An analysis of individuals' usage of bus transit in Bengaluru, India: Disentangling the influence of unfamiliarity with transit from that of subjective perceptions of service quality

L. Deepa
Research Scholar
Department of Civil Engineering
Indian Institute of Science (IISc), Bangalore 560012, India
Email: deepal@iisc.ac.in

Aupal Mondal
Research Scholar
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin, Austin TX 78712, USA
Email: aupal.mondal@utexas.edu

Aishwarya Raman
Director and Head of Research
Ola Campus, Prestige RMZ Star Tech - Wing C, No. 140, Industrial Layout,
Koramangala, Bengaluru 560095, India
Email: aishwarya.raman1@olacabs.com

Abdul Rawoof Pinjari (**corresponding author**)
Associate Professor
Department of Civil Engineering
Indian Institute of Science (IISc), Bangalore 560012, India
Ph: +91-80-2293 2043; Fax: +91-80-2360 0404
Email: abdul@iisc.ac.in

Chandra R. Bhat
Professor
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
Austin TX 78712, USA
Email: bhat@mail.utexas.edu

Karthik K. Srinivasan
Professor
Department of Civil Engineering, Indian Institute of Technology Madras
Chennai 600036, Tamil Nadu, India
Email: karthikks@iitm.ac.in

Ram M. Pendyala
Professor
Arizona State University, School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Email: ram.pendyala@asu.edu

Gitakrishnan Ramadurai
Associate Professor
Department of Civil Engineering, Indian Institute of Technology Madras
Chennai 600036, Tamil Nadu, India
Email: gitakrishnan@iitm.ac.in

AUTHOR STATEMENT

AUTHOR CONTRIBUTIONS: Study conception and design: Pinjari, A.R., Deepa, L., Bhat, C.R., Srinivasan, K; data collection: Raman, A.; Supervision of analysis: Pinjari, A.R.; Formal analysis, Validation, Investigation: Deepa, L., Mondal, A., Pinjari, A.R., C.R. Bhat., Srinivasan, K, Pendyala, R., Ramadurai, G., Raman, A.; Writing-Original draft preparation: Deepa, L., Pinjari, A.R.; Writing-reviewing and editing: Bhat, C.R., Srinivasan, K., Raman, A., Pendyala, R., Ramadurai, G.; All authors reviewed the results and approved the final version of the manuscript.

DECLARATIONS: The third author works for the Ola Mobility Institute (OMI). Research at OMI is funded by ANI Technologies Pvt. Ltd. (the parent company of brand Ola, which provides mobility services in many cities). However, ANI's financial considerations did not influence the work reported in this paper. The other authors are not aware of any competing interests to declare.

ACKNOWLEDGMENTS: The authors gratefully acknowledge the support provided by the Ministry of Human Resource Development (MHRD) of the Government of India through its Scheme for Promotion of Academic and Research Collaboration (SPARC) program in facilitating this research collaboration across various academic institutions. The survey data shared by Ola Mobility Institute is gratefully acknowledged. The last but one author acknowledges the support provided by the U.S. Department of Transportation through the Grant No. 69A3551747116.

ABSTRACT

This study presents an integrated model to shed light on the factors influencing individuals' likelihood and frequency of usage of bus transit in Bengaluru, India, with a focus on the role of individuals' subjective perceptions of service quality. Typically, subjective perceptions of transit service characteristics such as comfort, cleanliness, reliability, and safety are measured using Likert rating scale questions in travel surveys. A shortcoming with many such surveys is that the Likert rating scale questions do not include a "don't know" response category for the respondents to express their unfamiliarity and lack of opinion on the transit service. For this reason, some respondents who are not familiar with and do not have an opinion about the transit system are likely to choose the neutral response to Likert scale questions. At the same time, travelers who are familiar with and/or informed about the transit system may also choose the neutral response to state their opinion neutrality. As a result, some travelers' unfamiliarity with (and lack of opinion about) transit services may be confounded with the informed perceptions of those who are familiar with transit. This is because those who are unfamiliar with the transit system are less likely to use it and more likely to state neutral responses than those who are familiar with the system. Ignoring such influence of travelers' unfamiliarity can potentially distort the ordinal scale of Likert variables, result in biased parameter estimates and distorted implications about the influence of perceptions on transit usage. To address this concern, this study uses a generalized heterogeneous data model (GHDM) that allows a joint econometric analysis of the influence of individuals' perceptions of transit service quality on their likelihood of transit use and frequency of use and at the same time disentangle unfamiliarity from informed perceptions. The empirical results shed light on: (a) the role of individuals' demographic variables and subjective perceptions on their use and frequency of use of the bus transit system in Bengaluru, (b) the importance of separating unfamiliarity from informed opinions on transit service quality, (c) the need to include an option for respondents to reveal their unfamiliarity in Likert rating scale survey questions on perceptions, and (d) demographic segment-specific strategies for attracting new riders and enhancing ridership of current users of the bus transit system in Bengaluru.

Keywords:

public transit use in India**,** latent variable model, unfamiliarity, perceptions of service quality, Likert scale responses

1. INTRODUCTION

1.1 Background and objectives

Encouraging travelers' usage of public transport and enhancing their extent of usage is a key consideration in urban public transport planning and policymaking. To inform such policies, the role of objectively measurable variables such as travelers' demographic characteristics (e.g., age and gender) and transit service characteristics (e.g., travel times and costs) has been widely considered in models of mode choice and public transit usage (Cervero and Kockelman, 1997). However, a rich body of literature recognizes that the inclusion of individuals' subjective perceptions of transit service characteristics along with objective variables can enhance the understanding of travelers' mode choice behavior and transit usage (Kuppam *et al*., 1999; Johanson *et al*., 2006). In this regard, numerous studies have explored the influence of travelers' perceptions of service attributes (such as reliability, waiting time uncertainty, information, comfort, crowding, and cleanliness) on travelers' mode choice, public transit usage, and satisfaction as well as on their route-choice behaviour (Hensher *et al*., 2003; Srinivasan *et al*., 2007; Eboli and Mazzulla, 2011; de Ona *et al.,* 2013*;* Carrel and Walker, 2017; Shelat *et al*., 2021).

Generally, travelers' perceptions of transit service attributes are measured in travel surveys using questions that elicit responses on a Likert rating scale (Weinstein, 2000; Eboli and Mazzulla, 2010; Lai and Chen, 2011; Bordagaray, 2013; de Ona *et al.,* 2013). Originally proposed by Likert (1932), such questions are the predominant approach for measuring individuals' perceptions and attitudes – not only in transportation but also in many other social sciences. One way to incorporate Likert rating scale variables for individuals' perceptions of transit service characteristics in models of transit usage is to use them as explanatory variables directly in the choice model. This approach assumes that the Likert rating scale variables represent traveler's underlying perceptions of transit service characteristics that influence traveler choices. However, the Likert scale indicators may only be proxies or manifestations of underlying perceptions and are often associated with errors in measuring the perceptions (Ben-Akiva *et al*., 2002; Bhat and Dubey, 2014). Ignoring such measurement errors might lead to biased parameter estimation and distorted inferences of the influence of perceptions on transit usage. Therefore, to recognize the measurement errors, it is a common practice to use Likert scale variables and other indicators to construct latent, stochastic variables for travelers' perceptions. The latent variables are, in turn, employed as stochastic variables measuring travelers' subjective perceptions along with the traditionally used covariates (for objective factors) to explain transit usage (Ben-Akiva *et al*., 2002; Johansson *et al*., 2006; Vij *et al*., 2013; Das and Pandit, 2015; Sarkar and Mallikarjuna, 2018; Devika *et al*., 2020).

Typically, the Likert scale questions involve an ordinal rating scale for the respondents to indicate the extent to which they agree or disagree with a particular statement. The responses range from one extreme of strong agreement to another extreme of strong disagreement, with a neutral response in between. However, many travel surveys that elicit travelers' perceptions through Likert rating scale questions do not include a "don't know" response category for the respondents to express their unfamiliarity and lack of opinion on the transit service. For this reason, some respondents who are not familiar with, and do not have an opinion about, the transit system are likely to choose the neutral response to Likert scale questions (more on this in Section 1.3). At the same time, respondents who are familiar with and/or informed about the transit system may also choose the neutral response to state their opinion neutrality. In this context, it is hypothesized that some respondents' unfamiliarity with and lack of opinion about transit services may be confounded with the informed perceptions of those who are familiar with transit. This is because those who are unfamiliar with the transit system are less likely to use it (see Schmitt *et al*., 2014) and more likely to state neutral responses than those who are familiar with the system in the absence of a "don't know" option (Kalton *et al*., 1980; Strugis *et al*., 2014; Lam *et al*., 2010). Ignoring such influence of travelers' unfamiliarity can potentially distort the scale of Likert variables, result in biased parameter estimates and misleading implications about the influence of perceptions of service quality on transit usage.

In view of the above discussion, the objectives of the current study are the following:

(1) To analyze the influence of individuals' socio-demographic characteristics and subjective perceptions of transit service quality on their use and extent of usage of public transit -- while accounting for the possible role of individuals' unfamiliarity with (and lack of opinion about) the transit system -- with an empirical application in the city of Bengaluru, India.

(2) To propose an approach to use commonly collected Likert scale variables of transit service quality perceptions to disentangle the influence of individuals' unfamiliarity (and lack of opinion) from that of their informed perceptions on transit usage.

(3) To examine the effects of ignoring the influence of unfamiliarity in the analysis of service quality perceptions on transit usage and the extent of usage.

To achieve the above objectives, we employ a generalized heterogeneous data model (GHDM) proposed by Bhat (2015) that can be used to jointly analyze different types of outcome variables (aka, endogenous variables) – nominal variables, ordinal variables, binary variables, etc. – while considering the influence of latent variables representing subjective perceptions on the outcome variables. The primary endogenous variable of interest in this model is a multinomial discrete choice variable that classifies respondents into four categories – non-users, regular users, occasional users, and incidental users – to quantify their use and extent of use of public transit. The utility functions of these different levels of transit use are specified as functions of latent variables representing subjective perceptions on transit service quality. These latent variables are identified using ordinal Likert scale variables derived from survey questions on travelers' perceptions about service quality. Additionally, since such questions did not include a "don't know" response, we utilized the neutral responses to several Likert rating scale questions and non-use of transit as indicators to identify a latent variable for the respondents' unfamiliarity with (and lack of opinion about) the transit service. By doing so, the proposed approach helps in disentangling some individuals' unfamiliarity with (and lack of opinion on) transit services from the informed perceptions of other individuals who are more familiar with the system.

The proposed model is applied to analyze individuals' usage and the extent of usage of bus transit in a major metropolis called Bengaluru, India. Another simpler model that does not consider the role of individuals' unfamiliarity and lack of opinion about transit service on their transit use is also estimated to demonstrate the value of the proposed model. It is worth noting here that the literature on transit usage in Indian cities is limited (Srinivasan *et al*., 2007; Das and Pandit, 2015; Suman *et al*., 2017; Sarkar and Mallikarjuna, 2018; Devika *et al*., 2020). To the authors' knowledge, this is perhaps the first study to shed light on individuals' use and the extent of use of transit in an Indian city while also focusing on the role of subjective perceptions and the above-discussed methodological issues in using Likert rating scale variables to do so.

In the rest of the paper, Section 1.2 positions the current study in the context of the literature on transit usage and extent of usage. Section 1.3 discusses in detail the use of Likert scale variables and the confounding effect of unfamiliarity with informed opinion neutrality in Likert scale responses (and provides literature support on these issues). Section 2 provides a descriptive analysis of the empirical data used for the study. Section 3 explains the methodological framework used for this investigation. Empirical results are presented in Section 4, along with their behavioral interpretations and policy implications. Section 5 summarizes the paper with concluding remarks and directions for further research.

1.2 Transit usage and extent of usage

The literature abounds with studies of travel mode choice and transit usage. Most of these studies employ multinomial choice models of mode choice, binary choice models of whether a person uses transit or not, direct demand models of transit ridership, or regional travel demand models (e.g., four-step models or activity-based models) to understand the influence of various objective and subjective factors on transit usage and/or ridership. However, such approaches do not shed light on the factors influencing the extent to which an individual uses transit over a horizon of multiple days. While it is important to understand what makes an individual use transit service or not, it is also important to understand the factors that can help increase the individuals' frequency of usage of transit service. The former helps in formulating policies aimed at attracting new users, while the latter helps in increasing the patronage of existing users (de Ona *et al.*, 2013; Allen *et al.* 2018).[1] To do so, however, it is not sufficient to analyze individuals' mode choices on a single trip or on a single day. This is because a non-negligible portion of travelers is likely to be occasional users of transit. And most mode choice studies tend to underestimate the occasional users because their transit usage is typically underrepresented in daily activity or travel behavior datasets used for mode choice analyses.[2] Yet only a sparse set of studies in the literature focus on individuals' extent of usage of public transit (Srinivasan *et al*., 2007; De Vos and Witlox, 2017; Ingvardson and Nielsen, 2019; Bose and Pandit, 2020). And even if they do, most of these studies do not consider the influence of subjective perceptions of service characteristics on the extent of transit usage.

To distinguish individuals' choice to use transit from the extent of their usage, this study classifies individuals into four categories – non-users (those who do not use transit), regular users

[1] The issue of retaining and enhancing the patronage of existing users is relevant not only for public transit agencies, but also for emerging mobility modes such as ride-hailing services.

[2] Statistically speaking, occasional users of transit would not be underestimated if large samples of daily travel data are used. However, in situations where transit usage itself is rather small (e.g., many cities in the United States) and/or where the analysts must work with limited sample sizes due to resource limitations, occasional transit users are likely to be underrepresented.

(those who use transit at least three times a week), occasional users (those who use transit once or twice a week), and incidental users (those who use transit once or twice a month). Doing so helps in identifying which demographic segments are likely to belong to which category – non-users, incidental users, occasional users, and regular users – as well as inform the formulation of policies to attract new riders as well as enhance transit use by existing riders.

1.3 Likert rating scale responses and unfamiliarity

Typically, the possible responses to Likert scale questions involve an ordinal rating scale ranging from one extreme of strong agreement to another extreme of strong disagreement, with a neutral or moderate response in between (Likert, 1932). An ordinal rating scale of responses to Likert scale questions assumes that the underlying perception is a unidimensional continuum that can be mapped to one of the ordinal response categories. A shortcoming with this assumption is that all respondents are assumed to be familiar with and/or have an opinion about the transit service being examined. For example, the respondents who provide a neutral response are assumed to have formed a neutral opinion after considering their experience and/or the information they may have about the transit service (even if the information is from secondary sources such as news, social media, or word-of-mouth). This assumption ignores that not all respondents who chose the neutral response do so based on true opinion neutrality. However, some of the neutral respondents may not have much information/knowledge and do not have an opinion about the service. In the absence of a "don't know" option as a possible to response to questions about service quality, such respondents are more likely (than those who are familiar with or have an opinion on the service) to choose the neutral option as a way of saying they do not know or do not have an opinion. As documented in a few earlier studies (Kalton *et al*., 1980; Lam *et al*., 2010; Strugis *et al*., 2014; Sheela *et al*., 2018), such a response is not true opinion neutrality and does not act as a transition between negative and positive perceptions/opinions. For example, Strugis *et al*. (2014) administered follow-ups with respondents who initially selected the mid-point responses to Likert scale questions in their survey. Based on the analysis of the follow up responses they report the following:

> "*…the vast majority of responses turn out to be what we term 'face-saving don't knows' and that reallocating these responses from the mid-point to the don't know category significantly alters descriptive and multivariate inferences.*"

In the studies conducted by Kalton *et al*. (1980) and Lam *et al*. (2010), when the "don't know" alternative was provided along with a five-point scale that included a neutral option, the proportion of neutral responses decreased much more than the proportion of non-neutral responses. These results add further evidence that people uninformed about a topic are more likely to choose the neutral response if they do not have an option to express their unfamiliarity or lack of knowledge about the topic. In this context, Krosnick (1991) suggests that individuals who have little relevant knowledge in their memory about a topic to execute the cognitive procedures necessary during a survey process are likely to pick a neutral option as a way out (in the absence of a "don't know" option). Based on the evidence from these studies, our study hypothesizes that in response to Likert scale questions to measure perceptions of transit service, some respondents' unfamiliarity with (and lack of opinion on) transit services can be confounded with the neutral perceptions of those who have informed opinion about such services. Such confounding can potentially distort the ordinal scale of Likert variables, which, if ignored, can result in erroneous estimates of the influence of perceptions of service characteristics on transit usage. Therefore, it is important to disentangle the effect of travelers' unfamiliarity with (and lack of opinion on) transit services from the informed perceptions/opinions of those familiar with or have an opinion on the transit service. Besides, identifying those who are unfamiliar with and/or do not have an opinion on transit service may also help in devising suitable awareness campaigns.

2. EMPIRICAL DATA DESCRIPTION

The empirical context of the above-discussed analysis is a major metropolis called Bengaluru, one of the rapidly growing cities in India, with a population of more than 13 million served by a large bus transit network with more than 6,000 buses operating in the city, along with a metro rail system covering 42.3 km. The city has witnessed a declining mode share of public transit and particularly bus transit over the past several years (Badami and Haider, 2007; Manoj and Verma, 2015; Devika *et al.*, 2020). Therefore, it would be useful to understand the role of objective and subjective factors influencing Bengalurians' bus transit usage and its extent.

The empirical data for the current work comes from a national level survey titled "Ease of Moving Index" survey conducted by the Ola Mobility Institute in 20 cities across India in 2018. The survey questionnaire comprises 52 questions, split into the following four sections:

(a) *Know the respondent* section that elicits respondents' socio-demographic information,

(b) *Direct evaluation* section that elicits respondents' usual mode of travel, frequency of use of each mode, stated reasons for using or not using public transit,

(c) *Objective evaluation* section that elicits measurable mode-specific attributes such as travel time, travel cost, travel distance, and access modes used for travel, and

(d) *Comparative evaluation* section that elicits individuals' attitudinal information and perceptions of transit service characteristics on a Likert rating scale.

The sample data chosen for the current work comprises 2,413 respondents from Bengaluru. Table 1 presents a descriptive analysis of this sample. The dependent variable of interest – transit usage and the extent of usage – is a nominal variable with the following categories: regular users, occasional users, incidental users, and non-users (*i.e.*, those who never use bus transit). Most of the respondents in the sample (77%) belong to the non-user segment, whereas close to 9 percent of the respondents belong to occasional or incidental users. It is worth noting here that the mode choice data typically used to analyze travel mode choice often miss a non-negligible portion of the occasional or incidental users because the data include mode choices on only a day.

As can be observed from the table, the sample of respondents comprises a larger proportion of men than women, close to 70% of individuals with at least a bachelor's level of education, and a larger percentage of employed individuals or students than unemployed individuals or homemakers. Compared to the demographic makeup of Bengaluru's population, the sample has

Table 1. Data descriptive statistics

Dependent variable	**Count**	**Percentage**
(Market segments based on use and extent of use of transit)		
Regular users (who use transit 3 or more times in a week)	338	14.00%
Occasional users (who use transit once or twice in a week)	165	6.84%
Incidental users (who use transit once or twice in a month)	51	2.11%
Non-users (never use transit)	1859	77.04%

Exogenous variables	**Count**	**Percentage**	***Ordinal indicator variables***	**Count**	**Percentage**
Gender			*Indicators of Cleanliness*		
Female	983	40.74%	Extremely dirty	67	2.78%
Male	1353	56.07%	Somewhat dirty	422	17.49%
Transgender	77	3.19%	Neutral	1020	42.27%
Age			Somewhat clean	675	27.97%
<20 years	524	21.72%	Very clean	229	9.49%
20 – 40 years	1137	47.12%	*Indicators of Comfort*		
40 -60 years	524	21.72%	Extremely uncomfortable	115	4.77%
> 60 years	228	9.45%	Somewhat uncomfortable	562	23.29%
Educational qualification			Neutral	1060	43.93%
< 10^{th}	48	1.99%	Somewhat comfortable	452	18.73%
10 – 12th pass	464	19.23%	Very comfortable	224	9.28%
Graduate	992	41.11%	*Indicators of Reliability*		
Postgraduate	682	28.26%	Service is erratic throughout the day	750	31.08%
Doctoral and above	227	9.41%	Service is erratic for initial & last hrs.	939	38.92%
Employment status			Services are somewhat reliable	459	19.02%
Employed	811	33.61%	Services are reliable	265	10.98%
Student/ studying	705	29.22%	*Indicators of Safety*		
Unemployed	320	13.26%	Extremely unsafe	53	2.20%
Homemaker	577	23.91%	Somewhat unsafe	336	13.92%
Monthly income (INR per month)			Safe except at night	794	32.91%
< 15,000	48	5.92%	Somewhat safe	874	36.22%
15,000-30,000	295	36.37%	Very safe	356	14.75%
30,000-50,000	325	40.07%	*Indicators of Affordability*		
50,000-1,00,000	130	16.03%	Extremely expensive	53	2.20%
>1,00,000	13	1.60%	Somewhat expensive	458	18.98%
Household car ownership			Neutral	1033	42.81%
Zero car	1224	50.73%	Somewhat affordable	606	25.11%
One car	832	34.48%	Very affordable	263	10.90%
Two cars	312	12.93%			
Three cars or above	45	1.86%	*Neutral responses to Likert scale questions*		
Household two-wheeler ownership			Neutral to affordability	1033	42.81%
Zero two-wheeler	1185	49.11%	Neutral to cleanliness	1020	42.27%
One two-wheeler	868	35.97%	Neutral to comfort	1060	43.93%
Two two-wheelers	316	13.10%			
Three or above	44	1.82%			

an overrepresentation of the more educated and the employed or student segments. This may be one reason why the dependent variable shows a high percentage of non-users of public transit.

In the context of ordinal indicator variables of the respondents' perceptions of transit service, the right-side rows in Table 1 show the distributions of individuals' responses to questions on cleanliness, comfort, reliability, safety, and affordability. Note that the questions on cleanliness, comfort, and affordability had 5-point response scale that included the neutral response category. However, the question on perception of reliability had a 4-point response scale without the neutral category. And for the question on the perception of safety, although with a 5-point response scale, the middle category cannot necessarily be treated as neutral because this category treats public transportation as unsafe at all times during the night. As will be discussed later, the neutral responses to the questions on cleanliness, comfort, and affordability were used as binary indicators of the latent variable called unfamiliarity with the transit system. Therefore, the percentage of the neutral respondents to each of these questions are reported separately under the heading titled "*Neutral responses to Likert scale questions*" in the table. It is noteworthy that at least 42% of the individuals chose the neutral response to at least of these questions.

In this context, Figure 1 shows the percentages of transit users and non-users separately for neutral respondents and non-neutral respondents for each of the three questions used to obtain indicators for unfamiliarity. Interestingly, for any of these three questions, among those who gave a neutral response, the percentage of non-users of transit is close to 80%. On the other hand, among those who gave a non-neutral response, the percentage of non-users of transit is less than 75%. These statistics offer an inkling that those who state neutral responses to questions on transit service attributes are more likely to be non-users of transit than those who give non-neutral responses. Based on our discussion in Section 1.3, it is likely that some of the neutral respondents include those who are unfamiliar with transit services in the city. Their unfamiliarity may not only cause them to stay neutral to questions on their perceptions of transit service, in the absence of a "don't know" option, but also a reason for them not using transit. If so, it becomes important to separately account for the influence of unfamiliarity on transit usage to understand the influence of transit service perception variables on transit usage.

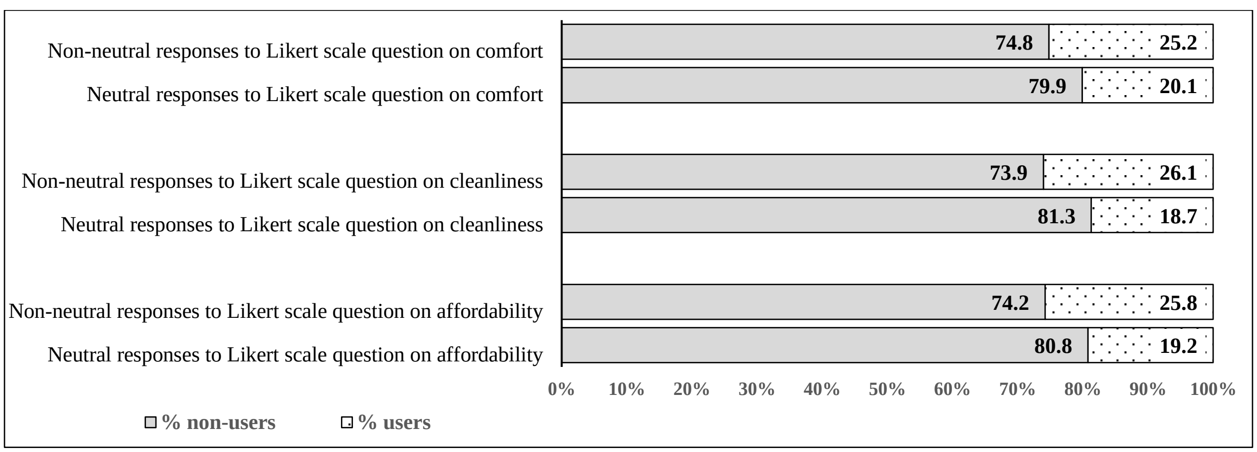

Figure 1. Percentage share of neutral and non-neutral responses among transit users and non-users

From a policy standpoint as well, it is useful to understand the influence of unfamiliarity with transit services on transit use because such unfamiliar respondents might respond differently after they gather information (Sheela and Mannering, 2019; Shelat *et al*., 2021).

3. MODELING FRAMEWORK

3.1 Structure of the Proposed Model

Figure 2 presents a schematic of the overall model structure, identifying the observed exogenous variables in a solid rectangle, observed endogenous outcomes in dashed rectangles, latent variables in ovals, structural relationships between variables in solid arrows, and measurement relationships in dashed arrows.

The primary endogenous variable of interest is a nominal variable representing an individual's usage and extent of usage of transit, categorized into the following classes: *non-user, incidental user, occasional user, and regular user*. This endogenous variable is influenced by exogeneous covariates (such as socio-demographic variables) as well as the individual's latent perceptions of the service characteristics of the transit system – cleanliness, comfort, reliability, safety, and affordability[3] – and their (un)familiarity with the transit system. The other observed endogenous variables, which are used as indicators to identify the latent variables, are represented in the dashed rectangles on the right side of the figure. These are: (a) four Likert rating scale variables of the individual's response to questions on perceptions of cleanliness, comfort, reliability, and safety and (b) three binary indicator variables representing if the individual provided a neutral response to Likert rating scale questions on affordability, cleanliness, and comfort.

We use the Generalized Heterogeneous Data Model (GHDM) developed by Bhat (2015) to jointly analyze the above-mentioned endogenous variables while also identifying the latent variables of interest. The model structure allows correlations among the latent variables, which are not shown in the figure to avoid clutter. The proposed model is described in the form of its components, as follows: (1) latent variable structural equations for individuals' perceptions of, and unfamiliarity with, the transit service, (2) measurement equations for the latent, perception variables, (3) measurement equations for the latent, unfamiliarity variable, and (4) the choice model for transit usage and extent of usage. Each of these components is described next.

[3] In the empirical models we explored, the latent construct for affordability did not show a statistically significant effect on the endogenous variable of interest (transit usage and the extent of usage). Therefore, we removed that latent construct from further consideration in the models (and in the subsequent discussion in the paper, including Figure 2). This helped conserve some space in the figure as well.

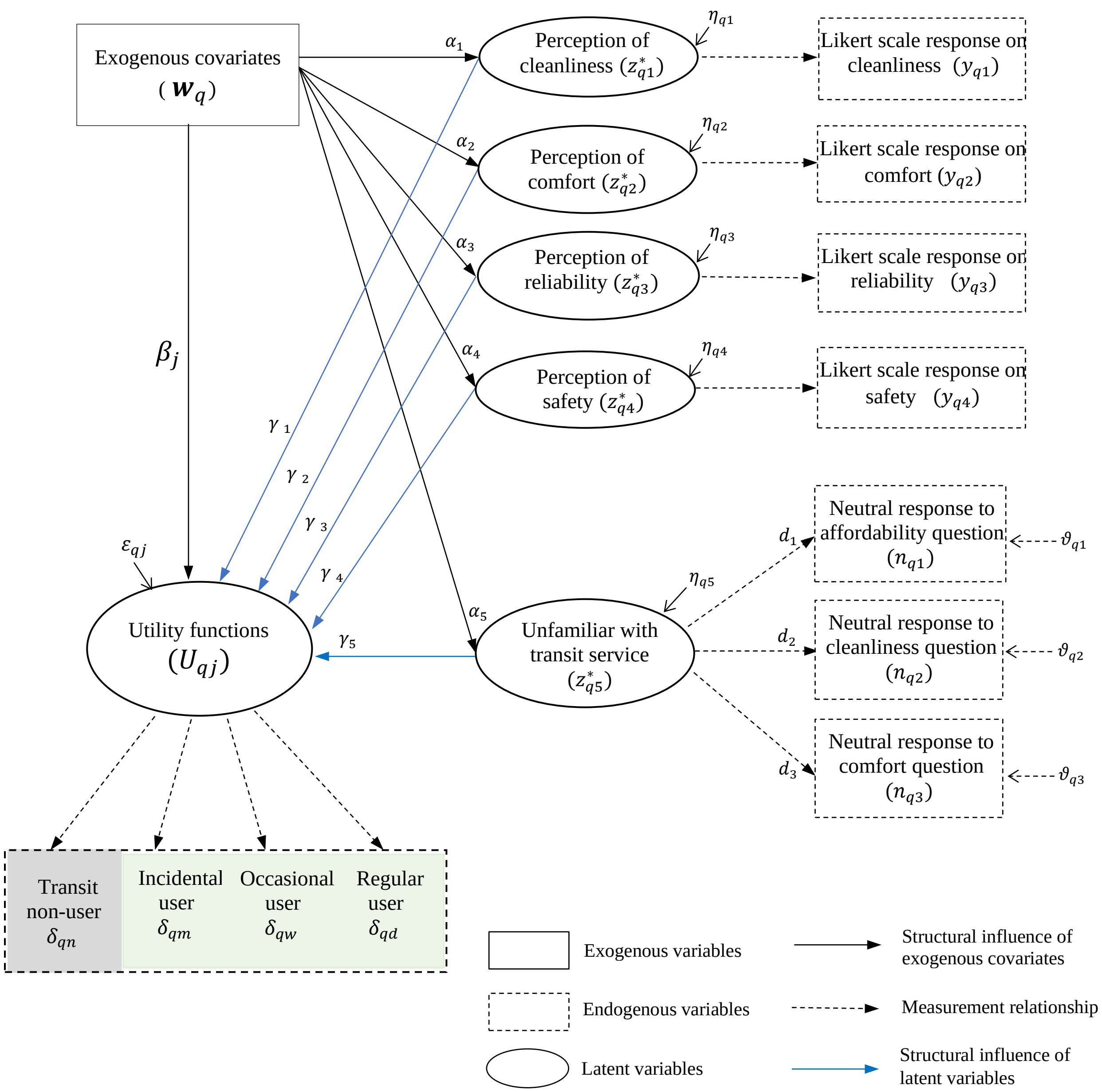

Figure 2. Joint framework to model individuals' perceptions of transit service, their unfamiliarity with the transit system, and their transit usage and extent of usage

3.1.1 Latent Variable Structural Equation Model (SEM)

As illustrated in Figure 2, the SEM component of the framework defines each latent variable for individuals' perceptions of, and unfamiliarity with, the transit system as linear functions of observed exogenous variables and stochastic error terms. In the following discussion, let l be the index for latent constructs and q be the index for individuals. The first four latent constructs ($l = 1,2,3,4$) used in the study represent perceptions on cleanliness, comfort, reliability, and safety respectively. The fifth latent variable ($l = 5$) represents unfamiliarity. In general, the structural equation for a latent construct z^*_{ql} can be written as:

$$z_{ql}^{*} = \boldsymbol{\alpha}_l{}' \boldsymbol{w}_q + \eta_{ql} \tag{1}$$

where $\boldsymbol{w}_q$ is a $(D \times 1)$ vector of observed covariates (excluding a constant), $\boldsymbol{\alpha}_l$ is the corresponding $(D \times 1)$ vector of coefficients to be estimated, and η_{ql} is a random error term assumed to be standard normal distributed. Although the same exogenous variable vector $\boldsymbol{w}_q$ is used for all latent variables in our notation, one can accommodate that some variables in $\boldsymbol{w}_q$ do not influence a latent variable z_{ql}^{*} by setting the corresponding element of the $\boldsymbol{\alpha}_l$ vector to zero.

3.1.2 Latent Variable Measurement Equation Model (MEM) for Perception Variables

The latent variables are not directly observed by the analyst. However, observed endogenous variables, including psychometric and other indicators, can be posited as manifestations of the latent constructs to allow their identification. Such indicators in the current study are responses to Likert rating scale survey questions regarding individuals' level of agreement on the extent of cleanliness, comfort, reliability, and safety of the transit service. The indicators being on a Likert scale, we use the familiar ordered response model framework where the latent variable for a specific perception (z_{ql}^{*}) is treated as a latent propensity variable that manifests in the form of the response (y_{ql}) to the corresponding Likert rating scale question. Specifically, the latter (y_{ql}) is mapped to the former (z_{ql}^{*}) through the threshold parameters (ψ_l^k) used in a typical ordered response modeling framework, as follows:

$$y_{ql} = r \quad if \quad \psi_l^{r-1} < z_{ql}^{*} < \psi_l^{r};\ \psi_l^{0} = -\infty,\ \psi_l^{R} = +\infty;\ \forall l = 1,2,3,4 \tag{2}$$

In the above equation, r denotes the five possible responses a survey respondent can choose for the Likert rating scale question, including the neutral response.

Most earlier applications (Allen *et al.* 2018; de Ona *et al.*, 2013; Ingvardson and Nielsen, 2019) of latent variable models, use a multiple indicator multiple cause (MIMIC) model formulation, in which multiple indicators are modeled as manifestations of each latent variable, and multiple causes (such as socio-demographic variables) affect latent variables. This is because the latent variables are viewed as nebulous constructs that are a combination of multiple facets represented by multiple indicators. In fact, if several such indicators are available to measure a few latent constructs, it is common to undertake factor analysis to reduce multiple indicators into fewer factors aka, latent variables. In contrast to such a MIMIC approach, it might appear from the above formulation that each of the four latent perception variables (perceptions of cleanliness, comfort, reliability, and safety of transit service) is identified using only a single indicator variable – individuals' response to the corresponding Likert rating scale questions regarding their level of agreement on the extent of cleanliness, comfort, reliability, and safety of transit service. However, as shown in Figure 2, and discussed later in Section 3.1.4, the utility functions of individuals' extent of transit use are also expressed as functions of the latent perception variables. Therefore, the observed nominal variable for the use and extent of transit use also serves as an extra indicator for these latent perception variables. Of course, if the empirical data had additional indicators measuring individuals' perceptions of service quality, it would be possible to extend the formulation to consider those additional indicators to identify latent perception variables (or conduct a factor analysis to reduce many indicators into fewer factors). However, since the empirical data did not have multiple indicators available specifically for each latent perception variable, and since the Likert scale questions directly asked individuals about transit service quality attributes, we interpreted latent propensity variables (z_{ql}^*) used to model the observed ordinal responses (y_{ql}) to the Likert scale questions as the latent perception variables.

3.1.3 Latent Variable Measurement Equation Model (MEM) for Unfamiliarity

Here we discuss the MEM formulation for the fifth latent construct, which is individuals' unfamiliarity with transit services. As discussed earlier, individuals' unfamiliarity with transit services can be confounded with informed opinions about perceptions on the quality of transit service. In the absence of a separate question on unfamiliarity, a latent construct labeled unfamiliarity with transit service is identified using individuals' neutral responses to the Likert rating scale questions on affordability, cleanliness, and comfort.

Let us define binary indicator variables n_{qi} ($i = 1,2,3$), where $n_{qi} = 1$ if the individual provides a "neutral" response to the corresponding Likert scale question, 0 otherwise. These neutral response indicators are modeled as binary outcome manifestations of the latent variable for unfamiliarity (z_{q5}^*), as follows:

$$n_{qi} = I\left[d_i{}'\, z_{q5}^* + \vartheta_{qi} > 0\right]; \quad i = 1,2,3\,. \tag{3}$$

In the above equation, $I\,[\,.\,]$ is an indicator function, z_{q5}^* is the latent construct "unfamiliar with transit service", d_i is the loading on z_{q5}^* by the i^{th} neutral indicator variable, and ϑ_{qi} a standard normal distributed measurement error term.

As will be discussed in the next section, the latent variable for individuals' unfamiliarity is used as an explanatory variable in the utility function for individuals not using transit. Therefore, the latent unfamiliarity variable is identified using the above-discussed neutral responses to Likert scale variables as well as the observed variable for individuals' non-use of transit (which is a part of the primary endogenous variable of interest in this study).

3.1.4 Choice Model

The choice model is for the primary endogenous variable of interest – transit usage and extent of usage – with a choice-set $\boldsymbol{J} = \{n, m, w, d\}$ whose elements denote the alternatives non-user, incidental user, occasional user, and regular user, respectively. The structural relationship in the choice model is encapsulated in the utility functions. Specifically, an individual's utility of an alternative 'j', U_{qj} is expressed as a linear function of observed exogenous covariates ($\boldsymbol{w}_q$), and latent constructs (z_{ql}^*) that include the perception variables and the unfamiliarity variable as follows:

$$U_{qj} = \beta'_j\, \boldsymbol{w}_q + \gamma_{1j}\, z_{q1}^* + \gamma_{2j}\, z_{q2}^* + \gamma_{3j}\, z_{q3}^* + \gamma_{4j}\, z_{q4}^* + \gamma_{5j}\, z_{q5}^* + \varepsilon_{qj} \tag{4}$$

In the above utility function, the error terms ε_{qj} are assumed to be multivariate normally distributed with mean zero and a covariance matrix identifiable in typical probit-based discrete choice models.

The measurement relationship assumes utility-maximizing behavior. To explain this, define δ_{qj} as the observed choice indicator for individual 'q' taking the value 1 if alternative 'j' is chosen, and 0 otherwise. The individual is assumed to choose alternative 'j' if it offers the highest utility, as in the equation below:

$$\delta_{qj} = I[U_{qj} > U_{qi}\,;\ \forall i \in \boldsymbol{J} \text{ and } i \neq j] \tag{5}$$

As discussed in earlier sections, the fact that the latent perception variables and the unfamiliarity variable are used to explain individuals' use and extent of use of transit, the corresponding observed variable ($\delta_{qj}\ \forall j \in \boldsymbol{J}$) serves as an extra indicator to help identify the latent variables.

3.2 Model System Estimation

The estimation of the parameters proceeds with setting up the joint likelihood function for all the observed endogenous variables and maximizing the likelihood function. In doing so, the parameters defining the latent variables as well as their relationships with the exogenous and endogenous variables are identified. To conserve space, we do not describe the likelihood setup here. Interested readers may refer to the paper by Bhat (2015) for details on the setup of the GHDM likelihood function and recent papers by Sharda *et al*. (2019) and Moore *et al*. (2020) for other similar applications of the GHDM framework. Note that the GHDM framework, in its original form, allows for the joint modeling of multiple endogenous outcomes of different types, including continuous, ordered, count, nominal, rank-ordered, and discrete-continuous choice outcomes. The specific formulation in this paper is similar to a traditional integrated choice and latent variable (ICLV) framework because the primary endogenous outcome is a nominal variable and the other endogenous outcomes are either ordinal or binary variables. However, the GHDM model employs multivariate normal distributions for the stochastic terms in the formulation, because the logit kernel based ICLV models typically encounter challenges during model estimation as the likelihood functions increase in their dimensionality of integration with the number of indicators (i.e., endogenous variables) used in the model. This issue is circumvented with the GHDM framework because the likelihood function involves multivariate normal cumulative distribution functions whose dimensionality does not rise with the number of indicators used in the model (see Bhat, 2015 for more details).

4. EMPIRICAL MODEL RESULTS AND FINDINGS

4.1 Structural equations for perceptions and unfamiliarity

Table 2 presents the parameter estimates of the structural equations for the latent variables – perceptions of transit service characteristics and unfamiliarity with transit service. As can be seen from this table, women tend to have greater negative perceptions regarding the cleanliness of buses than men. Interestingly, women and men have similar perceptions of comfort, reliability and safety of traveling in Bengaluru's bus transit system. Age-related parameter estimates suggest that those who are younger than 40 years have greater negative perceptions about the cleanliness, comfort, and reliability of traveling by bus than older individuals.

Table 2. Structural equations for perceptions and unfamiliarity

	Cleanliness	Comfort	Safety	Reliability	Unfamiliarity
Determinants of latent variables					
Female	-0.1004 (-2.80)	-	-	-	-
Age 20 years and less	-0.1030 (-1.66)	-0.0761 (-1.27)	-	-0.0798 (-1.18)	0.4410 (4.20)
Age between 20 and 40 years	-0.1219 (-2.50)	-0.0936 (-1.90)	-	-0.0435 (-0.85)	0.1020 (1.37)
Age between 40 and 60 years	-	-	0.1069 (2.02)	-	-
Age 60 years and more	-	-	0.0855 (1.27)	-	-
Education < 12^{th} grade	-	-0.0743 (-1.04)	-	-	-0.3530 (-4.15)
Education: Bachelors	-0.1551 (-3.50)	-0.1362 (-3.33)	-	-	-
Education: Postgraduate or more	-	-	-	-0.1831 (-3.97)	-
Current status is student	-	-	-	-0.1034 (-1.97)	-0.2180 (-2.51)
Household (HH) car ownership = 0	-	-	0.0439 (1.19)	-	-
HH car ownership > 1	-0.1649 (-3.04)	-0.1144 (-2.12)	-	-	-
HH two-wheeler ownership = 0	-	-	0.1797 (4.96)	-	-
Correlations among structural equations					
	Cleanliness	**Comfort**	**Safety**	**Reliability**	**Unfamiliarity**
Cleanliness	1	0.5690 (38.37)	0.5628 (37.35)	0.3542 (18.18)	0.1921 (4.38)
Comfort		1	0.6492 (51.82)	0.3804 (20.17)	0.1081 (2.65)
Safety			1	0.3387 (17.07)	0.2411 (5.87)
Reliability				1	0.2940 (7.00)
Unfamiliarity					1

Note: For each parameter estimate, t-statistic is presented in parentheses below the estimated value.

Interestingly, those older than 40 years tend to think of transit as safer than younger individuals. Also, those who have at least a post-graduate degree and those who are currently students tend to have greater negative perceptions than others on the reliability of bus transit, presumably because they may be facing greater consequences of '*not being on time*' than others. In terms of the influence of vehicle ownership on perceptions, those who do not own a car or a two-wheeler in their household tend to have a greater positive perception of safety of bus transit. This is likely a manifestation of cognitive dissonance (Ory and Mokhtarian, 2005), where perceptions of such individuals are possibly shaped out of their regular use of transit. On the other hand, individuals' perceptions of transit cleanliness and comfort tend to become more negative with increasing car ownership. This result is consistent with past studies on transit usage (Thompson *et al*., 2002; Ingvardson and Nielsen, 2019).

Now, we turn to the effects of exogenous variables on individuals' unfamiliarity with bus transit service. The parameter estimates indicate that individuals of age less than 40 years, people with more than 12th grade (high school) education, and those who are not students exhibit a greater level of unfamiliarity with bus transit service characteristics than their respective counterparts. While the reasons behind these results are not readily apparent, it is evident that demographic differences exist in the levels of (un)familiarity with transit services. To the extent that familiarity plays a role in the use of transit (more on this in the next section), it will be useful in the future to identify such differences and understand the reasons for the differences.

Next, we focus on the matrix of correlations among the latent variable equations (last set of rows in Table 2). It is apparent from this matrix that all latent variables representing perceptions of transit service cleanliness, comfort, safety, and reliability exhibit significant correlations. These correlations suggest the presence of various unobserved factors that can either improve or deteriorate individuals' perceptions of transit service characteristics. It will be a fruitful avenue in the future to identify those unobserved factors that might help improve traveler's perceptions of all four service dimensions – cleanliness, comfort, safety, and reliability. The latent variable for unfamiliarity also exhibits correlations with the latent perception variables, albeit the estimated correlation levels are lower than those among the perception variables (the corresponding t-statistic values are smaller as well). Even though the latent variable for unfamiliarity has been identified from the same set of Likert scale variables used for measuring the latent perception variables, the low levels of correlations between the unfamiliarity variable and perception variables suggest that the proposed approach works well for disentangling travelers' unfamiliarity with the transit service from their informed perceptions about the service. Since many travel surveys do not include a "don't know" option in Likert scale response nor include questions on travelers' unfamiliarity with a service, this approach can potentially be useful in several other applications involving the role of perceptions on travel behavior. Readers may note that to save space, we do not report the threshold values that map the latent perceptions to the observed Likert rating scale indicators. We also do not report the loadings of the unfamiliarity latent variable in the measurement equations for neutral response indicators. This is because these parameters do not offer any substantive interpretations and are not central to the focus of the current paper.

4.2 Direct influence of exogenous variables on use and intensity of use of bus transit

Table 3 reports the parameter estimates of the multinomial choice model for transit use and extent of usage (the transit non-usage category is the base in the analysis). In the context of the demographic variables, the data does not show significant gender differences in the use and extent of use of transit services in Bengaluru (more on gender differences in the following section). Individuals younger than 40 years are found to be more likely to use bus transit on a regular basis than others, while individuals above 40 years are more likely to use transit on an occasional basis relative to others. Individuals with lower than high school education are more likely to be incidental users than others. Interestingly, employed individuals are more likely to be either regular users or incidental users than others, but their likelihood of using transit on an occasional basis is not different from that of not using transit. As expected, students who choose transit are more likely to be regular users than others. In the context of car ownership, increasing car ownership levels show an increasingly deteriorating influence on the propensity of using transit at higher frequencies. This result confirms the usual belief that increasing car ownership levels is perhaps a major reason for non-usage or decreased usage of transit (Thompson *et al*., 2002).

4.3 Influence of latent perceptions on use and intensity of use of bus transit

Next, we turn to the influence of perception variables.[4] As expected, those who perceive transit to be clean, safe, and reliable are more likely to use it than those who do not have positive perceptions along these dimensions. Furthermore, the positive coefficients on the reliability perception variable – with their magnitudes increasing from the utility functions of incidental use to regular use – indicate that positive reliability perceptions not only make people more likely to choose transit but also make them use it more often. This result highlights the importance of planning and operational strategies that enhance the reliability of transit systems (and improve travelers' perceptions of reliability of transit service). Unexpectedly, the coefficient of the comfort perception variable appears with a negative sign in the regular use utility function. The reason behind this result is not clear and warrants further investigation. However, note from the interaction term of perception of comfort with individuals in the 20-40 years age group that enhancing perceptions of comfort in this middle-aged group seems to increase the likelihood of using transit occasionally. In other words, the young riders consider comfort as an essential service quality while making transit choice decisions. Similarly, from the interaction of the safety perception variable with gender, we find that enhancing safety perceptions among women increases their transit usage, a result consistent with earlier studies (Reed *et al*., 2000; Verma *et al*., 2017). It is worth noting here that the interaction of the safety perception variable with gender has a significant influence only in the occasional use and regular use utility functions but not in the incidental use utility function.

---

[4] In this context, note from Table 3 that the coefficients of the latent perception variables were retained even if they were not statistically significant at the 90% confidence level. This is because large sample sizes are needed to estimate the effects of latent variables with a high precision (Bhat and Dubey, 2014). Further, as will be discussed next, most of the parameter estimates provided behaviourally plausible interpretations on the influence of perceptions on transit usage. Therefore, we retained the parameter estimates as long as the corresponding t-statistic was at least 1.0, with an idea that future studies can investigate these effects with larger data samples.

Table 3. Choice model component for use and extent (frequency) of use of bus transit

Explanatory variables	Incidental use Parameter (t-stat)	Occasional use Parameter (t-stat)	Regular use Parameter (t-stat)
Constant	-3.0486 (-7.48)	-1.8584 (-5.97)	-2.1447 (-5.65)
Socio-demographic variables			
Gender (base: Male)			
Female	-	-	-
Age (base: 60 years and above)			
Age 20 years and less	-	-	0.6472 (3.62)
Age between 20 and 40 years	-	-	0.2333 (1.86)
Age between 40 and 60 years	-	0.2308 (2.01)	-
Educational qualification (base: PG or above)			
Educational qual. is less than 12^{th} standard	0.6454 (3.73)	0.2707 (2.49)	-
Educational qual. is graduation	-	-	-
Employment status (base: unemployed)			
Employed	0.4814 (3.23)	-	0.8475 (6.56)
Student	-	-	0.6437 (4.54)
Household car ownership (base: zero car)			
Household owns one car	-	-1.0268 (-6.07)	-1.4715 (-7.90)
Household owns more than one car	-0.9007 (-2.17)	-2.0288 (-5.15)	-2.2302 (-5.60)
Latent variables			
Perception of cleanliness	-	-	0.2180 *(1.12)
Perception of comfort	-	-	-0.4913 (-2.17)
Perception of safety	0.4227 (2.81)	-	0.3022 *(1.36)
Perception of reliability	0.2088 *(1.17)	0.5471 (3.73)	0.6576 (4.15)
Unfamiliarity with transit service	-0.8133 (-2.77)	-0.8133 (-2.77)	-0.8133 (-2.77)
Interacted variables			
Perception of comfort × Age btw 20 - 40	-	0.0826 *(1.24)	-0.1719 *(-1.44)
Perception of safety × Female	-	0.4700 (2.98)	0.4268 (2.09)
Performance measures			
Number of respondents	2413		
Log-likelihood at convergence	-20,416.39		
Predictive log-likelihood for the choice component	-1474.49		
Rho-square value of the choice model component	0.176		
Akaike Information Criterion (AIC) value	3124.97		
Bayesian Information Criterion (BIC) value	3432.36		

* not statistically significant at 90% confidence level but still retained in the model for behavioral interpretations

The implication of this result is that enhancing safety perceptions among women increases the frequency of their transit usage (i.e., women are more likely to become occasional or regular users than incidental users or non-users). Such a finding demonstrates the benefits of analyzing the extent of transit usage.

4.4 Influence of unfamiliarity

The last latent variable used to explain transit usage (and extend of usage) is unfamiliarity with the transit system. Consistent with the results from past studies (Thompson *et al*., 2002), the negative coefficient of this latent variable suggests that a greater level of unfamiliarity with transit services reduces an individual's likelihood of using transit. However, as can be observed from the equality of the coefficients across the different levels of transit usage, we did not find significant differences in the influence of the unfamiliarity on different levels of transit usage. The reasons for lack of differential effect of unfamiliarity on regular usage, occasional usage, and incidental usage are not clear. Future research should examine the reasons for this finding as well as accumulate additional empirical evidence on the influence of unfamiliarity on the different levels of transit usage.

To examine the repercussions of ignoring the influence of unfamiliarity, we estimated another empirical model without including the unfamiliarity variable to explain transit usage. The estimation results of this simpler model are not reported in a table format to conserve space, but the important findings are discussed here. Not accounting for the influence of unfamiliarity lead to a decreased influence of perception variables on transit usage and extent of usage. Specifically, in the model without the unfamiliarity variable, the coefficients of cleanliness and comfort did not turn out to be even marginally significant (the t-statistics were smaller than 1.0) whereas the safety perception coefficient had a smaller t-statistic (when compared to the corresponding coefficients in the model that includes the influence of unfamiliarity). Further, the reliability perception variable did not have a significant influence on the utility function of incidental (monthly) use. This is because, as discussed earlier, ignoring the influence of unfamiliarity leads to its confounding into the effect of perception variables on transit usage. Therefore, it is important to disentangle unfamiliarity and its influence from that of informed opinions in models of travel behavior. In addition, these results point to the importance of introducing an option for "don't know" as a response to Likert scale perception questions and additional questions to elicit individuals' unfamiliarity with the topics of Likert scale questions on perceptions and attitudes.

It is worth noting here that the survey questions on individuals' perception of reliability and safety did not include a neutral response option. As can be observed from Table 1, the reliability question had only a 4-point scale response and the middle category response to safety question cannot be categorized as neutral response. Nonetheless, the model with the unfamiliarity latent variable helped improve the estimates of the latent variables corresponding to the perceptions of reliability and safety on transit usage (as compared to the model without the unfamiliarity latent variable). This is because, once we were able to identify the latent construct for unfamiliarity using the neutral responses for the other three perception questions, the confounding of some respondent's unfamiliarity could be disentangled from the effect of other respondent's informed perceptions on their transit usage. As a result, the influence of the reliability and safety perceptions on transit usage became more discernible – even though the reliability and safety related Likert scale questions in the survey did not have a neutral response option.

4.5 Goodness-of-fit measures

The proposed model that attempts to separate the influence of unfamiliarity from that of perception variables in transit usage has a log-likelihood value of -20,416.39. The choice model component of this model has a log-likelihood value of -1,474.49. To understand the benefit of separating the influence of unfamiliarity from that of perceptions of transit service characteristics, the goodness-of-fit of this model is compared to that of a simpler model that does not include the latent variable for unfamiliarity in the framework (but includes all other perception latent variables). Since the specification of the other variables is not the same between the two models, the log-likelihood ratio could not be used for performing a direct statistical comparison. Therefore, we compare the Rho-square, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) metric values between the choice model components of the two models. The Rho-square, AIC and BIC values of the choice model component of the proposed model that considers unfamiliarity are 0.176, 3124.97, and 3432.36, respectively. The corresponding values for the model that does not consider unfamiliarity are 0.166, 3134.47, and 3574.41, respectively. These values suggest a better data fit of the model that considers unfamiliarity. Further, the average probability of correct prediction of the choice component for the model that considers unfamiliarity is 0.7214, whereas that for the model that ignores unfamiliarity is 0.6929, reinforcing the results from the other goodness-of-fit measures. In summary, the improved statistical performance of the model that includes unfamiliarity, and the findings presented earlier in Section 4.4 highlight the importance of separating the effect of unfamiliarity from that of perception variables.

4.6 Policy implications

The empirical model offers useful insights on who uses bus transit and who uses it more frequently than others. The model also sheds light on the role of subjective perceptions of service quality on individuals' transit usage. These insights can be used to customize policies aimed at increasing transit use among distinct demographic segments. For instance, the model estimation results indicate that men and women have similar perceptions about safety, comfort, and reliability of travel in Bengaluru's bus transit system. However, the results also suggest that women who have a more favorable perception about safety of the system tend to use it more regularly than others. A policy implication is that enhancing safety perceptions among women can help encourage them to use transit more regularly. Strategies to do so include: (a) enhancing safety measures both on-board and at transit stations, such as camera surveillance, GPS vehicle tracking, SoS buttons, night lighting, and night patrol, and (b) safety considerations in and around bus stops and related parking facilities, such as adequate lighting, safe walking access and egress, access to emergency telephones, and night patrol. It is worth noting here that the bus transit system in Bengaluru already has GPS vehicle tracking and is on the way toward installing video cameras and SOS buttons in the entire bus fleet. Such investments are likely to attract more women to use transit and use it more frequently than before.

Similarly, in the context of reliability, increasing magnitudes of the positive coefficients from the incidental use utility function to regular use utility function suggest that improving reliability of transit systems not only makes people more likely to choose transit but also makes them use it more often. These results highlight the potential benefits of planning and operational strategies to improve reliability of transit systems both for attracting new riders as well as enhancing the usage of current riders. Some reliability enhancement measures include enabling the availability of reliable information (e.g., through mobile apps and other means), transit signal priority systems, dedicated bus lanes or bus priority lanes, express fixed-route services, and bus-bunching control techniques that can help in enhancing schedule adherence.

The empirical model shows that individuals' likelihood of using transit is lower when they are unfamiliar with the service (than when they are familiar with it). Therefore, understanding the factors and implementing strategies that can help enhance people's familiarity with transit services can potentially attract more riders. In this context, Abdel-Aty *et al*. (1996) found that over one-third of the transit non-users they surveyed stated that they would be more likely to consider transit if information about transit systems were available. Therefore, strategies such as disseminating reliable information on-board and at bus stops, equipping mobile transit apps to help riders with personalized travel planning options, can potentially help in increasing familiarity and attracting new users (Thompson *et al*., 2002; Foote, 2004). Besides, publicity of transit systems through promotional outreach programs and information campaigns to enhance transit non-users' familiarity with the system may encourage them to utilize it. In this context, identifying those who are unfamiliar with transit service may also help in devising suitable awareness drives.

Finally, travel surveys might benefit from including a "don't know" option as a possible response to Likert rating scale questions on individuals' perceptions and additional questions to elicit individuals' level of (un)familiarity with the topic of the questions (e.g., transit system).

5. SUMMARY AND CONCLUSIONS

The present study uses a generalized heterogeneous data model (GHDM) that allows the joint econometric analysis of individuals' perceptions of transit service characteristics, their unfamiliarity with the transit system, and their likelihood of transit usage and extent of usage. The proposed methodology helps in disentangling individuals' unfamiliarity with transit service from their informed opinions (i.e., perceptions) on transit service characteristics such as cleanliness, comfort, safety, and reliability. Specifically, the proposed method employs individuals' neutral responses to several Likert rating scale questions as a set of binary indicators to identify a latent variable for individuals' unfamiliarity with the transit system. At the same time, the method uses the full range of responses (positive, negative, or neutral) for the Likert rating scale questions to inform individuals' perceptions of the quality of transit service characteristics. This method is applicable for isolating the role of perceptions from that of unfamiliarity in many empirical contexts where Likert rating scale questions are used to inform individual perceptions (and a "don't know" option is not available as a response to such questions).

The proposed framework is applied to analyze transit usage and the extent of usage in the city of Bengaluru in India using empirical data from a recent survey on the ease of mobility in the city. A comparison of empirical results from the proposed model with those from a model that did not consider unfamiliarity highlighted the importance of disentangling the influence of unfamiliarity from informed perceptions of service quality on transit usage. The model that did not separate the influence of unfamiliarity resulted in inferior model fit and underestimated influence of perception variables on individuals' transit usage. An implication of this result is that travel survey questionnaires might benefit from including a "don't know" option as a possible response, in addition to the usual responses that include a neutral category, to Likert scale rating questions eliciting individuals' attitudes and perceptions. Notably, studies from other fields such as education, psychology, and political science (Lam *et al*., 2010; Presser and Schuman, 1980; Liao, 1995) also suggest the inclusion of both "don't know" and "neutral" options in response to Likert scale questions. In addition, it will be useful to include additional questions to elicit information on individuals' level of (un)familiarity with a system (e.g., a transit system) they are asked to provide their opinions on. However, even if such additional questions are included in the surveys to elicit individuals' level of familiarity with the system, it is important to separate the influence of (un)familiarity from that of informed opinion neutrality in their responses to Likert rating scale questions.

Notably, the empirical model sheds light on the demographic differences in individuals' perceptions of transit service characteristics. The model also offers insights on the influence of individuals' demographic characteristics and their perceptions of transit service quality on their usage and extent of usage of bus transit in Bengaluru. These insights pave way toward identifying strategies or policies aimed at not only attracting new riders but also increasing transit usage by existing riders. Given the demographic differences between regular and irregular transit users as well as non-users (and the heterogeneity in perceptions of service quality), strategies to increase transit usage will likely be more fruitful if they are tailored to specific demographic groups. For example, Bengaluru's bus transit agency's plans to equip all buses with video cameras and SoS buttons for women's safety are likely to attract more women riders. Similarly, their efforts to enhance reliability of the transit system (such as driver information systems for controlling bus bunching) will likely attract more students and employed individuals to use bus transit.

The current study may be extended in the following ways. First, the modeling framework can be extended to recognize the bidirectionality of the relationship between individuals' unfamiliarity with a transit system and their and transit usage. The model proposed in the current study recognizes only a unidirectional causality in that unfamiliarity leads to non-usage of public transit. However, it is also likely that people who do not use transit would be unaware of its service characteristics and remain unfamiliar, perhaps because some common unobserved factors cause both unfamiliarity and non-usage. Second, it would be useful to explore if including an explicit "don't know" option as a response to Likert rating scale questions in travel surveys can help reduce the confounding effects discussed in this paper.

REFERENCES

Abdel-Aty, M., Kitamura, R., and Jovanis, P. (1996). Investigating Effect of Advanced Traveler Information on Commuter Tendency to Use Transit. *Transportation Research Record*, 1550:65-72.

Allen, J., Muñoz, J.C., and Ortúzar, J.de.D. (2018). Modelling service-specific and global transit satisfaction under travel and user heterogeneity. *Transportation Research Part A: Policy and Practice,* 113:509–528.

Badami, M. G., and Haider, M. (2007). An analysis of public bus transit performance in Indian cities. *Transportation Research Part A: Policy and Practice*, 41:961–981.

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

Bhat, C. R., and Dubey, S. K. (2014). A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B: Methodological*, 67:68-85.

Bhat, C.R. (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B: Methodological,* 79:50-77.

Bordagaray, M., dell'Olio, L., Ibeas, A., and Cecín, P. (2014). Modelling user perception of bus transit quality considering user and service heterogeneity. *Transportmetrica A: Transport Science*, 10:705-721.

Bose, T., and Pandit, D. (2020). Heterogeneity in perception of service quality attributes of bus transit across various user categories- A case of Kolkata. *Transportation Research Procedia*, 48:2784–2799.

Carrel, A., and Walker, J.L. (2017). Future mode choice intentions of transit riders as a function of past experiences with travel quality. *European Journal of Transport and Infrastructure Research*, 17:360-383.

Cervero, R., and Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 3:199-219.

Das, S., and Pandit, D. (2013). Importance of user perception in evaluating level of service for bus transit for a developing country like India: a review. *Transport Reviews: A Transnational Transdisciplinary Journal*, 33:402-420.

de Oña, J., de Oña, R., Eboli, L., and Mazzulla, G. (2013). Perceived service quality in bus transit service: A structural equation approach. *Transport Policy,* 29:219–226.

De Vos, J., and Witlox, F. (2017). Travel satisfaction revisited. On the pivotal role of travel satisfaction in conceptualising a travel behaviour process. *Transportation Research Part A:Policy and Practice*, 106:364–373.

Devika, R., Harikrishna, M., and Anjaneyulu, M. (2020). Influence of psychological factors in mode choice decision making: A structural equation modeling approach. *Transportation Research Procedia*, 48:2821–2830.

Eboli, L., and Mazzulla, G. (2010). How to Capture the Passengers' Point of View on a Transit Service through Rating and Choice Options, *Transport Reviews*, 30:435-450.

Eboli, L., and Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy,* 18:172–181.

Foote, P. (2004). Making Buses Better in Chicago: Strategic Implementation of Customer-Derived Performance Measures from 1995 to 2001. *Transportation Research Record*, 1884:18–26.

Hensher, D.A., Stopher, P., and Bullock, P. (2003). Service Quality: Developing a Service Quality Index in the Provision of Commercial Bus Contracts. *Transportation Research Part A: Policy and Practice*, 37:499–517.

Ingvardson, J., and Nielsen, O.A. (2019). The relationship between norms, satisfaction and public transport use: A comparison across six European cities using structural equation modelling. *Transportation Research Part A: Policy and Practice*, 126:37-57.

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*, 40:507-525.

Kalton, Graham, Julie Roberts, and David Holt. (1980). The effects of offering a middle response option with opinion questions. *The Statistician*, 65-78.

Krosnick, J. (1991). Response strategies for coping with the demands of attitude measures in surveys, *Applied Cognitive Psychology*, 5:214–236.

Kuppam, A. R., Pendyala, R. M. and Rahman, S. (1999). Analysis of the Role of Traveler Attitudes and Perceptions in Explaining Mode Choice Behavior. *Transportation Research Record,* 1676:68-76.

Lam, T., Allen, G. and Green, K. (2010). Is "neutral" on a Likert scale the same as "Don't know" for informed and uninformed respondents? Effects of serial position and labeling on selection of response options', Paper Presented at the Annual Meeting of the National Council on Measurement in Education, Denver, CA.

Lai, W. and Chen, C. (2011). Behavioral intentions of public transit passengers: The roles of service quality, perceived value, satisfaction and involvement. *Transport Policy*, 18:318-325.

Liao, T. F. (1995). The nonrandom selection of don't knows in binary and ordinal responses: Corrections with the bivariate probit model with sample selection. *Quality and Quantity*, 29(1):87-110.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology,* 140:55.

Manoj, M., and Verma, A. (2015). Activity–travel behavior of non-workers from Bangalore City in India. *Journal of Transport Geography*, 49:99-109.

Moore, M. A., Lavieri, P. S., Dias, F.F., and Bhat, C.R. (2020). On investigating the potential effects of private autonomous vehicle use on home/work relocations and commute times. *Transportation Research Part C: Emerging Technologies,* 110:166–185.

Ory, D.T., and Mokhtarian, P.L. (2005). When is getting there half the fun? Modeling the liking for travel. *Transportation Research Part A: Policy and Practice,* 39(2-3):97–123.

Presser, S. and Schuman, H. (1980). The measurement of a middle position in attitude surveys, *Public Opinion Quarterly*, 44:70–85.

Reed, T., Wallace, R., and Rodriguez, D. (2000). Transit Passenger Perceptions of Transit-Related Crime Reduction Measures. *Transportation Research Record*, 1731:130–141.

Sarkar, P. P., Mallikarjuna, C. (2018). Effect of perception and attitudinal variables on mode choice behavior: A case study of Indian city, Agartala. *Travel Behavior and Society,* 12:108-114.

Schmitt, L., Delbosc, A. and Currie, G. (2014). Lost in transit? Unfamiliar public transport travel explored using a journey planner web survey. *Transportation*, 42:1-22.

Sharda, S., Astroza, S., Khoeini, S., Batur, I., Pendyala, R.M., and Bhat, C.R. (2019). Do Attitudes Affect Behavioral Choices or Vice-Versa : Uncovering Latent Segments Within a Population. Presented at 98th Annual Meeting of the Transportation Research Board, Washington, D.C.

Sheela, P. V., Maness, M., Balusu, S. K., and Pinjari, A.R. (2018). When Neutral Responses on a Likert Scale Do Not Mean Opinion Neutrality: Accounting for Unsure Responses in a Hybrid Choice Modeling Framework. Presented at the 15th International Conference on Travel Behavior Research, Santa Barbara, CA.

Sheela, P. V., and Mannering, F. (2019). The effect of information on changing opinions toward autonomous vehicle adoption: An exploratory analysis. *International Journal of Sustainable Transportation*, 14:1-13.

Shelat, S., Cats, O., and Lint, J.W.C. (2021). Quantifying Subjective Beliefs regarding Waiting Time Uncertainty in Public Transport Networks. *Travel Behaviour and Society,* 25:209–222.

Srinivasan, K.K., Bhargavi, P.V., Ramadurai, G., Muthuram V., and Srinivasan S. (2007). Determinants of Changes in Mobility and Travel Patterns in Developing Countries, Case Study of Chennai, India. *Transportation Research Record*, 2038:42-52.

Srinivasan, K.K., Pradhan, G.N., and Naidu, G.M. (2007). Commute Mode Choice in a Developing Country: Role of Subjective Factors and Variations in Responsiveness Across Captive, Semicaptive, and Choice Segments. *Transportation Research Record*, 2038:53-61.

Sturgis, P., Roberts, C., and Smith, P. (2014). Middle alternatives revisited: how the neither/nor response acts as a face-saving way of saying 'I don't know'. *Sociological Methods and Research*, 43:15-38.

Suman, H. K., Bolia, N. B., and Tiwari, G. (2017). Comparing public bus transport service attributes in Delhi and Mumbai: Policy implications for improving bus services in Delhi. *Transport Policy*, 56:63–74.

Thompson, Brenda J., Perone, Jennifer S., and Gabourel, Kimberlee. (2002). Transit Non-User Survey: Restful Riding Rather than Stressful Driving. *NCTR report 416-08.4.* Office of Research and Special Programs, U.S. Department of Transportation, Washington, D.C.

Verma, M., Manoj, M., Rodeja, N., and Verma, A. (2017). Service Gap Analysis of Public Buses in Bangalore with respect to Women Safety. *Transportation Research Procedia,* 25:4322–4329.

Vij, A., Carrel, A., and Walker, J. L. (2013). Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transportation Research Part A: Policy and Practice*, 54:164-178.

Weinstein, A. (2000). Customer Satisfaction Among Transit Riders: How Customers Rank the Relative Importance of Various Service Attributes. *Transportation Research Record,* 1735:123-132.