

CHAPTER 3

RESEARCH METHODOLOGY

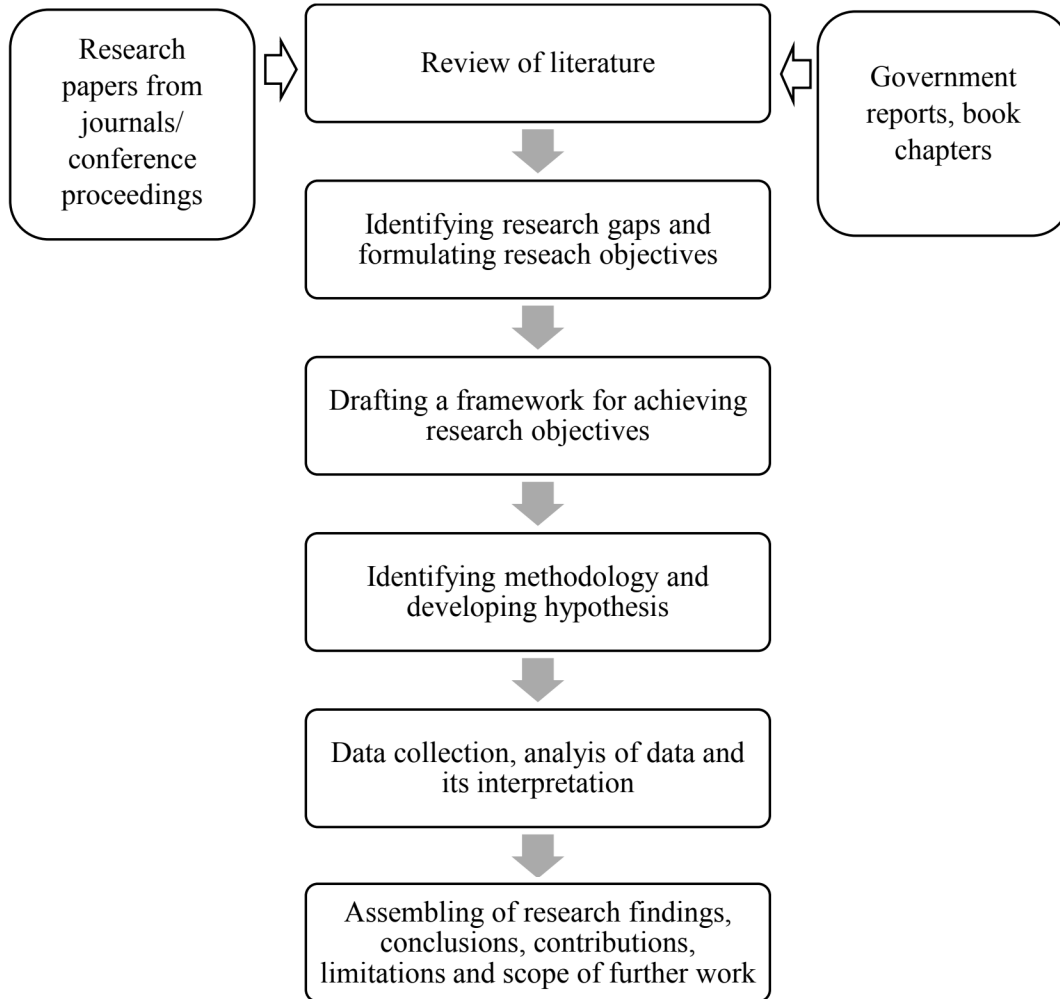
3.1 Introduction

As discussed in the previous chapter, based on a careful review of the existing literature, gaps in research were discovered. It is on the basis of this, that objectives of our study were identified. In this study we aim to examine travel behavior of individuals. We aim to explore the differences existing on the issue between men and women. This idea is expected to look upon from the varying perspectives of rural and urban regions in India. These frames the first objective of our thesis. In the second objective, we expect to investigate the differences arising in travel behavior between men and women due to differences in levels of socioeconomic status of individuals. Once the gender-based differences in travel behavior are made evident, this study also aims to propose certain policy measures which can aid in bridging the discovered gender gap in the issue. This formulates the third objective of our thesis. The contents of this chapter have been presented in the following manner. In the first section, we provide an outline of the research process of the thesis, followed by a section on discussion of the research framework developed under each objective of the study. The third section gives a brief overview of the research techniques employed for conducting the empirical work on the research objectives formulated. The last section concludes this chapter.

3.2 Research Process

The following figure depicts the research process adopted in the thesis:

Figure 1: Research Process



3.3 Development of Research Framework

As discussed in the previous chapters, after a careful analysis of existing literature, research gaps were identified based on which research objectives were developed. For every research objective, a research framework was outlined which acted as a basis for empirical investigation.

The research framework concerning each objective is as follows:

Objective 1: This objective aims to examine the gender differences in travel pattern in urban and rural settings in India. To achieve this, we capture area and gender-based differences in travel behavior in 27 states and eight union territories of India while also attempting to understand the role of socio-demographic factors of two Indian states with noticeable differences on travel choice using beta regression model. We also attempt to analyze this objective from the perspective of India's most populous state using OLS regression. This objective is then examined from the stance of a rural and urban area surveyed in Rajasthan using multinomial logit model.

Objective 2: The second objective of our thesis focusses on the role of income and socioeconomic factors on gender differences in travel pattern. To achieve this objective, surveyed data is analyzed using descriptive statistics complemented with binary logit model.

Objective 3: This objective targets at proposing policy suggestions for gender-sensitive transport planning by understanding gender differences in travel pattern. To accomplish this, we use univariate probit model, bivariate probit model and seemingly unrelated bivariate probit model along with descriptive statistics of the collected data.

3.4 An Outline of Research Design and Techniques Employed

3.4.1 Section-I

Overview

This section of the chapter gives a description of the research design and techniques employed for the first objective. We aim to understand area and gender-based differences in travel patterns in India. To achieve this objective, we divide the remainder of this section into four parts. The first part explains travel behavior differences between rural and urban areas and between men and women among 27 states and eight union territories of India. It then also tries

to examine the impact of certain socio-demographic factors on travel behavior of two selected states.

The second part of this section examines this objective of the thesis for one of the most populous states of India. The third and the fourth section explores this objective from the outlook of the surveyed locations of Rajasthan.

3.4.1.1 Part-I

A) Study areas

The first segment of this paper aims to establish the rural-urban and male-female differences in travel behavior. For this purpose, existing data on the issue of 27 Indian states eight union territories has been observed. Ladakh and Telangana have been excluded from the analysis due to data unavailability.

In the second section of this paper, we try to examine the factors influencing efficient travel behavior. For this a careful analysis has been made of the differences between states with respect to mode of transport use and its type and between efficient travel behavior displayed by them. We also attempted to take into consideration the existing demographic and sociodemographic differences between states, such as, the state's share of population in the total, their population densities, sex ratios, net state domestic products per capita, literacy rates, unemployment rates, poverty rates, infant mortality rates, etc. On the basis of these factors, we selected the states of Kerala and Uttar Pradesh. 14 districts of Kerala and 65 districts of Uttar Pradesh were studied. Due to data scarcity, the districts of Budaun, Jyotiba Phule Nagar, Kanshiram Nagar, Mahamaya Nagar, Ramabai Nagar, Sant Ravidas Nagar and Kanpur Dehat have been eliminated from the analysis.

B) Data sources and description of variables employed

The data on commuting patterns has been taken from 2011 Census of India titled ‘Other workers by distance from residence to place of work and mode of travel to place of work’ (Census of India, 2011g). This type of data was released for the very first time in 2016. It covers travel information of only those workers employed in activities other than agriculture. It enumerates total number of non-agricultural workers in a region using a specific mode within a specific distance category, separated by gender and area of residence. It is assumed that the respondents reported the main mode of travel used by them to travel the longest distance. Proportion of individuals displaying efficient travel behavior out of total number of individuals displaying efficient behavior is the dependent variable of our analysis. This variable depicts a fraction of those individuals who take up walking and cycling for short distances (up to 5 km) and those who take up public transport to travel longer distances (above 5 km) out of total number of individuals displaying the same in all the districts of both the selected states. Distance categories have been created on the basis of the study by Korzhenevych and Jain (2018).

‘State’ is a dummy variable which takes on a value of 1 if the state is Kerala and 0 if it is Uttar Pradesh.

‘GDVA pc’ is a continuous variable which depicts the gross district value added per capita of the studied districts of both the states. The data for this variable has been taken from the departments of Economics and Statistics of the respective states (Department of Economics and Statistics; Government of Kerala, 2011; Directorate of Economics and Statistics; Government of Uttar Pradesh, 2011).

‘Share of currently married individuals in total population’ is another explanatory variable of the study depicting the proportion of individuals who are currently married out of the total

population of individuals in the study area. This data has been taken from Census of India (2011d).

Another independent variable of the study is ‘share of working age individuals in total population’ which shows the proportion of individuals in the age group of 18-60 years out of the total population of individuals in the region. This data has been obtained from Census of India (2011j).

We also aimed to understand the impact of education on choice of travel behavior. For this, Census of India (2011c) data has been taken into consideration. This data accounts for the proportion of individuals having obtained an education less than matric out of the total population of individuals in an area.

‘Share of rural population in total’ discusses the proportion of rural population out of the total population of a district using (Census of India, 2011i, 2011l).

We also take into account the sex ratio of each state. ‘Share of females in total population’ uses Census of India (2011b, 2011a) data to discuss the influence of proportion of females in total population.

Population density which is another continuous explanatory variable in the study has been taken from Census of India (2011b, 2011a). It accounts for the number of individuals living per square kilometer in each district.

This study has the following limitations: (i) It examines the travel behavior of only a specific section of the total population, i.e., non-agriculture workers, which make up only about 42 per cent of the population (Census of India, 2011g). (ii) The data set that has been used for the analysis ignores the multimodal characteristic of certain trips. (iii) It does not calculate the combined effect of two or more socio-demographic factors on efficient travel behavior. (iv) It also does not assess the travel behavior in case of travel made for purposes other than work.

C) Methodology

To understand the existing disparities in different measures of travel behavior such as mode of transport and its type, distance travelled to, and a combination of these two, this study uses descriptive statistics such percentage and proportionate shares, and line graphs.

We then employ beta regression model to understand the influence of socio-demographic characteristics on proportion of efficient travel behavior displayed by regions. The idea behind developing this model is our variable of interest, i.e., efficient travel behavior, which is in proportions and the observations which are limited to an open interval (0, 1) (Ferrari and Cribari-Neto, 2004). The selected model uses the logit link function,

$$g(\mu) = \log \frac{\mu}{1 - \mu}$$

Beta regression model is a cross-sectional technique of modelling such data for which the observations are limited to the open interval (0, 1) (Ferrari and Cribari-Neto, 2004). Also, the data employed in this model is in continuous proportions (Douma and Weedon, 2019). It uses maximum likelihood method of estimation.

When using the logit link, the regression parameters of the model are interpretable in terms of the mean of the response. The conditional model for mean μ of the response y , given covariates X is assumed to be linear on logit transformed scale:

$$\text{logit}(\mu) = \log\left(\frac{\mu}{1 - \mu}\right) = \eta = X\beta$$

Where η is the linear predictor, β is a vector of parameters to be estimated and X is a matrix of covariate values. Using inverse logit function, the values from the linear predictor are transformed to obtain values which are interpretable on the scale of (0,1) observations.

$$\text{logit}^{-1}(\eta) = \frac{e^{\eta}}{1 + e^{\eta}}$$

3.4.1.2 Part-II

A) Data Sources

The data on home-to-work commuting have been taken from the table in the 2011 Census of India labeled ‘Other workers by distance from residence to place of work and mode of travel to place of work’ Census of India, (2011g). These data were released for the first time in 2016.

The data in this table covers information on only ‘other main workers.’ ‘Other workers’ are all workers who have been engaged in some kind of economic activity but are not cultivators, agricultural laborers, or employed in the household industry. These workers are referred to as non-agricultural workers in this paper. ‘Main workers’ are those workers who are employed for most of the reference period (i.e., 6 months or more) Census of India, (2011g).

This table from the Census contains information on the total number of trips undertaken in the rural and urban areas of the districts, segregated by choice of mode of transport, distance category, and gender.

The statistics for the total population, minorities, and literates were drawn from the Primary Census Abstract 2011 (Census of India, 2011h). The data for employment and the elderly were taken from Table B-1 titled ‘Main workers, marginal workers, non-workers and those marginal workers, non-workers seeking/available for work classified by age and sex’ (Census of India, 2011g).

This newly released Census data on commuting of non-agricultural workers from the place of residence to location of work in eight selected districts of Uttar Pradesh has been analyzed. Eight districts in Uttar Pradesh, namely, Agra, Allahabad, Ghaziabad, Jhansi, Kanpur, Lucknow, Meerut and Varanasi are selected. These eight districts not only are representatives

of the three major regions into which the state is divided, i.e., Western Uttar Pradesh, Eastern Uttar Pradesh and Central Uttar Pradesh, but also are those with highest levels of population and high rates of unemployment (Census of India, 2011m).

Data scarcity imposes the following limitations on the study: (a) it only explains commuting patterns within rural and urban areas, and fails to explain the same between the two areas. (b) Because this data was released for the first time, the study is found deficient in examining trends over time, making it unfit to predict future mobility patterns. (c) This study can be analyzed and disapproved on similar grounds as the census data, such as the underestimated employment values in the informal sector. (d) Other important attributes of commuting such as trip duration, trip cost, security aspects, and whether women travel with children and carry headloads, etc. are also not covered by the available data.

B) Methodology

Descriptive statistics

A summary statistic representing the overall pattern of home-to-work commuter trips is first presented by calculating the shares constituted by different distance categories and modes of travel, discriminating between rural and urban areas and between women and men.

Data on distance has been classified into the following categories on the basis of Korzhenevych and Jain (2018) : trips between 0 and 5 km are termed as 'short trips'; trips between 6 and 20 km are classified as 'intermediate trips'; and trips more than 20 km have been referred to as 'long trip.' Census data also has a category of 'no travel' which represents those workers who do not travel and work either in or nearby their place of residence.

Statistical analysis of the commuter data

This section is followed by descriptive statistics of commuting data. We then present a regression analysis with the focus on identifying the determinants of commuting patterns. The dependent variables used are shares of different modes of travel and distance categories.

As far as the identification of explanatory variables is concerned, existing studies conducted on India are used. In exploring the factors affecting urban mobility in India, Ahmad and Puppim de Oliveira (2016) observe the share of trips made with motorized modes to simultaneously increase with increasing income levels and population size and decreasing population density. NMT and public transport are more commonly used by households with low-income and groups who are considered to be comparatively socio-culturally disadvantaged. Rural-urban commuting increases with the increasing unemployment rate and increasing differences in urban and rural wages (Sharma and Chandrasekhar, 2014). In the city of Rajkot, the distance travelled was found to be positively related to income (Mahadevia and Advani, 2016). Elderly working people commute less than young working people (Sharma and Chandrasekhar, 2014). Therefore, population size, unemployment rate, population density, the share of minorities in the total population, the share of elderly people (more than 60 years of age) in the total population, and literacy rates are used as explanatory variables. Literacy rate is used as a proxy for average income. As in Korzhenevych and Jain (2018), it is assumed that people with lower levels of education get lower-paid jobs. In addition, dummies indicating the district and area type (rural or urban) have also been used as explanatory variables. Conventionally, literacy rates, the share of minorities and elderly is presumed to have a negative effect on private motorized vehicles and long trips. The coefficient of the dummy for area type is expected to distinguish commuting patterns between rural and urban areas.

Another category of the dependent variable used in the study is the indicator for 'predominance' of NMT and public transport (Banister and Gallent, 1998). Predominance NMT for short trips is expressed as the difference between the share of short (< 5 km) trips undertaken on foot or by bicycle and the share of similar trips undertaken in private motorized vehicles (moped, car, taxi). In a similar fashion, predominance of public transport for long (> 30 km) trips is explained as the difference between the shares of long trips by bus or train (public modes) and the share of such trips by private motorized vehicles (moped, car, taxi) (Korzhenevych and Jain, 2018).

We use ordinary least square (OLS) regressions to explain the relationship between dependent and explanatory variables.

Ordinary Least Squares (OLS) is a classical method first developed by Carl Friedrich Gauss in 1821. It is a least squares method used to estimate the values of unknown parameters in a linear regression model. This method of analysis estimates the relationship between a dependent variable and one or more independent variables.

Given the assumptions of the Classical Linear Regression Model (CLRM), the OLS estimates possess certain ideal properties which are contained in the Gauss-Markov theorem. An OLS estimator is said to be Best Linear Unbiased Estimator (BLUE) if the following hold true:

- It is a linear function of a random variable.
- It is unbiased in the sense that its expected value is equal to its true value.
- It is an efficient estimator in the sense that it has minimum variance in the class of all such linear unbiased estimators.

The relationship between explained and explanatory variables is contained in the following equation:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Where ε_i is the error term, and α , β are the true (but unobserved) parameters of the regression Y_i is the explained variable and X_i is the explanatory variable.

In this study, in the first part, Y_i is share of different modes, i.e., NMT, moped, car, bus for male and female commuters separately and X_i comprises urban dummy, minorities share, unemployment rate, population, population density, literacy rate, share of elderly and district dummy. In the next part, the three categories of trip lengths, i.e., short, intermediate and long for male and female commuters individually constitute Y_i and urban dummy, minorities share, unemployment rate, population, population density, literacy rate, share of elderly and district dummy are included in X_i . The third part determines Y_i as dominance NMT for short distance trips and dominance of public transport for long trips for male and female commuters and X_i as urban dummy, minorities share, unemployment rate, population, population density, literacy rate, share of elderly and district dummy.

3.4.1.3 Part-III

A) Introducing the study area

Rajasthan, located in northern India is the largest state by area in the country. It is the seventh largest by population having a population density of 200 square kilometres. The state is divided into 7 divisions and 33 districts. Jaipur division, which is one of the seven divisions, comprises of 5 districts including Jaipur and Jhunjhunu. The study area encompasses Jaipur city within Jaipur district and Sainipura village within Jhunjhunu district.

Rajasthan makes a part of this study because it is one of the largest states in India by area and population. The idea of understanding travel behaviour and mobility of individuals is better in a state which is not only large in size but also embodies a large number of people. On the contrary, in a small-sized state, inter-regional movement becomes frequent. Because the

understanding of mobility in this study is limited to intra-region, we thought it to be better to avoid such an issue. Within Rajasthan, Jaipur is not only a city with huge population but also the one with reported high level of crimes against women. This would facilitate a better understanding of the issue of gender differences in travel behaviour. The choice of the rural area is driven by the idea that any village which is in proximity to a city depicts travel behaviour which is very much influenced by the mere presence of the urban area. It is for this reason that a such a rural area is selected which neither very far-off from the urban area selected nor very close to any city.

The population of Rajasthan is 6,85,48,437 of which 30,73,350 are a part of Jaipur city and 1227 of Sainipura (Census of India 2011, 2011b) . The sex ratio of the state is 928, that of Jaipur city is 900 and of Sainipura is 851 (Census of India 2011, 2011b). Average literacy rate of the state is 66.1 per cent, of Jaipur city is 83.33 per cent and of Sainipura is 72.68 per cent (Census of India 2011, 2011b).

Jaipur which is the capital of Rajasthan is the largest city by area in the state. It is one of the top 10 most populous cities in India and has a population density of 6285.41 square kilometres (Smart Cities India, 2016). Approximately 10.62 per cent of the total population resides in slums (Census of India, 2011). Hinduism is a majority religion in the city with approximately 78 per cent followers (Census of India, 2011). It is one of the earliest planned cities of modern India and is a UNESCO World Heritage Site. It is a popular tourist destination in the country and forms a part of the West Golden Triangle. The economy of Jaipur is fuelled by tourism, gemstone cutting, manufacturing jewellery and luxury textiles and information technology.

The city buses in Jaipur are operated by Jaipur City Transport Services Limited of Rajasthan State Road Transport Corporation. The service operates more than 400 regular and low-floor buses. The total number of city buses, both government and private, in 2007 was 327 (Wilbur

Smith & MoUD, 2008). Jaipur Rapid Bus Transit System (BRTS) and 150 km 6 lanes ring road encircling the city is proposed to solve the problem of traffic. Though Jaipur Metro has commenced its operation in the city, our study finds limited use of the same. Usage of metro reduces carbon emission per person by a significant amount. Approximately 72 per cent of the roads in the city do not have a footpath. The remaining roads where footpaths existed were mostly parked with vehicles (Times News Network, 2019). The usage of bicycle is minimal in the city.

Sainipura is a medium-sized village in Chirawa tehsil of Jhunjhunu district in Rajasthan. This village is located in an area of approximately 3 square kilometres. There are a total of 199 families residing in the village. The average literacy rate of Sainipura is higher than that of Rajasthan. Approximately 55 per cent of the total population of the village is working population (Census of India, 2011b). The village has 'kutcha' roads and lack of public transport running inside. There is lack of efficient higher education and employment-giving institutions inside the village.

In India, approximately 81 per cent of the total automobile domestic sales in 2018-19 are of motorized two-wheelers (Society of Indian Automobile Manufacturers, 2018). We observe a somewhat similar trend in Jaipur and Sainipura where the use of motorized two-wheelers is abundant and frequent.

B) Data and methodology

For the purpose of sampling, one rural and one urban area from within the same division in Rajasthan is selected. Jaipur city represents the urban area and the rural area is represented by Sainipura. The sample, stratified by the proportion of sex ratio and rate of literacy, comprises of 600 individuals from the city and 150 individuals from the village. This study employs Slovin's formula to calculate sample size for the regions with 4 per cent and 7.5 per cent

sampling error for urban and rural areas, respectively. Each individual in the sample is in the age group of 18 to 60 years. From each surveyed individual, data on socioeconomic characteristics, asset ownership of the household and individual information was collected.

In the case of Jaipur, the study defines a trip as a one-way intra-city trip made by an individual between the age of 18 and 60 years, within the boundaries of the city. In the case of Sainipura, definition of the same trip extends to the boundaries of district Jhunjhunu and does not remain confined to the village alone because majority of the population travels outside the village daily given the unavailability of efficient required amenities. Access and egress trips are considered part of the main trip. The mode of transport used to travel the longest distance within the trip is considered the mode of the whole trip.

The individuals have been classified into six occupational groups. Housewives, students and unemployed comprise the non-working. Working people are those who are employed in any activity, agricultural or non-agricultural and comprise of self-employed, regular wage/salary earning and casual labor. We use Census of India (2011f) definition of work which defines it as participation in any economically productive activity with or without compensation, wages or profit. This participation can be physical and/or mental in nature, involves not only actual work but also effective supervision of work. It even includes part time help or unpaid work on farm, family enterprise or in any other economic activity.

This study has the following limitations: (i) the study does not differentiate between those who are self-employed, regular wage earning or are casual labors in agriculture and those in non-agriculture. (ii) It fails to evaluate the per capita trip rates in the city which help to capture the average of the total number of trips performed by an inhabitant in a day. (iii) The study also does not differentiate between different types of public transport used in the city, such as buses,

autorickshaws, etc. (iv) The study also does not differentiate between those who are drivers of motorized vehicles and those who are passengers.

3.4.1.4 Part-IV

A) About data

The primary data collected from the above-mentioned locations, i.e., Jaipur and Sainipura, has been further employed to understand the influence of socioeconomic characteristics on travel mode decisions and how these decisions differ on the basis of gender and area. For details about the study locations, refer 3.4.1.3 A.

For the purpose of sampling, one rural and one urban area from within the same division in Rajasthan is selected. The urban area is represented by Jaipur city and Sainipura represents the rural area. Stratified random sampling is used to select the respondents of the sample and the strata are created on the basis of sex ratio and literacy rate in the city. Within the strata, individuals are chosen randomly to minimize selection bias. The survey was conducted during March, 2020. The sample from the city comprises of 607 individuals and that from the village is composed of 151 individuals. The sample size of each of these regions is in proportion to their respective population. This study employs Slovin's formula to calculate sample size for the regions with 4 per cent and 7.5 per cent sampling error for urban and rural areas, respectively. It was assured that every surveyed individual is in the age group of 18 to 60 years. The data collected relates to information about socioeconomic characteristics, asset ownership of the household and individual information.

The study defines a trip as a one-way intra-city trip made by an individual between the age of 18 and 60 years, within the boundaries of the city. While this definition is more applicable for the urban area, for the rural area the boundaries of the district are assumed in the definition.

Access and egress trips are considered part of the main trip. The mode of transport used to travel the longest distance within the trip is considered the mode of the whole trip.

The explained variable of the study is the mode of transport that is opted for. In the urban area, this variable has six mode options, namely, walking, cycle, motorized two-wheelers, motorized four-wheelers, public transport and hired taxi. In the rural area, the number of options reduce to five as the use of hired taxi is almost absent in the region. In both the regions, the use of cycles appears to be minimal with women completely denying the option. Notably, the study does not differentiate between the use of a mode as a 'driver' or as a 'passenger'.

Socioeconomic variables such as level of education attained, work status and socioeconomic status of individuals constitute the explanatory variables of the study.

The variable of educational attainment has three categories. The first category is of individuals having no formal schooling or having studied till class 4, the second category comprises of those having completed schooling between class 5 and class 12, and those having attained education above class 12 comprise the third category.

The study takes into account Census of India (2011f) definition of work which defines it as participation in any economically productive activity with or without compensation, wages or profit. This definition of work also differentiates between workers and non-workers. The non-workers include unemployed, students, and housewives or home-makers who are attending to daily household chores. Workers, on the other hand, include those who are self-employed, earning regular wage/salary, working as main workers or casual labors in agricultural or non-agricultural activities.

To determine socioeconomic status, three categories of socioeconomic groups are determined adjusting monthly consumption expenditure for household members. The following categories are mutually exclusive and correspond to Q1 and Q3 values:

SEG1: This category comprises of individuals belonging to low socio-economic background. The monthly per capita consumption expenditure of people belonging to this group is less than or equal to INR 2500 in urban area and INR 1167 in rural area. In case of missing values for consumption expenditure, individuals whose household lives in houses which they have occupied without paying for any rent or own a kutchra house are adjusted in SEG1.

SEG2: This group corresponds to individuals belonging to middle socio-economic background where monthly per capita consumption expenditure is within the range of INR 2501 to INR 7000 in case of urban area and INR 1168 to INR 3333.333 in rural area. Individuals where the consumption expenditure data is missing, ownership of semi-pucca houses or where the household lives in rented houses are considered in SEG2.

SEG3: Individuals belonging to this group classify as those belonging to a high socio-economic background. The monthly per capita consumption expenditure of individuals in this group is above INR 7000 in urban area and above INR 3334 in rural area. For missing values, individuals whose household owns a pucca house are considered to be a part of SEG3.

Choice of mode of transport is considered to be significantly influenced by certain other factors such as, cost of travel (Vrtic *et al.*, 2010; Sharma, 2019) and whether or not the respondent owns a motorized vehicle. These variables are covered as control variables in the study.

This study has the following limitations: (i) it does not take into account the actual per capita income values for calculating the socio-economic categories. (ii) It fails to calculate the combined effect of two or more socioeconomic factors on choice of travel modes. (iii) It does not take into account the impact of travelling for different purposes and with different frequencies on choice of travel modes. Future studies should address these issues.

B) Methodology

The use of discrete choice models for travel mode choice determination has been observed in a number of previous studies (Li *et al.*, 2018; Liu, 2007; Nesheli *et al.*, 2016; Ng *et al.*, 2013; Omrani, 2015; Singh and Vasudevan, 2018; Srinivasan and Rogers, 2005; Szeto *et al.*, 2017). Logit models evaluate the probability of a certain event for a linear combination of variables. This study uses multinomial logit (MNL) framework to model the impacts of socioeconomic variables on choice of mode of transport of men and women in two different geographical locations. The dependent variable is a multiple response variable with six categories (walking, cycle, motorized two-wheeler, motorized four-wheeler, public transport and hired taxi) in urban area and five categories (walking, cycle, motorized two-wheeler, motorized four-wheeler and public transport) in rural area. Socioeconomic variables, which are the explanatory variables of the study, have been used as categorical variables. The MNL framework can be modelled as follows:

$$\ln \frac{P_{mode\ j|i}}{P_{walking|i}} = \beta_{mode\ j,walking} \cdot X_i$$

where $P_{mode\ j|i}$ denotes the probability that individual i will choose a particular mode j over walking, X_i is the vector of socioeconomic variables describing individual i and β displays the estimated coefficients associated with the vector X . Walking is taken as the reference alternative in the model.

A Multinomial Logit (MNL) model is a model used to predict the probabilities of several possible outcomes of a categorically distributed dependent variable, given a set of independent variables. This model is applicable when there are more than two categories of the dependent variable and these categories logistic regression cannot be ordered in any meaningful way. A MNL model is an extension of binary logistic regression in the way that it allows for more than

two categories of the outcome variable. Similar to a binary logit model, this model uses maximum likelihood estimation to calculate the probability of categorical membership. A MNL model has the following assumptions:

- Data are case-specific, i.e., each independent variable has a specific value for each case.
- The dependent variable cannot be predicted perfectly from the independent variable for any case.
- The different independent variables in the model need not necessarily be statistically independent of each other. Collinearity, however, is assumed to be relatively low.
- This model relies on the assumption of Independence of Irrelevant Alternatives (IIA). This assumption states that, for eg., the relative probabilities of taking a car or a bus to work does not change if a bicycle is added as an additional possibility.

3.4.2 Section-II

Overview

This section gives an overview of the research design and methodologies employed for the second objective of the thesis. In order to examine gender differences in travel behavior with respect to income groups, we split this section into three parts. The first part studies gender and class distinction in different measures of travel pattern. The second part examines the role of socioeconomic characteristics on gender differences in travel mode and the third part examines the same for purposes other than work.

Study area

Jaipur is the capital of north Indian state of Rajasthan, which is the country's largest state by area. Spread across an area of 484.64 sq. km (Smart Cities India, 2016), it is the largest city by area in the state. This city holds a place in the top ten populous cities of the country, with a

population of 30,73,350. A growth rate of 32.30 per cent in the population was observed in Census of India (2011) as against 2001. Approximately 83.33 per cent of the population of the city is considered to be literate and there are 900 females for every 1000 male (Census of India, 2011). Majority of the population of Jaipur follow Hinduism. This city is also one of the earliest planned cities of modern India. It is one of the popular tourist destinations in the country because of being a part of the West Golden Triangle tourist circuit and being a UNESCO World Heritage Site. The economy of the city is triggered by tourism, gemstone cutting, manufacturing jewellery and luxury textiles, and information technology.

Rajasthan makes a part of this study because it is one of the largest states in India by area and population. The idea of understanding travel behaviour and mobility of individuals is better in a state which is not only large in size but also embodies a large number of people. On the contrary, in a small-sized state, inter-regional movement becomes frequent. Because the understanding of mobility in this study is limited to intra-region, we thought it to be better to avoid such an issue. Within Rajasthan, Jaipur is not only a city with huge population but also the one with reported high level of crimes against women. This would facilitate a better understanding of the issue of gender differences in travel behaviour.

The Municipal Corporation of Jaipur is divided into eight zones encompassing 91 wards. The city buses are operated by Jaipur City Transport Services Limited of Rajasthan State Road Transport Corporation. The service operates more than 400 regular and low-floor buses. Wilbur Smith and MoUD (2008), report the total number of government and private city buses at 327 in 2007.

The problem of traffic in the city is proposed to be solved by the 150 km six lanes ring road encircling the city and Jaipur Rapid Bus Transit System (BRTS). Though Jaipur Metro has

commenced its operation, our study finds limited use of the same in the city. Usage of metro reduces carbon emission per person by a significant amount.

Approximately 72 per cent of the roads in the city do not have a footpath and the remaining roads where footpaths existed were mostly parked with vehicles (Times News Network, 2019). The usage of bicycles is minimal in the city.

In India, approximately 81 per cent of the total automobile domestic sales in 2018-19 are of motorized two-wheelers (Society of Indian Automobile Manufacturers, 2018). A report by Jaipur Metro Rail Corporation Ltd. (2018) finds that majority of the population, approximately 32 per cent uses two-wheelers in Jaipur. We observe a similar trend in the city where the use of motorized two-wheelers is ample and frequent.

3.4.2.1 Part-I

A) Data and Methodology

The sample comprises of 600 individuals stratified on the basis of sex ratio and average literacy rate in the city. The survey was conducted during March, 2020. This study employs Slovin's formula to calculate sample size for the region with 4 per cent sampling error. Each individual who is a part of the sample is between the age of 18 and 60 years. Data on socioeconomic characteristics, asset ownership of the household and individual information was collected from each individual.

The study defines a trip as a one-way intra-city trip made within the boundaries of the city. Access and egress trips are considered part of the main trip. The mode of transport used to travel the longest distance within the trip is considered the mode of the whole trip.

We also differentiate between the usage of motorized vehicles as a driver and as a passenger.

We categorize the individuals comprising the sample into three socio-economic groups using monthly per capita consumption expenditure. Three mutually exclusive categories, namely SEG1, SEG2 and SEG3 have been created using cut-off points for monthly per capita consumption expenditure as Q1 and Q3 values. In cases where the data for consumption expenditure was missing, which are approximately 6 per cent of the database, we use house ownership and house type as a proxy. We define the socio-economic groups as follows:

SEG1: This category represents individuals belonging to low socio-economic background with monthly per capita consumption expenditure \leq INR 2500. In case of missing values for consumption expenditure, individuals whose household lives in houses which they have occupied without paying for any rent or own a kutchra house are adjusted in SEG1.

SEG2: This group depicts middle socio-economic background with monthly per capita consumption expenditure between INR 2501-INR 7000. Ownership of semi-pucca houses or where the household lives in rented houses are considered in SEG2, in case of missing data.

SEG3: This category represents high socio-economic background of individuals with monthly per capita consumption expenditure $>$ INR 7000. For missing values, individuals whose household owns a pucca house are considered to be a part of SEG3.

This study has the following limitations: (i) it does not take into account the actual per capita income values for calculating the socio-economic categories. (ii) It fails to evaluate the per capita trip rates in the city which capture the average of total number of trips performed by an inhabitant in a day. (iii) The study also does not differentiate between different types of public transport used in the city, such as buses, autorickshaws, etc.

3.4.2.2 Part-II

A) Data

The study uses primary data collected from 600 individuals in Jaipur city. Stratified random sampling is used to select the respondents of the sample and the strata are created on the basis of sex ratio and literacy rate in the city. Within the strata, individuals are chosen randomly to minimize selection bias. Each individual who is a part of the sample is between the age of 18 and 60 years. This study employs Slovin's formula to calculate sample size for the region with 4 per cent sampling error. The questionnaire, which was used as a tool to collect data, is divided into two sections. The first section collects information related to respondent's socioeconomic characteristics while the second section intends to compute the travel behavior of individuals. A pilot study was conducted in the city with 60 respondents. After suggestions and modifications, final questionnaire was prepared. The survey was conducted during March, 2020, before the COVID-19 pandemic.

The study defines a trip as a one-way intra-city trip made by an individual between the age of 18 and 60 years, within the boundaries of the city. Access and egress trips are considered part of the main trip. The mode of transport used to travel the longest distance within the trip is considered the mode of the whole trip.

We categorize the modes of transport used by the individuals into sustainable modes such as walking, cycling or using public transport; and unsustainable modes such as motorized two-wheelers, motorized four-wheelers or hired taxis. The reason why we use the term 'sustainable' is the existing differences in carbon footprint of the two types. The modes which emit the lowest carbon per person are considered to be the ones which compromise the least with the needs of the future generations.

The socioeconomic factors, which are also the explanatory variables of the study, are the factors on which the data is further divided. These include work status of individuals, their socioeconomic status and educational attainment. The data in our study is also analyzed separately for the two groups of gender to clearly highlight the different impacts of these socioeconomic factors on travel mode choice of men and women.

We use Census of India (2011f) definition of work which defines it as participation in any economically productive activity with or without compensation, wages or profit. This definition of work also differentiates between workers and non-workers. The non-workers include unemployed, students, and housewives or homemakers who are attending to daily household chores. Workers, on the other hand, include those who are self-employed in non-agricultural activities, earning regular wage/salary or working as casual labors.

Individuals are also divided based on the education attained by them. Those having no formal schooling or having studied till class 4 make the first category, individuals having studied till class 5 to class 12 comprise the second category and those who have attained education above class 12 comprise the third category.

The socio-economic groups are defined as follows:

SEG1: We categorize the individuals belonging to low socio-economic background in this group. The monthly per capita consumption expenditure of people belonging to this group is less than or equal to INR 2500. In case of missing values for consumption expenditure, individuals whose household lives in houses which they have occupied without paying for any rent or own a kutchra house are adjusted in SEG1.

SEG2: This group corresponds to individuals belonging to middle socio-economic background where monthly per capita consumption expenditure is within the range of INR 2501 to INR

7000. Individuals where the consumption expenditure data is missing, ownership of semi-pucca houses or where the household lives in rented houses are considered in SEG2.

SEG3: Individuals belonging to this group classify as those belonging to a high socio-economic background. The monthly per capita consumption expenditure of individuals in this group is above INR 7000. For missing values, individuals whose household owns a pucca house are considered to be a part of SEG3.

Choice of mode of transport is considered to be significantly influenced by certain other factors such as, cost of travel (Vrtic *et al.*, 2010; Sharma, 2019) and some sociodemographic factors like age (Ji *et al.*, 2017; Sharma, 2019; Singh and Vasudevan, 2018) and marital status. These variables are incorporated as control variables in the study.

This study has the following limitations: (i) it does not take into account the actual per capita income values for calculating the socio-economic categories. (ii) It fails to calculate the combined effect of two or more socioeconomic factors on choice of travel modes. (iii) It does not take into account the impact of distance or purpose of travelling on choice of travel modes.

B) Methodology

Several studies have used discrete choice models for travel mode choice determination (Li *et al.*, 2018; Liu, 2007; Nesheli *et al.*, 2016; Ng *et al.*, 2013; Omrani, 2015; Singh and Vasudevan, 2018; Srinivasan and Rogers, 2005; Szeto *et al.*, 2017). Logit models estimate the probability of a certain event for a linear combination of variables. This study uses logit model to identify the impact of socioeconomic variables on sustainable mode choice of men and women separately. The dependent variable is a binary response variable having sustainable mode choice (walking/ cycle/ public transport) and unsustainable mode choice (motorized two-wheelers/ motorized four-wheelers/ hired taxi) as the two categories. For the respondent's travel decision of sustainable mode choice, this model takes a value of 1 with a probability of

p and when the choice is made for an unsustainable mode, a value of 0 with probability 1-p is assumed. Socioeconomic variables, which are the explanatory variables of the study, have been used as categorical variables. The equation of the binary logistic model can be as follows:

$$\ln\left(\frac{S}{1-S}\right) = \beta_0 + \sum_{i=1}^n (\beta_i X_i)$$

where S is the probability of outcome variable equal to 1, i.e., choosing a sustainable mode of transport for frequent travelling purposes, β_0 is the model constant, X_i is a categorical explanatory variable, and β_i is the coefficient of regression which is estimated.

A binary logit model is a model that uses a logistic function, which is the cumulative distribution function of logistic distribution, to model a binary dependent variable. The goal of this model is to estimate probability p for a given linear combination of independent variables. The dependent variable of logistic regression follows Bernoulli distribution, takes two possible values labelled 0 and 1. The logistic function is a sigmoid function, which takes any real input t , and outputs a value between zero and one.

3.4.2.3 Part-III

A) Data

For the purpose of sampling, sample stratified on the basis of sex ratio and literacy rate, is selected and each individual who is a part of the sample is between the age of 18 and 60 years. From each surveyed individual, data on socioeconomic characteristics, asset ownership of the household and individual information was collected. The survey was conducted during March, 2020.

The study defines a trip as a one-way intra-city trip made by an individual between the age of 18 and 60 years, within the boundaries of the city. Access and egress trips are considered part

of the main trip. The mode of transport used to travel the longest distance within the trip is considered the mode of the whole trip.

The modes of transport used by the individuals are categorized into two broad categories: less-polluting modes and more-polluting modes. Less-polluting modes include those modes of transport which cause less damage to the environment, produce a lower carbon footprint per person and contribute the least to the pollution levels in the city, for e.g., walking, cycle and public transport. The category of more-polluting modes is inclusive of all other modes which produce a higher carbon footprint per person and cause a greater damage to the environment. This category includes modes such as motorized two-wheelers, motorized four-wheelers or hired taxis.

Individuals in the sample are further divided on the basis of different socio-demographic factors, which also frame the explanatory variables of the study. The socio-demographic variables which are a part of the study include sex of individuals, their age, their level of educational attainment, their work status and the socioeconomic status to which they belong.

As for sex, only two categories of male and female are taken into consideration. The three age groups into which individuals are divided are 18-25 years which is the youngest group of individuals and includes the majority of students. The second group is of individuals who are of 26-39 years of age. They are neither the youngest nor the eldest and they comprise of a small percentage of students. The third category is of individuals belonging to the eldest age group. These are those individuals who are in the age of 40-59 years.

This study uses Census of India (2011f) definition of work where work is defined as participation in any economically productive activity with or without compensation, wages or profit. This definition of work also differentiates between workers and non-workers. The non-workers include unemployed, students, and housewives or home-makers who are attending to

daily household chores. Workers, on the other hand, include those who are self-employed in non-agricultural activities, earning regular wage/salary or working as casual labors.

Individuals are also classified on the basis of education attained by them. Those having no formal schooling or having studied till class 4 make the first category, individuals having studied till class 5 to class 12 comprise the second category and those who have attained education above class 12 comprise the third category.

Another explanatory variable in the study is the socio-economic status to which individuals belong. The three categories of socio-economic status are defined as follows:

SEG1: We group individuals belonging to low socio-economic background in this category. The monthly per capita consumption expenditure of people belonging to this group is less than or equal to INR 2500. In case of missing values for consumption expenditure, individuals whose household lives in houses which they have occupied without paying for any rent or own a kutchra house are adjusted in SEG1.

SEG2: This group corresponds to individuals belonging to middle socio-economic background where monthly per capita consumption expenditure is within the range of INR 2501 to INR 7000. Individuals where the consumption expenditure data is missing, ownership of semi-pucca houses or where the household lives in rented houses are considered in SEG2.

SEG3: Individuals belonging to this group classify as those belonging to a high socio-economic background. The monthly per capita consumption expenditure of individuals in this group is above INR 7000. For missing values, individuals whose household owns a pucca house are considered to be a part of SEG3.

While the primary focus of this study is to examine the role that sociodemographic factors play in choice of the type of mode of transport, one important variable which is expected to affect

this decision is cost of travel (Vrtic *et al.*, 2010; Sharma, 2019). Therefore, cost of travel for corresponding non-work purpose is assumed as a control variable in this study.

This study has the following limitations: (i) it does not take into account the actual per capita income values for calculating the socio-economic categories. (ii) It fails to calculate the combined effect of two or more socioeconomic factors on choice of travel modes. (iii) It does not examine the choice of mode of transport for different non-work purposes, separately.

B) Methodology

This study uses logit model to identify the impact of socio-demographic variables on choice of less-polluting modes of transport. 3.4.2.1 A explains logit models. The dependent variable is a binary response variable having less-polluting mode choice (walking/ cycle/ public transport) and more-polluting mode choice (motorized two-wheelers/ motorized four-wheelers/ hired taxi) as the two categories. For the respondent's travel decision of less-polluting mode choice, this model takes a value of 1 with a probability of p and when the choice is made for a more-polluting mode, a value of 0 with probability $1-p$ is assumed. Socio-demographic variables, which are the explanatory variables of the study, have been used as categorical variables. The equation of the binary logistic model can be as follows:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \sum_{i=1}^n (\beta_i X_i)$$

where P is the probability of outcome variable equal to 1, i.e., choosing a less-polluting mode of transport for travelling purposes other than work, β_0 is the model constant, X_i is a categorical explanatory variable, and β_i is the coefficient of regression which is estimated.

3.4.3 Section -III

Overview

This section of the chapter gives a description of the data and methodology for the third objective of our thesis. This section has been partitioned into two. The first part outlines the gap in travel behavior based on gender and then proposes policy suggestions for gender-sensitive transport planning and the opinion of public on these proposed measures. The second part of this section aims to understand the role of threat to safety in trading-off public transport with private modes.

Study area

Jaipur, capital of the north Indian state of Rajasthan, is one of the largest and the most populous cities of the country with an average literacy rate of 83.33 per cent and sex ratio of 900 (Census of India, 2011e). There are various zones and wards into which Jaipur Municipal Corporation is split up. About 10.62 per cent of the city's population lives in slums (Census of India, 2011e). While some wards have no slums, others have over 75 per cent of the population residing in them (Jawaid *et al.*, 2017).

The transportation system of the city is primarily road-based. More than 400 low-floor and regular city buses are operated by Jaipur City Transport Services Limited. In 2007, the total number of private and government city buses were 327 (Wilbur Smith and MoUD, 2008). Other than the service provided by JCTSL, auto-rickshaws, mini buses run by private operators, cycle-rickshaws, tempos and private vehicles also form a part of city's transportation infrastructure (Jawaid *et al.*, 2017). In 2018-19, more than 80 per cent of the total automobile domestic sales in India are of motorized two-wheelers (Society of Indian Automobile Manufacturers, 2018). A similar trend is observed in Jaipur where the use of motorized two-wheelers is ample and frequent.

The city faces the issue of traffic congestion and poor road condition (Dadhich, Goyal and Dadhich, 2018). Jaipur Rapid Bus Transit System (BRTS) and the six lanes ring road of about 150 km surrounding the city is proposed to solve the issue of traffic. Usage of metro lowers carbon emission per person by a notable amount. However, it is still in the nascent stage in the city with only limited usage.

While more than 70 per cent of the roads in the city are devoid of footpaths, others are largely parked with vehicles (Times News Network, 2019). Bicycle use is also minimal in the city.

3.4.3.1 Part-I

A) About data

This study is conducted in Jaipur city of Jaipur district of Rajasthan. Stratified random sampling technique is used and the data are stratified taking overall literacy rate and sex ratio of the city into account. 614 individuals comprise the sample and every respondent is between 18 and 60 years of age. This study employs Slovin's formula to calculate sample size for the region with 4 per cent sampling error. Information relating to asset ownership, household's socioeconomic characteristics and individual's travel behavior is collected. The survey was conducted during March, 2020.

This study interprets a trip as a one-way intra-city trip made within the city boundaries. This definition includes both access and egress trips. That mode of transport which is used to cover the longest distance within the entire trip is regarded the mode of the whole trip.

This study explains two binary response variables. The first one examines how long a person travels to, taking a value of 1 for short distances of less than 5 km and 0 for otherwise. The second one assesses the type of mode of transport that is used to travel this distance. This variable takes a value of 1 if non-motorized modes or modes which emit less carbon per person,

viz., public transport, walking and cycling are used and takes a value of 0 if the modes used are motorized, like motorized two-wheelers and motorized four-wheelers.

These dependent variables are explained by an independent variable, sex of the respondent. The set of variables which are expected to affect the two explanatory variables are accounted as control variables in the study. These include cost of travelling incurred, frequency of travelling, purpose of travelling, ownership of number of motorized two-wheelers and the socioeconomic status of respondents.

Sex is a binary response explanatory variable in the study which takes a value of 1 for male and 0 for female.

Corresponding roughly to Q1 and Q3 values, three categories of cost variable are created. Individuals which incur no cost of travelling make up the first group, those that incur a comparatively lower cost of less than or equal to INR 20 make up the second group, and individuals incurring high cost of more than INR 20 make up the third group of cost variable.

Frequency of travelling, which is another independent variable in the study, is binary variable. If a person travels frequently, this variable takes a value of 1, in the other case it takes a value of 0. Frequent travelling is determined on the number of trips made within a week. If an individual travels five times or more within a week, he is assumed to be travelling frequently.

Another explanatory variable of the study is purpose of travelling. If travel is made for work purposes, this binary independent variable takes a value of 1, otherwise it takes a value of 0.

The number of motorized two-wheelers that is owned by a household is also assumed to affect the choice of type of mode taken and the distance travelled. It is a continuous variable.

Regulating the monthly per capita consumption expenditure values, three mutually exclusive categories, corresponding to Q1 and Q3 values are created. This variable, regarded as socioeconomic status of an individual, is another explanatory in the study.

SEG1: This category is of individuals with low socio-economic background. Those who are a part of this group have monthly per capita consumption expenditure less than or equal to INR 2500. For those individuals whose consumption expenditure values are missing, living in houses occupied rent-free or owning a kutchra house are adjusted in SEG1.

SEG2: This category correlates to respondents with middle socio-economic background. The monthly per capita consumption expenditure of the respondents in this category is in the range of INR 2501 to INR 7000. For missing values of consumption expenditure, individuals living in rented houses or owning semi-pucca houses accounted in SEG2.

SEG3: This classification is of individuals belonging to a high socio-economic background. Respondents with monthly per capita consumption expenditure above INR 7000 frame a part of this group. In case of missing values, individuals whose household owns a pucca house are regarded as a part of SEG3.

This study has the hereunder limitations: (i) Socio-economic status categories that are created are calculated without considering the actual per capita income values. (ii) It does not account for the combined effect of two or more explanatory variables on both choice of type of mode of transport and the distance covered. (iii) It does not take into account the impact of ownership of motorized four-wheelers and driving license. (iv) It also is unsuccessful at taking into account the impact of other sociodemographic factors, like, the level of education attained, occupation, marital status, age, etc. of the respondents. Future studies should address these issues.

B) Methodology

This study first investigates if any relationship exists between dependent and independent variables using chi-squared test of association and univariate probit regression. Then, bivariate probit models and seemingly unrelated bivariate probit models are made use of.

Before fitting any model, chi-squared test of association is used for checking existence of relationship between the response and explanatory variables. A univariate probit model is then fitted for distance travelled and mode used, individually, to check the impact of each independent variable on the dependent variable, when fitted with other independent variables. This is done keeping in mind that some variables which might be insignificant from test of association, might be significant when fitted together with other explanatory variables. Chi-squared test and univariate probit model show if there exists any kind of relationship between the regressed variables. Further, we jointly model the two dependent variables using first a bivariate probit model and then a seemingly unrelated bivariate probit model. It is then that we choose the model that best fits the data. In seemingly unrelated bivariate probit model, only significant variables are included for both dependent variables.

(i) Univariate Probit Model

A probit model is usually accredited to Chester Bliss (Bliss, 1934). A probit model is a regression model where the dependent variable can take only two possible outcomes. The purpose of this model is to estimate the probability that an observation with specific characteristics will fall into a specific category. This model is a popular binary response model and provides solution to the same set of problems as a logistic regression model. It uses maximum likelihood method of estimation. This model takes the following form:

$$\Pr(Y = 1 | X) = \Phi(X^T \beta)$$

where Y is a binary distance variable that takes a value of 1 for short distances of less than 5 km and 0 for otherwise. X represents a vector of explanatory binary variable sex and other control variables, Pr denotes probability, Φ is Cumulative Distribution Function (CDF) of the standard normal distribution. The parameters β are typically estimated by maximum likelihood.

Next, Y becomes a binary mode choice variable which takes a value of 1 if non-motorized modes or modes which emit less carbon per person, viz., public transport, walking and cycling are used and takes a value of 0 if the modes used are motorized, like motorized two-wheelers and motorized four-wheelers. X represents a vector of explanatory binary variable sex and other control variables, Pr denotes probability, Φ is Cumulative Distribution Function (CDF) of the standard normal distribution. The parameters β are typically estimated by maximum likelihood.

(ii) Bivariate probit model

We assume that Y_{1i}^* and Y_{2i}^* are two latent variables. Y_{1i}^* measures how far an individual travels, i.e., if it is near or far and Y_{2i}^* measures the type of mode of transport that the individual makes use of. Because these two latent variables cannot be directly observed; the two-equation model is specified as:

$$Y_{1i}^* = \beta_1 X_{1i} + \varepsilon_{1i}$$

$$Y_{1i}^* = \begin{cases} 1 & \text{if } Y_{1i}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where X_{1i} indicates observed explanatory and control variables explaining choice of distance category travelled, β_1 denotes the parameter associated to each explanatory and control

variable and ε_{1i} refers to the random error term. Similarly, choice of type of mode of transport is measured as:

$$Y_{2i}^* = \beta_2 X_{2i} + \varepsilon_{2i}$$

$$Y_{2i}^* = \begin{cases} 1 & \text{if } Y_{2i}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where X_{2i} indicates observed independent variables explaining the type of mode of transport that is chosen, β_2 denotes the parameter linked to each explanatory and control variable and ε_{2i} refers to the random error term.

It is worth noting that the two dependent variables are explained by the same exogenous variables. On that account, the error terms of the two models are dependent and distributed as a bivariate standard normal, such that:

$$E(\varepsilon_{1i}) = E(\varepsilon_{2i}) = 0$$

$$\text{var}(\varepsilon_{1i}) = \text{var}(\varepsilon_{2i}) = 1$$

$$\rho = \text{cov}(\varepsilon_{1i}, \varepsilon_{2i})$$

If the error terms of both the equations are affected by analogous components, $\varepsilon_{ji} = \mu_i + \eta_{ji}$, then albeit they are expected to be normally distributed, they will be dependent on the value of μ_i . We propose a Wald test to identify if the two models must be estimated jointly, with null hypothesis $\rho=0$.

While in the binary probit model, there is only one binary dependent variable Y and therefore only one latent variable Y^* is used. However, in bivariate probit models there are two binary dependent variables and Y_1 and Y_2 implying two latent variables Y_1^* and Y_2^* . It is then assumed that variable that is observed takes on the value 1 if and only if its underlying continuous latent variable takes on a positive value:

$$Y_1 = \begin{cases} 1 & \text{if } Y_1^* > 0, \\ 0 & \text{otherwise} \end{cases}$$

$$Y_2 = \begin{cases} 1 & \text{if } Y_2^* > 0, \\ 0 & \text{otherwise} \end{cases}$$

With

$$\begin{cases} Y_1^* = X_1\beta_1 + \varepsilon_1 \\ Y_2^* = X_2\beta_2 + \varepsilon_2 \end{cases}$$

And

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} | X \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

(iii) Seemingly unrelated bivariate probit model

Seemingly Unrelated Regression Equations (SURE) is model which is a generalization of linear regression model that comprises of various equations with each equation having its own explained variable and a set of explanatory variables. It is called ‘seemingly unrelated’ because each equation is a linear regression on its own and can be interpreted individually. Certain researchers are of the view that the term ‘seemingly related’ would be more apt given the fact that the error terms are supposed to be correlated across equations.

This study considers the application of seemingly unrelated bivariate probit model which is applicable when two equations are to be jointly estimated and one of the independent variables of another equation is the dependent variable of the previous one.

The latent variable for this model is as follows:

$$Y_{1i}^* = \beta_1 X_{1i} + \varepsilon_{1i}$$

$$Y_{2i}^* = \beta_2 X_{2i} + \varepsilon_{2i}$$

The estimation procedure for this model is the same as the bivariate probit model, other than incorporating yY_1 .

(iv) Descriptive statistics for responses to suggested policy measures

Once it is confirmed that there exist gender differences in travel behavior, responses of individuals in the sample to most likely policy changes are recorded. Descriptive statistics, such as column graphs and pie charts, are used for this purpose.

3.4.3.2 Part-II

A) About data

For sampling, area of Jaipur city in Jaipur district of Rajasthan is chosen. The sample is stratified taking into consideration the overall literacy rate and sex ratio of the city. Every individual who frames a part of the sample belongs to the age group of 18 to 60 years and the total sample comprises of 600 individuals. This study employs Slovin's formula to calculate sample size for the region with 4 per cent sampling error. The collected data relates to information about ownership of assets, socioeconomic characteristics of the household and individual information. The survey was conducted during March, 2020.

A trip, as defined in the study, is a one-way intra-city trip made by individuals in the specified age group, within the boundaries of the city. Access and egress trips form a fraction of the main trip. That mode of transport which is used to cover the longest distance within the entire trip is regarded the mode of the whole trip.

There are two explained variables in the study, both being binary response variables. The first dependent variable studies if public transport is used for frequent traveling purposes. This variable takes a value of 1 if public transport is used by an individual and 0 otherwise. The second dependent variable examines the reason why the individuals in the sample would prefer using a private mode of transport over a public transport, both as a driver or as a passenger.

This variable takes a value of 1 if safety is the reason why they would make that trade-off and 0 if they make that decision due to any other reason, such as easy availability of private mode, rude and cheating staff, insufficient and irregular services of public transport, etc.

Socio-demographic variables, such as age, marital status, gender, work status, level of educational attainment and socioeconomic status of individuals comprise the explanatory variables of the study.

Gender is a binary response independent variable in the study which takes a value of 0 for male and 1 for female.

Age variable is divided into three categories. Individuals within the age group of 18-25 years make the first category, those between 26-39 years are the second category and the eldest group of 40-59 years make the third category.

Marital status is another independent variable of the study which assumes two categories: 1 for married individuals and 0 for others (single/divorced or separated/widowed).

Education variable, in our study, assumes three categories. The first is of individuals having minimal or no formal schooling or those having studied till standard 4, the second comprises of those having completed schooling between standard 5 and standard 12, and the third category of those having attained education above standard 12.

Census of India (2011f) defines work as participation in any economically productive activity with or without compensation, wages or profit. Using this definition, work status, which is another independent variable in the study, is taken into account. It also differentiates between non-workers and workers. Non-workers include students, unemployed, and home-makers or housewives who are handling daily household chores. On the other hand, workers include the ones who are earning regular wage/salary, are self-employed, working as main workers or casual labors in agricultural or non-agricultural activities.

Adjusting monthly household consumption expenditure for each member in the house, we create three categories of another socio-demographic independent variable in the study, socioeconomic status. The following categories are mutually exclusive and correspond to Q1 and Q3 values:

SEG1: This category comprises of those associated with low socio-economic background. Individuals who are a part of this group have monthly per capita consumption expenditure less than or equal to INR 2500. For those individuals whose consumption expenditure values are missing, living in houses occupied rent-free or owning a kutchra house are adjusted in SEG1.

SEG2: This category is of those belonging to middle socio-economic background. The monthly per capita consumption expenditure of those within this group is in the range of INR 2501 to INR 7000. For missing values of consumption expenditure, living in rented houses or owning semi-pucca houses is considered in SEG2.

SEG3: Those corresponding to this bracket classify as the ones belonging to a high socio-economic background. Individuals in this group have monthly per capita consumption expenditure above INR 7000. In case of missing values, individuals whose household owns a pucca house are regarded as a part of SEG3.

This study has the hereunder limitations: (i) Categories of socio-economic status are calculated without considering the actual per capita income values. (ii) It does not calculate the combined effect of two or more socio-demographic factors on both use of and rationale for not preferring public transport. (iii) It does not take into account the impact of travelling for different purposes for different distance categories and frequencies on choice of opting for public transport. (iv) This study primarily highlights the role of safety in accessing public transport and fails to provide equal weightage to other important factors, such as irregularity of services, rude and cheating staff, etc. Future studies should address these issues.

B) Methodology

This study first checks for existence of a relationship between explained and explanatory variables using chi-squared test of association and univariate probit models. Finally, bivariate probit models and seemingly unrelated bivariate probit models are used.

Before fitting any model, chi-squared test of association is used for checking existence of relationship between the response and explanatory variables. A univariate probit model is then fitted for the two response variables, individually, to check the impact of each independent variable on the dependent variable, when fitted with other independent variables. This is done keeping in mind that some variables which might be insignificant from test of association, might be significant when fitted together with other explanatory variables. Chi-squared test and univariate probit model show if there exists any kind of relationship between the regressed variables. Further, we jointly model the two dependent variables using first a bivariate probit model and then a seemingly unrelated bivariate probit model. It is then that we choose the model that best fits the data. In seemingly unrelated bivariate probit model, only significant variables are included for both dependent variables.

(i) Univariate Probit Model

3.4.3.1 B explains univariate probit models. This model takes the following form:

$$\Pr(Y = 1 \mid X) = \Phi (X^T \beta)$$

where Y is a binary variable that takes a value of 1 if public transport is used by an individual and 0 otherwise. X represents a vector of sociodemographic factors which are the explanatory variables of the study, Pr denotes probability, Φ is Cumulative Distribution Function (CDF) of

the standard normal distribution. The parameters β are typically estimated by maximum likelihood.

In the next case, Y is that binary dependent variable examines the reason why the individuals in the sample would prefer using a private mode of transport over a public transport, both as a driver or as a passenger. This variable takes a value of 1 if safety is the reason why they would make that trade-off and 0 if they make that decision due to any other reason. X represents a vector of sociodemographic factors which are the explanatory variables of the study, Pr denotes probability, Φ is Cumulative Distribution Function (CDF) of the standard normal distribution. The parameters β are typically estimated by maximum likelihood.

(ii) Bivariate Probit Model

Let us assume that Y_{1i}^* and Y_{2i}^* are two latent variables. Y_{1i}^* measures the usefulness of public transport to an individual and Y_{2i}^* measures safety as a reason for not preferring public transport over private. These two latent variables cannot be directly observed; therefore, we specify the two-equation model as:

$$Y_{1i}^* = \beta_1 X_{1i} + \varepsilon_{1i}$$

$$Y_{1i}^* = \begin{cases} 1 & \text{if } Y_{1i}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where X_{1i} indicates observed independent variables explaining usefulness of public transport to an individual, β_1 denotes the parameter associated to each explanatory variable and ε_{1i} refers to the random error term. Similarly, safety as a reason for not preferring public transport over private is measured as:

$$Y_{2i}^* = \beta_2 X_{2i} + \varepsilon_{2i}$$

$$Y_{2i}^* = \begin{cases} 1 & \text{if } Y_{2i}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where X_{2i} indicates observed independent variables explaining insecurity as reason for not preferring public transport over private vehicle, β_2 denotes the parameter linked to each explanatory variable and ε_{2i} refers to the random error term.

We observe that the two variables are explained by the same exogenous variables, therefore, the error terms of the two models are dependent and distributed as a bivariate standard normal, such that:

$$E(\varepsilon_{1i}) = E(\varepsilon_{2i}) = 0$$

$$\text{var}(\varepsilon_{1i}) = \text{var}(\varepsilon_{2i}) = 1$$

$$\rho = \text{cov}(\varepsilon_{1i}, \varepsilon_{2i})$$

If the error terms of both the equations are influenced by analogous components, $\varepsilon_{ji} = \mu_i + \eta_{ji}$, then although they are expected to be normally distributed, they will be dependent on the value of μ_i . We propose a Wald test to check if the two models must be estimated jointly, with null hypothesis $\rho=0$.

While in the binary probit model, there is only one binary dependent variable Y and therefore only one latent variable Y^* is used. However, in bivariate probit models there are two binary dependent variables and Y_1 and Y_2 implying two latent variables Y_1^* and Y_2^* . It is then assumed that variable that is observed takes on the value 1 if and only if its underlying continuous latent variable takes on a positive value:

$$Y_1 = \begin{cases} 1 & \text{if } Y_1^* > 0, \\ 0 & \text{otherwise} \end{cases}$$

$$Y_2 = \begin{cases} 1 & \text{if } Y_2^* > 0, \\ 0 & \text{otherwise} \end{cases}$$

With

$$\begin{cases} Y_1^* = X_1\beta_1 + \varepsilon_1 \\ Y_2^* = X_2\beta_2 + \varepsilon_2 \end{cases}$$

And

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} | X \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

(iii) Seemingly unrelated bivariate probit model

This study considers the application of seemingly unrelated bivariate probit model which is explained in 3.4.3.1 B.

The latent variable for this model is as follows:

$$Y_{1i}^* = \beta_1 X_{1i} + \varepsilon_{1i}$$

$$Y_{2i}^* = \gamma Y_1 \beta_2 X_{2i} + \varepsilon_{2i}$$

The estimation procedure for this model is the same as the bivariate probit model, other than incorporating γY_1 .

3.5 Conclusion

This chapter gives a description of the research methodology employed in this study. It begins with illustrating the research process, then gives a description of the research framework employed to reach each of the objectives. This is followed by a brief discussion over the research design, description about the data source, sample statistics and research techniques, selected for carrying out the empirical work. With this groundwork, the next chapter presents and discusses the analysis of the empirical results.



This document was created with the Win2PDF "print to PDF" printer available at <http://www.win2pdf.com>

This version of Win2PDF 10 is for evaluation and non-commercial use only.

This page will not be added after purchasing Win2PDF.

<http://www.win2pdf.com/purchase/>