European Transport \ Trasporti Europei (2024) Issue 99, Paper n° 4, ISSN 1825-3997

https://doi.org/10.48295/ET.2024.99.4

Modelling traveler mode shift behavior from personalized two-wheelers to public transit for work trips : A stated preference approach for Indian metropolitan city

Dr.Nipa A Desai[1]* , Dr.G J Joshi[2]

[1] *Professor, Neotech Institute of Technologie, Vadodara, Gujarat India*
[2] *Professor, Sardar Vallabhbhai National Institute of Technology, Surat, India*

Abstract

This study focuses on assessing potential thrust areas that can attract personalized vehicle based daily work trips to mass transit. The proposed study specifically targets to Tier-II medium sized Indian metropolises for the estimation of a probability mode shift from personalized 2-wheelers' to public transport on the introduction of hypothetically designed improved mass transit services. Data pertaining to study were collected through a stated preference approach that involved home-interview and online survey methods. The Latent variable enriched binary logit mode-shift behaviour model was calibrated using the Statistical Package for the Social Sciences platform and validated using hold-back data. The utility function developed using coefficients obtained through calibrated model can be further used as a user friendly tool for predicting pobability of mode shifts.The current research is an attempt to incorporate latent variables in context of Indian metropolitan area which has rarely been attempted by any researcher.

Keywords: Travel modal shift, Latent variables, stated preference approach, binary logit model, modeshift model.

1. Introduction

In developing countries such as India due to the growing demand for transportation and the lack of a satisfactory and efficient public transit system, potential transit users in the metropolitan area are forced to use personalized travel modes, in particular, the motorised 2-wheelers to fulfil their day-to-day travel needs. To deal with this alarming scenario and mitigate the problems linked to the increase in the number of vehicles on urban roads various strategies are in place out of those, shifting of personal vehicle users to other environmentally friendly transport modes is one of the key transit planning strategies adopted worldwide and found suitable for Indian condition, according to the study carried out by Rastogi and Krishnarao (2009) for Chennai city. Summary of all these studies carried out in the past, concluded the serious need for diverting personalized vehicle users to mass transportation modes through improved mass transit facilities, curtailed usage of personalized travel modes with strict policy interventions and bringing behavioural changes.

The prevailing study is focusing on estimation of a probability mode shift from personalized 2-wheelers' to public transport on the introduction of hypothetically designed improved mass transit services with following specific objectives. 1. To identify the most influential variables associated with travel mode choice and mode shift decisions of personalized two-wheeler users for their work trips in the context of Indian metropolitan areas. 2. To develop a model that can predict behaviour of personalized two-wheeler users' towards shifting to mass transit for their work trips with respect to the increased attractiveness of public transport services.

The paper is further organized as follows. Section 2 focuses on review of literature pertaining to the area of research and various mathematical techniques and modelling approaches used for model development. Section 3 emphases on the methodology adopted for the present study and specifications used for modelling transport mode shift behavior. It describes about the selection of cordon line, delineation of the study area, constricting research area, calculation of sample size and sampling distribution prior to data collection. It also describes about design of data collection tools and technique used for actual data collection. The collected data were converted into compatible format using binary coding and further utilized for development of the mode shift behavior model. Section 4 demonstrates the results of calibrated models. Important findings and conclusions from the statistical summary of data and analysis of developed models is explained in section 5 and section 6 respectively. On the basis of study findings, recommendations for policy implications are given in section 7, while section 8 focuses on study limitations and scope of future work.

2. Review of literature and study of modelling approaches used in previous studies

In order to explore varied approaches and modelling techniques in the domain of travel mode choice and mode shift, an exhaustive review of existing literature has been carried away. Popular fundamental techniques used to analyze the mode choice behaviour of an individual or group of people in a particular context are diversion curve used by Kanafani (1983), regression analysis adopted by Neels and Mather (1987), discriminant analysis explored by Louviere and Hensher (1983) and utility based choice modelling techniques in the area was explored by Koppelman and Bhat (2006). In utility based modelling techniques logit and probit regression methods are most widely used, however, as documented by Vedagiri and Arasan (2009) interpretation and specification is straight forward in the logistic approach rather than probit approach.

Findings of Sutomo et al. (2003), Yang (2003), Johansson (2006) and Domarchi and González (2008) in the transport-modelling field suggested that only a socioeconomic attributes and a trip related attributes are not sufficient to characterize travellers and make forecasts about their travel behaviour. Researchers have shown the importance of latent variables representing attitudes, perceptions, and behaviour into choice models and the suitability of psychometric scales for measurement of latent variables. Ben-Akiva et al. (2002) presented a behavioural framework and related logical formulations for modelling these latent variables. Morikawa et al. (2002) included comfort and convenience in their analyses of mode choice and modeled latent variables through the attitudinal indicators. Vredin et al. (2005) studied the problem of unobservable preferences in mode choice models and developed latent variable enriched discrete choice models and shown improvement in model performance with introduction of latend variables. Temme et al. (2008) have also given valuable insight in summarizing integrated choice and latent variable models. Till today, no research attempts are made for the integration of latent categories in mode choice and mode shift studies for the prevailing travel scenario in Indian metropolitan cities. Hence, non-inclusion of behavioural aspect in previously developed prediction models focusing on Indian metropolitan cities was considered as one of the major research gaps.

3. Methodology and data collection

According to the accessible literature, for collection of data two different approaches can be utilized i.e revealed preference (RP) and stated preference (SP). In case of present research, as previous records were not available, commuters' preferences for hypothetical scenarios were collected using stated preference approach. Based on literature and detailed study about the advantages and disadvantages of various choice modelling approaches especially when only two transport modes are involved binary logistic regression modelling technique was identified as the most suitable and appropriate for the present study. Studies carried out in past have also highlighted the significance of including attitudinal and behavioural aspects in travel mode choice and mode shift studies to achieve better results, therefore, the variables included in the study are not limited to direct, measurable variables such as travel time and cost, but also include latent variables. For measurement of latent variables psychometric scale was developed from significant latent variable indicators and user responses for each indicator was measured using ranking scale. Large number of latent variables were converted into latent classes using principal component analysis. Details about how the principal component analysis is performed, its results and structuring of latent classes can be referred from Desai and Joshi (2016).

Following sub sections 3.1, 3.2 and 3.3 describe the approaches used for fixation of cordon line for the present study, research area refinement through logical elimination of certain parameters, sampling specifications used for the survey execution, design of survey instruments, the process adopted for actual data collection and specifications used for development of the mode shift behaviour model.

3.1 Study area delineation, refinement of research areas and sampling.

The study was carried out for Vadodara city which is the third largest and one of the most developing cities in the State of Gujarat, with a population of more than 2.1 million. The reason behind selecting Vadodara city was the observed decline in Vadodara city public bus transport share during the last two decades, which was similar to most of the tier –II metro cities in India. Till date, majority of the studies carried out in India are associated with tier-I metro cities and results obtained through those studies are not directly applicable to medium sized tier-II cities, because of the huge difference in city population, employment condition and lifestyle. Necessary information pertaining to study area was collected from Municipal Corporation, city development authority and regional transport office through inventory survey and previous studies. A study carried out in Vadodara city by Desai and Vashi (2007) concluded that more than 85% of Vadodara city commuters were diverted to personalized vehicles or other intermediate transportation modes due to inefficient fleet size and poorly managed city bus services. A study to evaluate overall operation and management scenario of Vadodara city public bus transportation was taken-up by Desai and Joshi (2011) and their findings shows that in spite of various reforms in bus conditions, route alterations and change made in operational body during last decade, a desired reversal shift in public transport share was not observed. Recent share of Vadodara public bus transport services was observed only as 1% of the total population, which is very less compared to the cities that have achieved remarkable milestones in attaining their sustainable urban transportation goals worldwide.

Study area and scope of work was defined based on potential thrust areas chosen as per pilot survey results. Detailed explanation about study area delineation and income group formation can be reffered from Desai and Joshi (2016). Three level filtration process was adopted for selection of most influential variables for development of mode shift models. Initially, the least important variables were omitted based on variable ranking results. The highly ranked top five variables were identified for final survey design, out of which travel time (transit access time, waiting time and in-vehicle time), travel cost and a number of transfers were directly measurable variables but comfort & convenience and safety & security were latent variables. For collection of data, the city is divided into 10 administrative wards forming 4 zones for execution of the main survey based on similar ward system adopted by the Vadodara Municipal Corporation for routine administrative activities. Travellers' residing and working within the administrative boundary of Vadodara Municipal Corporation were included in the sample frame. Probability sampling using stratified random sampling technique was used for data collection. For determination of sample size the method suggested by Watson (2001) was adopted for the calculation of sample size. The willingness to shift study was carried out on hypothetically designed improved public bus services and hence stated preference survey instrument was designed for collection of user responses. For data collection, participants were approached by survey team and responses were recorded through personal interview method. Partial data was also collected through online forms considering convenience of respondents. Total 850 survey forms were distributed zone wise amongst working class people from Lower Income Group (LIG), Middle Income Group (MIG) and Upper Middle Income Group (UMIG) categories as per sample size calculation. Out of 850 forms, 783 were collected back. The incomplete forms or forms observed with illogical response were identified and rejected during data validation. Various statistical tests were also applied to identify extreme outliers and to check the correctness of data entry. Finally, 657 valid responses of 2-Wheeler users were used for further data analysis and modelling. Collected responses were used for development of mode shift behaviour models and further validation using Statistical Package for Social Sciences (SPSS) platform. The results obtained in SPSS platform were also validated using Gnu Regression, Econometrics and Time-series Library (GRETL) software and found satisfactory.

The scope of this paper is limited to the mode shift behaviour analysis LIG, MIG and UMIG travellers who are residing within the limit of Vadodara Urban development Authority (VUDA) and using a 2-wheeler for their daily work purpose travel.

3.2 Design of stated preference survey and data collection

Importance of associated variables responsible for mode shift behaviour was measured through variable importance ranking (stated preference) survey, which was carried out between 2- wheeler users of different income groups. Based on ranking survey results the variables with lower importance ranking were omitted and only the highly influential variables with higher importance ranking were retained for design of willingness to shift survey. Out of all selected variables, travel time (transit access time, waiting time and in-vehicle time), travel cost and number of transfers were directly measurable variables, but comfort - convenience and safety - security were latent variables which were converted into latent classes using principal component analysis. Details about how the principal component analysis was performed and how the structuring of latent classes made can be referred from Desai and Joshi (2016).

To examine the effect of hypothetical improvements brought in the attributes of public bus services, a stated preference questionnaire was prepared. A hypothetical scenario was created stating improvements in the above selected variables associated with urban area work trip travel. Three different attraction categories of reduction in travel time, reduction in travel cost, reduced number of transfers, improvements in comfort - convenience, improvements in safety - security is prepared.

During execution of stated preference survey, the respondents were asked to compare improved attributes of public bus services with their existing personalized 2-wheeler mode while stating their intentions to shift for their work purpose trips. Here, 2-wheeler user commuters were directly offered the difference in the value of attributes between two modes (Improved hypothetical bus services and existing public bus services) and asked to express their willingness to shift or otherwise with each unit change in the attribute value of public bus service. For the execution of willingness to shift survey the city was divided into 10 administrative wards forming 4 zones as suggested by the Vadodara Municipal Corporation for decentralization of activities. Probability sampling using stratified random sampling technique was used for data collection.

To collect willingness to shift responses, survey forms were distributed zone wise amongst working class people of lower, middle and upper middle income group categories. The incomplete forms or forms observed with illogical response were identified and rejected during data validation. Various statistical tests were also applied to identify extreme outliers and to check the correctness of data entry. After the data validation process 657 valid responses of 2-wheeler users were used for development of the mode shift behaviour model. Detailed information about the overall procedure followed for database development of the present study is referred from Desai and Joshi (2016).

3.3 Descriptive Statistics of 2-Wheeler User samples

Valid responses of 2-wheeler users were analysed to know the distribution pattern of collected samples. Table 1 shows descriptive statistics of 2-wheeler users' segregated samples. The average age of 2-wheeler user was observed as 35 years and the average household size was around 4 persons per household out of which around 2 persons were found working. Most of the households were found with ownership of two or more motorized 2-wheelers.

Table 1 : Descriptive Statistics of 2-Wheeler User samples

Variables		AGE	HHSIZE	HHWORK	HHMINC	NO_2W	NO_CARS	DISTWP	TIONEWAY	COSTDAY
Number of observations	Valid	657	657	657	657	657	657	657	657	657
Mean		35	4	2	34102	2	0	7	17	40
Median		32	4	2	22500	2	0	7	15	40

Mode		25	4	2	22500	2	0	10	15	50
Std. Deviation		11	1	1	19423	1	1	3	8	13
Minimum		22	1	0	12000	1	0	1	2	10
Maximum		64	12	6	75000	4	2	33	50	75
Percentiles	25	26	3	1	22500	1	0	5	10	30
	50	32	4	2	22500	2	0	7	15	40
	75	42	5	2	42500	2	1	9	22	50

Note : AGE- (In years) , HHSIZE- Household size in number of persons) , HHWORK- Number of working persons per household , HHMIC – Household monthly income in INR (Indian Rupee), NO_2W – Number of 2-wheelers owned per household , NO_CAR – Number of cars owned per household, DISTWP – Distance to workplace in kilometres, TIONEWAY- Time of one way trip in minutes, COSTDAY – Daily cost of travel for one-way trip in INR(Indian Rupee)

In a particular household similarity between ownership of 2-wheeler and a number of working persons in a family shows the dependency of working people on personalized modes for their work trips. The average trip length of 2-wheeler users for their daily work trips was observed around 7 km with an average travel time of 17 minutes and average spending of around 40 INR/Day for their daily work trip.

3.4 Dataset : Input data format for Modelling

The major objective of this study is to estimate the disaggregate probability of commuters' towards shifting to suggested improved mass transit services as a function of socio-economic, trip-related and transit service improvement categories. Necessary information pertaining to formation of income group categories and fixating of classes can be referred from Desai and Joshi (2016). This section describes about numeric coding defined for each class of independent variables, refer variables from serial number 1 to 17 as shown in Table -2 and dependent variable Willingness to Shift (WILLINGSHIFT), refer variable at serial number -18 as shown in Table -2. Here, a dependent variable is the variable that changes as a result of the changes in the values of independent variable. Based on the defined numeric coding the collected responses were converted into numeric form data sheet to be used for model calibration. The numeric coding used for model calibration and model validation for different classes of an individual variable can be referred form column titled numeric coding for variable classes in Table-2.

Table 2 : Independent and dependent variable codes used in SPSS software

Sr. No	Variable Description	Variable code	Type	Unit of measurement	Numeric coding for variable classes
1	Gender	GEN	Nominal	-	{1 – Male , 2- Female }
2	Age of respondents in years	AGE	Scale	Years	None
3	Household size	HHSIZE	Scale	Numbers	None
4	Number of working persons per house hold	HHWORK	Scale	Numbers	None
5	Income group categories	INGCAT	Ordinal	Income Class	{1 - LIG, 2- MIG, 3- UMIG, 4-HIG }
6	Type of dwelling unit	TYPDU	Ordinal	Housing Class	{1 - kaccha house, 2- Apartment, 3-

					row house, 4- tenement, 5-others}
7	Number of 2-Wheelers per household	NO2W	Scale	Numbers	None
8	Number of cars per household	NOCAR	Scale	Numbers	None
9	Visiting multiple destination during work trip	MLTDES	Nominal		{1- yes, 2- No}
10	Distance of work place from home	DSTWRK	Scale	Kilometers	None
11	Willingness to shift w.r.to reduction in bus stop waiting time	WAITT	Ordinal	Minutes	{1- not more than 15 minutes, 2 - not more than 10 minutes , 3- not more than 5 minutes, 4- Case do not attract}
12	Willingness to shift w.r.to reduction in transit access time	ACCTIME	Ordinal	Minutes	{1- 6-9 minutes , 2 - 3-6 minutes, <3 - minutes , 4- It does not attract me at all}
13	Willingness to shift to reduction in in-vehicle travel time	RTT	Ordinal	Minutes	{1- up to 10 minutes, 2 - 10-15 minutes, 3- >15 minutes, 4- Case do not attract}
14	Willingness to shift w.r.to reduction in travel cost	RTC	Ordinal	Indian Rupee	{1- up to 7 INR/day, 2 - 7 to 15 INR/day , 3- >15 INR/day, 4- Case do not attract} (Note : INR- Indian rupee)
15	Willingness to shift w.r.to number of bus interchanges	INTCHG	Ordinal	Numbers	{1- Not more than 2 interchanges, 2 Not more than 1 interchange, 3- Prefer only direct bus , 4- It does not attract me at all}
16	Willingness to shift w.r.to enhanced level of comfort - convenience	CCLVL	Ordinal	-	{1 - Level-I (Basic), 2 – Level- II (Average), 3- Level-III(Highest), 4- Does not attract me at all }
17	Willingness to shift w.r.to enhanced level of safety - security	SSLVL	Ordinal	-	{1 – Level-I (Basic), 2 – Level- II (Average), 3- Level-III(Highest), 4- Does not attract me at all }
18	Willingness to shift from personalized mode to mass transit with respect to enhanced level of bus services or otherwise	**WILLINGSHIFT (Dependent Variable)**	Nominal	-	{1 -Yes, 2- No}

Notes: Variable coding shown in Table 2 is developed by researchers exclusively for the present study which is further used for data processing. This coding may be adopted for future studies as per suitability.

Final data set with 657 valid responses of 2-wheeler users stating their willingness to shift to public transport for various hypothetically designed scenarios with 13797 data points (7 different categories in which improvements were proposed for public bus services X 3 levels of each category X 657 valid responses of 2-wheeler users i.e 7X3X657 = 13797) was processed for model development. Table 2 Shows input data form which was converted into binary coding and used for development of Logit model using SPSS software.

Here, different levels of comfort - Convenience and Safety - security are derived based on psychometric survey and converted into levels using principal component analysis (PCA) as explained in earlier sections and further to be referred from Desai and Joshi (2016) . CCLevel-I is referred as normal and clean buses with good support for standing and stress-free riding. CCLevel-II is a Level-I added with low boarding platform less crowded buses with more seating space, curtain/sun film, improved riding quality, added routes, sheltered bus stops, and published schedule on stops. CCLevel-III is Level II added with air-conditioned buses, spacious and comfortable seats with more sitting space, smart ticketing and priority reservation system, wash room availability at the main terminals. Similarly, SSLevel- I is safely driven buses with sheltered bus stops for security of belongings. SSLevel II is a Level-I added with fire safety alarm, safely located bus stops with lighting facility. SSLevel III is Level-II added with CCTV in buses, a queue formation mechanism for safe boarding and alighting, appointment of security officers at terminals and strict checking for drunken driving.

3.5 Model specification

As discussed earlier, the present study is focusing on shifting personalized 2-wheeler users to proposed public transport bus services considering the difference in attributes of existing and improved bus services. The model was developed using a binary logistic regression approach which is based on utility maximization theory supported with an assumption that the decision maker's preference for switching travel mode is captured by a value, called the utility, and the decision maker selects an alternative in a choice set with the highest utility. The choice set is a set of alternatives that are available to a person to select from. In our case there were only two alternatives available in a choice set to choose from, so the problem was of binary nature. In the utility-based binary choice model, the term decision maker is an individual who is directly connected with choice or who is taking the decision to choose or not to choose an alternative, in our case decision maker is personalized 2-wheeler user respondents. The term attribute means set of characteristics perceived by each alternative in the choice set, in our case attributes offered to decision makers are decisive variables related to their travel using personalized 2-wheeler and perceived improvements in the variables associated with enhanced public bus services. The term decision rule is the process used by the decision maker to evaluate alternatives in the choice set to determine a choice, in our case utility maximization is the decision rule.

When attribute of one of the alternatives is changing there is a change in utility value. These changes in utility values are different for different respondents for the same amount of change in an attribute value. In this scenario an individual is expected to switch his/her travel mode, in our case from personalized 2-wheeler to enhanced level of public transport bus services if utility of improved bus mode (U_{bus}) will achieve higher value than utility of personalized 2-wheeler (U_{2-w}) for an individual traveller ($U_{bus} \geq U_{2-w}$) So, in this situation mode shift occurs when the difference in utility is greater than zero.

$(U_{bus} - U_{2-w} \geq 0) = (U_{diff} \geq 0)$.

Here utility difference is assumed to be made up of two components deterministic components (V_{diff}) and random components (ε_{diff}). The systematic component or deterministic term (V_{diff}) representing observed effect which is assumed to be measured by linear in parameter specification. Hence, $V_{diff} = A_0 + A_1X_1 + A_2X_2 + \cdots + A_nX_n$

Here, A_0, A_1, A_2………, An are the model parameters to be estimated and X_1, X_2………X_n are variables influencing the mode shift.

The random term (ε_{diff}) representing unobserved effect affecting choice. In the present study, logistic regression approach is used and hence the random term (ε_{diff}) is assumed to be distributed as logistic distribution.

Thus, the probability of a shift from personalized 2-wheeler to enhanced level of public transport is,

$$P_{shift} = pr(U_{bus} \geq U_{2-w}) \quad (1)$$

$$P_{shift} = pr(V_{bus} + \varepsilon_{bus} \geq V_{2-w} + \varepsilon_{2-w}) \quad (2)$$

here $(V_{diff} \geq 0) = (V_{bus} - V_{2-w} \geq 0)$ and $(\varepsilon_{diff} \geq 0) = (\varepsilon_{bus} - \varepsilon_{2-w} \geq 0)$

As per standard logistic regression framework probability of mode shift can also be represented as

$$P_{shift} = \frac{e^{V_{diff}}}{1 + e^{V_{diff}}} \quad (3)$$

$$P_{shift} = \frac{e^{A0 + A1X1 + A2X2 + A3X3 + \ldots\ldots\ldots\ldots AnXn}}{1 + e^{A0 + A1X1 + A2X2 + A3X3 + \ldots\ldots\ldots\ldots AnXn}} \quad (4)$$

4. Development of mode shift behaviour model

For prediction of mode shift behaviour, it is necessary to estimate and validate the model parameters using survey data. First of all, preferences given by 2-wheeler users towards shifting to improved public bus services considering the difference in attributes of existing and improved bus services were converted into binary data coding and calibrated using SPSS platform.

4.1 Model Calibration

Preferences given by commuters for various hypothetical cases were analysed and calibrated to obtain statistically tested and validated mode shift behaviour model. The objective was to estimate the aggregate probability of commuters' towards shifting to suggest improved mass transit services as a function of socioeconomic, trip related and other public transport improvement categories. For calibration of model, first of all entire data set was divided into 2/3rd and 1/3rd part for model calibration and validation respectively. Random number generation technique was used for data division so that selection bias can be avoided. Table 3 is showing case processing summary adopted in SPSS software for separation of data for model calibration and validation. Selected cases are those which are used for calibration of developed model and unselected cases are randomly separated samples which are not used for model calibration, but after the development of model the unselected set of data is used for testing and validation of the model.

Table 3 : Case Processing Summary for 2-Wheeler User Samples

Un-weighted Cases		N	Percent
Selected Cases	Included in calibration	474	72.1
	Missing Cases	0	.0
	Total	474	72.1
Unselected Cases	Used for model validation	183	27.9
Total		657	100.0

Notes: Data collected through actual survey and used for model development. The above table shows data count used for model calibration and model validation. Here selected cases mean data points used for model calibration and unselected cases mean holdout data points which are not used for model calibration. N means total number of observations. The results obtained from calibrated model is validated using holdout data points and performance consistency of calibrated model is checked using unselected cases.

Table 4 : Results of Model-1 (Calibrated as constant only model)

(a) Model parameter estimation for constant only model

Number of observations used for model parameter estimation : 474 (Selected cases)

	B	S.E.	Wald	df	Sig.	Exp(B)
Constant	1.197	0.109	120.967	1	0.000	3.309

(b) Observed and predicted probability summary for constant only model

		Predicted					
		Selected Cases			Unselected Cases		
		WILLING SHIFT			WILLING SHIFT		
Observed		0	1	% Correct	0	1	%Correct
WILLINGSHIFT	0	0	110	0.0	0	38	0 .0
	1	0	364	100.0	0	145	100.0
Overall Percentage				76.8			79.2

(c) Test statistics for constant only model

Hosmer and Lemeshow Test results

Chi-square	df	Sig.	-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
0.000	0	0.00	396.134	0.219	0.332

Notes : The values shown in the above table is derived through calibration of constant only model calibrated using collected data.

The initial model calibrated in SPSS is known as a constant only model. The constant only model was calibrated without inclusion of dependable variables. The calibration results of initial model, also known as a base model which is presented as Model-1 in Table 4. All the independent variables as shown in Table -2, were added one by one into the base model using the forward likelihood method until the model converges. The finally converged models with all the significant variables were compared with constant only model and various statistical tests were performed to check, whether adding various independent variables have increased the prediction capacity of the model in comparison with the constant only model or not.

4.2 Model parameter estimation and validation

Model 2 as presented in Table 5 was developed by adding different variables to the constant-only model. In comparison, of the converged Model 2 with base model it was found that the calibrated model has shown significant improvement in the values of test parameters compared to the base model, hence inclusion of various independent variables for development of prediction model was logical. Table 5 shows significant variables extracted after model calibration and validation by the forward log likelihood method. Variables with significance value < 0.05 were retained and non-significant variables were omitted during model calibration process. Visit to multiple destinations during the journey to work, bus stop waiting time, transit access time, highest level of comfort - convenience and highest level of safety - security were identified as the most significant variables during model calibration. The sign of significant variables was analysed and found logical in calibrated model.

Table 5 : Results of model 2 (Calibrated with inclusion of variables)

(a) Model parameter estimation

Number of observations used for model calibration : 474 (Selected Cases)

Variable	B	S.E.	Wald	df	Sig.	Exp(B)
MLTDES	-0.588	0.285	4.265	1	0.039	0.555
WAITT_3	0.808	0.357	5.120	1	0.024	2.244
ACCTIME_1	0.793	0.333	5.654	1	0.017	2.209
CCLVL_3	1.504	0.366	16.880	1	0.000	4.499
SSLVL_3	1.501	0.387	15.067	1	0.000	4.486
Constant	-1.003	0.576	3.028	1	0.082	0.367

(b) Test statistics

Hosmer and Lemeshow Test results					
Chi-square	df	Sig.	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
2.491	6	.0869	360.602	0.276	0.417

(c) Observed and predicted probability summary

		Predicted					
		Selected Cases : 474 (Model calibration)			Unselected Cases : 183 (Model Validation)		
		WILLING SHIFT			WILLING SHIFT		
		0	1	% Correct	0	1	%Correct
WILLINGSHIFT	0	60	50	54.5	23	15	60.5
	1	18	346	95.1	14	131	90.3
Overall Percentage				84.0			84.2

Notes : The values shown in the above table is derived through calibration of model. Data used for model calibration is collected through actual survey.

In model 2, the negative sign of variable coefficient MULTIDES reflects the inverse relationship between willingness to shift and visit multiple destinations. It can be concluded from model results that more is the number of destinations to be visited by 2-wheeler user commuter during their daily work trips, less is their wiliness to shift to public transport. A positive sign of variables WAITT3 also found the logical stating positive impact of the variable reduction in the bus stop waiting time on their willingness to shift. The positive sign of waiting time at bus stop variable shows, the higher is the reduction in bus stop, wait time, more will be the number of 2-wheeler users attracted to shift to enhanced public bus services. Here, it is to be noted that respondents were offered three different categories of reduced bus stop waiting time as WAITT1(Up to 15 minutes), WAITT2 (Up to 10 minutes) and WAITT3 (Up to 5 minutes), but only WAITT3 is found significant. It shows that bus stop waiting time more than 5 minutes is not found attractive to 2-wheeler users while taking mode shift decision. A positive sign of variable (ACCTIME1) also shows the positive relationship between reduced transit access time as per hypothetical design and increase in percentage willingness to shift. Here, commuters were offered three different categories of transit access time as ACCTIME1 (6-9 minutes) , ACCTIME2 (3-6 minutes) and ACCTIME3 (<3 minutes), but only ACCTIME1 is found significant. It shows that 6-9 minutes transit access time is acceptable to most of the respondents even though they were offered, lesser transit access time as an option. Similarly, the positive values of CCLVL3 and SSLVL3 signify an increase in wiliness to shift with an increase in the level of comfort-convenience and safety-security. Here, it is to be noted that highest level of improvements in both the latent categories is found significant, which shows respondents inclination towards journey comfort - convenience and safety - security.

Out of all significant variables in Model 2, CCLVL3 and SSLVL3 are having high values of odds ratio (Wald statistics) which shows its higher influence in models while making mode shift decisions. This also indicates that the latent class variables of comfort-convenience and safety-security were able to partially substitute effect of reduction in travel time, travel cost and other variables which were found significant in studies carried out by other researchers without considering latent variables.

Various statistical tests were performed to check model accuracy. It was observed that the -2 Log Likelihood statistics for constant only model (Model 1) was 396.134 and for the calibrated model (Model 2) was 360.602. The -2 Log Likelihood statistic measures how better the model predicts the decisions, the smaller the statistic, better the model. Here, Model 2 developed with the inclusion of variables, shows the lesser value of -2 Log Likelihood representing better fit of the converged model compared to base mode. The value of Hosmer- Lemeshow (H&L) test statistics for calibrated model (Model 2) was observed as 0.869 which is greater than 0.5 shows good fit. Cox & Snell R^2 and Nagelkerke R^2 were obtained to evaluate model fitness and improvements in model performances, It is normally used to choose the best fit model out of different models. Values of Cox & Snell R^2 also improved from 0.219 to 0.276 and Nagelkerke R^2 improved from 0.332 to 0.417 for calibrated model 2 which shows that with the inclusion of significant variables in base model its performance is enhanced.

In addition to R^2 and other measures of model fit, the most popular and appropriate measure to check model accuracy level is a case-to-case summary of observed and predicted values. Here, for Model 2 the prediction success classification summary as shown in Table 5(c) also indicates a reasonable fit in comparison with constant only models with improvement in prediction success percentage from 74.6 % in the base model (Model 1) to 84% in a calibrated model (Model 2). For calibrated model (Model 2) percentage of observed and predicted probability values given in Table 5(c) are almost similar for selected cases and unselected cases which shows that 1/3rd of data retained for model validation are producing similar results as 2/3rd of data used for model calibration. Thus, the model is validated and can be adopted for mode shift prediction.

4.3 Development of utility function and probability to shift analysis

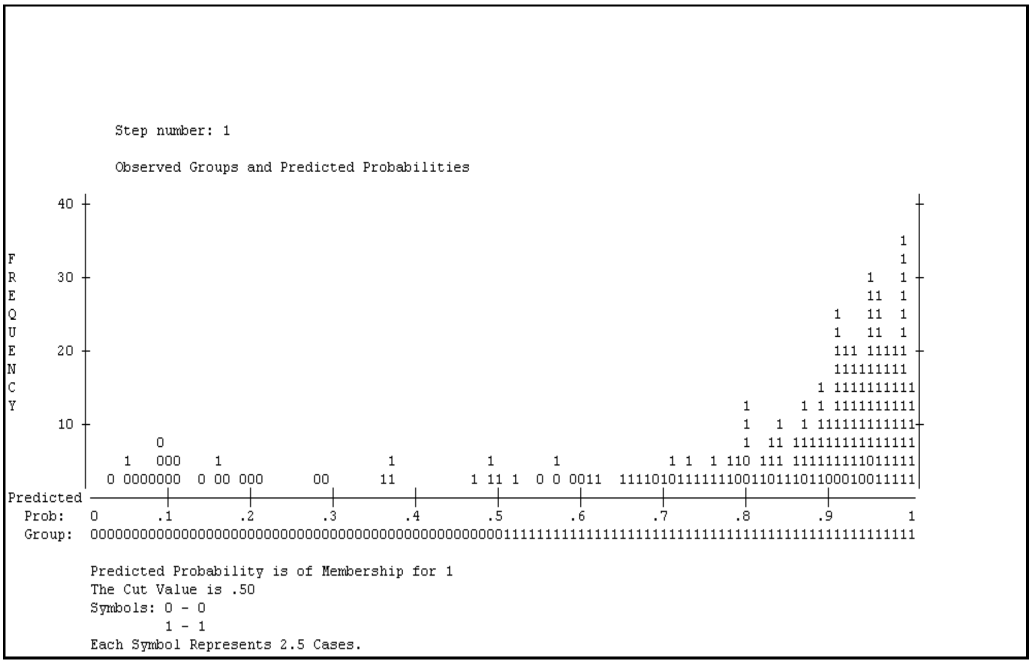
The utility function developed using significant variables and variable coefficients obtained during calibration is as shown in equation 5.

$$V_{diff} = -1.003 - 0.588*MLTDES + 0.808*WAITT\_3 + 0.793*ACCTIME\_1 + 1.504*CCLVL\_3 + 1.501*SSLVL\_3 \quad \ldots(5)$$

For prediction of mode shift probability various cases are developed by adding different constrains to the utility function. Willingness to shift probability values can be calculated by substituting variable values in the calibrated model 2 represented as equation 6.

$$P_{shift} = \frac{e^{-1.003-0.588*MLTDES+0.808*WAITT\_3+0.793*ACCTIME\_1+1.504*CCLVL\_3+1.501*SSLVL\_3}}{1+e^{-1.003-0.588*MLTDES+0.808*WAITT\_3+0.793*ACCTIME\_1+1.504*CCLVL\_3+1.501*SSLVL\_3}} \quad (6)$$

The developed aggregate mode shift probability model as shown in equation 6 can be further utilized for prediction of probability of an individual towards shifting to public transport through the development of various mode shift probability curves by considering various permutations and combinations of above significant variables. Figure 1 shows predicted probability value plot calculated using a calibrated model.



Notes : This probability prediction graph is plotted in SPSS software using actual results of calibrated model.

Figure 1 : Graphical Representation of Predicted Probabilities Showing Willingness to Shift (Model 3)

The values of predicted probability for each case can be calculated and tabulated for detailed analysis. If the predicted probability value or set of values is 1(Willing to shift and falling between the range 0.5 to 1 on horizontal axis means it is a correct prediction but the cases falling between the range 0.5 to 1 with predicted probability values as 0 are wrongly predicted cases. Likewise, when predicted probability value or set of values is 0 (Not willing to shift) and lying between the range 0.5 to 0 on horizontal axis means it is a correct prediction but the cases falling between the range 0.5 to 0 with predicted probability values 1 are wrongly predicted cases. A model with the maximum predicted probability cases lying between the extreme ranges 0.9 to 1 and 0.1 to 0 with correct prediction values 1 and 0 respectively is considered more accurate. It is observed from the plot, that while evaluation case to case summary majority of the predicted probability values were correctly predicted while calculated using the developed model.

5. Discussion

Important findings obtained from the study are summarized as under.

i. Detailed analysis of data shows that more than 50% of 2-wheeler users' have shown their willingness to shift to public transport for up to 10-minute reduction in travel time and up to 15 INR/day reduction in travel cost which confirms the findings of Vedagiri and Arasan (2009). It is also observed that instead of, reduction in travel time and travel cost, lesser transit access time attracts more commuters to shift to public transport. More than 70% of commuters' have shown their preferences for wait time up to 10 minutes or less and direct buses. At the highest level of comfort - convenience and safety - security more than 80% of commuters' have shown their willingness to travel by public transport buses. In this study, such higher preference given to comfort-convenience variable by commuters justify the findings of Morikawa et al. (2002).

ii. Based on model calibration results, visiting multiple destinations during the journey to work, bus stop waiting time, transit access time, highest level of comfort - convenience and highest level of safety - security were identified as the most influential variables with a significance value greater than 0.05 and logical signs of the parameters which justifies the findings of Sutomo et al. (2003), Yang (2003), Johansson (2006) and Domarchi and González (2008) who have suggested that only a socioeconomic attributes and a trip related attributes are not sufficient to characterize travellers and make forecasts about their travel behaviour.

iii. The mode shift behaviour logit model calibrated using 2-wheeler user willingness to shift responses for their work purpose trips was found statistically significant with Hosmer- Lemeshow (H&L) test statistics values > 0.5. Model prediction success percentage for calibrated model was similar to regenerated model developed using holdout data which has established the validity of the developed model and confirmed the results obtained by Vedagiri and Arasan (2009).

iv. A utility equation obtained through model calibration can be used as statistically tested and validated tool to predict the probable mode shift from personalized 2-wheeler to enhanced level of public bus for a wide range of variable values in prevailing urban from and traffic condition. The model results can be used for strategic planning and decision-making targeting shift from personalized to public mode of transport.

v. In converged model out of all significant variables, the highest level of comfort - convenience (CCLVL3) and highest level of safety - security (SSLVL3) were having high values of odds ratio (Wald statistics) which shows higher influence of latent variables in making mode shift decisions which justifies the findings of Morikawa et al. (2002) who included comfort and convenience in their analyses of mode choice and modelled latent variables through the attitudinal indicators.

vi. The present research is an attempt to incorporate latent variables in the travel behaviour study of urban commuters' especially in the context of Indian metropolitan area which is rarely attempted by any researcher. The detailed description of the study presented in this article will be useful for further extension and developments in the area of latent variable enriched mode shift behaviour models.

6. Conclusion

From the above study and discussion, it is concluded that the studies carried out in the past were missing the perceived importance of latent variables especially in the prevailing scenario of Tier-II Indian cities. The research outcomes available in existing literature, focused solely on socioeconomic or trip-related variables that were directly measurable, while analysing travellers' mode choice and potential for mode shift. Those studies have identified reduction in travel time and a reduction in travel cost as the most influential variables and most of the improvement strategies were planned with primary focus on reduction in travel time and travel cost. On the other side, for medium sized metro cities (Tier-II) like Vadodara, when the study is attempted with inclusion of latent variables the importance of other service parameters was observed noticeably higher than reduction in travel time and costs. The importance given to latent variables comfort-convenience and safety-security was found to be very high in all the income segments of personalized 2-wheeler users. It is concluded from the above study results that corrective measures adopted to deal with the present challenge of enhancing modal share of public transport buses for different categories of metro cities (Tier-I, Tier-II, and Tier-III) especially in Indian scenario of diversified culture, tradition, religion, lifestyle, and habits need to be designed separately. Sustainable transportation strategies designed without considering city size, daily travel need of residents, existing road network type, existing traffic pattern, and overall transportation situations in the city will not stand effective.

7. Recommendations for policy implications

Vadodara is a city with an urban area population of more than 2.1 million. The road network of Vadodara city is a combination of rectangular grid and radial pattern with an average trip length of 6.66 km with varying road widths of 12 meters in city centres, to 32 meters on radial routes and up to 40 meters of ring roads. However, due to on-road vehicle parking in a haphazard manner, on the majority of road networks effective road width is considerably less than the actual one. The city has grown horizontally due to which population density on any identified mass transit corridor is very less and urban form is unsupportive of exclusive mass transit corridors. Based on the findings from mode shift analysis, it is recommended that for medium sized metro cities (Tier-II) with similar characteristics as Vadodara, planners should initially focus on achieving economical, stress free, comfortable and secure travel. Their initial focus should be to achieve, lower transit access time, lower the bus stop waiting time and reducing the number of interchanges by investing on fleet size and bringing improvements in other service parameters as per commuters' preference.

8. Limitations and Scope of Future Work

The limitations of above study are listed below; further studies in the areas listed in study limitations can be attempted.

i. The scope of above study is limited to Vadodara city only and hence, before using the mode shift probability prediction model for other similar areas testing of model for geographic transferability is recommended.
ii. The present study is limited to, work trips only, however the meticulously designed research tool and modelling approach adopted in the present research can be directly used for development of models for other purpose trips.
iii. Detailed study showing trade-offs between mode shift probability and changes in the value of individual variable using elasticity analysis can be attempted as a scope of future work.

References

Arasan, V.T. and Perumal, V., 2009. Modelling modal shift from personal vehicles to bus on introduction of bus priority measure. In Proceedings of the Eastern Asia Society for Transportation Studies Vol. 7 (The 8th International Conference of Eastern Asia Society for Transportation Studies, 2009) (pp. 256-256). Eastern Asia Society for Transportation Studies.

Ben-Akiva, M. and Morikawa, T., 1990. Estimation of switching models from revealed preferences and stated intentions. Transportation Research Part A: General, 24(6), pp.485-495.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T. and Polydoropoulou, A., 2002. Integration of choice and latent variable models. Perpetual motion: Travel behaviour research opportunities and application challenges, 2002, pp.431-470.

Chiu Chuen, O., Karim, M.R. and Yusoff, S., 2014. Mode choice between private and public transport in Klang Valley, Malaysia. The Scientific World Journal, 2014.

Desai, N. and Vashi, B.D., 2007. Optimization & privatization of city bus network using GIS: a case study of Vadodara city in Gujarat state. In Indian Road Congress (IRC) (Vol. 69).

Desai, N. and Joshi, G.J., 2011. Urban public bus transportation scenario and observed modal shift in last three decades using revealed preference data: A case study of Vadodara city in Gujarat state, National conference Sutrims-11 organised by Sardar Vallabbhai National Institute of Technology, Volume 2, pg 113-124

Desai, N.A. and Joshi, G.J., 2013. Geographical and cultural variations in factors affecting travel mode choice: an analysis of travellers' behaviour. Indian journal of transport management, 37(3).

Desai, N.A. and Joshi, G.J., 2016. Database Development Approach and Survey Design for Travel Mode Shift Behavior Study with Respect to Mass Transit in a Metropolitan Context. Transportation in Developing Economies, 2(2), pp.1-16.

Domarchi, C., Tudela, A. and González, A., 2008. Effect of attitudes, habit and affective appraisal on mode choice: an application to university workers. Transportation, 35(5), pp.585-599.

Johansson, M.V., Heldt, T. and Johansson, P., 2006. The effects of attitudes and personality traits on mode choice. Transportation Research Part A: Policy and Practice, 40(6), pp.507-525.

Kanafani, A. (1983). Transportation demand analysis, The national academy of science, engineering and medicine, Highway Capacity manual, McGraw-Hill publication, New York, NY USA.

Koppelman, F.S. and Bhat, C., 2006. A self-instructing course in mode choice modeling: multinomial and nested logit models.

Louviere, J.J. and Hensher, D.A., 1983. Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event. Journal of Consumer research, 10(3), pp.348-361.

McFadden, D., 1986. The choice theory approach to market research. Marketing science, 5(4), pp.275-297.

Morikawa, T., Ben-Akiva, M. and McFadden, D., 2002. Discrete choice models incorporating revealed preferences and psychometric data. In Advances in Econometrics. Emerald Group Publishing Limited.

Neets, K. and Mather, J.O.S.E.P.H., 1987. Modeling Mode Choice in New Jersey. Transportation Research Record, 1139, pp.20-27.

Rastogi, R. and Rao, K., 1999. A policy framework for a sustainable multi-mode urban transportation planning.

Rastogi, R. and Rao, K.K., 2002. Survey design for studying transit access behavior in Mumbai City, India. Journal of transportation engineering, 128(1), pp.68-79.

Rastogi, R. and Krishna Rao, K.V., 2003. Travel characteristics of commuters accessing transit: Case study. Journal of Transportation Engineering, 129(6), pp.684-694.

Rastogi, R. and Krishna Rao, K.V., 2009. Segmentation analysis of commuters accessing transit: Mumbai study. Journal of transportation engineering, 135(8), pp.506-515.

Rastogi, R., 2010. Willingness to shift to walking or bicycling to access suburban rail: Case study of Mumbai, India. Journal of urban planning and development, 136(1), pp.3-10.

Sutomo, H., Istiyanto, B., & Matsumoto, S., 2003. Psychological factors affecting travel mode choice (case: bus-lane plan for Yogyakarta, Indonesia). Journal of the Eastern Asia Society for Transportation Studies, 5.

Temme, D., Paulssen, M. and Dannewald, T., 2008. Incorporating latent variables into discrete choice models—a simultaneous estimation approach using SEM software. Business Research, 1(2), pp.220-237.

Vedagiri, P. and Arasan, V.T., 2009. Estimating modal shift of car travelers to bus on introduction of bus priority system. Journal of transportation systems engineering and information technology, 9(6), pp.120-129.

Vinayakam, K. and Sekar, S.P., 2013. Rural to urban migration in an Indian metropolis: Case study Chennai city. IOSR Journal of Humanities and Social Science, 6(3), pp.32-35.

Vredin Johansson, M., Heldt, T. and Johansson, P., 2005. Latent variables in a travel mode choice model: Attitudinal and behavioural indicator variables. Department of Economics, Uppsala University.

Watson, J., 2001. How to determine a sample size: tipsheet no. 60. Unpublished raw data, Penn State Cooperative Extension, Penn State, University Park, PA. Retrieved from http://www. extension. psu. edu/evaluation/pdf/TS60. pdf.

Yang, C.W., Duann, L.S., Chen, Y.C. and Fang, W.P., 2003. Intercity bus choice models with choice set generation and heterogeneity. Journal of Eastern Asia Society for Transportation Studies, 5, pp.1518-1533.

Table of abbreviations

ACCTIME	Willingness to shift w.r.to reduction in transit access time
AGE	Age in years
CCLVL	Willingness to shift w.r.to enhanced level of comfort – convenience
CCTV	Closed Circuit Television
CI	Confidence Interval
COSTDAY	Daily cost of travel for one-way trip in INR (Indian Rupee)
DISTWP	Distance to workplace in kilometres
GEN	Gender
GRETL	Gnu Regression, Econometrics and Time-series Library
GSRTC	Gujarat State Road Transport Corporation
HHMIC	Household monthly income in INR (Indian Rupee)
HHSIZE	Household size in number of persons
HHWORK	Number of working persons per household
H & L	Hosmer- Lemeshow
INGCAT	Income group categories
INTCHG	Willingness to shift w.r.to number of bus interchanges
LIG	Lower Income Group
ME	Margin of Error
MIG	Middle Income Group
MLTDES	Visiting multiple destination during work trip
NO2W	Number of 2-wheelers owned per household
NOCAR	Number of cars owned per household
PPP	Public Private Partnership
RP	Revealed preference
RTC	Willingness to shift w.r.to reduction in travel cost
RTT	Willingness to shift to reduction in in-vehicle travel time
SD	Standard Deviation
SP	Stated preference (SP)
SPSS	Statistical Package for the Social Sciences
SSLVL	Willingness to shift w.r.to enhanced level of safety - security
TIONEWAY	Time of oneway trip in minutes
TYPDU	Type of dwelling unit
UMIG	Upper Middle Income Group
VTPL	Vitcos Transportation Private Limited
VUDA	Vadodara Urban development Authority
WAITT	Willingness to shift w.r.to reduction in bus stop waiting time
WILLINGSHIFT	Willingness to Shift