Journal of Transportation Engineering

Modelling Automobile Driver's Toll-Lane Choice Behaviour at a Toll Plaza

--Manuscript Draft--

Manuscript Number:	TEENG-1181R3
Full Title:	Modelling Automobile Driver's Toll-Lane Choice Behaviour at a Toll Plaza
Article Type:	Technical Paper
Keywords:	Toll Plaza, Discrete Multinomial Choice Model, Multinomial Logit Model
Abstract:	A toll-lane selection process at a toll plaza can be looked at as an outcome of the choice processes of individual drivers. In this paper an attempt is made to develop a random utility based discrete multinomial choice model for the behaviour of automobile drivers while selecting toll-lanes at a toll plaza. Specifically, a multinomial Logit model is developed and calibrated using disaggregate level choice data from three toll plazas with different geometry and rates of arrival of vehicles. The calibrated Logit models from the different sites when statistically compared show that a generic model, applicable to all the sites, is possible. Such a generic model is also developed. The use of the proposed model can improve the analysis of traffic flow at toll plazas and ultimately lead to more effective designs of these facilities.
Corresponding Author:	Partha Chakroborty Indian Institute of Technology Kanpur Kanpur, INDIA
Corresponding Author E-Mail:	partha@iitk.ac.in
Order of Authors:	Avinash Dubedi
	Partha Chakroborty
	Debasis Kundu
	K. Harikishan Reddy
Opposed Reviewers:	

To:
Prof. Chris T. Hendrickson
Editor in Chief
ASCE Journal of Transportation Engineering

From:
Partha Chakroborty
Rajeeva and Sangeeta Lahri Chair Professor
Department of Civil Engineering
Indian Institute of Technology, Kanpur
Kanpur 208016, INDIA

March 17, 2012

Dear Prof. Hendrickson:

I am attaching the revised manuscript for MS TEENG-1181. The third round of revisions have been carried out keeping in view the third round of comments given by the reviewer.

I have also included a detailed point-by-point response to the third round comments of the reviewer.

We sincerely hope that you will find this version worthy of publication in the ASCE Journal of Transportation Engineering.

Regards

Partha Chakroborty

Modelling Automobile Driver's Toll-Lane Choice Behaviour at a Toll Plaza

Avinash Dubedi [1], Partha Chakroborty [2], Debasis Kundu[3] and K. Harikishan Reddy[4]

Abstract

A toll-lane selection process at a toll plaza can be looked at as an outcome of the choice processes of individual drivers. In this paper an attempt is made to develop a random utility based discrete multinomial choice model for the behaviour of automobile drivers while selecting toll-lanes at a toll plaza. Specifically, a multinomial Logit model is developed and calibrated using disaggregate level choice data from three toll plazas with different geometry and rates of arrival of vehicles. The calibrated Logit models from the different sites when statistically compared show that a generic model, applicable to all the sites, is possible. Such a generic model is also developed. The use of the proposed model can improve the analysis of traffic flow at toll plazas and ultimately lead to more effective designs of these facilities.

Keyword: Toll Plaza, Discrete Multinomial Choice Model, Multinomial Logit Model.

Introduction

A toll plaza is a structure built on a highway where a vehicle has to pay toll. A vehicle has to either stop or slow down to make the payment. The toll plazas are generally on expressways built to allow uninterrupted and high speed traffic flow. Yet,

[1]Former Graduate Student, Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India, e-mail: dubediavinash@gmail.com, Ph:+91-9004828881

[2]**Corresponding author**, Professor, Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India, e-mail: partha@iitk.ac.in, Ph:+91-512-2597037

[3]Professor, Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur 208016, India, e-mail: kundu@iitk.ac.in, Ph:+91-512-2597141

[4]V.R.Techniche Consultants (P) Ltd, Noida, U.P. 201301, INDIA, Ph:+91 9810365075, e-mail: vrtechniche@yahoo.co.in

by their very design, the toll plazas create interruptions to the flow and are potential bottlenecks. Although over the years different technologies have helped reduce the interruptions that toll plazas cause, there are still a large number of toll plazas around the world where vehicles need to stop and often queue in order to pay toll.

The importance of properly designing toll plazas cannot be overstated. If under-designed or improperly designed, these facilities can act as major bottlenecks. Understanding of the queueing process at toll plazas is essential to the effective design of a toll plaza. The queueing process at a toll plaza, among other things, is impacted by how an approaching driver chooses a toll-booth lane (or simply, a toll-lane) from among the many available to him/her at the plaza. Unfortunately, despite the obvious importance of studying the choice process of drivers at toll plazas, not much effort has been spent in collecting real-world data on and modelling of the toll-lane choice process. This paper makes an attempt to fill this gap.

The selection of a toll-lane by a driver, it is assumed, is an outcome of a choice process where an individual evaluates the alternatives (or toll-lanes) and then selects one based on some criteria. Given this assumption, the paper attempts to develop a random utility based discrete multinomial choice model for the choice behaviour of drivers approaching a toll plaza. For this purpose extensive data on vehicles' (or drivers') choices at different toll plazas are collected. First, these revealed preference data are used to arrive at multinomial Logit models of the choice behaviour at toll plazas with many toll-lanes. The maximum likelihood technique (MLE) is used to estimate the scaled parameters of the Logit models.

Next, the estimated parameters for different sites are statistically compared using the technique proposed by Swait and Louvierre (1994). This is done to see whether the data can be combined (after appropriate scaling, where required) so that a generic Logit model, applicable to all the sites, can be developed. Indeed, such a model could be developed and has been developed in this paper. The paper also proposes two procedures through which calibrated Logit models can be evaluated.

Literature Review and Motivation

The toll-lane choice behaviour of drivers at toll plazas has a direct bearing on the queueing process at the plaza and hence (i) on the design of toll plazas and (ii) on devising efficient operating strategies. Yet, an extensive search of the literature did not yield any rigorous study which tries to develop a model of the toll-lane choice behaviour of drivers at toll plazas from real world observations. In fact, one of the most recent studies on lane selection at toll plazas (Mudigonada et al. 2009) suggests that one should try and develop a Probit or Logit model of toll-lane choice behaviour based on detailed data.

Some authors like Mudigonada et al. (2009) and Gulewicz and Danko (1994), while not attempting to develop a choice model based on observations, have used some observed data to underline the habits of drivers while making toll-lane choices at toll plazas. For example, Gulewicz and Danko (1994) states that drivers have affinity towards toll-lanes or booths of lower queue length and those which require lower number of lane changes (from the approach lane).

Other papers on toll plaza design, while simulating traffic through a toll plaza, use ad hoc rules as models of the toll-lane choice process. For example, Zarrillo and Radwan (2009) points out that SHAKER (a model to study queueing process at toll plazas) gives the analyst an option to choose one driver type from four possible types while simulating traffic through a toll plaza. These types are defined based on how a driver will behave while approaching a toll plaza. Drivers can be of the type who choose a toll-lane which has the smallest queue length in terms of number of vehicles or in terms of length units, and so on. The Toll Plaza Simulation model developed by Lin and Su (1994) assumes that drivers select a toll-lane which has the shortest equivalent queue length. This equivalent queue length is defined by Lin and Su (1994) as the actual queue length plus a quantity which depends on the number of lane changes required to reach that particular toll-lane and the type of booth (based on the mode of payment and type of vehicles served). In a later paper, Lin (2001) while referring to, what seems like the same simulation model as in Lin and Su (1994), mentions the use of “utility functions to measure the relative desirability of available toll lanes.” Although Lin indicates that these are functions of number of lane changes and queue length, Lin does not give any further details; neither the functional forms nor the process of obtaining them are presented.

Astarita et al. (2001) assumes that drivers choose a toll-lane by evaluating each lane by the number of vehicles queued in that lane and the number of vehicles the driver expects will join that lane before he/she does. Ozbay et al. (2005) assumes that the drivers select a toll-lane with the shortest queue length and the selection also depends on the directions from which they enter and exit the freeway. Correa et al. (2004) assumes that drivers choose the toll-lane with shortest queue length. Others like Junga (1990) and Al Deek et al. (2000) propose other kinds of ad hoc rules as models of the toll-lane choice process.

From the review of the literature three points emerge (i) in order to study and design toll plazas, one needs to have a model of the driver’s toll-lane choice behaviour, (ii) queue length (in terms of vehicles and length) and number of lane changes are amongst the most important factors that impact toll-lane choice behaviour, and (iii) there are no models of driver’s toll-lane choice behaviour based on real-world observations of the choice process. The motivation of this paper is to fill this void by developing a multinomial Logit model of the toll-lane choice behaviour of drivers based on a rigorous analysis of real-world data on the behaviour of drivers approaching a toll plaza. It is believed that such a model will help in developing a more realistic queueing model for the toll plaza and ulimately lead to a better understanding of the flow process at a plaza. This improved understanding will lead to better design of toll plazas and their operating strategies.

The scope of the paper includes (i) collecting data at multiple toll plazas on how drivers choose toll-lanes, (ii) studying the data in order to determine what are the important factors which seem to impact this behaviour and (iii) developing a discrete multinomial choice model, specifically, a multinomial Logit model, to describe the choice behaviour. This paper, however, restricts itself to studying the choice behaviour of only automobile drivers.

In the next section, a discussion on the Logit model, its calibration and other associated statistical operations are presented. The section also introduces the notation used in this paper.

Multinomial Logit Model

The probability that individual (or driver) n selects alternative (i.e., toll-lane) i, P_{ni} is given by the multinomial Logit model as,

$$P_{ni} = \frac{\exp\left(\dfrac{V_{ni}}{\theta}\right)}{\sum_{\forall j} \exp\left(\dfrac{V_{nj}}{\theta}\right)} \tag{1}$$

where, V_{ni} is the systematic utility derived from alternative i by individual n and the random part of the utility is distributed as the Gumbel distribution with scale parameter, θ. Note that, in this analysis, the number of alternatives in the choice set (i.e., number of toll-lanes available to an approaching driver) is not restricted to two and can be any value greater than or equal to two.

V_{ni} is a function of attributes which the analyst uses to represent an alternative's utility to an individual; it is assumed to be deterministic. This paper, in keeping with the general practice, assumes V_{ni} as a function linear in parameters. Further, V_{ni} is not alternative (or toll-lane) specific since it is assumed that a driver evaluates every toll-lane using the same yardstick. Later, in the section titled "Data Analysis," this assumption is also corroborated using the collected data. It may be noted that under this assumption, the constant term can be omitted from V_{ni} specifications without any loss of generality. In other words, under such an assumption, the constant term is not required to be used. This is so because, under the present assumption, the constant term has no impact on P_{ni}. Here, V_{ni} is written as,

$$V_{ni} = \boldsymbol{\beta}^{\star T}\mathbf{X_{ni}} \tag{2}$$

where, $\mathbf{X_{ni}}$ is a vector of variables or attributes which are used to represent the utility derived by individual n from alternative (or toll-lane) i and $\boldsymbol{\beta}^{\star}$ is a vector of parameters which represent the effect of each variable or attribute on V_{ni} irrespective of the alternative. Using MLE and data on the choices and the corresponding attribute values one can estimate the scaled parameter $\boldsymbol{\beta} = (1/\theta)\boldsymbol{\beta}^{\star}$. The significance of the estimated parameters can be tested using the standard t-test. Further, different models for V_{ni} (when fitted to the same data set) can be compared using the Mcfadden's pseudo ρ^2 statistic (referred here as ρ^2_{Mp}), given as

$$\rho^2_{\mathrm{Mp}} = 1 - \frac{LL(\hat{\boldsymbol{\beta}})}{LL(\mathbf{0})} \tag{3}$$

where, $LL(\hat{\boldsymbol{\beta}})$ is the value of log likelihood function at the estimated parameters and $LL(\mathbf{0})$ is the value of log likelihood function when the parameters are taken as zeros. For a detailed discussion on Logit models and its calibration one may refer to Ben Akiva and Lerman (1985), Cascetta (2001), Train (2002), or Washington et al. (2011).

Often data on the choice process is obtained from different sources. In such cases it becomes important to test whether the Logit model parameters obtained from different data sets are statistically different. Such an analysis can help one obtain parameter estimates which are more generic and applicable for a wide variety of data sources. In order to test whether the parameters of the Logit model estimated from different data sets are statistically different the procedure developed by Swait and Louvierre (1994) is used. The procedure is briefly described here.

Comparing the parameters estimated from two different data sets, say, s1 and s2 implies testing the following null hypothesis

$$H_0 : \boldsymbol{\beta_{s1}} = \boldsymbol{\beta_{s2}} \quad \text{and} \quad \theta_{s1} = \theta_{s2} \tag{4}$$

where, si represents the data set and $\boldsymbol{\beta}$ and θ are as described earlier. This null hypothesis is tested by testing the following two sub-hypothesis sequentially.

$$H_{0A} : \boldsymbol{\beta_{s1}} = \boldsymbol{\beta_{s2}} = \boldsymbol{\beta_A} \tag{5}$$

$$\text{and, } H_{0B} : \theta_{s1} = \theta_{s2} = \theta \tag{6}$$

Notice, if one fails to reject both the sub-hypotheses then H_0 cannot be rejected. Initially H_{0A} is tested by assuming that the scale parameters of the two data sets are different. The test proceeds by combining the two data sets into one by appropriately scaling one of the data sets to nullify the impact of the scale parameter. As both the scale parameters cannot be estimated simultaneously, one of the scale parameters (say, θ_{s1}) is assumed to be unity and the other scale parameter (θ_{s2}) is assumed to be a relative scale parameter.

By maximizing the Log likelihood (LL) function for the combined data set one gets the maximum likelihood estimates of θ_{s2} and $\boldsymbol{\beta_A}$. In order to determine the validity of H_{0A} one uses the standard likelihood ratio test (see, for example, Swait and Louviere, 1994; Ben Akiva and Lerman, 1985). In this case, if LL_1, LL_2, and $LL_{H_{0A}}$ are the maximum LL function values while estimating the parameters using data set s1, data set s2 and the combined data set, respectively, the statistic is given as

$$\lambda_A = -2[LL_{H_{0A}} - (LL_1 + LL_2)] \tag{7}$$

If H_{0A} is rejected, by implication H_0 is rejected. If, however, one fails to reject H_{0A}, one proceeds to test H_{0B} by combining the data sets assuming $\theta_{s2} = \theta_{s1}$ (note that θ_{s1} is assumed to be unity). Next the parameters are estimated using the combined data set and maximum LL value, $LL_{H_{0B}}$ is noted. As before, this LL function value is compared with the case where the assumed restriction of H_{0B} is not imposed using the likelihood ratio test. In this case, the statistic is given as

$$\lambda_B = -2[LL_{H_{0B}} - LL_{H_{0A}}] \tag{8}$$

If H_{0B} is rejected, H_0 is also rejected; otherwise H_0 is not rejected. If one fails to reject H_0 then one can combine the data sets (without scaling) and estimate the parameters. These estimates take into account the information from both the data sets and therefore are, in some sense, more generic. Further, these estimates could be thought of as estimates of the true parameter since the scale parameters can be assumed to be unity.

Experiment Methodology

Based on the review of the literature and the initial observations at toll plazas the following data are thought to be relevant for the purposes of obtaining a reasonable model of the choice process of drivers:

- the toll-lane (or toll-booth-lane) a driver finally chooses,
- the queue lengths at all the toll-lanes when the driver is about 150 m away from the plaza (the distance of 150 m was decided based on observations which suggested that drivers tend to choose a toll-lane when they are about 100 m to 200 m from the plaza),
- the number of heavy vehicles in the queues when the driver is about 150 m away from the plaza,
- the approximate lateral location of the driver (vehicle) when the vehicle is about 150 m away from the plaza; lateral location means the position of a vehicle along the transverse direction of the road, and
- the type of vehicle being driven by the driver.

At every site multiple video cameras are used to record the movement of vehicles. From the video recordings data on the five aspects mentioned earlier are extracted and recorded in data sheets as values of variables VT_n, CB_n, AL_n, q_{ni}, hv_{ni} and l_{ni}; where,
VT_n is the vehicle-type of the n^{th} vehicle,
CB_n is the toll-lane vehicle (driver) n chooses,

AL_n is the approach lane of the n^{th} vehicle (the lateral location of an approaching vehicle about 150 m upstream of the toll booth is noted in terms of a discrete variable indicating the location; this discrete variable is referred to as approach lane),

q_{ni} is the queue length on toll-lane i when vehicle n is about 150 m away from the toll plaza,

hv_{ni} is the number of heavy vehicles on toll-lane i when vehicle n is about 150 m away from the toll plaza, and

l_{ni} is the number of lanes vehicle n will have to cross to reach CB_n from AL_n (the difference between AL_n and CB_n gives l_{ni}).

It may be noted that AL_n records the approximate lateral location of an approaching vehicle. This location is determined by dividing (on the video) the road area into virtual lanes by extending the toll-lane markings 150 m upstream as shown in Figure 1. This extension is done on the video and not on the actual road.

Data is collected at three different sites. All the three sites are on interstate expressways. Site 1 is on an urban expressway while the other two are on semi-urban / rural expressways. Schematics of these sites are shown in Figure 1. In the figure VCL represents the location where multiple video cameras were placed at a reasonable height (12-15 m) above the road surface. The cameras were oriented so as to capture the region between the approach lanes and the toll booths; care was taken to ensure that the approach lanes and the queueing at the plaza were clearly visible on the recordings. As can be seen from the figure, Toll-lanes 1 and 2 at Site 1 have automatic toll collection booths and Toll-lane 4 of Site 2 and Toll-lane 3 of Site 3 were closed for toll collection during the period of data collection. All the toll plazas are of the barrier kind with no exit or entry ramps within a couple of kilometers of the plaza in either direction. Data was collected only for the manual toll payment booths; i.e., vehicles which chose Toll-lanes 1 and 2 at Site 1 were removed from the data set. The reasons for not using this data are (i) very few vehicles used these toll lanes, (ii) Sites 2 and 3 did not have such toll booths, and (iii) the location of and the queues at these booths were such that ignoring this data did not create any impact on the analysis.

============ Figure 1 here ============

Table 1 provides descriptive statistics of the data that was collected. Since the data on choices made by automobile drivers are only analyzed here the descriptive statistics pertain to only those cases where the approaching vehicle is an automobile. The total number of observations on the choice behaviour of automobile drivers at Sites 1, 2, and 3 are 450, 185, and 152, respectively. Further note that the total number of alternatives (or toll-lanes) that drivers could choose from at Sites 1, 2, and 3 are 5, 3, and 3 respectively.

============ Table 1 here ============

Data Analysis

In this section the data obtained from the three sites is analyzed to determine

- whether there is any apparent preference towards certain toll-lanes, and
- whether and how factors such as q_{ni}, l_{ni} and hv_{ni} impact the choice of a toll-lane by a driver.

A quick analysis of the data from Site 1 shows that each of the toll booths (or toll-lanes) 3 to 7 are chosen more or less the same number of times. For example, the toll-lane chosen the lowest number of times is chosen 19% of the time while the toll-lane chosen the highest number of times is chosen 22% of the time. This indicates that drivers do not exhibit a preference towards any of the toll-lanes at this site. Similar is the case for the other two sites. This is as expected since in the absence of entry or exit ramps in the vicinity of the toll plazas there is no reason why certain toll-lanes will be preferred over others. This observation also corroborates the earlier assumption that an approaching driver evaluates every toll-lane using the same yardstick.

In order to study the impact of various factors on toll-lane choice, the toll-lanes are ranked by queue length (with Rank 1 being the set of toll-lanes which had the least queue length when an automobile was approaching the toll plaza, Rank 2 being the set which had the second to least queue length, and so on), ranked by number of heavy vehicles in the queue in much the same way as before, and classified according to the number of lane changes required by an approaching automobile to reach that toll-lane. Figure 2 (a) plots the proportion of times toll-lanes belonging to different ranks (when ranked with respect to queue length) are chosen by drivers at Site 1. Figure 2 (b) provides a similar plot for Site 1 drivers when the ranks are with respect to heavy vehicles in the queues. Figure 2 (c) plots the selection proportions of toll-lanes belonging to different classes when classified as per the number of lane changes required. All the figures show that the proportion reduces as the value on the abscissa increases. Similar is the case for the other two sites. This implies that (i) a toll-lane with lesser queue length is more preferred, (ii) a toll-lane which has the lesser number of heavy vehicles in the queue is more preferred, and (iii) a toll-lane which requires lesser number of lane changes is more preferred. These findings are also in consonance with many earlier suggestions on factors that influence the choice behaviour. For example, as also discussed in the literature review section, Gulewicz and Danko (1994) suggest that queue length and number of lanes one has to cross are important factors while Zarrillo and Radwan (2009) suggest that queue length in length units is also important (it may be noted that the number of heavy vehicles in a queue significantly impact the length of the queue).

============ Figure 2 here ============

As a part of exploratory analysis and having identified the important influencing variables, different types of functional forms for V_{ni} are fitted to the Site 1 data with 450 data points. The following reasonably simple functional form provided a higher ρ^2_{Mp}:

$$V_{ni} = \beta_1^\star \sqrt{q_{ni}} + \beta_2^\star \sqrt{hv_{ni}} + \beta_3^\star l_{ni}^2 \tag{9}$$

Hence, this form is suggested as a model and is used for further detailed statistical analysis. It may be noted that this functional form implies that the effect of an additional queued vehicle on V_{ni} when the queue length is small is more than when the queue length is large. Similar is the effect of an additional heavy vehicle in the queue. The reducing elasticities seem to be an appropriate description of how a driver might view these factors. On the other hand, it is felt that the impact on V_{ni} of a unit increase in the lane change value increases with the number of lane changes. The increasing elasticity seems to be an appropriate description of how a driver might view this factor.

Before proceeding it may be noted that the utility function assumes that drivers do not evaluate different alternatives differently; i.e., the impact of a particular variable, say queue length, on the utility is the same irrespective of whether the queue length is for toll-lane i or toll-lane j. This implies that there is only one utility function for all the alternatives.

Calibration and Evaluation

In this section detailed results from the calibration of the model with data from each of the three sites are presented. The performance of the calibrated models are evaluated through two processes devised as a part of this study. This section is divided into two subsections; the first presents calibration results and the last presents results from the evaluation process. In a later section the calibrated models from the three sites are compared to see whether any statistical difference exists between them. As it turns out, none exists, and hence that section also concentrates on developing a "generic choice model" based on all three data sets.

Calibration results

This section presents the calibrated models for each of the three sites.

============ Table 2 here ============

Table 2 presents the calibrated values of the parameters as well as the $\mathrm{LL}(\hat{\boldsymbol{\beta}})$ and ρ^2_{Mp} values. The values provided parenthetically in Table 2 are the t-statistic values for the corresponding parameters. Further, the $\mathrm{LL}(\hat{\boldsymbol{\beta}})$ values in each of the three panels are the values of the log likelihood function at its maximum (i.e., at the estimated values of the parameters). Before discussing the results in Table 2, the reader is encouraged to revisit section on experiment methodology for a description of the sites and the data. The number of observations from each of the three sites indicates a large sample. Hence, the critical t-value at 95% level of confidence can be taken as 1.96 for all the cases. As can be seen, all the parameters are statistically significant and the ρ^2_{Mp} value is 0.2 in each of the three cases. It may noted that the estimated parameters are the scaled parameters and not the true parameters.

Evaluation of the calibrated models

The ρ^2_{Mp} value does not provide a direct indication as to how much of the observations are "explained" by the model and should not be used in the way R^2 is used in regression. As Train (2002) writes, ρ^2_{Mp} "has no intuitively interpretable meaning" and typically this value is to be used to compare the relative performance of competing models on the same data set. Further, the predictions from the model are probabilities of choosing different toll-lanes while the observations denote specific choices. In some sense, therefore, the predictions and the observations are non-comparables. Yet, using the predictions of the model, one must try and get a feel for how well the model represents the underlying choice process. In this paper two evaluation processes are proposed. These are described in the following as Evaluation Process I (EP-I) and Evaluation Process II (EP-II).

Evaluation Process I (EP-I)

In this process, for each of the observed drivers the probability of choosing each of the available toll-lanes is determined using the calibrated choice model; that is, for the observed values of the parameters, P_{ni} are determined for all n and i. Those toll-lanes which have similar probabilities of choice (say between 0 and 0.1 or between 0.1 and 0.2 and so on) irrespective of the driver, n are put in the same set or bin; for example, the set or bin "0.0-0.1" contains all $i's$ such that $0.0 \leq P_{ni} \leq 0.1$ for any n. For each set or bin, the number of times (from observations) toll-lanes of that set is actually chosen by the drivers is noted. From this and total number of toll-lanes in the set (i.e., the cardinality of the set), the proportion of time toll-lanes of that particular set are actually selected is determined. For each of the sites and when the approaching vehicle is an automobile, Figure 3 shows the selection proportions against the indicative probability value for a set. As is expected, as the predicted probability of choosing a toll-lane (indicated by the bin to which that toll-lane belongs) increases so does the proportion of time these toll-lanes are actually selected. This in some way indicates that the proposed model can discern which toll-lane is more attractive as compared to others. This process, however, is not to be confused with the goodness-of-fit measurement referred to as "percent right" (Ben-Akiva and Lerman 1985) which, if applied here, would only look at how many times a toll-lane which has the highest probability is actually chosen.

============ Figure 3 here ============

Evaluation Process II (EP-II)

This process, like EP-I, relies on creating different sets of toll-lanes based on their choice probabilities, P_{ni}. In EP-II, for every n, the toll-lane i which has the highest P_{ni} is put in a set called Rank 1. Similarly, for every n, the toll-lane, i which has the second highest P_{ni} is put in a set called Rank 2, and so on. Like EP-I, the number of times toll-lanes belonging to a particular set, say Rank k, are actually chosen is determined. These numbers for different sites when the approaching vehicle is an automobile are shown in Table 3. As is expected, "the number of times selected" is higher for higher ranks. Next, EP-II evaluates the discrete choice model by determining whether the observed number of choices are a natural outcome of a process which assumes the underlying V_{ni} model to be true. This is achieved by assigning a toll-lane i to vehicle n based on the calculated $P_{ni}s$ through a stochastic simulation process. From repeated simulations one can determine the distribution of the number of times toll-lanes belonging to a particular rank is chosen. Similar distributions can be obtained for every rank. In the present case, these distributions are determined from thousand independent simulations; from the distributions, 95% confidence intervals for every Rank, k are determined for each of the three sites. These intervals (when the approaching vehicle is an automobile) are shown in Table 3 just below the observed number of times booths of a given rank were chosen. As can be seen, the observed number of times booths of a given rank are chosen are contained in the 95% confidence intervals obtained under the assumption that the proposed model is true. This indicates that the underlying V_{ni} model is a reasonable description of the observed choice process.

============ Table 3 here ============

Comparison of Calibrated Models

The previous section presented calibrated toll-lane choice models for automobiles using data from three different sites. Not surprisingly, the calibrated values of the parameters vary from site to site. However, the question is whether the differences in parameter values are an artefact of the stochastic variations in the data or are they indicative of a difference in driver behaviour. An associated question is whether the differences, even if statistically significant, arise because of a difference in the scale parameters (note that the estimated parameters presented in the section titled "Calibration and Evaluation" are scaled parameters) or are these differences indicative of a difference in driver behaviour.

============ Table 4 here ============

This section presents results from the comparison of the calibrated V_{ni} models developed in the previous section. The process used for comparison is outlined in the section titled "Multinomial Logit Model." The results from the comparison study are presented in Table 4. The first column in the table indicates the calibrated models being compared. Columns 3, 4, and 5 provide the parameter values obtained from the MLE exercise using the restriction of the hypothesis listed in Column 2. The sixth column gives $\hat{\theta}_{s2}$ under H_{0A}. In the three panels of Table 4, s2 refers to Site 2 in first panel and Site 3 in second and third panels. The LL_h provides the maximum log likelihood value using the restriction listed under Column 2. Column 8 gives the likelihood ratio values calculated using Equations 7 or 8, depending on the situation. The LL values required in the equations are provided in Tables 2 and 4. The last column gives the critical χ^2 value at 95% level of confidence for the degree of freedom provided parenthetically.

As can be seen from Table 4, one fails to reject hypothesis H_{0A} in each of the three pairwise comparisons. Further, one fails to reject hypothesis H_{0B} while comparing calibrated models for Sites 1 and 2 and Sites 2 and 3. However, this hypothesis is rejected for the pairwise comparison of calibrated models for Sites 1 and 3. These results indicate that the parameter values may be assumed to be same in all the three calibrated models; however, the scale parameter for Site 3 is different from that of Site 1. Further, scale parameters for Sites 1 and 2 can be taken as unity. Based on these observations one can combine the three data sets after appropriately scaling the third data set (i.e., under the assumption that the scale parameters for Site 1 and Site 2 are both unity and that of Site 3 is some value θ_{s3}).

Using the combined data set the parameters of the V_{ni} model in Equation 9 is again estimated. The estimated values are: $\hat{\beta}_1 = -2.15$, $\hat{\beta}_2 = -1.27$ and $\hat{\beta}_3 = -0.2$ ($\hat{\theta}_{s3}$ is estimated to be 0.65).

In order to build confidence in these estimates the estimated value of $\widehat{\theta_{s3}}$ is used as a scaling factor for data set from Site 3 only. The parameters of V_{ni} are estimated using this scaled data set. If the assumptions that $\theta_{s1} = \theta_{s2} = 1$ and the relative scale parameter of Site 3 is 0.65 are valid then the estimates of $\boldsymbol{\beta}$ obtained from only the scaled Site 3 data should be statistically no different from the ones obtained using the combined data set. The parameters estimated by using only the scaled Site 3 data are $\hat{\beta}_1 = -1.74$, $\hat{\beta}_2 = -1.57$ and $\hat{\beta}_3 = -0.21$. Bootstrapping technique (for more details see Efron and Tibshirani 1993) is used to obtain distributions for these $\hat{\beta}_i$ values and their 95% confidence intervals. The 95% confidence intervals obtained from bootstrapping for $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are $[-2.54, -1.21]$, $[-2.24, -1.13]$ and $[-0.32, -0.12]$, respectively. As can be seen, each of these confidence intervals contain the corresponding $\hat{\beta}_i$ value obtained using the combined data. This lends further credence to the claim that the $\hat{\boldsymbol{\beta}}$ values from the combined data are representative of the choice behaviour observed in each of the three sites. Thus, a generic function for the V_{ni} for automobiles can be written as:

$$V_{ni}^{gen} = -2.15\sqrt{q_{ni}} - 1.27\sqrt{hv_{ni}} - 0.2l_{ni}^2 \ . \tag{10}$$

The generic model is evaluated using EP-I and EP-II described earlier. The plot for the EP-I is shown in Figure 4. It shows expected behaviour. EP-II, as described in the previous section, is applied using the model in Equations 10. The results (similar to those in Table 3) are presented for the generic model in Table 5. Results on Site 1 are presented seperately from Sites 2 and 3 as the number of toll-lanes at Site 1 are different from the number at Sites 2 and 3. As can be seen from the table the observed selection proportions are contained in the 95% confidence intervals obtained under the hypothesis that the proposed generic model is true. This indicates that the underlying generic model (V_{ni}^{gen}) is a reasonable description of the observed toll-lane choice behaviour of automobile drivers at all the three sites.

============ Table 5 here ============

This and the previous two sections present a comprehensive analyses of the data culminating in the calibrated generic model for the toll-lane choice behaviour of automobile drivers. From this analyses the following points emerge:

- q_{ni}, hv_{ni}, and l_{ni} are variables which significantly impact the choice of toll-lanes by drivers,
- the negative sign of the parameters in Table 2 and Equation 10 indicate, as is expected, that each of the variables create disutility; i.e., increase in their values make the toll-lane less attractive,
- on comparing the magnitudes of the parameters in Equation 10 and noting the general range in which $\sqrt{q_{ni}}$, $\sqrt{hv_{ni}}$, and l_{ni}^2 lie it can be said that queue length is by far the most important determinant and number of lane changes is not as important.

Extension: Modelling Drivers' of Heavy Vehicles

A study similar to the one presented here was done for the case when the approaching vehicle is a heavy vehicle. The detailed results from the study are not presented here because the number of observations were fewer and more studies on parameter transferability among models for drivers of different vehicle types need to be done. Here, only the generic function, developed based on the data from all the three sites, for the V_{ni} for drivers of heavy vehicles (referred to as $V_{ni}^{gen,hv}$) is presented. The method used for developing the generic model for heavy vehicle drivers is the same as the one used for developing a similar model for automobile drivers. The calibrated $V_{ni}^{gen,\,hv}$ is:

$$V_{ni}^{gen,\,hv} = -2.63\sqrt{q_{ni}} - 0.21 l_{ni}^2 \, . \tag{11}$$

It should be noted that the number of heavy vehicles in the queue is no longer a statistically significant variable. The reason for this could be that for a driver of a heavy vehicle existence of other heavy vehicles in the queue is not as important as it is for a driver of an automobile. Another point that needs to be mentioned is that the calibrated generic model for heavy vehicles was tested using the same tests that were used for the model for automobiles. The calibrated model presented in Equation 11 performed acceptably.

Conclusion

The main objective of this work is to develop a random utility based discrete choice model for the toll booth (lane) choice behaviour of driver's approaching a toll plaza. Specifically, a Logit model is developed. After initial analysis of the data obtained from three sites, a utility function is proposed for the case when the approaching vehicle is an automobile. The parameters of the utility function are calibrated using the data on the revealed preferences of drivers (in terms of chosen toll booth) at three different sites (toll plazas). On statistically comparing the calibrated models it is found that all the data sets can be combined to determine a single set of parameters for the proposed utility function. This is referred to as the generic model for the toll-lane choice behaviour of drivers. The generic model for automobiles is given in Equation 10.

It is believed that with the development of a generic model for automobile drivers' toll-lane choice behaviour a step has been taken to fill the gap in the literature identified earlier in this paper. The use of this driver behaviour description, in any of the existing models which study toll plazas, has the potential to improve the way in which queueing analysis is done for the toll plazas. That, in efect, will improve the design of toll plazas.

It is, however, felt that a lot more work needs to be done in order to more completely understand the behaviour of drivers while choosing toll-lanes. In this regard, among other things, research effort needs to be directed in the following two areas. First, athough some results based on limited data were presented for the behaviour of drivers of heavy vehicles, more work on studying the behaviour of drivers of other types of vehicles needs to be carried out. The same line of analysis as presented here can be used for such studies. Of course, in addition to studying whether data from various sites can be clubbed to develop a generic model for drivers of other types of vehicles, studies on whether statistically significant difference exists between drivers of different types of vehicles also need to be carried out. Second, attempt should be made to include "taste" variations or unobserved heterogeneity among drivers choosing toll-lanes. Random coeficient multinomial Logit models (i.e, Mixed Logit models) can be thought of as a possible way to include such inter-driver variation. One may refer to Washington et al. (2011) and Hensher and Greene (2003) for more details on Mixed Logit models.

References

[1] Al-Deek, H.M., Mohamed, A.A. and Radwan, E.A. (2000). "New model for the evaluation of traffic operations at electronic toll collection plazas." *Transportation Research Record*, 1710, 1-10.

[2] Astarita, V., Florian, M. and Musolino, G. (2001). "A Microscopic Traffic Simulation Model for the Evaluation of Toll Station Systems." *Proceedings of IEEE Intelligent Transportation Systems Conference*, August 25-29, 692-697.

[3] Ben-Akiva, M., and S. Lerman. (1985). "Discrete Choice Analysis: Theory and Application to Travel Demand." MIT Press, Cambridge, Massachusetts.

[4] Cascetta, E. (2001). "Transportation Systems Engineering: Theory and Methods." Kluwer Academic Publishers, Netherlands.

[5] Correa E., C. Metzner, and N. Nino. (2004). "TollSim: Simulation and Evaluation of Toll Stations." *International Transactions in Operational Research*, 121-138.

[6] Efron, B., and R. Tibshirani. (1993). "An Introduction to the Bootstrap." Chapman and Hall, New York.

[7] Gulewicz, V., and Danko, J. (1994). "Simulation-based approach to evaluating optimal lane-staffing requirements for toll plazas." *Transportation Reseach Record*, 1484, 33-39.

[8] Hensher, D.A., and Greene, W.H. (2003). "The Mixed Logit Model: The State of Practice." *Transportation*, Vol. 30, No. 2, 133-176.

[9] Junga, A. J. (1990). "A Multi-Purpose Toll Collection Plaza Model." *Proceedings of the 1990 Winter Simulation Conference, Eds. O. Balci, R. Sadowski, and R. Nance*, 793-797.

[10] Lin, F.B., and Su, C.W. (1994). "Level of service analysis of toll plaza on freeway mainlines." *Journal of Transportation Engineering*, Vol. 120, No.2, 119-127.

[11] Lin, F.B. (2001). "Delay model for planning analysis of mainline Toll plazas." *Transportation Research Record*, 1776, 69-74.

[12] Mudigonda, S., Bartin, B., and Ozbay, K. (2009). "Microscopic Modeling of lane selection and lane changing at toll plazas." *Presented at 88th Annual Meeting of the Transportation Research Board*, Washington, D.C., USA.

[13] Ozbay, K., Mudigonda, S., and Bartin. B. (2005). "Development and calibration of an Integrated Freeway and toll plaza Model for New Jersey Turnpike using Paramics microscopic simulation toll." *Proceedings of the 8th International IEEE Conference on ITS, Vienna*, 13-16.

[14] Swait, J., and Louviere. J. (1994). "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." *Journal of Marketing Research*, Vol. XXX, 305-314.

[15] Train, K. (2002). "Discrete Choice Methods with Simulation." Cambridge University Press, California.

[16] Washington, S. P., Karlaftis, M. G., and Mannering, F. L. (2011). "Statistical and Econometric Methods for Transportation Data Analysis, 2nd edition." CRC Press, Boca Raton.

[17] Zarillo, M. L., and Radwan, A. E. (2009). "Methodology Shaker and the Capacity analysis of five toll plazas." *Journal of Transportation Engineering*, Vol. 135, No.3, 83-93.

List of Tables

Table 1: Descriptive statitics of the data on automobile drivers' choice behaviour

Table 2: Results from the calibration of v_{ni}

Table 3: Observed number of times toll-lanes in sets Rank k are selected and 95% confidence intervals for the selection frequency obtained using simulation with the calibrated V_{ni} as the choice process model

Table 4: Pairwise comparison of calibrated models from all the three sites

Table 5: Observed number of times toll-lanes in sets Rank k are selected and 95% confidence intervals for the selection frequency obtained using simulation with the calibrated V_{ni}^{gen} as the choice process model

List of Figures

Figure 1: Schematic of (a) Site 1, (b) Site 2, and (c) Site 3.

Figure 2: Plot of proportion selected versus (a) rank based on queue length, (b) rank based on heavy vehicle, and (c) class based on lane difference.

Figure 3: Plot of proportion selected versus predicted probability obtained using site-specific models for automobiles.

Figure 4: Plot of proportion selected versus predicted probability obtained using generic model for automobiles.

Table 1: Descriptive statitics of the data on automobile drivers' choice behaviour

Variable	Descriptive statistic	Location		
		Site 1	Site 2	Site 3
q_{ni}	minimum	0	0	0
	maximum	17	8	14
	mean	6.2	2.3	3.8
	std. dev.	2.5	1.6	3.8
hv_{ni}	minimum	0	0	0
	maximum	8	4	7
	mean	2.7	0.7	1.1
	std. dev.	1.5	0.9	1.6
l_{ni}	minimum	0	0	0
	maximum	6	2	3
	mean	1.7	0.8	1.2
	std. dev.	1.3	0.6	1.1

Table 2: Results from the calibration of v_{ni}

Parameter & Other statistics	Calibrated using data from		
	Site 1	Site 2	Site 3
$\hat{\beta}_1$	-2.37	-1.96	-1.12
t-statistic	(11.44)	(6.29)	(4.52)
$\hat{\beta}_2$	-1.23	-1.21	-1.01
t-statistic	(10.48)	(5.42)	(5.74)
$\hat{\beta}_3$	-0.21	-0.23	-0.14
t-statistic	(11.96)	(2.34)	4.09
LL($\hat{\boldsymbol{\beta}}$)	-561.6	-155.6	-137.6
ρ^2_{Mp}	0.2	0.2	0.2

Table 3: Observed number of times toll-lanes in sets Rank k are selected and 95% confidence intervals for the selection frequency obtained using simulation with the calibrated V_{ni} as the choice process model

Site	Obs. num. of times toll-lanes from set Rank k is selected [95% confidence interval from simulation]				
	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
1	218	114	64	40	14
	[196,237]	[93,129]	[52,81]	[29,51]	[9,24]
2	126	38	21	-	-
	[99,123]	[38,62]	[16,32]	-	-
3	88	43	21	-	-
	[77,99]	[30,51]	[16,32]	-	-

Note: Sites 2 and 3 each have three toll-lanes

Table 4: Pairwise comparison of calibrated models from all the three sites

Comparison between	Hypothesis (h)	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\theta}_{s2}$	LL_h	λ_h	Critical $\chi^2_{df,0.05}$ (df)
Site 1	H_{0A}	-2.32	-1.26	-0.21	0.89	-717.8	1.0	9.5 (4)
Site 2 (s2)	H_{0B}	-2.24	-1.23	-0.2	-	-718.1	0.7	3.8 (1)
Site 1	H_{0A}	-2.24	-1.28	-0.21	0.62	-701.6	4.8	9.5 (4)
Site 3 (s2)	H_{0B}	-1.94	-1.17	-0.19	-	-705.9	8.7	3.8 (1)
Site 2	H_{0A}	-1.83	-1.3	-0.21	0.7	-293.9	1.4	9.5 (4)
Site 3 (s2)	H_{0B}	-1.52	-1.1	-0.17	-	-295.6	3.3	3.8 (1)

Table 5: Observed number of times toll-lanes in sets Rank k are selected and 95% confidence intervals for the selection frequency obtained using simulation with the calibrated V_{ni}^{gen} as the choice process model

Site	Obs. num. of times toll-lanes from set Rank k is selected [95% confidence interval from simulation]				
	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
1	218	112	62	45	13
	[193,234]	[92,130]	[52,83]	[30,52]	[10,25]
2 & 3	206	91	40	-	-
	[196,229]	[70,101]	[29,51]	-	-

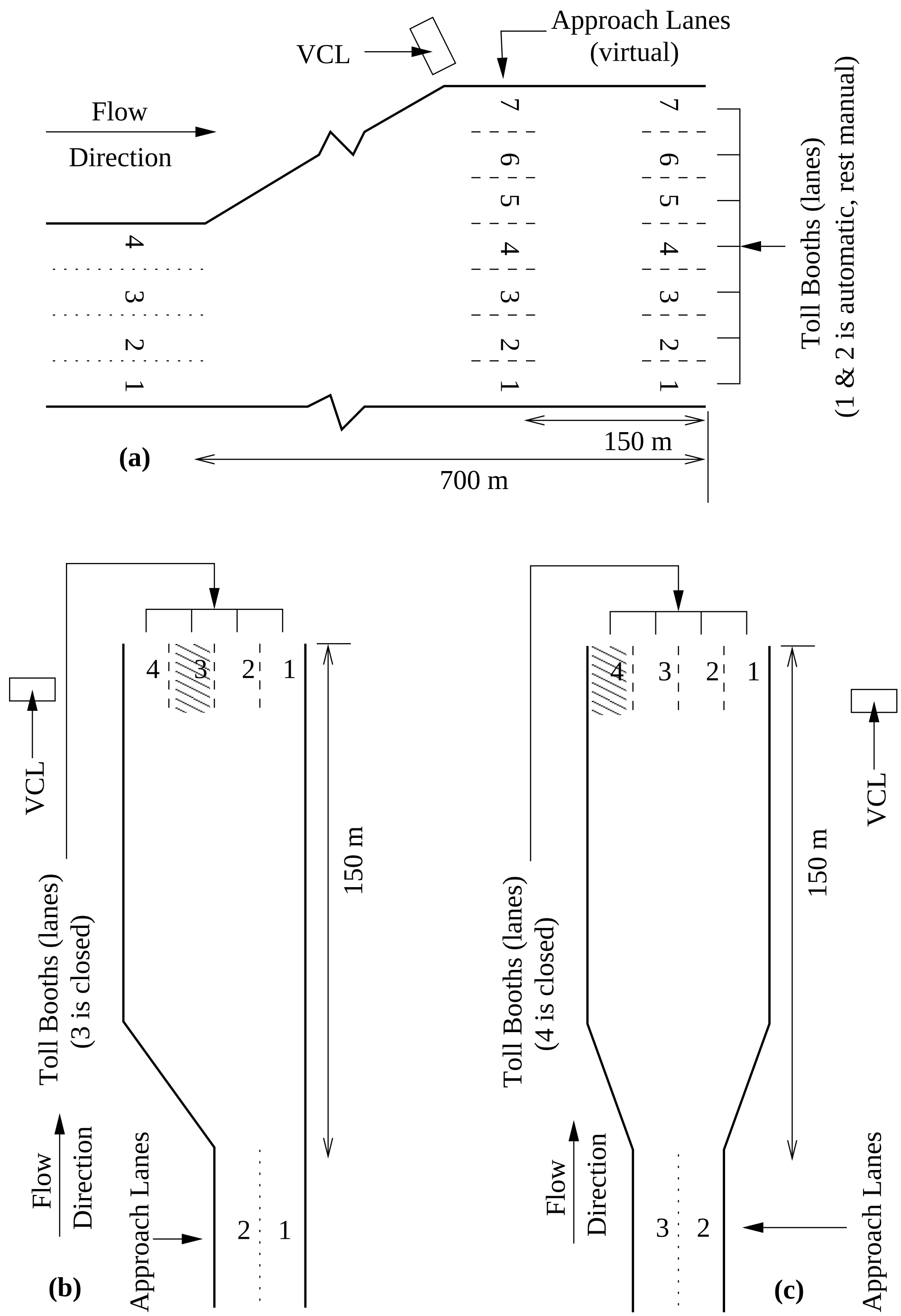

Figure 1: Schematic of (a) Site 1, (b) Site 2, and (c) Site 3.

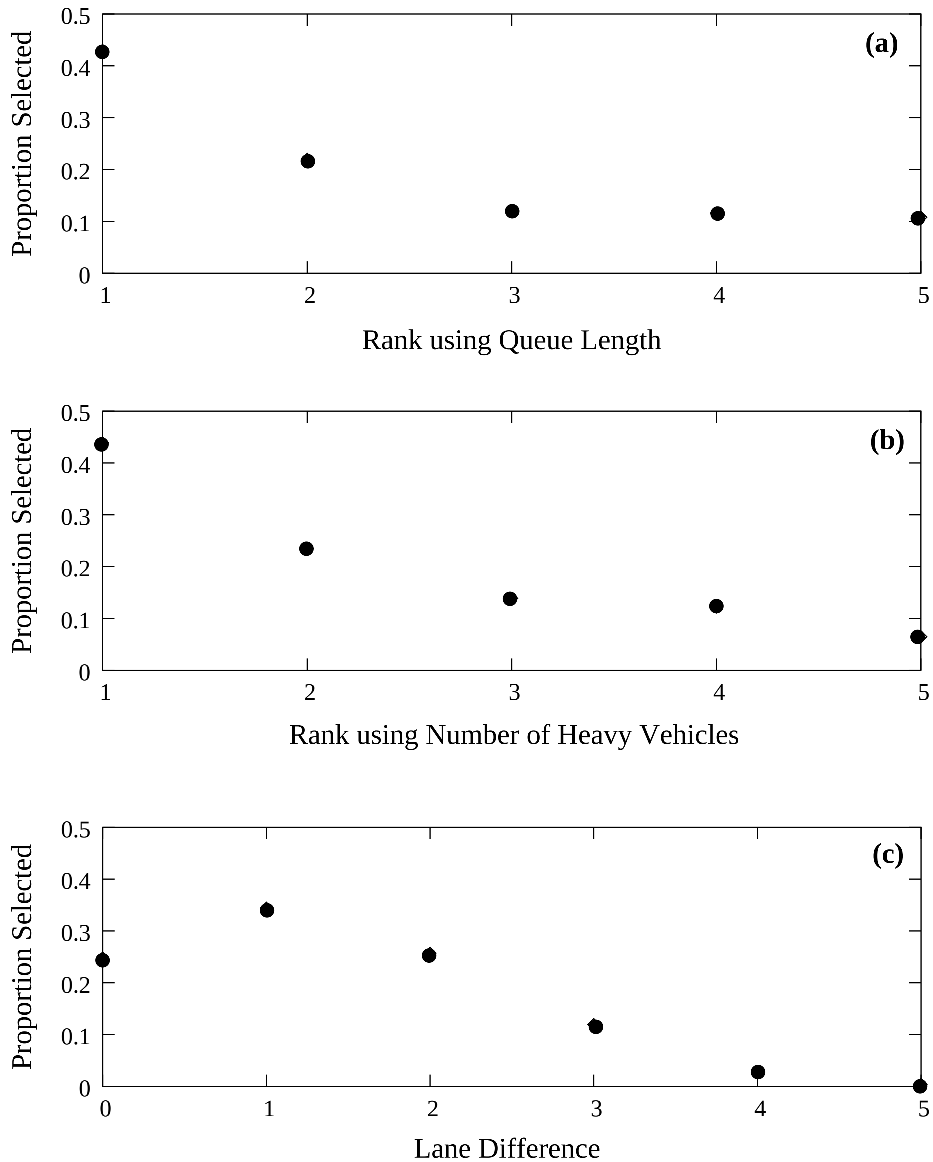

Figure 2: Plot of proportion selected versus (a) rank based on queue length, (b) rank based on heavy vehicle, and (c) class based on lane difference.

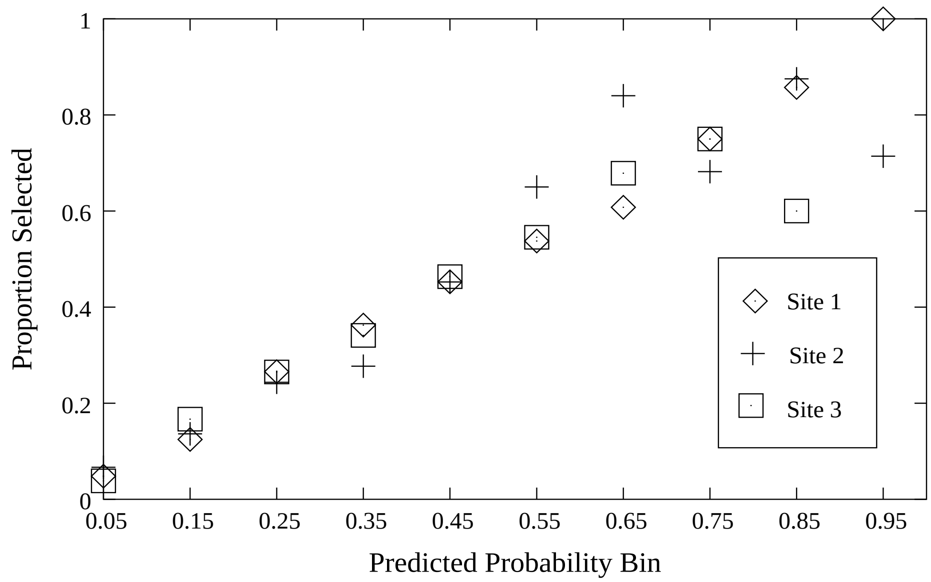

Figure 3: Plot of proportion selected versus predicted probability obtained using site-specific models for automobiles.

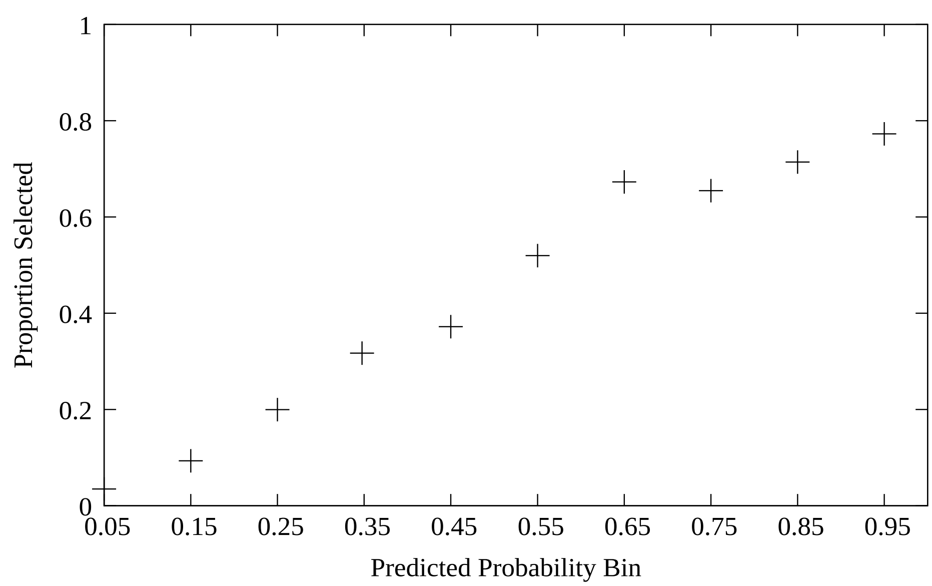

Figure 4: Plot of proportion selected versus predicted probability obtained using the generic model for automobiles.

AMERICAN SOCIETY OF CIVIL ENGINEERS
COPYRIGHT TRANSFER AGREEMENT

Publication Title: JOURNAL OF TRANSPORTATION ENGG.

Manuscript/~~Chapter~~ Title: MODELLING AUTOMOBILE DRIVER'S TOLL-LANE CHOICE BEHAVIOUR AT A TOLL PLAZA

Author(s) - **Names and addresses of all authors** A. DUBEDI, CIVIL ENGG., IIT KANPUR, INDIA
P. CHAKROBORTY, CIVIL ENGG., IIT Kanpur, INDIA
D. KUNDU, MATHS & STATISTICS, IIT KANPUR, INDIA
K. H. REDDY, V.K. TECHNICHE CONS. (P) LTD., NOIDA, INDIA

The author(s) warrants that the above-cited manuscript is the original work of the author(s) and has never been published in its present form.

The undersigned, with the consent of all authors, hereby transfers, to the extent that there is copyright to be transferred, the exclusive copyright interest in the above-cited manuscript (subsequently called the "work") in this and all subsequent editions of the work, and in derivatives, translations, or ancillaries, in English and in foreign translations, in all formats and media of expression now known or later developed, including electronic, to the American Society of Civil Engineers subject to the following.

- The undersigned author and all coauthors retain the right to revise, adapt, prepare derivative works, present orally, or distribute the work provided that all such use is for the personal noncommercial benefit of the author(s) and is consistent with any prior contractual agreement between the undersigned and/or coauthors and their employer(s).

- In all instances where the work is prepared as a "work made for hire" for an employer, the employer(s) of the author(s) retain(s) the right to revise, adapt, prepare derivative works, publish, reprint, reproduce, and distribute the work provided that such use is for the promotion of its business enterprise and does not imply the endorsement of ASCE.

- No proprietary right other than copyright is claimed by ASCE.

- An author who is a U.S. Government employee and prepared the above-cited work does not own copyright in it. If at least one of the authors is not in this category, that author should sign below. If **all** the authors are in this category, check this box □and sign here:__________. Please return this form by mail.

SIGN HERE FOR COPYRIGHT TRANSFER [Individual Author or Employer's Authorized Agent (work made for hire)]

PARTHA CHAKROBORTY
Print Author's Name

Print Agent's Name & Title

Partha Chakroborty
Signature of Author (in ink)

Signature of Agency Rep (in ink)

Date: DEC., 30, 2011

Note: If the manuscript is not accepted by ASCE or is withdrawn prior to acceptance by ASCE, this transfer will be null and void.

Response to Third Round of Reviewer's Comments on MS TEENG-1181
(Modelling Driver's Booth Choice Behaviour at a Toll Plaza by Dubedi, Chakroborty, Kundu and Reddy)

We thank Reviewer 1 for reading the manuscript again and providing comments. We have gone through the comments carefully and made changes wherever necessary. The following lists our detailed response to the comments of the reviewer.

Response to Reviewer 1's comments

Comment - I: Heavy-vehicle analysis should be added back

Response:

In response to the concern raised by the reviewer (in the second round of review) on removing the analysis with heavy vehicles we had said:

> "We can, however, include only the calibrated equations for heavy vehicles in an appropriate section if the reviewer believes that such an inclusion is necessary."

Now, that the reviewer feels that we should add the analysis back, we have, in keeping with our response cited above, introduced a small section ahead of the concluding section. In this section we have presented the generic model for heavy vehicles and also indicated its limitations. In light of this inclusion we have also modified two sentences in the conclusion.

Comment - II: Issue related to the opinion of the reviewer that "the lane that the vehicle is moving before entering the toll plaza - in a sense - accounting for the effect of the missing coefficients."

Response:

We are uncertain about what the reviewer means. Hence, we have not been able to provide a discussion on this issue.

Comment - III: About references on "the random parameters issue."

Response: We thank the reviewer for suggesting many reference material which can be included. We have now included Washington et al. (2011) and Hensher and Greene (2003) as references on this topic. These references have been decided based on the list provided by the reviewer and the scope of this paper. We have also removed the Ben Akiva and Lerman (1985) reference while discussing the random parameters model.