping purposes in the dusk hours.
- The model developed shows that online grocery shopping decreases the grocery shopping trip generation. This will provide an insight to those entrepreneurs who want to build start-ups on online grocery delivery from the local grocery shops and malls.

This study can be extended to the shopping of all products to predict the shopping trip generation in India. The impacts of trip generation due to shopping activity on environment (in terms of emission of harmful pollutants) can be examined.

REFERENCES

[1]   "https://www.ibef.org/download/E-Commerce-October-2020.pdf,"," [Online].

[2]   P. L. Mokhtarian, I. Salomon and S. L. Handy, "The impacts of ict on leisure activities and travel: A conceptual exploration," Transportation, 33(3), pp. 263-289, 2006.

[3]   M. Senbil and R. Kitamura, "Simultaneous Relationships Between Telecommunications and Activities," in 10th International Conference on Travel Behaviour Research, Lucerne, 2003.

[4]   F. Zhang, K. J. Clifton and Q. Shen, "Reexamining ICT Impact on Travel Using the 2001 NHTS Data for Baltimore Metropolitan Area," 2005.

[5]   D. Wang and F. Y. T. Law, "Impacts of Information and Communication Technologies (ICT) on time use and travel behaviour: a structural equations analysis," Transportation, vol. 34, pp. 513-527, 2007.

[6]   S. Choo, I. Kim and H. Lee, "Exploring Relationships between Information and Communication Technology (ICT) Use and Travel,"

Journal of the Eastern Asian Society for Transportation Studies, vol. 8, 2010.

[7]   S. Farag, M. Djist and M. Lanzendorf, "Exploring the Use of E-Shopping and Its Impact on Personal Travel Behavior in the Netherlands," Transportation Research Record 1858, pp. 03-3058, 2003.

[8]   L. W. Hiselius, L. S. Rosqvist and E. Adell, "Travel behaviour of online shopper in Sweden," Transport and Telecommunication, vol. 16, pp. 21-33, 2015.

[9]   T. B. Joewono, A. K. Tarigan and M. Rizki, "Segmentation, Classification, and Determinants of In-Store Shopping Activity and Travel Behaviour in the Digitalisation Era: The Context of a Developing Country," Sustainability, 2019.

[10]   S. Farag, T. Schwanen and M. Dijst, "Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping," in 45th Congress of the European Regional Science Association: "Land Use and Water, Amsterdam, 2005.

[11]   Z. F, X. Du, J. Cao and P. L. Mokhtarian, "The association between spatial attributes and e-shopping in the shopping process for search goods and experience goods: Evidence from Nanjing," Journal of Transport Geography 66, pp. 291-299, 2018.

[12]   G. Xi, F. Zhen, X. Cao and F. Xu, "The interaction between e-shopping and store shopping: empirical evidence from Nanjing, China," Transportation Letters, 2018.

[13]   K. Shi, J. D. Vos, Y. Yang and F. Witlox, "Does e-shopping replace shopping trips? Empirical evidence from Chengdu, China," Transportation Research Part A 122, pp. 21-33, 2019.

[14]   L. P. C and M. V. L. Anjaneyulu, "Modeling the Choice of Tele-work and its Effects on Travel Behaviour in Indian Context," in Procedia - Social and Behavioral Sciences 104, 2013.

[15]   L. P. C and M. V. L. Anjaneyulu, ""Modeling the Impact of ICT on the Activity and Travel Behaviour of Urban Dwellers in Indian Context," Transportation Research Procedia 17, pp. 418-427," 2016.

[16]   L. P. C and M. V. L. Anjaneyulu, "Networkwide Impact of Telework in Urban Areas: Case study of Bangalore,India," Journal of Transportation Engineering, Part A: Systems, 2017.