Modeling Driving Decisions with Latent Plans

by

Charisma Farheen Choudhury

Bachelor of Science in Civil Engineering
Bangladesh University of Engineering and Technology (2002)

Master of Science in Transportation
Massachusetts Institute of Technology (2005)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

The Field of Civil and Environmental Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2007

Signature of Author _____
Department of Civil and Environmental Engineering
August 27, 2007

Certified by _____
Moshe E. Ben-Akiva
Edmund K. Turner Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by _____
Daniele Veneziano
Chairman, Departmental Committee for Graduate Students

Modeling Driving Decisions with Latent Plans

by

Charisma Farheen Choudhury

Submitted to the Department of Civil and Environmental Engineering
On August 17, 2007 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Field of Civil and Environmental Engineering

Abstract

Driving is a complex task that includes a series of interdependent decisions. In many situations, these decisions are based on a specific plan. The plan is however unobserved or latent and only the manifestations of the plan through actions are observed. Examples include selection of a target lane before execution of the lane change, choice of a merging tactic before execution of the merge. Change in circumstances (e.g. reaction of the neighboring drivers, delay in execution) can lead to updates to the initially chosen plan. These latent plans are ignored in the state-of-the-art driving behavior models. Use of these myopic models in the traffic simulators often lead to unrealistic traffic flow characteristics and incorrect representation of congestion.

A modeling methodology has been formulated to address the effects of unobserved plans in the decisions of the drivers and hence overcome the deficiency of the existing driving behavior models and simulation tools. The actions of the driver are conditional on the current plan. The current plan can depend on previous plans and be influenced by anticipated future conditions. A Hidden Markov Model is used to address the effect of previous plans in the choice of the current plan and to capture the state-dependence among decisions. Effects of anticipated future circumstances in the current plan are captured through predicted conditions based on current information. The heterogeneity in decision making and planning capabilities of drivers are explicitly addressed.

The methodology has been applied in developing driving behavior models for four traffic scenarios: freeway lane changing, freeway merging, urban intersection lane choice and urban arterial lane changing. In all applications, the models are estimated with disaggregate trajectory data using the maximum likelihood technique. Estimation results show that the latent plan models have a significantly better goodness-of-fit compared to the 'reduced form' models where the latent plans are ignored and only the choice of actions are modeled.

The justifications for using the latent plan modeling approach are further strengthened by validation case studies within the microscopic traffic simulator MITSIMLab where the simulation capabilities of the latent plan models are compared against the reduced form models. In all cases, the latent plan models better replicate the observed traffic conditions.

Thesis Supervisor: **Moshe E. Ben-Akiva**
Title: Edmund K. Turner Professor of Civil and Environmental Engineering

Thesis Committee

Moshe E. Ben-Akiva (Chair)
Edmund K Turner Professor
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology

Nigel H. M. Wilson
Professor
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology

Patrick Jaillet
Department Head and Edmund K Turner Professor
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology

Haris N. Koutsopoulos
Associate Professor
Department of Civil and Environmental Engineering
Northeastern University

Joan L. Walker
Assistant Professor
Department of Geography and Environment & Center for Transportation Studies
Boston University

Tomer Toledo
Senior Lecturer
Department of Civil and Environmental Engineering
Technion – Israel Institute of Technology

Acknowledgements

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Traffic congestion is a major problem in urban areas that adversely affects mobility, air quality and safety. According to the Urban Mobility Report (Schrank and Lomax 2005), congestion caused 3.7 billion vehicle-hours of delay and 2.3 billion gallons of wasted fuel in major US cities alone, resulting a total loss more than $63 billion. California Air Resources Board estimates that emissions are 250% higher under congested conditions than during free-flow conditions (Schiller 1998). Increased driving stresses resulting from congestion have led to aggressive driving and unsafe driving behaviors (NHTSA 1997). All these factors cause direct economic losses due to delays and accidents, and indirect economic losses due to increased stress, health and environmental impacts. Moreover, with the rapid growth of population and car ownership, the extent of traffic congestion is spreading both spatially and temporally. These concerns make congestion alleviation a major transportation priority.

Congestion reduction primarily involves increasing the roadway capacity: either through building new roads to increase the physical capacity or by improving the operational capacity of the existing network by adapting optimum traffic management and control strategies. Additional congestion management mechanisms include demand management techniques and planning measures to reduce urban sprawl. The optimum strategy often includes the combination of multiple measures of congestion reduction and is difficult to deduce theoretically. Field tests of these congestion management techniques are also generally prohibitively expensive and not feasible.

Microscopic traffic simulation tools, which mimic individual drivers to deduce real world traffic situations, are ideal tools to analyze and test different congestion

management strategies in a controlled environment. These tools analyze traffic phenomena through explicit and detailed representation of the behavior of individual drivers. Driving behavior models are thus an important component of the microscopic traffic simulation tools. These models include route choice models, speed/acceleration models and lane changing models. Speed/acceleration models describe the movements in the longitudinal direction and lane changing models describe drivers' lane selection and gap acceptance behaviors.

Driving decisions are influenced by a wide range of factors. These include neighborhood conditions, features of the vehicle and characteristics of the driver, attributes of the network, overall traffic situation etc. The relative speed, position and type of vehicles in the vicinity of the driver have a direct effect on the lane changing and acceleration decisions. The features of the vehicle like acceleration and deceleration capabilities and the characteristics of the driver, such as the path-plan and schedule, the network knowledge and driving capabilities can also significantly influence driving behavior. The speed and acceleration of the driver can also be affected by the network attributes: grade, curvature, surface quality and speed limit for example. Further, in the same network, drivers can behave differently in different traffic situations. In particular, the level of congestion can have a significant impact on driving decisions. For example, in heavily congested situations, there can be significant cooperation among the drivers; they are likely to be more alert and conscious about their actions, and their driving decisions can involve substantial planning and anticipation. It is essential to address these factors in the corresponding driving behavior models for proper simulation of congested traffic.

The existing driving behavior models address many of these factors: either fully or partially. The effects of neighborhood conditions on the decisions of the driver in particular have received considerable attention from researchers. However, in most cases the models do not adequately capture the sophistication of driver behavior and the causal mechanism behind their observed decisions. Specifically, the existing models represent instantaneous decision-making and assume drivers to be myopic. These shortcomings are more evident in congested and incident affected scenarios where the observed driving behavior is actually the result of a conscious planning process. These plans may evolve

dynamically and an initially chosen plan may not be executed in the end. The plans are however unobserved and only the actions (e.g. maneuvers like acceleration, lane changes, route choice etc.) are observed. The behavioral predictions based only on myopic considerations are therefore bound to contain significant noise as a result of the models' structural inability to uncover underlying causal mechanisms. Implementation of these models in traffic micro-simulation tools can lead to unrealistic traffic flow characteristics: underestimation of bottleneck capacities and incorrect representation of congestion (Abdulhai *et al.* 1999, DYMO 1999). This was reflected in the findings of the Next Generation Simulation (NGSIM) study on Identification and Prioritization of Core Algorithm Categories where congested, oversaturated and flow breakdown scenarios have been identified by the users as weak points of traffic micro-simulation tools (Alexiadis *et al.* 2004). Using these tools to evaluate congestion management planning and policy scenarios can result in bias in the analysis.

Therefore, in order to properly simulate congested scenarios in a microscopic simulator, it is essential to develop more realistic driving behavior models that will capture the complexity of human decision making processes.

1.2 Planning in Driving Decisions

According to the NGSIM Core Algorithm Analysis Report (Hranac *et al.* 2004a), travel decisions can be classified into the following categories based on the time scale of application (shown in Figure 1.1):

1. Pre-trip traveler decisions: These strategic decisions are taken before starting a trip and constitute the pre-trip plan of the traveler. Examples include, deciding whether or not to travel, selecting the time of departure, destination, mode of transportation and route etc.

2. Strategic en-route traveler decisions: Once the pre-trip decisions are made, the traveler either executes the originally selected plan without any change, or makes one or more modifications to the initial plan. This category of decisions includes modification of destination, mode or route, parking choice etc.

The decisions in category 1 and 2 take over 30 seconds (and in most cases much longer) to make and execute.

3. Tactical route execution decisions: This category deals with traveler decisions that take between 5 and 30 seconds to make and execute. While executing a route from an origin to a destination, a series of tactical maneuvers are performed by drivers based on sub-goals generated from a variety of factors. Examples include, maintaining a desired travel speed, making up lost time from a previous delay, avoiding large trucks, pre-positioning to get into the appropriate lane, etc. These broad set of route execution decisions result in a combination of lower-level tactical plans.
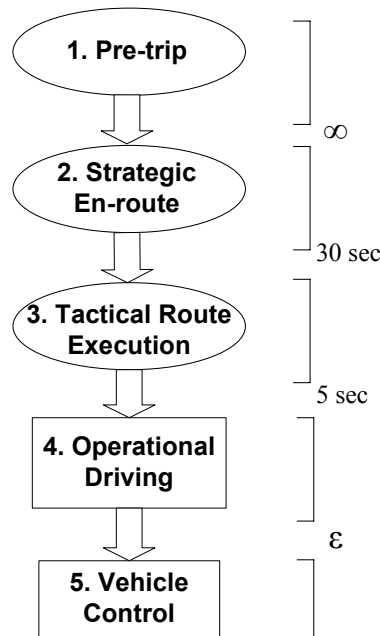
Figure 1.1: Classification of traveler behavioral algorithms
(adapted from NGSIM Core Algorithm Analysis Report, 2004)

4. Operational driving decisions: The operational behaviors of travelers include decisions to control their vehicle at a time scale of less than five seconds. These include lane shifting, gap acceptance for executing a lane change or for maneuver at an unsignalized intersection, acceleration/deceleration, queue discharge behavior etc.

5. Vehicle control decisions: This category deals with driver decisions related to controlling the vehicle at a nanoscopic time-scale level, steering the wheel of the vehicle or pressing the accelerator for example.

Driving behavior models encompass the tactical route execution and operational driving decisions. It should be noted that only the actions associated with the operational

driving decisions and sometimes the vehicle control decisions are observed. The strategic and tactical plans that lead to that action are generally unobserved or latent.
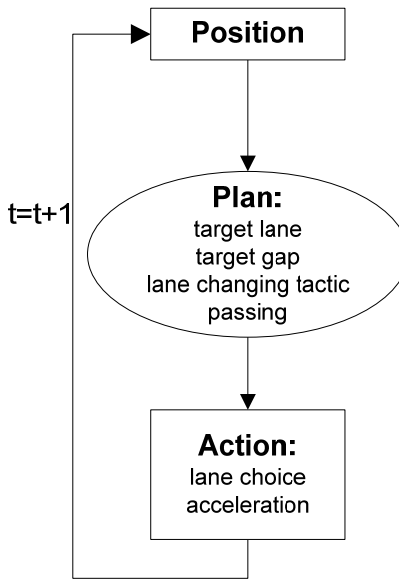
Figure 1.2: General framework of driving behavior

A general framework of the driving behavior model is presented in Figure 1.2. As seen in the figure, in the initial position, the driver makes a plan: selecting a target lane for example. Depending on the traffic situation and the driver characteristics, the plan can consist of various additional levels: the choice of target gap, the choice of tactic for execution of the lane change, choice of gaps for making a passing maneuver etc. The choice of action depends on the choice of plan and consists of lane choice and acceleration decisions. The chosen action is reflected in the updated position of the driver.

An example of choice of plans of the driver is shown in Figure 1.3. The pre-trip and en-route strategic plans of the driver (illustrated in Figure 1.1) may lead to the tactical plan to reach a target lane to take an exit for example. The subsequent actions of the driver involve looking for an acceptable gap to maneuver to the target lane in order to execute the plan. In this process, the driver may also target forward or backward gaps and adjust the acceleration to avail those gaps. In congested situations, where normally acceptable gaps may not be available, the chosen plan can also involve selection of an alternate lane changing tactic (e.g. courtesy or forced gap acceptance). The chosen plan is

unobserved and manifests itself through the chosen lane actions and accelerations. However, the plans may be updated due to situational constraints and contextual changes and the observed actions may not be the ones that were originally intended. Failure to change to the target lane, for example, may lead to an observation of no change from the current lane.
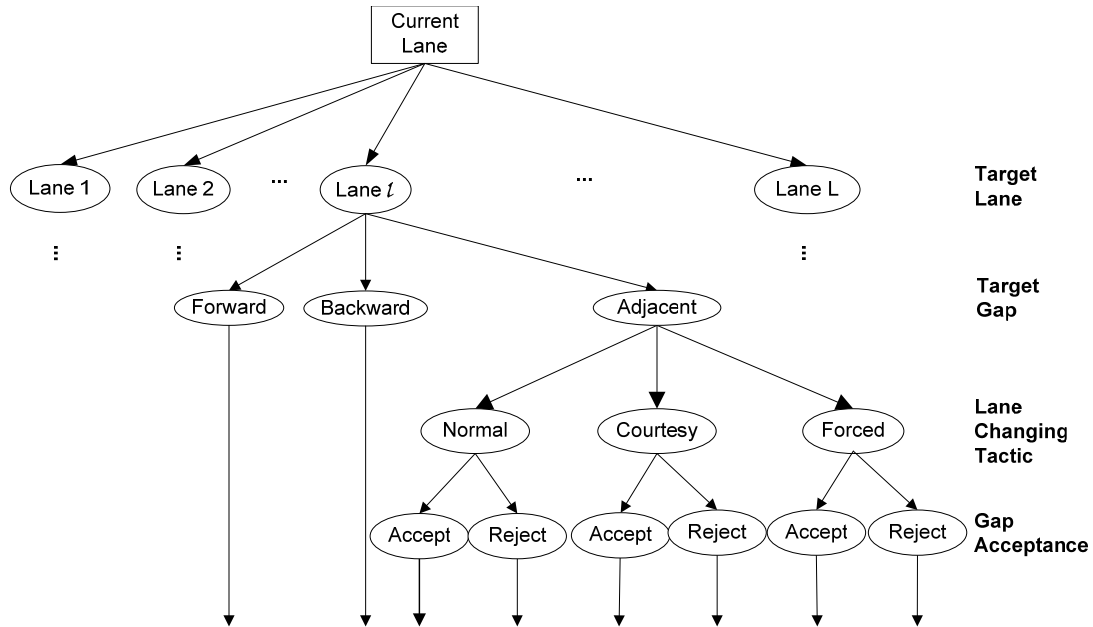
Figure 1.3: Framework of choice of plan

Further, the strategic and tactical plans and actions can take place in a dynamic environment where a driver's goals, resulting plans, and external conditions are all subject to change. The driver may consider several alternatives to come up with a plan, but the actions that he/she ends up executing might be different from those initially planned. This evolution in plans could be due to several factors. First, situational constraints or contextual changes might lead to revision of the plan. For example, an unusual level of congestion might lead a driver to revise the planned time of travel or route. Or non-cooperation of a driver in the target lane may lead to reevaluation of the lane changing tactic to that lane. Second, the driver's current plans are influenced by the past experiences so that as the history evolves, the plan can also evolve. For example, the choice of an action with an unfavorable outcome might lead one to abandon the plan that led to this action in future choice situations. Third, drivers might eventually adapt to

conditions in their environment so that they might exhibit inertia in the choice of their plans and actions. For instance, drivers may have a preference to stay in the current lane.

There can be considerable difference in aggressiveness, driving skills, intelligence and planning ability of drivers. Drivers may also have different levels of familiarity with the network. These driver-specific characteristics (generally unobserved) can have significant impact on the latent plans.

The strategic and tactical choices comprising the latent plans can also be influenced by the geometric and traffic attributes. The effect of latent path-plan for example may be more evident in an urban arterial with closely spaced turns compared to a freeway network where exits are far apart. Similarly, there can be higher propensity to target a distant lane if there is a large difference in level of service (LOS) among different lanes. Again, the underlying plan for executing a lane change in a congested freeway can differ significantly from the choice of plan in an uncongested situation where acceptable gaps are readily available.

Thus the inclusion of the effect of plans in the behavioral framework is more important in certain scenarios. Examples include urban arterials, traffic situations with significant congestion and/or high differential in level of service, work zones, incident spots etc.

1.3 Modeling Approach

The models presented in this thesis address the planning behaviors described in the previous section in the behavioral framework of drivers to increase the reliability of microscopic traffic simulations. The methodology for modeling behaviors with unobserved or latent plans is developed first and then demonstrated through empirical studies of lane changing behaviors of drivers in different scenarios. The overall model development approach is summarized in this section.

1.3.1 Theoretical Framework

Drivers are assumed to conceive plans that are unobserved (latent) and execute actions based on the plans (as shown in Figure 1.2). These latent plans are defined by the chosen target/tactic of the driver. The actions are represented by driving maneuvers. The

interdependencies and causal relationships between the choice of plan and choice of action of the same driver are captured through individual-specific latent variables.

The plans depend on past decisions as well as anticipated future conditions. The interdependencies between successive plans lead to state-dependence in the decisions. A Hidden Markov Model (HMM) based methodology is adapted to capture the dynamics of the plans.

The heterogeneity in planning capability and aggressiveness of the drivers is also captured in the model framework. Two different approaches: a discrete latent class based technique and a continuous latent 'plan-ahead' distance based approach, have been proposed and demonstrated to address the heterogeneity among drivers in terms of planning. The aggressiveness of the driver is captured through continuous latent variables that enter successive decisions across all choice dimensions of the same driver (agent effect).

1.3.2 Empirical Studies

As discussed in the previous sections, the decisions leading to the selection of plan, and the choice of action given the selected plan, differ depending on the driving scenario and the effect of planning is more evident in urban arterials, congested and incident affected traffic situations, traffic streams with high differential in level of service etc. This was also reflected in the findings of the NGSIM study on Identification and Prioritization of Core Algorithm Categories (Alexiadis *et al.* 2004), where the urban arterial lane selection, oversaturated freeway behavior, freeway lane changing and weaving section behaviors topped the list of prioritized scenarios chosen for improvement. Based on the priority ranking of this NGSIM study and guided by data availability (Hranac *et al.* 2004b), four lane selection scenarios have been selected to demonstrate empirically the effect of latent planning in observed driving decisions. These selected scenarios are as follows:

- Freeway lane changing,
- Freeway merging ,
- Urban intersection lane choice, and
- Urban arterial lane changing within sections.

The general decision framework is the same in all cases: latent plans followed by observed actions. However, the type of plan and the causal relationship among plans and actions of drivers can differ depending upon the scenario and is often dictated by the level of congestion. For example, in a relatively uncongested freeway lane changing situation, if acceptable gaps are readily available, the target gap is always the adjacent gap and the lane changing tactic is always normal. Therefore, the target gap choice and lane changing tactic selection levels are redundant and the latent plan is manifested only through the selection of target lanes. On the other hand, in freeway on-ramp merges in congested situations, the target lane is always the rightmost lane of the mainline and the target gap is restricted to the adjacent gap (due to maneuverability constraints). The latent plan in such situations thus constitutes only the choice of merging tactic. Again, in urban intersection lane choice and lane changing in urban arterial sections, the motivation behind the lane selection and the implementation of the latent plans differ significantly from the freeway scenarios.

The models in all scenarios have been developed using the process shown in Figure 1.4, which involves using both disaggregate and aggregate data. Disaggregate data, which are detailed vehicle trajectories at a high time resolution are used in the model estimation phase. In this phase, the model is specified and explanatory variables, such as speeds and relations between the subject vehicle and other vehicles are generated from the vehicle coordinates extracted from the trajectory data. Parameters of all model components: the plan selection, the plan transition (for the state-dependent case) and the action choice are estimated jointly using a maximum likelihood technique to match observed lane changes of the drivers that occurred in the trajectory data (panel data).

In this study, the statistical estimation software GAUSS (Aptech Systems 2003) has been used to program the log-likelihood for the model estimation. The likelihood function is not globally concave. For example, if the signs of all the coefficients of the individual-specific error term are reversed, the solution is unchanged due to its symmetric distribution function. To avoid obtaining a local solution, different starting points are used in the optimization procedure. Statistical tests are performed to refine the models and to determine the best model specifications. It may be noted that the estimation

approach does not involve the use of any traffic simulator, and so the estimated models are simulator independent.
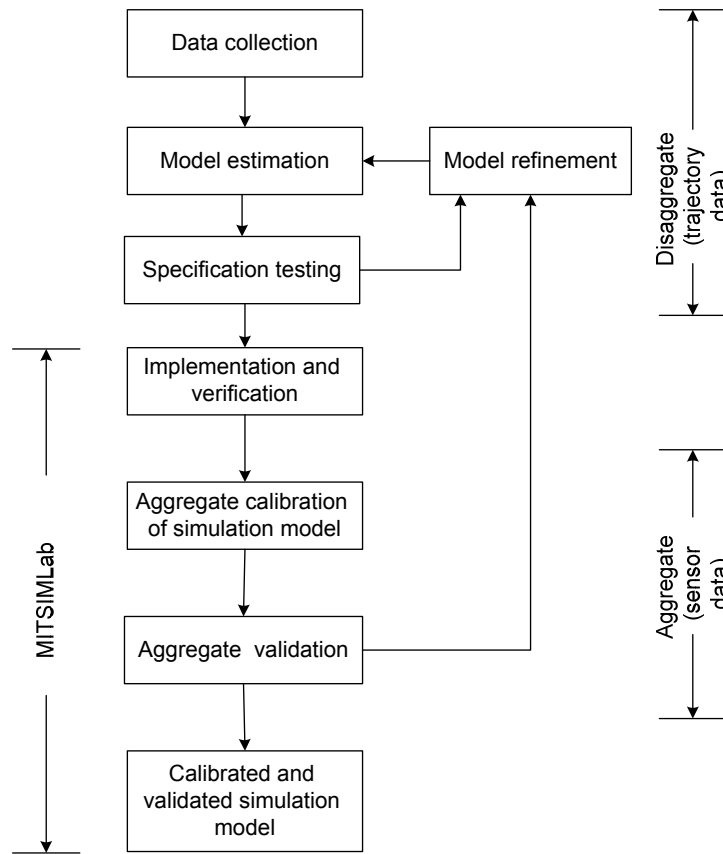
Figure 1.4: Model development framework

The value of inclusion of the latent plans is demonstrated in two ways:

- Goodness-of-fit of the estimated model
- Model validation using simulation runs

The latent plan models are compared against corresponding 'reduced form' models that have no latent plan mechanisms. Both models are estimated with the same data. These reduced form models however cannot be viewed as nested within the latent plan models. Therefore 'adjusted' goodness-of-fit measures are used to statistically compare the non-nested models.

In model validation, the simulation capabilities of the latent plan models are compared against the replications of the reduced form models. The validation results demonstrate the benefits that can be derived from using the modified models. For this, the

improvements must be demonstrated within a microscopic traffic simulator using data that has not been used for model estimation. The microscopic traffic simulator incorporates not only the lane changing models being studied, but also other driving behavior models, such as acceleration models. MITSIMLab (Yang and Koutsopoulos, 1996) has been used for validation of the models presented in this thesis. A brief description of MITSIMLab and its model components is presented in Appendix A. In the validation case studies, aggregate data has been used.

The key parameters of the behavior models of the simulator need to be adjusted before the validation runs. These parameters are often identified through sensitivity analysis where the impact of an individual factor on the overall predictive quality of the simulator is measured by allowing the corresponding parameter to change while keeping all other parameters at their original values. Part of the aggregate data is first used for aggregate calibration of behavioral parameters of MITSIMLab as well as for estimating the travel demand on the case study network. This aggregate calibration problem is formulated as an optimization problem, which seeks to minimize a function of the deviation of the simulated traffic measurements from the observed measurements and of the deviation of calibrated values from their a-priori estimates (Toledo and Koutsopoulos 2004). The formulation is detailed in Appendix B.

The remaining part of the aggregate data (not used for calibration of the model) is used for the validation runs. The measures of performances are calculated from the remaining validation data and compared with the corresponding outputs from the simulator for both the proposed and the reduced form models. The measures of performances include sensor speeds and flows, the distribution of vehicles among the lanes, frequency and locations of lane changes etc.

1.4   Thesis Contributions

The objective of the thesis is to improve the simulation of congested traffic situations by developing more realistic driving behavior models that capture the unobserved plans behind the observed driving maneuvers. A latent plan based modeling approach for driving behaviors is proposed that differs significantly from the state-of-the art driving

behavior modeling procedures which adopt a 'black-box' approach based on a limited field of view and instantaneous decision making of drivers.

The effectiveness of the new approach has been demonstrated in the thesis through modeling lane changing behaviors in different scenarios (freeway lane selection, freeway merging, urban intersection lane choice and urban arterial lane changing). The usefulness of capturing the underlying causal mechanism in each scenario has been presented through comparison of goodness-of-fit of estimation results and validation case studies within traffic simulators. In both cases, the latent plan models outperform the corresponding reduced form models that do not have any latent mechanism establishing the supremacy of the approach.

The developed lane changing models have bridged some of the significant gaps in the existing simulation tools. The specific contributions of each empirical study are listed below:

- In freeway lane changing scenario, the new lane changing model with explicit choice of target lane gives the flexibility to accommodate lane changing behavior with exclusive lanes (e.g. High Occupancy Vehicle Lanes, High Occupancy Tolled Lanes, and Heavy Vehicle Lanes etc.). These lanes are characterized by high level of service differential. Traditional modeling approaches tend to fail in such situations. The new model, with its agility to address choice of distant targets, performs substantially better.

- In the freeway merging model, latent plans in terms of lane changing tactics of the driver (normal, courtesy and forced) are integrated in a combined decision framework for the first time. The combined decision framework gives the flexibility to model the transition between the three merging tactics. This enables the model to better capture merges that occur earlier in the merge section.

- The urban intersection lane choice and arterial lane changing models constitute the first rigorously estimated behavior models for urban arterials. These models replace the existing rule-based lane assignment models used for modeling urban arterial lane choices.

Thus, implementation of the new models in micro-simulation tools can contribute to simulation of more realistic traffic flow and better representation of congestion, and hence result in better planning and policy analysis tools.

1.5   Thesis Outline

The remainder of this thesis is organized in six chapters. In Chapter 2, a literature review on state-of-the-art driving behavior models is presented. Chapter 3 provides the generic model structure for latent plan models and presents the modeling methodology. The application of the latent plan models in different scenarios: freeway lane changing, freeway merging and lane selection in urban arterials (both intersection lane choice and lane changing within sections) are presented in Chapter 4, Chapter 5 and Chapter 6 respectively. Each chapter presents the detailed model structure, description of the data used for model development, and the model estimation and validation results. Comparison of the latent plan models against reduced form models are also shown in each chapter: both in terms of goodness-of-fit of model estimation and in terms of simulation capabilities within MITSIMLab. Finally, conclusions and directions for further research are summarized in Chapter 7.

Chapter 2

Literature Review

Existing literature on driving behavior models focus on several key aspects: longitudinal maneuvers or acceleration and lateral movement decisions involving lane selection and gap acceptance. These behaviors have been modeled both as disjoint models and integrated models combining multiple aspects. The significant disjoint and integrated driving behavior models are described below with their overall limitations highlighted in the end.

2.1 Lane Changing Models

The first lane changing model intended for micro-simulation tools was introduced by Sparmann (1978). In this model, a distinction is made between the desire to change lanes and the execution of the lane change. The model also distinguishes between changes to the nearside (in the direction of the exit) and to the offside (in the direction away from the exit). Changes to the nearside are motivated by not having obstructions in that lane. Changes to the offside are motivated by an obstruction in the current lane (e.g. slow vehicles) and/or better conditions on the offside lane. The model implements psycho-physical thresholds on the relative speed and spacing to define obstructions to which drivers will respond. The possibility of execution of a lane change is determined by the space available in the selected lane.

Gipps (1986) developed a rule based zone dependent model that addresses the necessity, desirability and safety of lane changes. Drivers' behavior is governed by two basic considerations: maintaining a desired speed and being in the correct lane for an intended turning maneuver. The distance to the intended turn defines which zone the

driver is in and which of the considerations are active. When the turn is far away it has no effect on the behavior and the driver concentrates on maintaining a desired speed. In the middle zone, lane changes are only considered to the turning lanes or lanes that are adjacent to those. Close to the turn, the driver focuses on being in the correct lane and ignores other considerations. The zones are defined deterministically, ignoring heterogeneity among drivers and variations in the behavior of a driver over time. When more then one lane is acceptable, the conflict is resolved deterministically by a priority system considering locations of obstructions, presence of heavy vehicles and potential speed gain. The limitation of the rule based models is that the lane selection rules are evaluated sequentially, and therefore less important considerations are only evaluated if more important ones did not yield a lane choice. The deterministic rule priority system thus ignores trade-offs among the considerations (e.g. drivers would always avoid lanes with heavy trucks and avoid lanes away from their exit, even if these lanes offer immediate speed advantage and overtaking provisions etc.). No framework for rigorous estimation of the model parameters has been proposed.

Several micro-simulators implement lane changing behaviors based on Gipps' model. In CORSIM (Halati *et al.* 1997, FHWA 1998) lane changes are classified as either mandatory (MLC) or discretionary (DLC). MLC is performed when the driver must leave the current lane (e.g. in order to use an off-ramp or avoid a lane blockage). DLC is performed when the driver perceives that driving conditions in the target lane are better, but a lane change is not essential. A similar distinction between MLC and DLC is also considered by SITRAS (Hidas and Behbahanizadeh 1999), Yang and Koutsopoulos (1996), Ahmed (1999) and Zhang *et al.* (1998).

In SITRAS (Hidas and Behbahanizadeh 1999), downstream turning movements and lane blockages may trigger either MLC or DLC, depending on the distance to the point where the lane change must be completed. In this model, MLC is also performed in order to obey lane-use regulations. DLC is performed in an attempt to obtain speed or queue advantage, defined as the adjacent lane allowing faster traveling speed or having a shorter queue. Model parameters were not rigorously calibrated and no framework to perform this task has been proposed.

Unlike the deterministic rule based models, in Yang and Koutsopoulos' model (1996), lane selection is based on a random utility, which captures trade-offs between the various factors affecting this choice (e.g. speed advantage, the presence of heavy vehicles and merging traffic). In Ahmed's model (1999), a more rigorous discrete choice framework is used to model the lane changing decisions in three steps: decision to consider a lane change, choice of a lane and acceptance of gaps in the chosen lane. The model framework is presented in Figure 2.1 with unobserved decisions shown in ovals.
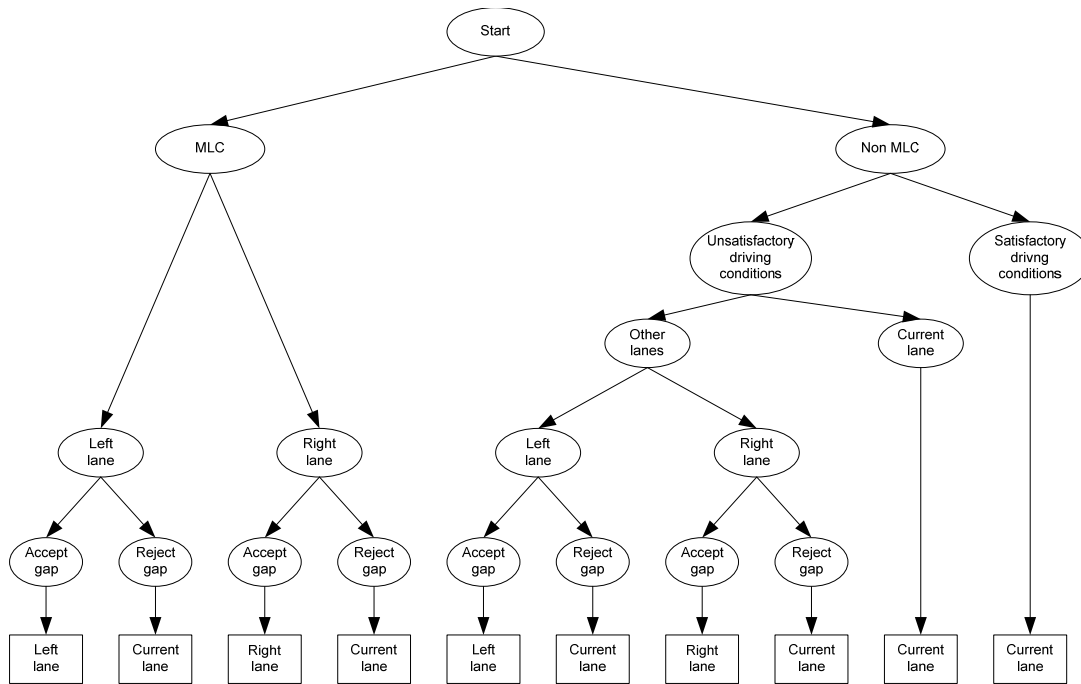
Figure 2.1: Structure of the lane changing model proposed by Ahmed (1999)

When an MLC situation applies, the decision whether or not to respond to it depends on the time delay since the MLC situation arose. DLC is considered when MLC conditions do not apply or the driver chooses not to respond to them. The driver's satisfaction with conditions in the current lane depends on the difference between the current and desired speeds. If the driver is not satisfied with driving conditions in the current lane, neighboring lanes are compared to the current one and the driver selects the most desirable lane. Lane utilities are affected by the speeds of the lead and lag vehicles in these lanes relative to the current and desired speeds of the subject vehicle. Gap acceptance models (detailed in the next sub-section) are used to model the execution of the lane changes. The parameters of this model are estimated using second-by-second

vehicle trajectory data. The model however does not explain the conditions that trigger MLC situations and the parameters of the MLC and DLC components of the model have been estimated separately. The MLC model has been estimated for the special case of vehicles merging to a freeway, under the assumption that all vehicles are in MLC state. The DLC model has been estimated with offside lane changing data collected from a freeway section (to ensure that the lane changes are discretionary).

Zhang *et al.* (1998) use similar definitions of MLC and DLC and the gap acceptance logic. The authors validate the model but do not suggest a framework for its calibration.

The separation between MLC and DLC in the above mentioned models imply that there are no trade-offs between mandatory and discretionary considerations. For example, a vehicle on a freeway that intends to take an off-ramp will not overtake a slower vehicle if the distance to the off-ramp is below a threshold, regardless of the speed of that vehicle. Furthermore, in order to implement MLC and DLC models separately, rules that dictate when drivers begin to respond to MLC conditions need to be defined. This point is however unobservable, and judgment based heuristic rules, which are often defined by the distance from the point where the MLC must be completed, are used.

Toledo *et al.* (2003) developed an integrated lane shift model that allows joint evaluation of mandatory and discretionary considerations. In this model, the relative importance of MLC and DLC considerations vary depending on explanatory variables such as the distance to the off-ramp. This way the awareness to the MLC situation is more realistically represented as a continuously increasing function rather than a step function. The structure of the model is shown in Figure 2.2.

The model consists of two levels: choice of lane shift and gap acceptance decisions for execution of the lane change. Variables that capture the need to be in the correct lanes and to avoid obstacles and variables that capture the relative speed advantages and ease of driving in the current lane and in the lanes to the right and to the left are all incorporated in a single utility model that captures the trade-offs among these variables. Estimation results indicate that path-plan related variables play an important goal in the lane changing behavior of drivers. Path-plan effects are captured by a group of variables like the distance to the point where drivers have to be in specific lanes and the number of lane changes that are needed in order to be in these lanes. The parameters of the lane shift

and gap acceptance models have been estimated jointly using second by second trajectory data collected from a freeway situation.
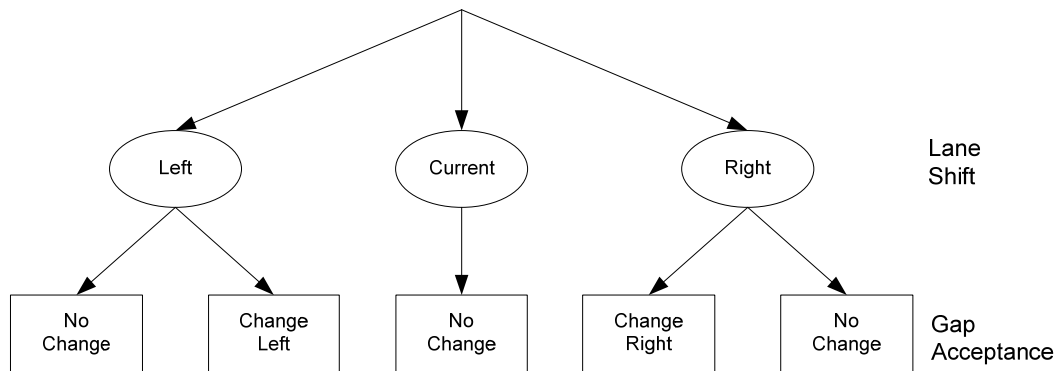
Figure 2.2: Structure of the lane shift model proposed by Toledo et al. (2003)

Most of the existing lane changing models have been developed for freeway scenarios. Wei *et al.* (2000) developed a deterministic rule based model for a two-lane urban arterial based on observations from Kansas City, Missouri. Lane selection is determined by the location and direction of intended downstream turns and classified as mandatory, preemptive or discretionary. Drivers who intend to turn at the next intersection are in an MLC situation and try to move to the correct lane. Drivers who intend to turn farther downstream try to move to the lane that connects to their planned path and attempt preemptive lane changes. Vehicles already in the correct lane may undertake a discretionary passing maneuver (double lane change to the other lane and back) in order to gain speed advantage only if the maneuver is perceived to be possible. The model requires that both the adjacent gap in the other lane and the gap in the current lane between the subject's leader and its leader are acceptable for passing maneuvers to take place.

Hunt and Lyons (1994) used neural networks as an alternative method of modeling driver behavior within road traffic systems. Their main approach makes use of a learning vector quantization classification type of neural network. A driver is assumed to make a decision based on vehicle movements within a zone of influence, i.e., the activity within a certain distance behind the vehicle and a certain distance in front. Their model uses visual pattern based input to describe the driving environment around the vehicle about to make

a lane change. The model is calibrated by exposure to a large number of representative example inputs and corresponding decisions or answers.

2.2   Gap Acceptance Models

Gap acceptance models have been studied in the context of intersection crossing and within merging and lane changing models. The definitions of terms used in this section are illustrated in Figure 2.3 with an example of a lane changing scenario.
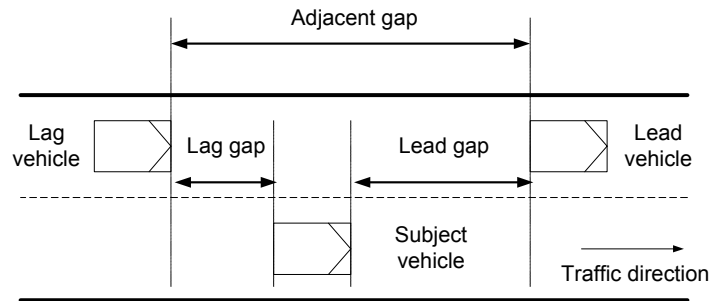
Figure 2.3: Relation between subject, lead and lag vehicles

Gap acceptance models are formulated as a binary choice problem. The driver either accepts or rejects the available gap, based on comparison of the gap with an unobserved critical gap (minimum acceptable gap). This can be expressed as follows:

$$Y_{nt} = \begin{cases} 1 & \text{if } G_{nt} \geq G_{nt}^{cr} \\ 0 & \text{if } G_{nt} < G_{nt}^{cr} \end{cases} \qquad (2.1)$$

Where,

Y_{nt} = *choice indicator variable with value 1 if the gap is accepted and 0 otherwise*

G_{nt} = *available gap*

G_{nt}^{cr} = *critical gap*

The definition of critical gap varies among different models. In Highway Capacity Manual (1997), the critical gap for a two-way stop controlled intersection, is defined as the minimum time interval in the major-street traffic stream that allows intersection entry to one minor-stream vehicle. In CORSIM (Halati *et al.* 1997), critical gaps are defined through risk factors. The risk factor is defined by the deceleration a driver will have to apply if the leader brakes to a stop. The risk factors are calculated for every lane change based on the relative speed and position of the lead and lag vehicles and compared to an

acceptable risk factor, which depends on the type of lane change to be performed and its urgency. Yang and Koutsopoulos (1996) and Ahmed (1999) define critical gaps as minimum space gaps.

For critical gap, Herman and Weiss (1961) assume an exponential distribution, Drew *et al.* (1967) assume a log-normal distribution, and Miller (1972) assumes a normal distribution. Daganzo (1981) proposes a framework to capture critical gap variation in the population as well as in the behavior of a single driver over time. He uses a multinomial probit formulation appropriate for panel data to estimate parameters of the distribution of critical gaps. Mahmassani and Sheffi (1981) assume that the mean critical gap is a function of explanatory variables, and so could capture the impact of various factors on gap acceptance behavior. They estimate the model for a stop controlled intersection and find that the number of rejected gaps (or waiting time at the stop line) , which captures drivers' impatience and frustration has a significant impact on critical gaps. Madanat *et al.* (1993) use total queuing time to capture impatience. Cassidy *et al.* (1995) differentiate the first gap from subsequent gaps, and gaps in the near lane from gaps in the far lane. These variables significantly improve the fit of the model. Other parameters that may affect critical gaps include the type of maneuver, speeds of vehicles on the major road, geometric characteristics and sight distances, the type of control in the intersection, the presence of a pedestrian, police activities, and daylight conditions (e.g. Brilon 1988, 1991, Adebisi and Sama 1989, Saad *et al.* 1990, Hamed *et al.* 1997). However, most of the discussion is qualitative and addresses macroscopic characteristics rather than microscopic driver behavior.

In congested situations, acceptable gaps are often not available and more complex gap acceptance phenomena may be observed. For example, drivers may change lanes through courtesy of the lag driver in the target lane or decide to force their way in and compel the lag driver to slow down. Existing microscopic traffic simulators, such as AIMSUN, Paramics and VISSIM, use basic or modified versions of their normal gap acceptance models to model freeway merging behavior (TSS 2004, Quadstone 2004, PTV 2004). These models consider gaps created by adjacent vehicles, and in some cases model reduced gap acceptance thresholds during congested conditions, but they do not explicitly consider the anticipatory aspect of cooperation among drivers and aggressive

merges by impatient drivers. Further discussion about gap acceptance in merging conditions is presented in Section 2.4.

2.3   Acceleration Models

Acceleration models can be broadly classified into two groups: car following models and general acceleration models. Car following models describe the behavior of drivers reacting to the behavior of their leaders and the general acceleration models include behaviors in both car following and non car following situations.

The concept of car following was first proposed by Reuschel (1950) and Pipes (1953). Pipes assumes that the follower wishes to maintain safe time headway of 1.02 s from the leader. This value was derived from a recommendation in the California Vehicle Code. Using Laplace transformations, he develops theoretical expressions for the subject's acceleration given a mathematical function that describes the leader's behavior.

Researchers at the GM Research Laboratories introduced the sensitivity-stimulus framework that is the basis for most car following models to date. According to this framework a driver reacts to stimuli from the environment. The response (acceleration) the driver applies is lagged to account for reaction time and is given as follows:

$$response_n(t) = sensitivity_n(t) \times stimulus_n(t - \tau_n) \tag{2.2}$$

Where,

   t=time of observation

   τ_n=reaction time for driver n

The reaction time includes perception time (time from the presentation of the stimulus until the foot starts to move) and foot movement time. The GM models assume that the stimulus is the leader relative speed (the speed of the leader less the speed of the subject vehicle) and the response is linear. Over the years, several extensions to the GM model were proposed to overcome its limitations (Chandler *et al.* 1958, Gazis *et al.* 1959, 1961, May and Keller 1967, Ozaki 1993). Herman and Rothery (1965) and Bexelius (1968) hypothesized that drivers follow vehicles in front of their leader as well as the immediate leader and assumed different sensitivities to the relative speed with respect to each of these leaders.

Lee (1966) developed a variation of the GM model that takes into account the past observations of the driver in the current acceleration decision by means of considering the relative leader speed over a period of time rather than the instantaneous speeds. The mathematical model is expressed as follows:

$$a_n(t) = \int_0^t M(t - t') \Delta V_n^{front}(t') dt' \qquad (2.3)$$

Where,

$\Delta V_n^{front}(t)$=relative speed of front vehicle at time t

$M(.)$=memory (or weighting) function, which represents the way the driver

acts on information that has been received over time.

Lee proposed several functional forms of the memory function and analyzed the stability of the resulting response to periodic changes in the leader speed. Darroch and Rothery (1972) empirically estimated the shape of the memory function using spectral analysis.

Helly (1961), Bekey *et al.* (1977), Gabard *et al.* (1982), Koshi *et al.* (1992) developed acceleration models assuming that the driver tries to attain some desired measure, for example: minimizing both the leader relative speed and the difference between the actual and desired space headway.

Gipps (1981) developed the first general acceleration model that applies to both car following and free flow conditions. The maximum applicable acceleration is determined based on two constraints: the desired speed may not be exceeded and a safe headway must be kept. Models with similar structure are developed by Benekohal and Treiterer (1988) and Hidas (2002). Yang and Koutsopoulos (1996), Ahmed (1999) and Zhang (1998) extended these models by including additional driving regimes (e.g. emergency regime, uncomfortable car following regime etc.).

Multiple driving regimes require definition of boundaries to determine which regime the driver is in. For example, headway thresholds are used to determine whether a vehicle is in the car following or free-flow regimes. However, in most of the above models (except Ahmed 1999), these thresholds are modeled deterministically. Similarly, reaction time is explicitly represented in acceleration models, but is often assumed to be deterministic and assigned arbitrary values.

Moreover, many of these model developments do not involve rigorous estimation of model parameters. Most models either completely ignore the issue of estimation or assume values for some parameters and use ad-hoc procedures to determine values for others.

2.4 Combined Models

Several models have been developed that incorporate multiple model components in a single framework and capture the planning behavior of the drivers to some extent.

Hidas (2002) developed a merging model with components essential for lane changing under congested traffic conditions. In this model, if a vehicle cannot merge by normal gap acceptance, it evaluates the flow conditions in the target lane, and attempts to set an acceleration which may lead to a more favorable situation for lane changing. These decisions constitute the lane changing plan of the driver. Hidas (2005) extended this model and included cooperative merging by explicit modeling vehicle interactions using intelligent agent concepts. In the extended model, drivers in a merging scenario have individual goals and they interact and cooperate with each other to solve the conflicting goals. Lane change maneuvers are classified as free, forced and cooperative based on the relative gaps between the leader and follower. In free lane changes there is no noticeable change in the relative gap between the leader and follower during the whole process, indicating that there is no interference between the subject and the following vehicle. In forced lane change, the gap between the leader and follower is either constant or narrowing before the merge, but starts to widen after the subject vehicle enters, indicating that the subject vehicle has forced the follower to slow down. In cooperative lane change the gap between the leader and follower is increasing before the entry point and starts to decrease afterwards, indicating that the follower has slowed down to allow the subject vehicle to enter. However, it is postulated in this model that each vehicle involved in a lane changing maneuver has perfect information about the lane changing plans of other vehicles and vehicles are able to communicate with each other in order to cooperate, coordinate and resolve conflicts. Video data was used to develop the model, but details of the calibration methodology were not available.

Several other models have been developed specifically to model the cooperative and/or forced lane changing plans of the driver (Ahmed 1996, Wang *et al.* 2005). Ahmed (1996) estimated a forced merging model that captures drivers' lane changing behavior in heavily congested traffic as shown in Figure 2.4. A driver is assumed to evaluate the traffic environment in the target lane to understand whether the driver's right of way is established and a forced merge is possible. If a driver intends to merge in front of the lag vehicle and right of way is established the decision process ends and the driver gradually moves into the target lane. Once the forced merging has started the driver is assumed to remain in this state, persisting till the merge to the target lane is completed. However, the model assumes that once a driver initiates a forced merge, he/she completes it. There is no gap acceptance level after the decision to initiate a forced merge is taken. In other words, the probability of completion of the merge is 1 if the driver has initiated a force merge. Normal lane change and voluntary cooperation among drivers is ignored.
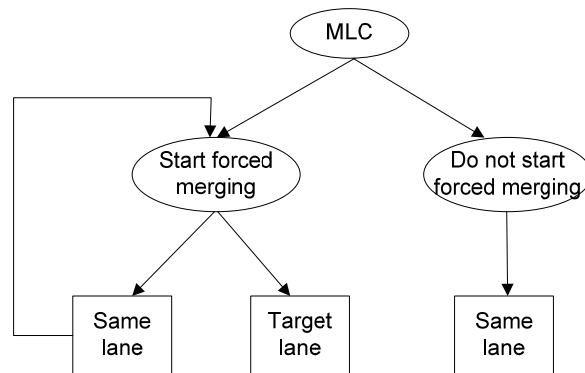
Figure 2.4: The forced merging model structure proposed by Ahmed (1999)

Wang *et al.* (2005) consider the merging plan of the driver with the possibility of courtesy from the lag driver in the mainline. The model framework is presented in Figure 2.5. The probabilities of the lag driver providing courtesy are drawn from binomial distributions with parameters calibrated using video observations. The merging vehicle selects a target gap and accelerates or decelerates to adjust speed and position with respect to that gap. The merge is executed if the target gap is acceptable. The model however ignores the possibility to force merge and if the merging vehicle has not found an acceptable gap before reaching the end of the merging lane, the vehicle is removed

and a merge failure is registered. Moreover, heterogeneity among drivers is not explicitly considered in this model.
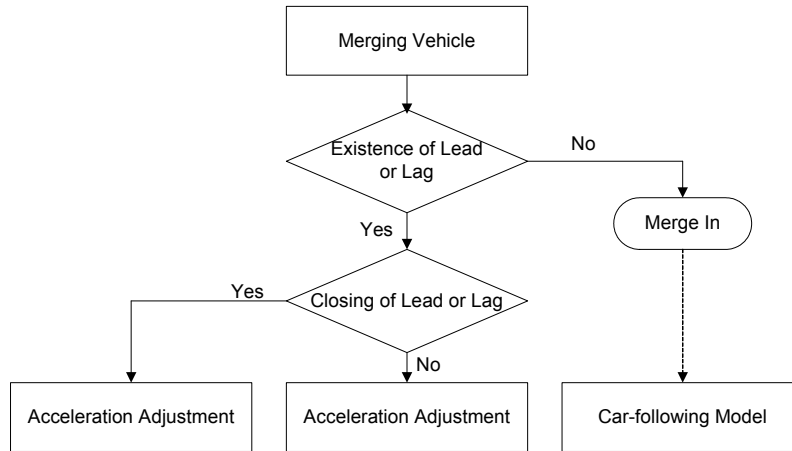
Figure 2.5: The merging model structure proposed by Wang et al. (2005)

Toledo (2002) presented a framework based on the concepts of a short-term goal and short-term plan for a driver. Driving behavior consists of three main elements: the short-term goal, the short-term plan and the driver's actions. The short-term goal is defined by the driver's target lane. The driver constructs a short-term plan, which is defined by the target gap in the target lane that the driver wishes to use in order to accomplish the goal. The accelerations and lane changes are the driver's actions used to execute the short-term plan. The conceptual framework of the model is illustrated in Figure 2.6.
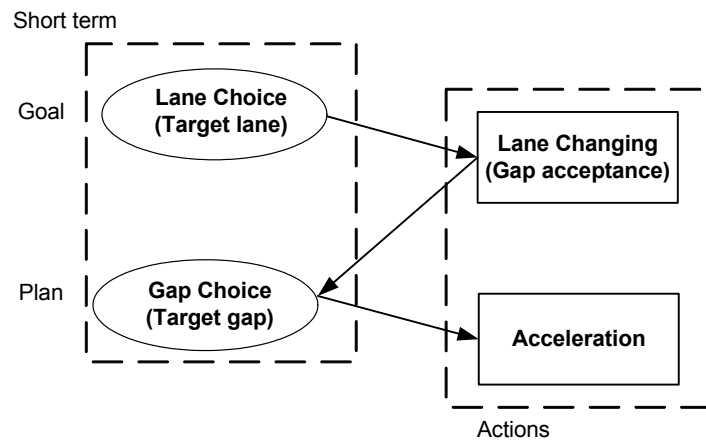
Figure 2.6: Conceptual framework for the driving behavior process (Toledo 2002)

When the adjacent gap is rejected by the driver, the driver creates a short-term plan by choosing a target gap in the target lane traffic. The alternatives in the target gap choice set include available gaps in the vicinity of the subject vehicle. A gap which may not be acceptable at the time of the decision may still be chosen in anticipation of becoming acceptable in the future.

However, due to the computational difficulty of modeling all possible combinations of states of the short-term goal and short-term plan (which are unobserved), a partial short-term plan was hypothesized. It is assumed that the driver executes one step of the short-term plan, re-evaluates the situation and decides the next action to be taken. Thus, it is assumed that a driver formulates a plan at every instant and the effect of previous plans is not fully captured. The structure of the combined lane changing and acceleration model proposed by Toledo is presented in Figure 2.7.
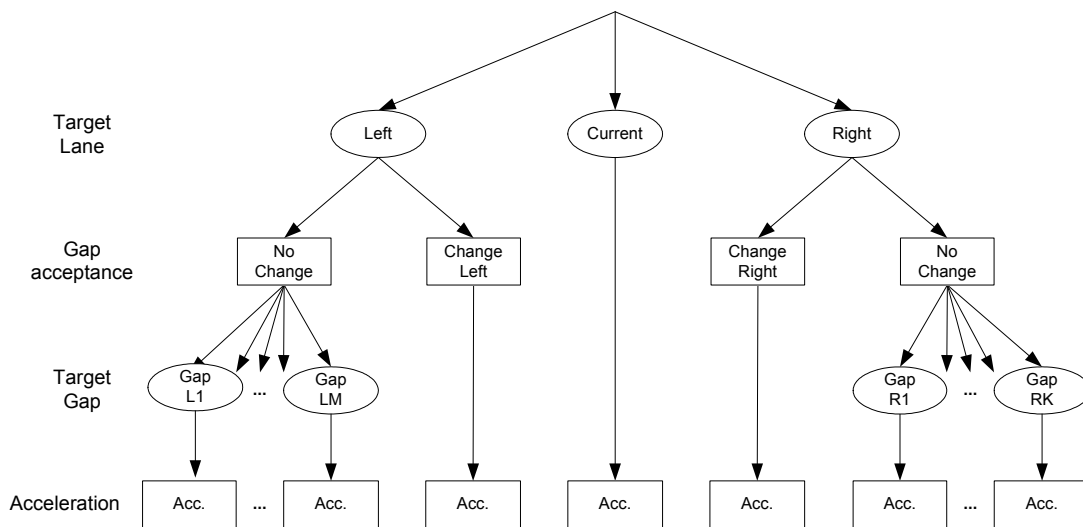
Figure 2.7: Structure of the driving behavior model (Toledo 2002)

The model captures both lane changing and acceleration behaviors. The driver selects the best lane among the current and adjacent lanes and if a lane shift is required, looks for an acceptable gap to make the lane change. Drivers who wish to change lanes but cannot change lanes immediately, select a short-term plan to perform the desired lane change. Short-term plans are defined by the various gaps in traffic in the target lane. Drivers adapt their acceleration behavior to facilitate the lane change using the target gap. The scope of the partial short-term plan thus only captures variables associated with the immediate

surroundings. For example, the choice set for lane selection only includes the current and adjacent lanes, and lanes beyond the adjacent lanes do not affect the lane selection. Similarly, the choice set for target gap selection only includes the adjacent and immediate forward and backward gaps. The model therefore does not address the sequence of maneuvers to achieve a distant target and can fail if there are significant differences in level of service among different lanes.

Rao (2006) formulated a theoretical framework for a dynamic programming based approach to modeling lane changing decisions where expectations of future conditions are explicitly addressed. The solution of the dynamic program takes the form of an optimal decision rule that specifies drivers' optimal utility based decisions as a function of their current information. The computational complexity of applying such a model however prohibited model estimation.

Webster *et al.* (2007) proposed a tactical lane change model using the forward search algorithm. The completed forward search tree enumerates a complete set of subject vehicle maneuver sequences, and each sequence is evaluated in terms of how it improves the distance traversed over the planning horizon. The model however makes several simplifying assumptions. For example, the decisions are based only on distance traversed and effects of path-plan; the inertia in the decision making process and effects of other variables are ignored. Also, it imposes restrictions on lateral movements of other vehicles and ignores the heterogeneity of the planning horizon of drivers. It is mentioned that the model parameters are calibrated with trajectory data using simulation runs but the computational burden associated with the forward search is not detailed.

2.5   Limitations of Existing Models

It is apparent from the critique in the previous sections that although there have been many advances in driving behavior models over the years, the existing models still have significant limitations as described below:

Tactical and strategic planning

Most models assume that drivers make instantaneous decisions based on current traffic conditions. In reality, drivers may conceive a plan and perform it over a length of

time. The notion of planning is ignored in most of the existing models. The few models that address the effect of planning in driving decisions have a limited extent and/or make simplifying assumptions. For example, as described in Section 1.2, the planning process is likely to be affected by strategic trip planning and navigation decisions such as selected trip schedule and path. Drivers may adjust their speeds according to the trip schedule; they may pre-position themselves in correct lanes to follow their path. The effect of path-plan is considered in some of the lane selection models but the effect of the trip schedule has not been incorporated in the existing models.

Anticipation

Anticipation of future conditions has a significant effect on the plans involving the driving decisions. Drivers tend to anticipate the downstream traffic conditions, the behavior of other vehicles etc. and make their decisions to facilitate their plans. Drivers familiar with the network can pre-position in specific lanes in order to avoid delays caused by turning or merging traffic. Drivers may avoid following a bus or delivery vehicle that is likely to make frequent stops. This is more evident in congested and incident affected traffic conditions where consideration of the anticipated conditions can substantially minimize travel delays. The effect of anticipation in strategic driving decisions has not been adequately represented in most of the existing models.

Interdependence

The decisions of a driver over time and choice dimensions are interdependent. For example, a driver's gap acceptance and acceleration decisions can depend on his/her initial decision to change lanes. Interdependencies among decisions, particularly over the time dimension for the same driver are not captured in detail in most of the existing models. For example, the persistence of drivers to follow their originally chosen plans, which can lead to state-dependence, has been ignored in the state-of-the-art models.

Choice set

In most cases, existing models explain driving behaviors using variables related to the subject's immediate driving neighborhood, such as the relative speeds and positions of

neighboring vehicles in the adjacent lanes. But in reality drivers are not myopic and are likely to select their targets based on a broader set of factors.

Mixed traffic

Mixed traffic streams, with vehicles having distinct differences in size and speed sharing the same right of way, exhibit behavior significantly different from homogeneous, lane based traffic streams and are generally characterized by 'weak lane discipline'. The state-of the art driving behavior models have focused on modeling homogeneous lane based traffic conditions and are not applicable in heterogeneous traffic conditions.

Heterogeneity among drivers

The heterogeneity in driver behavior is ignored in most of the existing models, mostly due to data limitations. The heterogeneity in aggressiveness and reaction time of the drivers has been considered in some of the models through estimated distributions. But heterogeneity exists in many other aspects of driving and includes traits of the driver like intelligence, planning capability, risk averseness etc. The effects of socio-economic characteristics of the driver (e.g. age, education, driving experience etc.) on driving behavior have also not been explored.

2.6   Summary

With these limitations, application of the state-of-the-art models in a simulation environment can result in unrealistic traffic flow characteristics. This can result in errors in the corresponding analysis and bias planning and policy decisions. According to the NGSIM study for Identification and Prioritization of Core Algorithm Categories (Alexiadis *et al.* 2004), the scenarios with highest priority include urban arterial lane selection, oversaturated freeway behavior, freeway lane distribution and decisions at a weaving section. As discussed in Section 1.2, a common link between all these scenarios is that the decisions in all these cases involve significant planning and anticipation by the drivers.   Success in bridging the existing gaps in the traffic simulators therefore depends on an efficient modeling technique to address the plans behind the observed decisions.

Chapter 3

Modeling Methodology

This chapter presents a general methodology and framework for modeling behaviors with unobserved or latent plans. The planning behavior of decision makers have been modeled by researchers in many different fields. A short review of these research methodologies are also presented in this chapter.

The chapter is structured as follows: we first present an overview of approaches that are used in different fields to capture the planning behavior of individuals. The features of latent plan models are then presented. The general model frameworks are presented next: first for a basic case with only serial correlation and no state-dependence, and then extended to include state-dependence. The chapter concludes with comparisons of the modeling methodology with the state-of-the-art discrete choice modeling approaches.[1]

3.1 Modeling Planning Behavior

The problems regarding modeling planning and decision making under uncertainty have been addressed by researchers in many different fields, including artificial intelligence, economic analysis, operations research and control theory.

Artificial intelligence planning algorithms are concerned with finding the course of action (plans or policies) to be carried out by some agent (decision maker) to achieve its goals. In the classical case, the aim is to produce a sequence of actions that targets to guarantee the achievement of certain goals when applied to a specified starting state. Decision-theoretic planning (DTP) (Feldman & Sproull 1977) is an attractive extension of the classical artificial intelligence planning paradigm that selects courses of action that

---

[1] Earlier versions of parts of this chapter have been presented in Ben-Akiva *et al.* (2006, 2007a and 2007b)

have high expected utility. These models capture the risks and tradeoffs of different plans rather than guaranteeing the achievement of certain goals. However, in many practical cases, calculation of expected utility involves evaluation of numerous possible plans and it is usually not feasible to search the entire space of plans to find the maximum utility plan. With increasing planning horizon, computing the expected utility of a single plan can also be prohibitively expensive since the number of possible outcomes from the plan can be very large (Blythe 1999). Some other assumptions in artificial intelligence planning algorithms such as complete knowledge of the initial state and completely predictable effects of actions have also been challenged by researchers, for instance, in conditional planning (Peot and Smith 1992) and probabilistic planning (Kushmerick *et al.* 1994).

Dynamic programming techniques have been applied to model the planning behavior in partially observable settings (Smallwood and Sondik 1973). In cases with partially observable current states, past observations can provide information about the system's current state and decisions are based on information gleaned in the past. The optimal policy thus depends on all previous observations of the agent. These history-dependent policies can grow in size exponentially with the length of the planning horizon. While history-dependence precludes dynamic programming, the observable history can often be summarized adequately with a probability distribution over the current state, and policies can be computed as a function of these distributions (Astrom, 1965).

Markov Decision Processes (MDP) (Bellman 1957) assume that current state transitions and actions depend only on the current state and are independent of all previous states. This significantly improves the computational tractability. MDP have two kinds of variables: state variables s_t and control variables a_t. According to Rust (1994) a decision-maker can be represented by a set of primitives (U, p, β) where $U(s_t, a_t)$ is a utility function representing the preferences at time t, $p(s_{t+1} \mid s_t, a_t)$ is a Markov transition probability representing the subjective beliefs about uncertain future states, and $\beta \in (0,1)$ is the rate at which the individual discounts utilities in future periods. Recent research on DTP has explicitly adopted the MDP framework as an underlying model (Barto *et al.* 1995, Boutilier and Dearden 1994, Boutilier *et al.* 1995,

Dean *et al.* 1995, Simmons and Koenig 1995, Tash and Russell 1994), allowing the adaptation of existing results and algorithms for solving MDPs from the field of operations research to be applied to planning problems. The tradeoffs using MDP based utility discounting methods have been reviewed in detail by Rao (2006).

In the artificial intelligence context, the utility of a plan is based on the reward and cost values associated with the actions constituting the plan (Boutelier *et al.,* 1999). Boutelier *et al.* describe two approaches for calculating the utility function: the time-separable approach and the additive approach. In the time-separable approach, the utility is taken to be a function of costs and rewards at each stage, where the costs and rewards can depend on the stage t, but the function that combines these is independent of the stage, most commonly a linear combination or a product (see Luenberger 1973 for details). The addition of rewards and action costs in a system with time-separable value is illustrated in Figure 3.1, where at time t, the cost (C_t) is a function of the previous state (s_{t-1}) and previous action (a_{t-1}) and the reward R_t is a function of the current state (s_t). A value function is additive if the combination function is a sum of the rewards and cost function values accrued over the history of stages. Thus, in both cases, the derivation of the utility functions associated with the plans and actions do not involve any rigorous calibration framework.
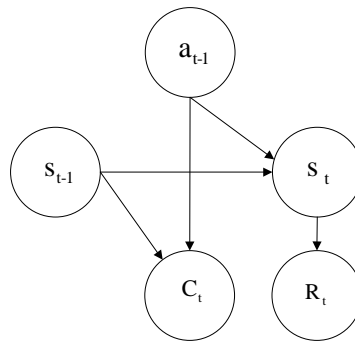
Figure 3.1: Framework for reward and action costs (Boutilier *et al.* 1999)

Baum and Petrie (1966) proposed the Hidden Markov Model (HMM) framework where the system being modeled is assumed to be a Markov process with unknown parameters. The challenge in this framework is to determine the hidden parameters from the observable parameters. This is illustrated in Figure 3.2 where latent plans l affect observed actions j and evolve over time t.
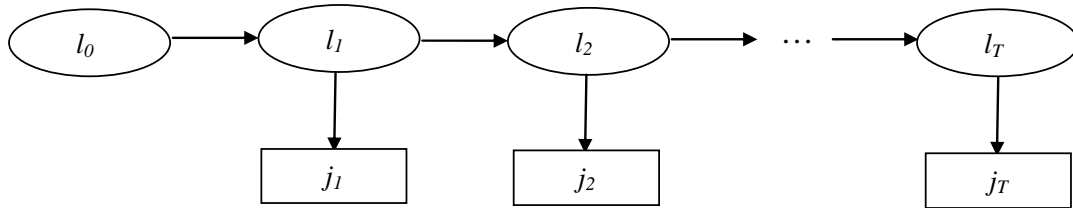
Figure 3.2: First-order Hidden Markov Model (adapted from Bilmes 2002)

The HMM framework has been used in various applications including speech recognition (Rabiner 1989, Baker 1975, Jelinek 1976), machine translation (Vogel *et al.* 1996), bioinformatics (Koski 2001), and the evolution of health and wealth in elderly people (Ribeiro 2002, Ribeiro *et al.* 2003). However, its use in these applications has generally been to model certain processes that do not involve behavioral states. In other words, these applications do not involve choice or decision-making of individuals.

To summarize, planning models in different research fields address the dynamics of planning through various approaches. While the assumptions and perspectives adopted in these areas differ in substantial ways, Markovian approaches are widely used to capture the model dynamics in a tractable manner. However, these models do not focus much on the behavioral aspect of choice or decision making and the methods reviewed in this section are not directly applicable to modeling the evolution of the unobserved driving decisions. But they form the basis of the modeling methodology proposed in the next section.

3.2   Latent Plan Models

The general framework of latent plan models is schematically shown in Figure 3.3. At any instant, the decision maker makes a plan based on his/her current state. The choice of plan is unobserved and manifested through the choice of actions given the plan. The actions are reflected in the updated states.
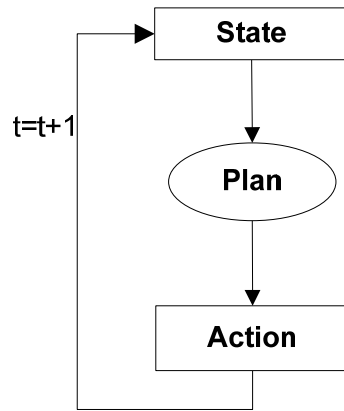
Figure 3.3: General decision structure

The key features of the latent plan model are as follows:

1. Individuals choose among distinct plans (target/tactic). Their subsequent decisions are based on these choices. The chosen plans and intermediate choices are latent or unobserved and only the final actions (maneuvers) are observed.

2. Both the choice of plan and the choice of action conditional on the plan can be based on the theory of utility maximization. The interdependencies and causal relationships between the successive decisions of an individual result in serial correlation among the observations.

3. The observed actions of the individuals depend on their latent plans. The utility of actions and the choice set of alternatives may differ depending on the chosen plan.

4. The choice of the plan at a particular time may depend on previous plans. For example, persistence and inertia effects may affect the choice whether or not to continue to follow the original plan or to shift to an alternative one. Thus, the choice of plans can lead to state-dependence in the decision process.

5. The current plan can also depend on anticipated future conditions and may include expected maximum utility (EMU) derived from the decisions involved with the execution of the plan.

In the following subsections, we first present the basic latent plan model that is applicable for cases without state-dependence (only serial correlation). These include situations involving one-time decisions, as well as panel observations where the subsequent choices of plans (conditional on individual-specific characteristics) are

independent. The basic model is then extended to explicitly capture the state-dependence between subsequent plans and actions.

3.2.1   Latent Plan Model without State-dependence

In this section the basic latent plan model framework is presented. This framework only addresses the serial correlation among the decisions of the individual across time and choice dimensions but do not address the state-dependence among subsequent plans. That is, conditional on individual-specific characteristics, the successive plans of individuals are assumed to be independent. The overall model formwork is presented in Figure 3.4. Variables or choices in rectangles are observable, while those in ovals are unobservable or latent.
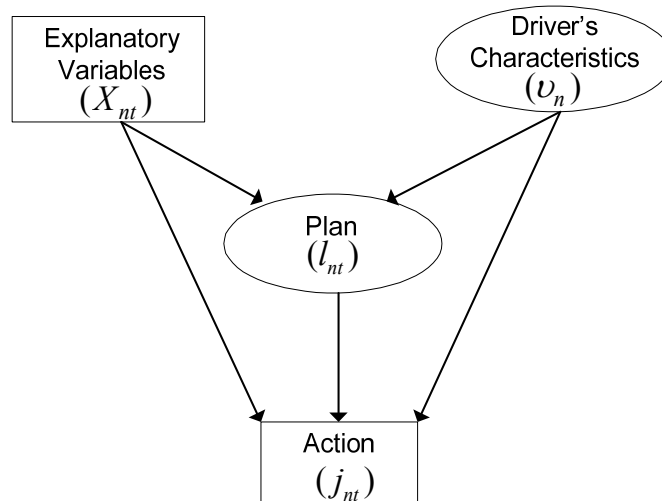
Figure 3.4: Latent plan model without state-dependence

The plan of an individual n at any instant t (l_{nt}) is influenced by explanatory variables and individual-specific characteristics. The attributes of the alternatives (X_{nt}) are generally observed but the individual-specific characteristics associated with the individual (υ_n) are generally unobserved or latent. For example, in case of lane selection behavior, attributes of the alternatives (target lanes) like average speed, density, lead and lag vehicle characteristics etc. are observed and driver characteristics like aggressiveness, driving skills, planning horizon etc. are latent. These latent variables can be discrete or continuous. Characteristics of the driver such as planning capability, for example, can be

represented by discrete classes of drivers (e.g. drivers who plan-ahead and drivers who do not). Continuous latent variables include attitudes, perceptions and personality traits of the individual (e.g. impatience, aggressiveness, planning horizon etc.). The actions of the individuals depend on the chosen plan as well as the observed and latent explanatory variables. These individual specific variables remain the same for all decisions of the same individual across time and choice dimensions (agent effect). However, it is assumed that actions (j_n) and plans (l_n) of individual n (conditional on υ_n) are independent over time. This assumption is relaxed in Section 3.2.2.

The general model framework is presented in Figure 3.5. This framework consists of two levels: choice of plan and choice of action conditional on the plan. The selection of the plan (indexed by l) in the upper level drives the selection of an action (indexed by j). The action choice sets and corresponding utilities, shown in the lower level, may vary depending on the plan.
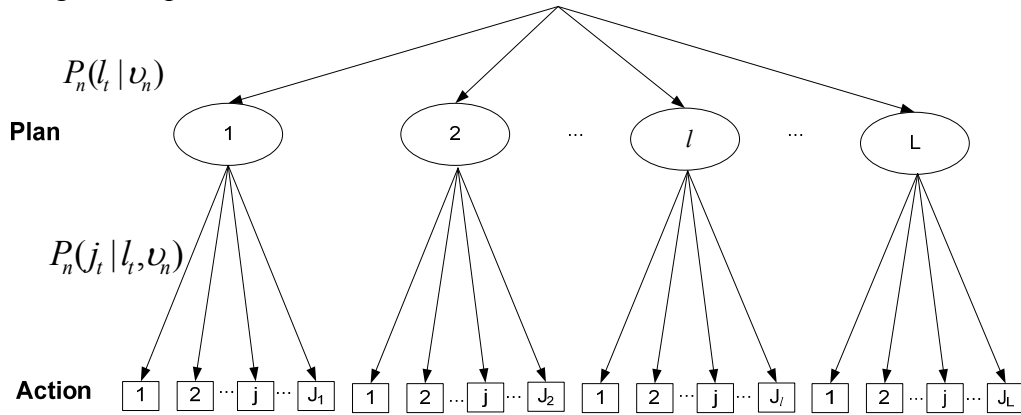
Figure 3.5: Basic model framework (without state-dependence)

Probability of a Trajectory

The trajectory of an individual includes a series of observed actions. For driving behavior models, this corresponds to a series of lane actions and acceleration decisions of the driver.

Let,

$P_n\left(l_t \mid \upsilon_n\right)$ = probability of individual n selecting plan l at time t conditional on individual-specific characteristics

$P_n\left(j_t \mid l_t, \upsilon_n\right)$ = probability of individual n selecting action j at time t given plan l conditional on individual-specific characteristics

$P_n (j_t | \upsilon_n)$ = probability of action j by individual n at time t conditional on individual-specific characteristics

L_n = the set of plans in the choice set of individual n

T_n = number of consecutive observations of individual n

At time t for individual n, the probability of observing a particular action j is the sum of probabilities that he/she is observed to execute action j given that the selected plan is l, over all plans in the choice set of the individual.

$$P_n (j_t | \upsilon_n) = \sum_{l \in L_n} P_n (j_t | l_t, \upsilon_n) P_n (l_t | \upsilon_n) \tag{3.1}$$

Assuming that actions (j_n) and plans (l_n) of individual n (conditional on υ_n) are independent over time (relaxed in next section), the probability of observing his/her sequence of decisions can be expressed as follows:

$$P_n (j_1, j_2, ..., j_{T_n} | \upsilon) = \prod_{t=1}^{T_n} \sum_{l=1}^{L_n} P_n (j_t | l_t, \upsilon_n) P_n (l_t | \upsilon_n) \tag{3.2}$$

The unconditional choice probabilities of observing the sequence of decisions by individual n are given by the following equation:

$$P_n (j_1, j_2, .. j_T) = \int_{\upsilon} P_n (j_1, j_2, .. j_{T_n} | \upsilon) f (\upsilon) d\upsilon \tag{3.3}$$

Where, $f(\upsilon)$ is the distribution of the individual-specific random term (e.g. aggressiveness).

Specification

The probabilities of choice of plan and action can be calculated using a utility-based choice framework. The specifications of these utilities are discussed below.

Choice of Plan

The choice of a plan can be based on utility maximization and may include expected maximum utility (EMU) derived from the decisions involved with executing that plan.

The utility of latent plan l for individual n at time t can be expressed as follows:

$$U_{lnt} = U\left(X_{lnt}, I_{lnt}, \upsilon_n, \varepsilon_{lnt}\right)$$

$$I_{lnt} = \mathrm{E}\left(\max\left(U_{1\,lnt}, U_{2\,lnt}, \ldots, U_{j\,lnt}, \ldots, U_{J_l\,lnt}\right)\right)$$

(3.4)

Where,

X_{lnt} =attributes of plan l for individual n at time t, a subset of X_{nt}

I_{lnt} =expected maximum utility from actions associated with plan l of individual

  n at time t

U_{jlnt} =utility of action j under plan l to individual n at time t

υ_n = individual-specific random effect

ε_{lnt} =random utility component of plan l for individual n at time t

Choice of Action

The observed choices/actions depend on the chosen plan. The choice set, as well the functional form of the utility of an action j may vary depending on the chosen plan. The utility of action j under plan l can be expressed as follows:

$$U_{jlnt} = U\left(X_{jlnt}, \upsilon_n, \varepsilon_{jlnt}\right)$$

(3.5)

Where,

X_{jlnt} = atttributes of action j and plan l at time t, a subset of X_{nt}

υ_n = individual-specific random effect

ε_{jlnt} = random utility component of action j and plan l at time t

The conditional probabilities of selecting plan $\left(P_n\left(l_t \mid \upsilon_n\right)\right)$ and action $\left(P_n\left(j_t \mid l_t, \upsilon_n\right)\right)$ are based on the utilities discussed above (U_{lnt} and U_{jlnt} respectively). The specification of the probabilities will depend on the assumptions made regarding the distribution of the random utility components of U_{lnt} and U_{jlnt}. For example, if the random components are independently and identically extreme value distributed, then the kernel of the choice model will be logit.

3.2.2 Latent Plan Models with State-dependence

In the model with explicit consideration of state-dependence, the previous assumption regarding independence of successive plans of individuals (conditional on individual-

specific characteristics) is relaxed. Selection of plan l by individual n at time t in this case is influenced by his/her previously chosen plans and actions leading to state-dependence in the choice process. The overall framework of latent plan models with state-dependence is presented in Figure 3.6.

As shown in the figure, in the general case, the plan at time t is influenced by previous plans $(l_{n1}, l_{n2}, ..., l_{n\,t-1})$ and previous actions $(j_{n1}, j_{n2}, ..., j_{n\,t-1})$ in addition to the current attributes of the alternatives and individual-specific characteristics. The observed choices/actions depend on the previously chosen plans and actions as well as the current plan, attributes of the alternatives and individual-specific characteristics.
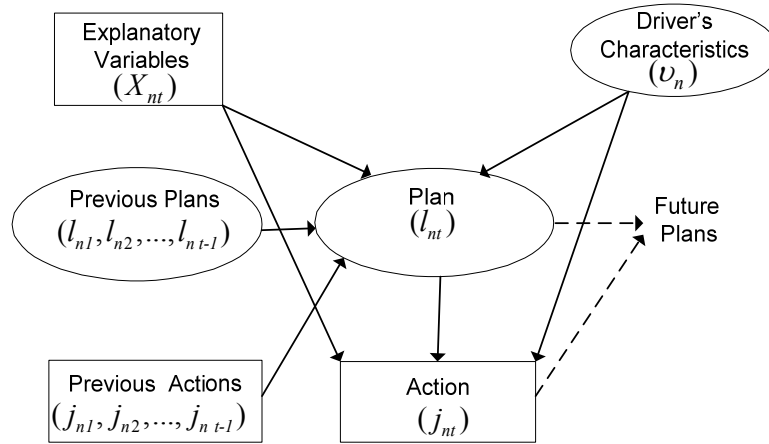
Figure 3.6: Model framework of latent plan models with state-dependence

Probability of Trajectory

As in the case presented before, the trajectory of an individual includes a series of observed actions. But in this case the conditionality of current plans and actions on previous plans and action are considered.

Let,

$P_n\left(l_t \mid l_{1:t-1}, j_{1:t-1}, \upsilon_n\right)$ = conditional probability of individual n selecting plan l at time t

$P_n\left(j_t \mid l_{1:t}, j_{1:t-1}, \upsilon_n\right)$ = conditional probability of individual n selecting action j at time t

$P_n\left(j_t \mid \upsilon_n\right)$ = conditional probability of action j by individual n at time t

L_n = plans in the choice set of individual n

Where, $1\!:\!t$ is shorthand for $1, 2, ..., t\text{-}1, t$.

At time t for individual n, the probability of observing a particular action j is the sum of probabilities that he/she is observed to execute action j given that the selected plan is l, over all sequences of plans that could have led to plan l.

$$P_n\left(j_t \mid j_{1:t-1}, \upsilon_n\right) = \sum_{(l_1,\ldots,l_t)} P_n\left(j_t \mid l_{1:t}, j_{1:t-1}, \upsilon_n\right) P_n\left(l_t \mid l_{1:t-1}, j_{1:t-1}, \upsilon_n\right) \tag{3.6}$$

The number of possible sequences in the summation of Equation 3.6 is $|l|^T$, where $|l|$ denotes the maximum cardinality of the set of discrete plans over all decision instances. Except for degenerate cases with a very small choice set of plans or a very short observation period, modeling all possible sequences is thus prohibitively expensive.

Application of a first order Hidden Markov Model (HMM) (Baum and Petrie 1966, Baum 1972) based solution approach simplifies the problem of estimating the model with a large number of latent plans and/or observation periods. HMM is represented graphically in Figure 3.7, in which the upper level represents the evolution of the plans from an initial plan at time 0 (denoted as l_0) to a final plan at time T denoted as l_T. The plan at every time period is determined only by the plan at the previous time period (first-order Markov model) and may be affected by the action taken in the previous time period (experience). The lower level represents the observed actions. An action at a given time period is determined only by the plan during the same time period. Also, the dynamics in the observed actions are explained by the dynamics in the underlying latent or unobserved plans (Hidden Markov Model).
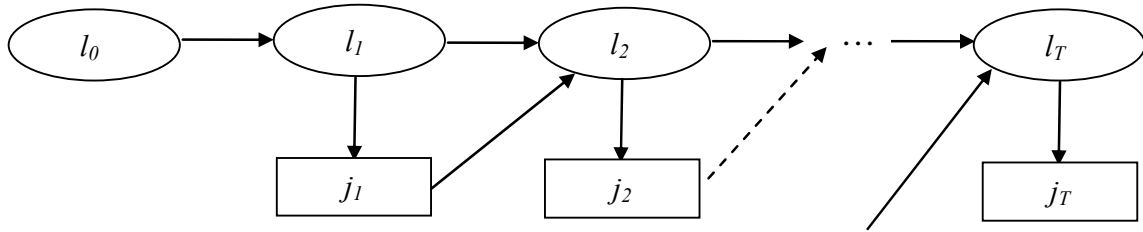
Figure 3.7: First-order Hidden Markov Model
(latent plans l affect observed actions j and evolve over time t)

The first order HMM assumption thus enables us to simplify the choice of plan and choice of action. This can be expressed as follows:

Plans: The plan at a given time period depends only on the plan of the previous time period and all previous actions. The expression for the choice probability of a plan in the current time period, under the above assumptions, is as follows:

$$P_n\left(l_t \mid l_{1:t-1}, j_{1:t-1}, \upsilon_n\right) = P_n\left(l_t \mid l_{t-1}, j_{1:t-1}, \upsilon_n\right) \tag{3.7}$$

Actions: The dynamics in the observed actions are caused by the dynamics in the latent plans. That is, the effects of past plans and past actions affect the current actions through the choice of current plan and there is no direct causal effect of past plans and past actions on the current actions. Therefore, conditional on the plan, the action observed at a given time period is independent of the plans and actions observed at previous time periods; it is only dependent on the current plan.

$$P_n\left(j_t \mid l_{1:t}, j_{1:t-1}, \upsilon_n\right) = P_n\left(j_t \mid l_t, \upsilon_n\right) \tag{3.8}$$

The model framework is presented in Figure 3.8.
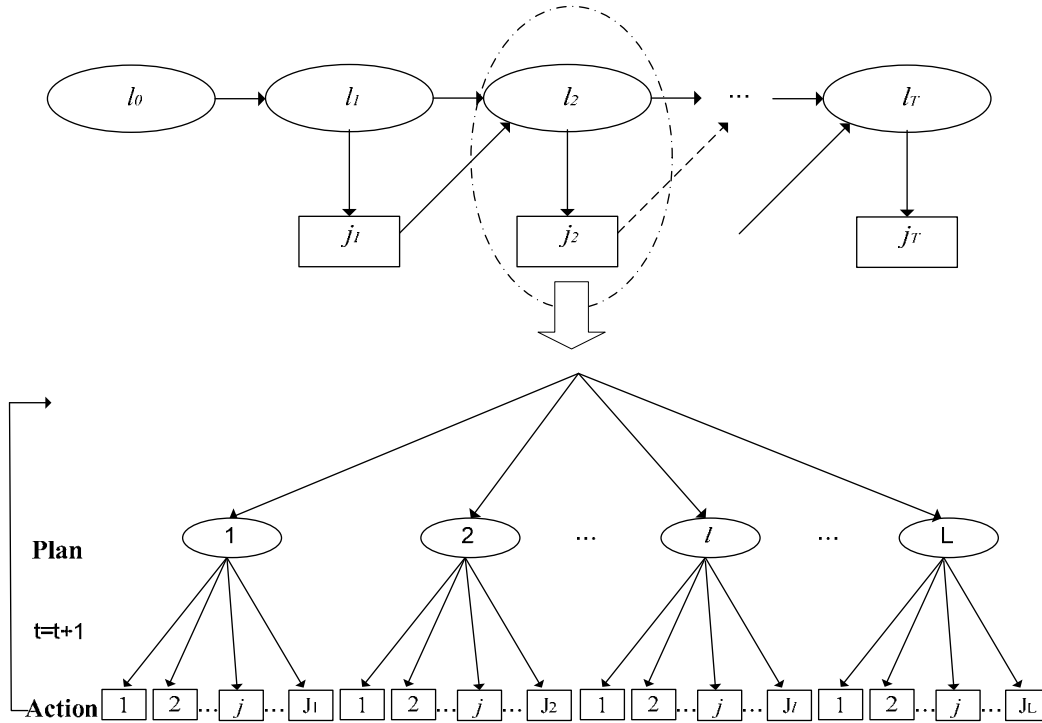
Figure 3.8: Model framework with state-dependence

Under these assumptions, the probability of observing a particular action j at time t can be expressed as follows:

$$P_n\left(j_t \mid j_{1:t-1}, \upsilon_n\right) = \sum_{(l_1,\dots,l_t)} P_n\left(j_t \mid l_t, \upsilon_n\right) P_n\left(l_t \mid l_{t-1}, j_{1:t-1}, \upsilon_n\right) \tag{3.9}$$

The joint probability of a sequence of actions of an individual n over a time horizon T_n can be expressed as follows:

$$P_n\left(j_1,\dots,j_{T_n} \mid \upsilon\right) = \sum_{(l_1,\dots,l_{T_n})} P_n\left(j_1 \mid l_1, \upsilon_n\right) \cdots P_n\left(j_{T_n} \mid l_{T_n}, \upsilon_n\right) P_n\left(l_1 \mid l_0, \upsilon_n\right) \cdots P_n\left(l_{T_n} \mid l_{T_n-1}, j_{1:T_n-1}, \upsilon_n\right) \tag{3.10}$$

$$= \sum_{l_{T_n}} P_n\left(j_{T_n} \mid l_{T_n}, \upsilon_n\right) \sum_{l_{T_n-1}} P_n\left(l_{T_n} \mid l_{T_n-1}, j_{1:T_n-1}, \upsilon_n\right) P_n\left(j_{T_n-1} \mid l_{T_n-1}, \upsilon_n\right) \cdots \sum_{l_1} P_n\left(l_2 \mid l_1, j_1, \upsilon_n\right) P_n\left(j_1 \mid l_1, \upsilon_n\right) P_n\left(l_1 \mid l_0, \upsilon_n\right)$$

Where, the initial plan l_0 is assumed to be fixed or, if random, can be assumed to be handled through specific methods designed for dealing with initial conditions problems in this context (see for example Wooldridge 2003). The above simplification reduces the order of complexity for computing the probability from $O\left(|l|^T\right)$ to $O(lT)$, where $|l|$ denotes the maximum cardinality of the set of discrete plans over all decision instances.

The unconditional choice probabilities of observing the sequence of decisions are given by:

$$P_n\left(j_1, j_2, \dots j_T\right) = \int_{\upsilon} P_n\left(j_1, j_2, \dots j_{T_n} \mid \upsilon\right) f(\upsilon) d\upsilon \tag{3.11}$$

Where, $f(\upsilon)$ denotes the distribution of the individual-specific random effect.

Specification

The probabilities of choice of plan and action can be calculated using a utility-based choice framework. The specifications of these utilities are discussed below.

Choice of Plan

With HMM assumptions, the choice of the plan at time t in the state-dependent case depends on the choice of plan in the previous time period $\left(l_{n\,t-1}\right)$ and all previous

actions $\left(j_{n\ 1:t-1} \right)$. As in the case without state-dependence, the choice of the plan can be a function of attributes of the plans and individual-specific characteristics, and may include expected maximum utility (EMU) derived from the decisions involved with executing that plan. The utility of latent plan l for individual n at time t can therefore be expressed as follows:

$$U_{lnt} = U\left(X_{ln\,t}, l_{n\,t-1},\ j_{n\ 1:t-1}\ , I_{lnt}, \upsilon_n, \varepsilon_{lnt} \right)$$
$$I_{lnt} = \mathrm{E}\left(\max\left(U_{1\,lnt}, U_{2\,lnt}, ..., U_{j\,lnt}, ..., U_{J_l\,lnt} \right) \right)$$

(3.12)

Where,

X_{lnt} =attributes of plan l for individual n at time t

$U_{j\,lnt}$ =utility to individual n from action j at time t under plan l

I_{lnt} =expected maximum utility from actions associated with plan l of individual

n at time t

υ_n = individual-specific random effect

ε_{lnt} =random utility component of plan l for individual n at time t

Choice of Action

According to the HMM assumption, the action observed at a given time period depends on the current plan. The plan and action of previous time periods affect the current action through the current plan. The utility of action j under plan l can therefore be expressed as follows:

$$U_{jlnt} = U\left(X_{jln\,t}, l_{nt}, \upsilon_n, \varepsilon_{jlnt} \right)$$

(3.13)

Where,

X_{jlnt} = atttributes of action j under plan l at time t

υ_n = individual-specific random effect

ε_{jlnt} = random utility component of action j and plan l at time t

The specification of the conditional probabilities of plan $\left(P_n\left(l_t \mid l_{t-1}, j_{1:t-1}, \upsilon_n \right) \right)$ and action $\left(P_n\left(j_t \mid l_t, \upsilon_n \right) \right)$ will depend on the assumptions made regarding the distribution of the random utility components of U_{lnt} and U_{jlnt}. For example, if the random components

are independently and identically extreme value distributed, then the kernel of the choice model will be logit.

3.3 Comparison with Other Discrete Choice Modeling Approaches

The latent plan choice model presented in the previous section have similarities with existing discrete choice models that are commonly used to model choice behavior from multidimensional choice sets (see Ben-Akiva and Lerman, 1985 and the recent update in Ben-Akiva and Bierlaire, 2003).

From the structural point of view, the latent plan models resemble the cross-nested logit (CNL) model (McFadden 1978), where an alternative can share unobserved utility components from different nests (Figure 3.9).
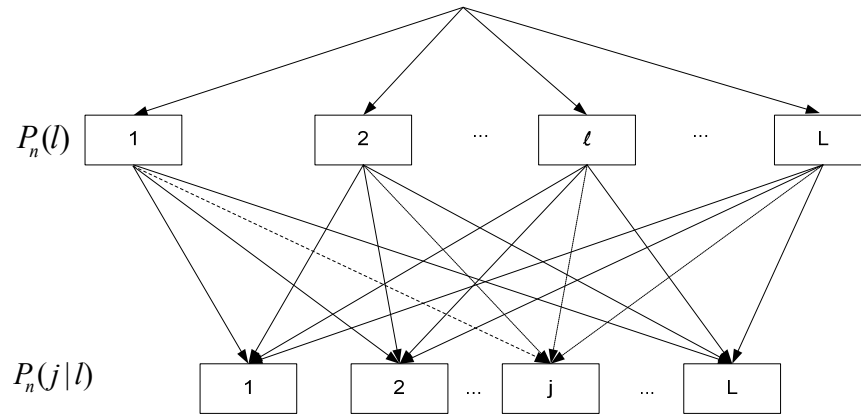
Figure 3.9: Cross-nested logit model

For the two-dimensional case presented in Figure 3.9, the probability of selecting an alternative in the lower level can be expressed as follows:

$$P_n(j) = \sum_{l=1}^{L} P_n(j \mid l) P_n(l) \tag{3.14}$$

Where,

$P_n(l) = $ *probability of choosing l*

$P_n(j \mid l) = $ *probability of choosing j given l*

L=number of alternatives in upper level

A CNL model for the plan and action case thus assumes that the marginal probability of choosing a particular action can be obtained by summing the joint probabilities of that action and each plan leading to the action over all plans. However, in CNL models, the systematic utilities of an alternative at the lower level are independent of the upper nest. That implies that the utility associated with alternative j given l can be expressed as follows:

$$U_{jln} = U\left(X_j, X_n, \varepsilon_{jln}\right)$$ (3.15)

Where,

$X_j =$ *attributes of alternative j*

$X_n =$ *characteristics of decision-maker*

$\varepsilon_{jln} =$ *random utility component*

Thus in CNL, the choice of an action is unaffected by the chosen plan that led to that particular action. In latent plan models, on the other hand, the utilities of the alternatives at the execution level depend on the plan that led to that decision. Moreover, CNL models cannot capture the choice of individuals in complex situations where observable choices are affected by dynamic planning.

Another existing discrete choice model that is similar to the latent plan model is the Latent Class Choice Model (LCCM) where the factors 'generating' the heterogeneity among individuals can be conceptualized as discrete or categorical constructs (Kamakura and Russell 1989, Gopinath 1995). The latent class choice model can be expressed as follows:

$$P_n\left(j\right) = \sum_{l=1}^{L} P_n\left(j \mid l\right) P_n\left(l\right)$$ (3.16)

Where,

$P_n\left(l\right) =$ *class-membership model*

$P_n\left(j \mid l\right) =$ *class-specific choice model*

$L =$ *number of classes*

The class-specific choice models are characterized by heterogeneity in taste variation and/or choice sets associated with the class. If individual n belongs to class l, his/her utility associated with alternative j is as follows:

$$U_{jln} = U\left(l, X_j, X_n, \varepsilon_{jln}\right) \tag{3.17}$$

Where,

X_j = *vector of attributes of alternatives*

X_n = *vector of characteristics of decision-maker*

However, the class-membership models are based only on characteristics of the individuals and not on other variables that influence their attitude. The utility associated with the probability of class-membership can be expressed as follows:

$$U_{ln} = U\left(X_n, \varepsilon_{ln}\right) \tag{3.18}$$

The membership of an individual in a class is thus static and do not change over time with change in situations. The latent plan models on the other hand, are estimated with panel data and the unobserved factor (the latent plan) can vary dynamically with change in situation based on neighborhood variables. The latent plan models thus have a more flexible structure and can therefore be inferred as an extension of LCCM that is applicable in a dynamic case.

3.4   Summary

A general methodology and framework for modeling behaviors with unobserved or latent plans has been presented in this chapter. The action at any time depends on the plan at that time. For situations where the subsequent choices of plans conditional on individual-specific characteristics are independent, the plan at any time can be affected only by the attributes of plans, expected utilities of executing the plan and the characteristics of the individual. However, in the state-dependent case, the current plan can also depends on previous plans and actions as well as attributes of different plans, expected utilities of executing the plans and the characteristics of the individual. The computational tractability of the state dependent model is attained by using the HMM approach. The HMM assumptions imply that the current plan depends only on the plan and action of the previous time step, the attributes of alternative plans and the characteristics of the individual.

Structurally, the proposed latent plan model has similarities with CNL and LCCM. The model comparison reveals that latent plan models can be viewed as a hybrid of these models extended to a dynamic setting.

Chapter 4

Freeway Lane Changing

In this chapter, the latent plan involving the lane changing decision of a driver in a freeway is presented. The overall decision framework consists of the two stages presented in the introductory chapters: choice of latent plans followed by selection of action to execute the plan. However, the detailed structure is formulated based on the geometric configuration and traffic attributes that characterize a typical freeway lane selection scenario.

The chapter is organized as follows: the background of the research is presented in Section 4.1. In Section 4.2, the structure of the latent plan lane changing model is proposed. The details of the model estimation are presented in Section 4.3. This section includes description of the data used to estimate the model parameters, the likelihood function and the estimation results. This section also includes statistical comparison of the goodness-of-fit of the latent plan model and a reduced form model (estimated with the same data). The aggregate validation results are presented in Section 4.5. The calibration and validation exercises within the microscopic traffic simulator MITSIMLab are presented in this section followed by a summary of the validation results within the commercial simulators. The chapter concludes with a summary of the findings.[2]

---

[2] The model presented in this chapter has been developed as part of the NGSIM program of FHWA. The results presented in this chapter have been reported in Choudhury (2005), Toledo *et al.* (2005) and Choudhury *et al.* (2006, 2007). The validation exercises in AIMSUN, Paramics and VISSIM have been performed by TSS (Barceló *et al.* 2006), Quadstone (Speirs 2006) and PTV (Vortisch and Rössel 2006) respectively.

4.1 Background

A driver in a freeway is likely to choose the lane that he/she perceives to be the best and construct a tentative plan to move to that target lane. However, because of the neighboring vehicles, it may not be possible to execute this plan immediately. A lane change occurs in the direction implied by the chosen target lane only if the available gaps are acceptable. The plan that is the choice of the target lane is therefore unobserved and the observed actions are the gap acceptance decisions in the direction of the target lane. In highly congested situations, where acceptable adjacent gaps are not readily available, the plan may also include selection of target gaps and involve alternative lane changing tactics (e.g. courtesy/forced gap acceptance). However, the focus of this chapter is modeling a freeway lane changing scenario with moderate congestion where the target gap is always the adjacent gap and the lane changing tactic is normal gap acceptance. In such situations, the lane changing maneuver of drivers is a two stage process:

- Choice of target lane (plan)
- Decision to accept available gaps and make the lane change (action)

This is illustrated with a hypothetical scenario of a four lane road in Figure 4.1. In this example, Lane 1 is a High Occupancy Vehicle (HOV) lane with significantly higher level of service compared to the other lanes. The lane utilities may be affected by various variables but for simplicity it has been assumed in this example that the lane utilities are fully captured by the average speed. It is further assumed that the subject driver (driver A), is eligible to enter the HOV lane. Driver A is therefore likely to choose Lane 1 as the target lane and look for gaps in Lane 2 to reach Lane 1 eventually. If the available gap is acceptable, the driver is observed to make a lane change to Lane 2. If the gap is not acceptable, he/she is still observed in the current lane (Lane 3). Therefore, an observation of lane change to Lane 2 can result from the plan to move to either Lane 2 or Lane 1. An observation of no lane change can be due to the fact that Lane 3 is indeed the best available lane or another lane is the target lane but maneuver in that direction is not possible. Thus the observed lane action can result from many possible plans.
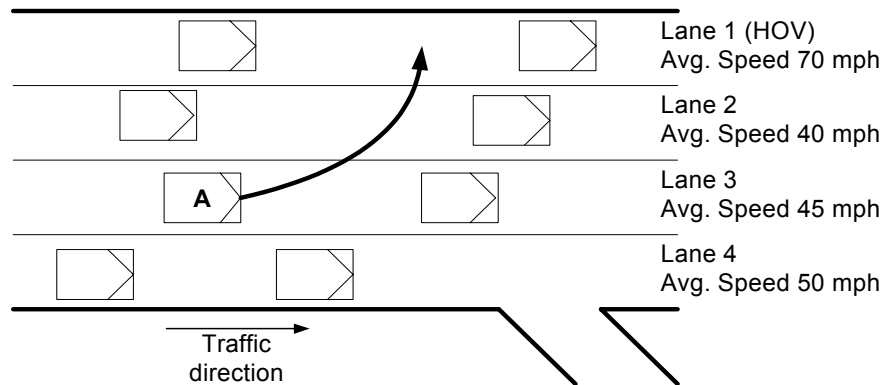
Figure 4.1: Illustration of myopic behavior in existing lane changing models

As mentioned in the literature review in Chapter 2, most lane changing models (e.g., Gipps 1986, Yang and Koutsopoulos 1996, Zhang *et al.* 1998, Ahmed 1999, Hidas and Behbahanizadeh 1999, Hidas 2002, Toledo *et al.* 2003) are based on the assumption that drivers evaluate the current and adjacent lanes and choose a direction of change (or not to change) based on the attributes of these lanes only. The lane choice set is therefore dictated by the current position of the vehicle, and in multi-lane facilities would be restricted to a subset of the available lanes. Thus, existing models lack an explicit tactical choice of a target lane, which may require a sequence of lane changes from the current lane. Instead, these myopic models can only explain one lane change at a time. The need to improve the existing freeway lane selection model is also reflected in the findings of the NGSIM study on Identification and Prioritization of Core Algorithm Categories, where development of freeway lane selection model was ranked as third in importance by model developers and users (Alexiadis *et al.* 2004).

This deficiency of existing models is most evident in situations where there are large differences in the attributes of the available lanes. An example of this is facilities with HOV lanes or other types of exclusive lanes, where a particular lane may be significantly more attractive compared to other lanes. Eligible vehicles may make several lane changes in order to get to the exclusive lane. However, in existing models since only the adjacent lanes are considered for each lane change, the influence of a non-adjacent exclusive lane may not be captured. To illustrate this, consider the hypothetical situation presented in Figure 4.1. With existing models, the driver only compares the current lane (Lane 3) with

the left lane (Lane 2) and the right lane (Lane 4). Based on the lane speeds, Lane 4 is the most desirable of the three and the model will indicate that the driver will try to change to this lane. However, a more plausible model would be that based on the average lane speeds the driver chooses the HOV lane (Lane 1) as the most desirable lane. Thus, driver A is likely change to Lane 2 to reach Lane 1 eventually. In other words, the driver is likely to move to a 'worse' adjacent lane (Lane 2) as the means of getting to a 'lot better' target lane further away (Lane 1).

4.2   Model Structure

The discussion in the previous section demonstrates the need to introduce an explicit choice of target lane in the lane changing model framework. The target lane is the lane the driver perceives as the best lane to be in considering a wide range of factors and goals. These factors may include attributes of specific lanes as well as variables that relate to the spatial relations between the subject vehicle and neighboring vehicles, the driver's path-plan and driver-specific characteristics. The choice of the immediate direction for changing lanes is determined by the direction from the current lane to the target lane.

Examples of the structure of this lane changing model are shown in Figure 4.2. The decision structure shown on the top (Figure 4.2a) is for the driver of a vehicle that is currently in the third lane (Lane 3) in a four-lane road. Lanes 1 and 2 are on its left, and Lane 4 is on its right. At the highest level, the driver chooses the target lane. In contrast with existing models, the choice set constitutes all four lanes in the road (Lanes 1, 2, 3 and 4). If the target lane is the same as the current lane (Lane 3 in this case), no lane change is required (No Change). Otherwise, the direction of change is to the right if the target lane is Lane 4, and to the left if the target lane is Lane 1 or Lane 2. If the target lane choice dictates a lane change, the driver evaluates the gaps in the adjacent lane corresponding to the direction of change and either accepts the available gap and moves to the adjacent lane (Change Right or Change Left) or rejects the available gap and stays in the current lane (No Change). The bottom decision structure (Figure 4.2b) is for the driver of a vehicle in Lane 1 in a similar setting.

The model hypothesizes two levels of decision-making: the target lane choice and the gap acceptance. The target lane choice and the direction of immediate lane change that is implied by the selected target lane are latent. Only completed lane changes (or No Changes) are observed. In the figure latent choices are shown as ovals and observed choices are represented as rectangles.
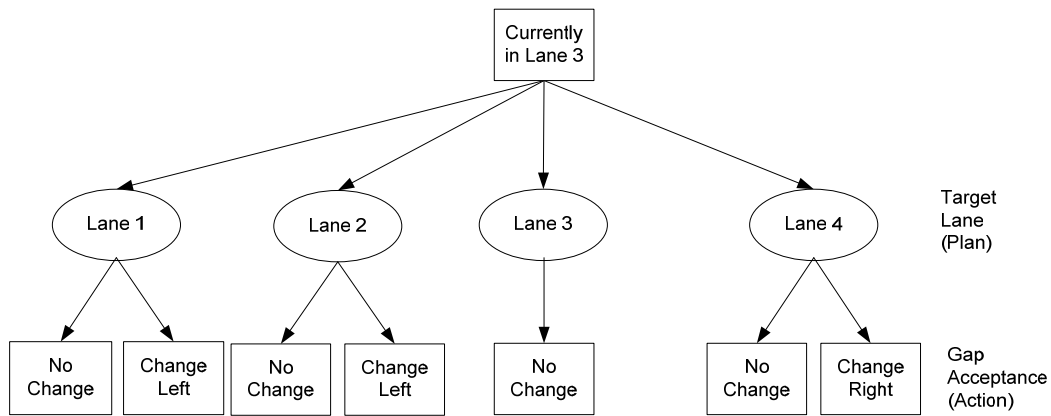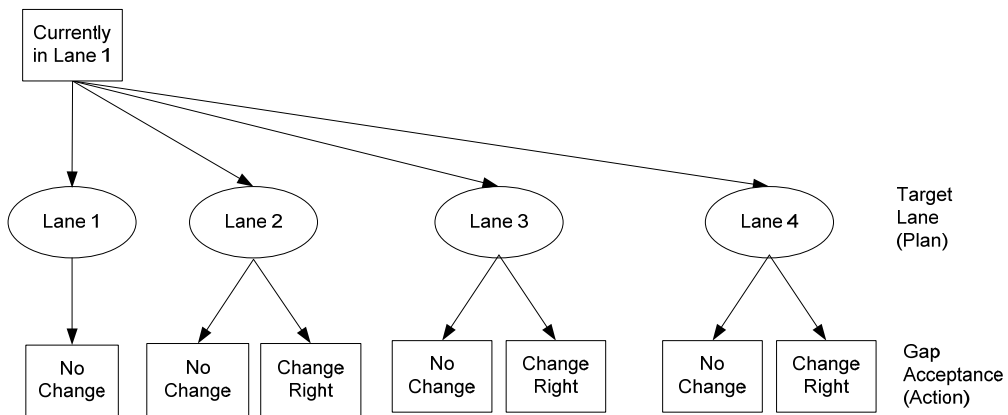
Figure 4.2: Examples of the structure of the proposed lane changing model

We now describe in detail the specification of the models to explain the two choices drivers make within the latent plan lane changing model: the target lane choice and the gap acceptance.

66

4.2.1 Choice of Plan: The Target Lane Model

At the highest level of lane changing, the driver chooses the lane with the highest utility as the target lane. The target lane choice set constitutes all the available lanes in the roadway.

The total utility of lane l as a target lane to driver n at time t can be expressed as follows

$$U_{lnt} = V_{lnt} + \varepsilon_{lnt} \qquad \forall l \in L_n \tag{4.1}$$

Where,

V_{lnt} = *systematic component of the utility*

ε_{lnt} = *random utility component of target lane l for individual n at time t*

L_n = *choice set of target lane of driver n*

The systematic utilities can be expressed as follows:

$$V_{lnt} = V(X_{lnt}, \beta, \alpha^l, \upsilon_n) \quad \forall l \in L_n \tag{4.2}$$

Where,

X_{lnt} = *explanatory variables that affect the utility of lane l*

β = *corresponding vector of parameters*

υ_n = *individual-specific random effect (e.g. aggressiveness):* $\upsilon_n \sim N(0,1)$

α^l = *parameter corresponding to individual specific random effect for lane l*

The choice of the target lane implies whether or not the current lane of the driver is the most preferred lane and if not, which adjacent lane the driver needs to move to get to the target lane. The target lane utilities of a driver may be affected by the following:

- Lane attributes

- Neighboring vehicle attributes

- Path-plan

General lane attributes, such as the density and speed of traffic in the lane, traffic composition (e.g. percentage of heavy vehicles) etc. can affect the target lane utilities. Apart from these, particular lanes may have special lane-specific attributes that enter the utility function of that particular lane. For example, the exclusive lane-specific variables are included in the utility if the lane in consideration is an exclusive lane. If the driver is eligible to enter the lane, the exclusive lane is likely to have a very high utility for that

driver. On the other hand, if the driver is not eligible to move to a particular lane, a very high disutility is likely to be associated with that particular lane for that specific driver. Thus, for a single occupancy vehicle, the HOV lane is likely to have a high disutility capturing the penalty associated with moving to that lane violating the law. Similarly, for high occupancy tolled (HOT) lanes, the associated value of tolls can enter the utility of the exclusive lane for drivers of single occupancy vehicles.

The variables associated with the surrounding vehicles, such as speed, spacing and type of the neighboring vehicles may affect the driver's target lane choice. For example, if the front vehicle in the current lane has a very low speed compared to the driver's desired speed, the current lane is likely to be less preferred by the driver, even if the average speed in that lane is higher than that of the other lanes. It may be noted that the value of these neighboring variables is denoted by the current position of the vehicle.

The driver usually has a pre-defined destination and schedule (e.g. desired arrival time) for the trip and chooses a path accordingly. These path-plan variables have an important effect on target lane choice. Variables in this group may include distance to a point where the driver needs to be in a specific lane and the number of lane changes required from the target lane to the correct lanes. For example, if the driver is very close to the exit that he/she needs to take to follow the path, he/she is less likely to choose a lane further away from the rightmost lane as the target lane.

Drivers have different intrinsic preferences, aggressiveness and level of inertia for example. All else being equal, driver heterogeneity can lead to different target lane choices by different drivers.

Thus the systematic utility of a lane can have up to five components at any instant:

- Utility component comprising the generic characteristics of the lane;
- Utility component comprising the exclusive/special characteristics of the lane;
- Utility derived from the relative position of the lane with respect to the current lane;
- Utility component derived from the path-plan of the driver;
- Utility component derived from the individual-specific characteristics of the driver which can have different specifications: linear or non-linear (e.g. interaction with other variables in the utility).

Assuming a linear specification of the individual-specific characteristics, the total systematic utility of lane l for individual n at time t can be expressed as follows:

$$V_{lnt} = V_{lnt}^g + V_{lnt}^s + V_{lnt}^c + V_{lnt}^p + \alpha^l \upsilon_n \quad \forall l \in L_n \tag{4.3}$$

Where,

V_{lnt}^g = *general systematic utility component of the lane l*

V_{lnt}^s = *exclusive/special lane-specific utility component*

V_{lnt}^c = *utility component of lane l that depends on the current lane of the vehicle*

V_{lnt}^p = *utility component of lane l from the path plan p of the vehicle*

It may be noted that V_{lnt}^s is equal to zero if lane l is not an exclusive lane, has a positive value if driver n is eligible to use the exclusive lane and a negative value if there is a cost associated with using that lane (amount of toll, penalty associated with moving to that lane violating the law etc.).

Different choice models are obtained depending on the assumption made about the distribution of the random term ε_{lnt}. Assuming that these random terms are independently and identically extreme value distributed, choice probabilities for target lane l, conditional on the individual-specific error term (υ_n) are given by a logit model:

$$P_n(l_t \mid \upsilon_n) = \frac{\exp(V_{lnt} \mid \upsilon_n)}{\sum_{l' \in L_n} \exp(V_{l'nt} \mid \upsilon_n)} \quad \forall l, l' \in L_n \tag{4.4}$$

The choice of the target lane dictates the direction of lane change, if one is required. If the current lane is chosen as the target lane, no change is needed. Otherwise, the change will be in the direction from the current lane to the target lane. For example, in Figure 4.2a, the current lane is Lane 3. If the target lane is Lane 3, no change is needed. If the target lane is Lane 4, a lane change to the right is needed. If the target lane is Lane 1 or Lane 2, a lane change to the left is needed.

4.2.2 Choice of Action: The Gap Acceptance Model

The direction of immediate lane changing is determined as a consequence of the chosen target lane indicated by the target lane selection model. Next, the driver evaluates the gaps in the corresponding adjacent lane to decide whether or not the desired lane

change can be undertaken. Conditional on the target lane choice, the gap acceptance model indicates whether a lane change is possible or not using the existing gaps.

The adjacent gap in the target lane is defined by the lead and lag vehicles in that lane as shown in Figure 4.3. The lead gap is the clear spacing between the rear of the lead vehicle and the front of the subject vehicle. Similarly, the lag gap is the clear spacing between the rear of the subject vehicle and the front of the lag vehicle. It may be noted that one or both of these gaps may be negative if the vehicles overlap.
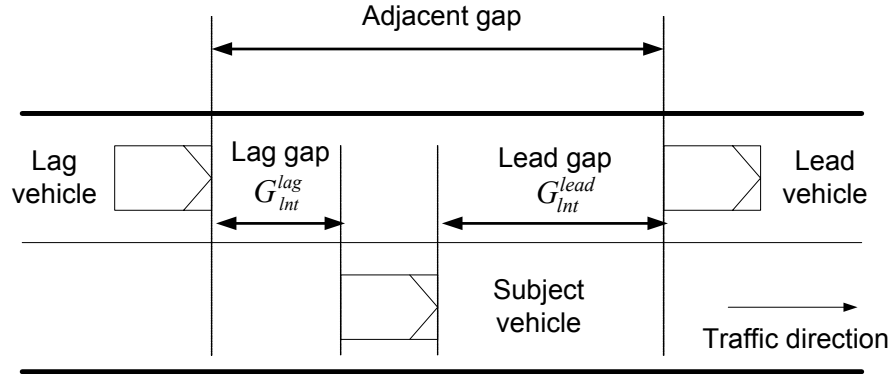
Figure 4.3: Definitions of the lead and lag vehicles and the gaps they define

The structure of the gap-acceptance model is based on the one proposed, estimated and validated by Ahmed (1999) and later by Toledo (2002).The model assumes that if the adjacent gap in the target lane is acceptable the driver performs the lane change and does not consider any other gaps. This assumption is consistent with satisficing behavior theory (Simon 1955), which states that human behavior is not optimizing, but is satisficing: if an available option (i.e. using the adjacent gap to change to the target lane) is satisfactory the driver does not try to find a better one. The driver therefore compares the available lead and lag gaps to the corresponding critical gaps, which are the minimum acceptable space gaps. An available gap is acceptable if it is greater than the critical gap. Critical gaps are modeled as random variables. Their means are functions of explanatory variables (Mahmassani and Sheffi 1981). The individual-specific error term captures correlations between the critical gaps of the same driver over time. Critical gaps are assumed to follow lognormal distributions to ensure that they are always non-negative and have been expressed as follows (Ahmed 1999, Toledo 2002):

$$G_{lnt}^{g\,cr} = \exp(\beta^T X_{lnt}^g + \alpha^g \upsilon_n + \varepsilon_{lnt}^g) \qquad g \in \{lead, lag\} \qquad (4.5)$$

Where,

$G_{lnt}^{g\,cr}$ = critical gap g in the direction of target lane l, measured in distance
units (e.g. meters)

X_{lnt}^{g} = explanatory variables that affect the critical gap g in the direction
of target lane l

β = coefficients of explanatory variables

α^{g} = coefficients of individual-specific latent variable υ_{n} for gap acceptance

ε_{lnt}^{g} = random term: $\varepsilon_{lnt}^{g} \sim N\left(0, \sigma_{g}^{2}\right)$

The gap acceptance model assumes that the driver must accept both the lead gap
and the lag gap to change lanes. The probability of changing lanes at time t, (lane action
$j_{t}=1$), conditional on the individual-specific term υ_{n} and the choice of target lane l_{t}, is
therefore given by:

$$
\begin{aligned}
P_{n}(j_{t}=1\,|\,l_{t},\upsilon_{n}) &= P_{n}(accept\ lead\,|\,l_{t},\upsilon_{n})P_{n}(accept\ lag\,|\,l_{t},\upsilon_{n}) \\
&= P_{n}\left[G_{lnt}^{lead} \geq G_{lnt}^{cr\ lead}(\upsilon_{n})\right]P_{n}\left[G_{lnt}^{lag} \geq G_{lnt}^{cr\ lag}(\upsilon_{n})\right]
\end{aligned} \tag{4.6}
$$

Based on the assumption that critical gaps follow lognormal distributions (ε_{glnt} is normally
distributed), the conditional probabilities that gap $g \in \{lead, lag\}$ is acceptable is given
by:

$$
\begin{aligned}
&P_{n}\left[G_{lnt}^{g} > G_{lnt}^{g\,cr}(\upsilon_{n})\right] \\
&= P_{n}\left[\ln\left(G_{lnt}^{g}\right) > \ln\left(G_{lnt}^{g\,cr}(\upsilon_{n})\right)\right] \quad = \Phi\left[\frac{\ln\left(G_{lnt}^{g}\right) - \left(\beta^{T}X_{lnt}^{g} + \alpha^{g}\upsilon_{n}\right)}{\sigma_{g}}\right] \\
&\varepsilon_{lnt}^{g} \sim N(0, \sigma_{g}^{2})
\end{aligned} \tag{4.7}
$$

$\Phi[\cdot]$ denotes the cumulative standard normal distribution.

Probability of no lane changes at time t ($j_{t}=0$), conditional on the individual-specific
term υ_{n} and the choice of target lane l_{t}, is therefore given by the following equation:

$$
\begin{aligned}
P_{n}(j_{t}=0\,|\,l_{t},\upsilon_{n}) &= 1 - P_{n}(j_{t}=1\,|\,l_{t},\upsilon_{n}) \\
&= \left[1 - \Phi\left(\frac{\ln\left(G_{lnt}^{g}\right) - \left(\beta^{T}X_{lnt}^{g} + \alpha^{g}\upsilon_{n}\right)}{\sigma_{g}}\right)\right]
\end{aligned} \tag{4.8}
$$

Gap acceptance is affected by the interaction between the subject vehicle and the lead
and lag vehicles in the adjacent lane. This may be captured by variables such as the

relative speed of the subject vehicle with respect to the lead and lag vehicles, type of lag vehicle etc. In case of mandatory lane changes acceptable gaps can also be a function of the distance to the mandatory lane changing point and/or the associated delay. For example if the driver needs to take an exit to follow the path, acceptable gaps can reduce as the driver approaches the exit or has become impatient after waiting for a suitable gap for a considerable time.

4.3 Model Estimation

4.3.1 Data

Study Area

The dataset used in this study was collected in 1983 by FHWA in a four-lane section of Interstate 395 (I-395) Southbound in Arlington, Virginia (Figure 4.4).

Figure 4.4: The I-395 data collection site

It is 997 meters in length, one of the longest sites for which trajectory data is available, and includes an on-ramp and two off-ramps. The section is shown schematically in Figure 4.5. An hour of data at a rate of 1 frame per second was collected through aerial photography of the section. A detailed technical description of the systems and technologies used for data collection and reduction is found in FHWA (1985). The dataset, smoothed by Toledo (2002) using the local regression procedure developed by

Cleveland (1979) and Cleveland and Devlin (1988), contains observations of the position, lane and dimensions of every vehicle within the section every 1 second.

This dataset is particularly useful for estimation of the proposed lane changing model since the geometric characteristics of the site, with two off-ramps and an on-ramp, initiate a lot of weaving and lane changing. Though there are no exclusive lanes, the drivers are free to select the lane with the highest utility as the target lane and make subsequent lane changes depending on availability of gaps along the stretch of collection site. The ramps within the site provide path-plan information for the various drivers. However, the path-plan beyond the section is not observable. Characteristics of the drivers such as aggressiveness and level of driving skill are also unobserved.
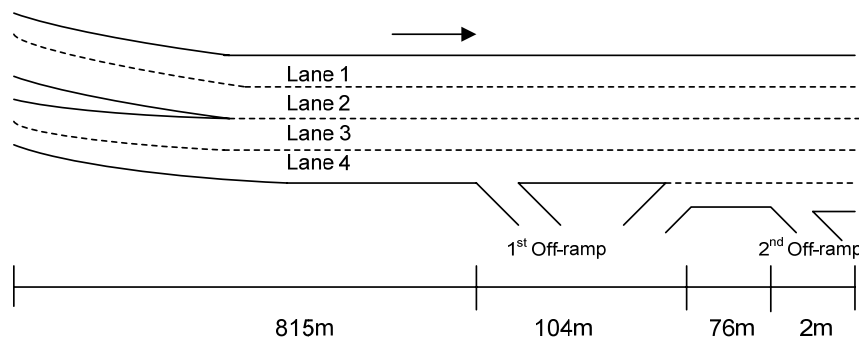
Figure 4.5: Schematic diagram of the I-395 data collection site
(not to scale)

Characteristics of Estimation Dataset

The vehicle trajectory data of the various vehicles in the section and the speeds and accelerations derived from these trajectories are used to generate the required variables. The resulting estimation dataset includes 442 vehicles for a total of 15632 observations at a 1 second time resolution. On average a vehicle was observed for 35.4 seconds (observations). All the vehicles are first observed at the upstream end of the freeway section. At the downstream end, the majority of traffic (76%) remains in the freeway. The 8% and 16% of vehicles, which exit the section using the first and second off-ramps (Figure 4.5) respectively, are useful to capture the effect of the path-plan on driving behavior.

Lane-specific variables including lane density, lane speed, and percentage of heavy vehicle have been calculated from the raw dataset. The lane-specific variables across the different lanes are summarized in Table 4.1.

Table 4.1: Lane-specific variables

Variable	Lane 4	Lane 3	Lane 2	Lane 1	Segment
Average Density d/s, veh/km/lane	28.41	28.29	28.64	26.56	29.22
Average Density u/s, veh/km/lane	29.86	30.06	30.52	28.29	
Average Speed , m/sec.	14.22	15.79	16.23	17.50	15.75

The same dataset was used by Toledo (2002) in estimating the integrated lane driving behavior model. The detailed characteristics of the dataset documented by Toledo are summarized below:

Speeds in the section range from 0.4 to 25.0 m/sec. with a mean of 15.6 m/sec. Densities range from 14.2 to 55.0 veh/km/lane with a mean of 31.4 veh/km/lane. The level of service in the section is D-E (HCM 2000). The vehicles the subject interacts with and the variables related to these vehicles are shown in Figure 4.6.
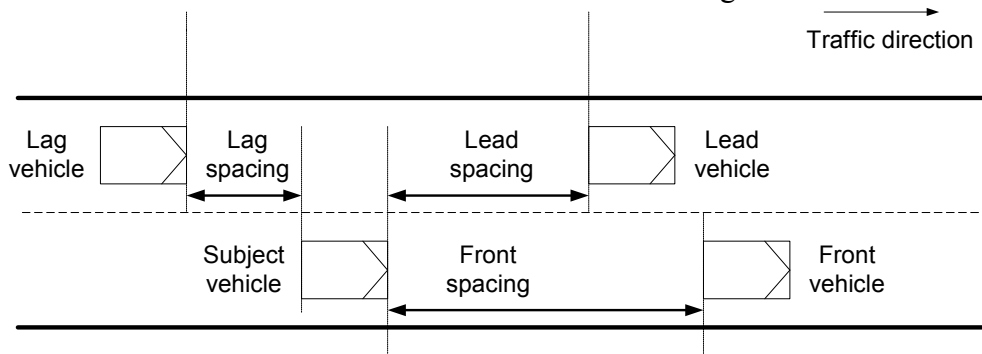
Figure 4.6: The subject, front, lead and lag vehicles and related variables

Relative speeds with respect to various vehicles are defined as the speed of these vehicles less the speed of the subject. Tables 4.2 and 4.3 summarize statistics of the variables related to the subject vehicle and the vehicle in front.

Table 4.2: Statistics of variables related to the subject vehicle

Variable	Mean	Std Dev	Median	Minimum	Maximum
Speed (m/sec)	15.6	3.1	15.8	0.4	25.0
Acceleration (m/sec2)	0.05	1.21	0.05	-3.97	3.99
Positive	0.96	0.76	0.78	0	3.99
Negative	-0.93	0.75	-0.74	-3.97	0
Density (veh/km/lane)	31.4	6.5	30.8	14.2	55.0

Table 4.3: Statistics of relations between the subject and the front vehicle

Variable	Mean	Std Dev	Median	Minimum	Maximum
Relative speed (m/sec)	0.2	1.7	0.2	-8.6	9.7
Spacing (m)	26.6	21.2	20.4	1.4	250.5
Time headway (sec)	2.0	1.4	1.7	0.3	27.3

The distributions of speed, acceleration, density and time headway are shown in Figure 4.7.
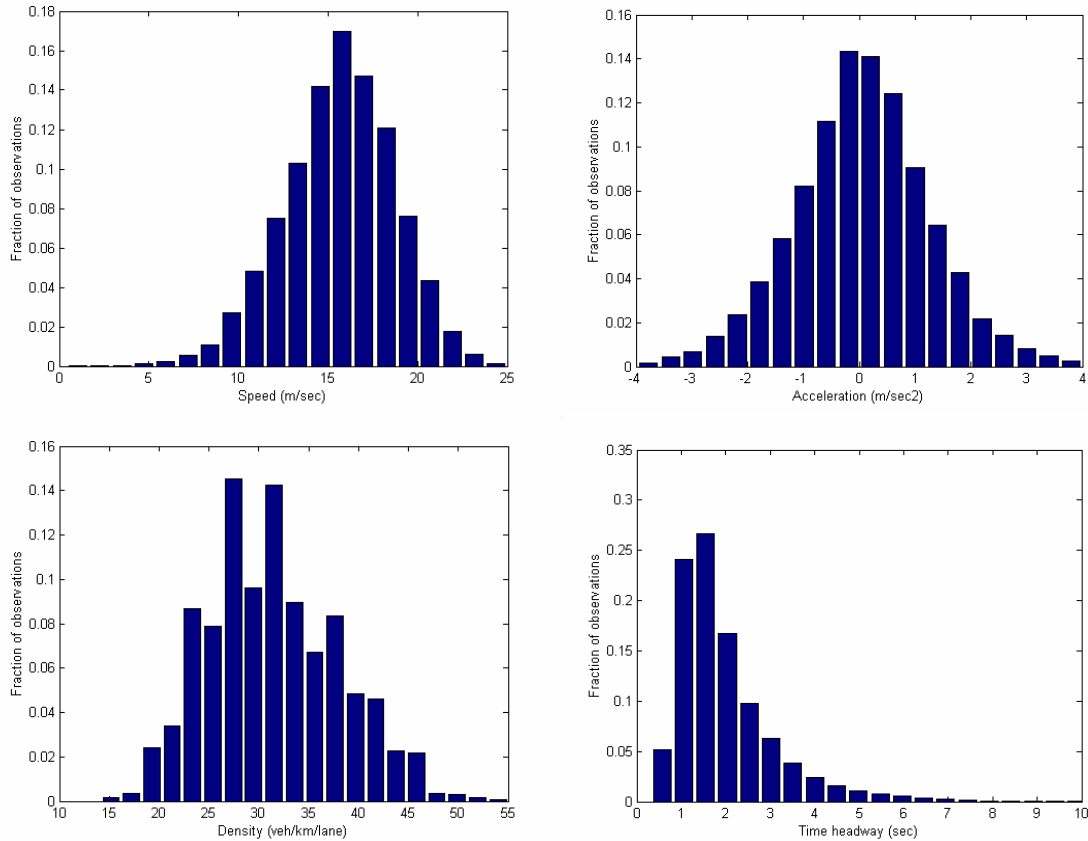
Figure 4.7: Distributions of speed, acceleration, density and time headway

Lane selection and gap acceptance behaviors are captured by observing lane changes performed by the drivers. An important factor in these behaviors is drivers' desire to follow their path. In this dataset drivers have three possible destinations, each with a corresponding path-following behavior:

- Exiting the section at the first off-ramp.
- Exiting the section at the second off-ramp.
- Staying in the freeway at the downstream end of the section.

The distribution of observed lane changes by direction (right, left) and by destination is described in Table 4.4. It is worth noting that many of the vehicles that exit the section through the off-ramps are observed in the right-most lane at the upstream end of the section. This indicates that they may have started considering the path-plan constraint earlier. As a result the coefficients of explanatory variables related to the path-plan may be biased towards aggressive behaviors since the more timid drivers are discounted in the dataset.

Table 4.4: Distribution of lane changes by direction and destination

Destination	Right	Left
Total	123	74
Freeway	71	71
1st ramp	12	0
2nd ramp	40	3

The relations between the subject and the lead and lag vehicles in the right and left adjacent lanes affecting the gap acceptance and gap choice behaviors of the driver are presented in Table 4.5. This table summarizes statistics of the accepted lead and lag gaps (i.e. the gaps vehicles changed lanes into) both for the accepted gaps and for the entire dataset (both accepted and rejected gaps). Statistics for the entire dataset are presented in parentheses.

Table 4.5: Statistics describing the lead and lag vehicles

Variable	Mean	Std Dev	Median	Minimum	Maximum
Relations with lead vehicle					
Relative Speed (m/sec)	0.2	2.6	0.5	-17.3	8.1
	(0.0)	(2.9)	(0.1)	(-17.5)	(15.5)
Lead spacing (m)	22.2	21.9	14.1	0.04	117.9
	(19.6)	(39.9)	(13.0)	(-18.1)	(268.9)
Relations with lag vehicle					
Relative Speed (m/sec)	-0.4	2.2	-0.3	-6.7	5.2
	(0.0)	(2.7)	(0.0)	(-15.0)	(14.1)
Lag spacing (m)	23.1	20.6	16.6	1.7	110.1
	(18.6)	(23.0)	(12.0)	(-18.1)	(232.6)

Accepted lead gaps vary from 0.04 to 117.9 meters, with a mean of 22.2 meters. Accepted lag gaps vary from 1.7 to 110.1 meters, with a mean of 23.1 meters. No significant differences were found between the right and left lanes. Relative speeds are

defined as the speed of the lead (lag) vehicle less the speed of the subject. Statistics for the entire dataset are also shown in parenthesis. With these statistics, negative spacing values indicate that the subject and the lead vehicle partly overlap (this is possible because they are in different lanes). As expected, the mean accepted gaps are larger than the mean gaps in the traffic stream. Similarly, lead relative speeds in accepted gaps are larger than the mean of the dataset and lag relative speeds are smaller in the entire dataset (i.e. on average, in accepted gaps the subject vehicle is slower relative to the lead vehicle and faster relative to the lag vehicle compared to the entire dataset).
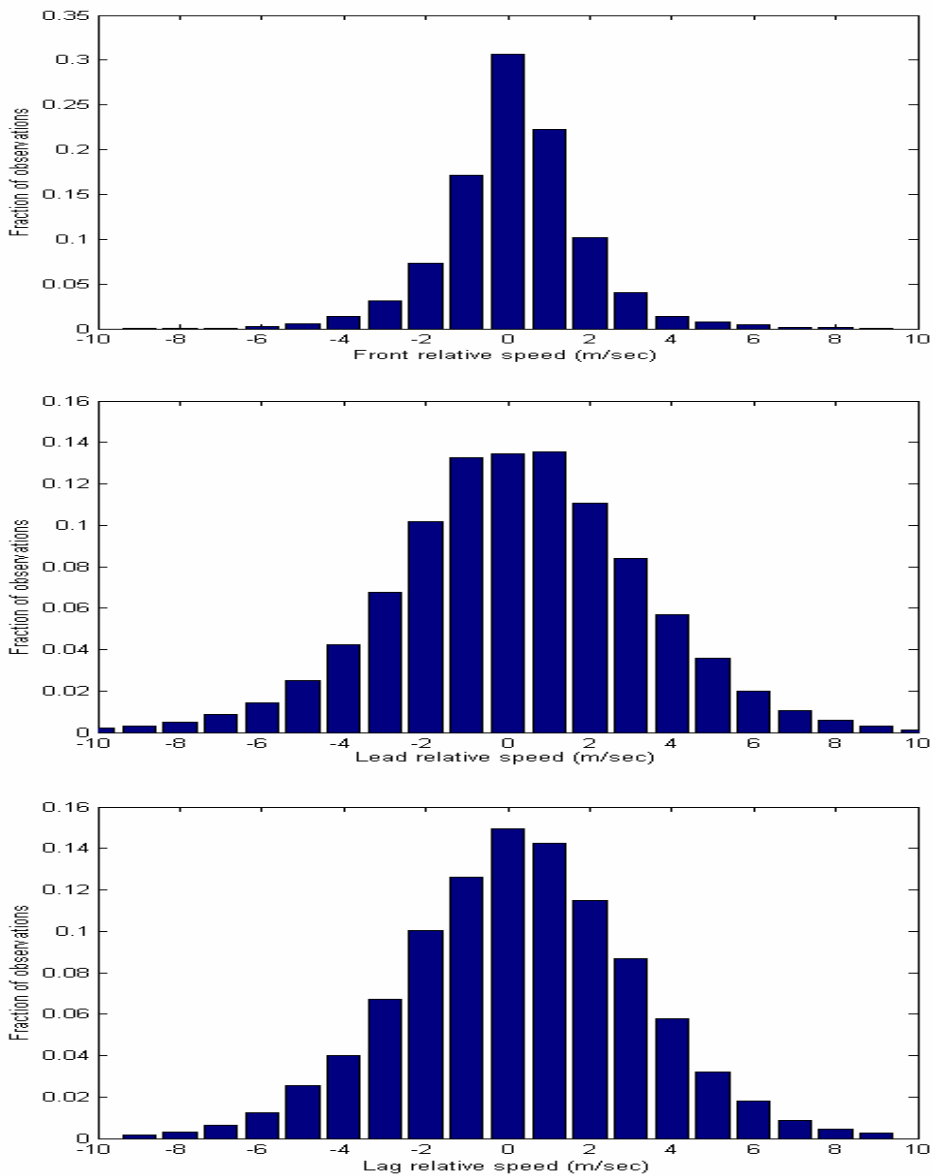
Figure 4.8: Distributions of relative speed with respect to front, lead and lag vehicles

The distributions of relative speeds and spacing, with respect to the front, lead and lag vehicles are shown in Figures 4.8 and 4.9 respectively.
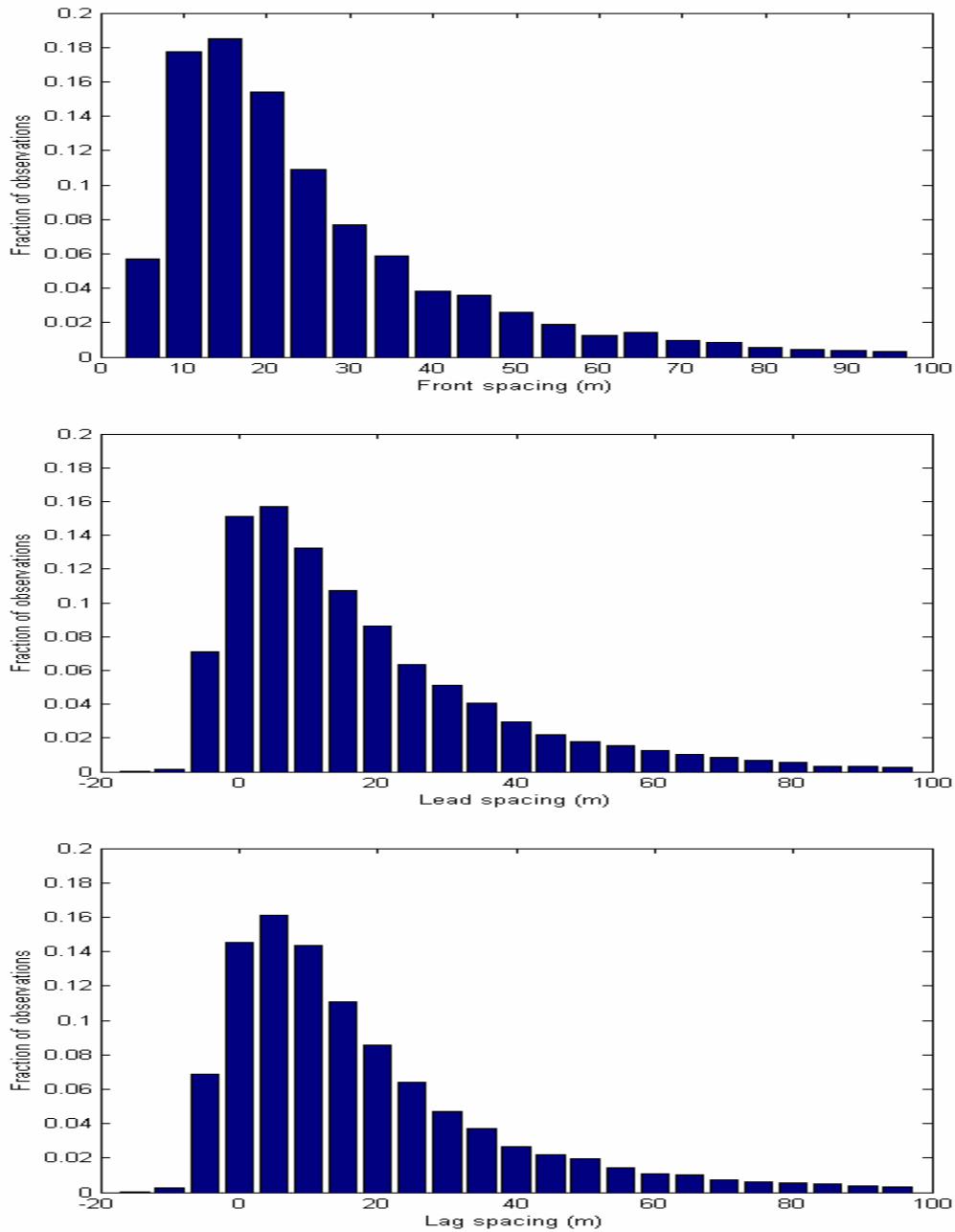
Figure 4.9: Distributions of spacing with respect to the front, lead and lag vehicles

4.3.2 Likelihood

In this section, the likelihood function used to model the trajectory of the driver is presented. Important explanatory variables affecting the target lane choice are those

related to the path-plan. For vehicles exiting the freeway within the data collection section, the remaining distance to the exit (d_{nt}^{exit}) is observed. However, for vehicles exiting the freeway downstream of the observed section, this information is not likely to be observed for some of the vehicles. In order to capture the effect of these variables, a distribution of the distances from the downstream end of the road section being studied to the following exit points (s_n) is estimated. The alternatives considered are the first, second and subsequent exits. For a driver taking the 1st downstream exit, the definition of the remaining distance to the exit is illustrated in Figure 4.10.
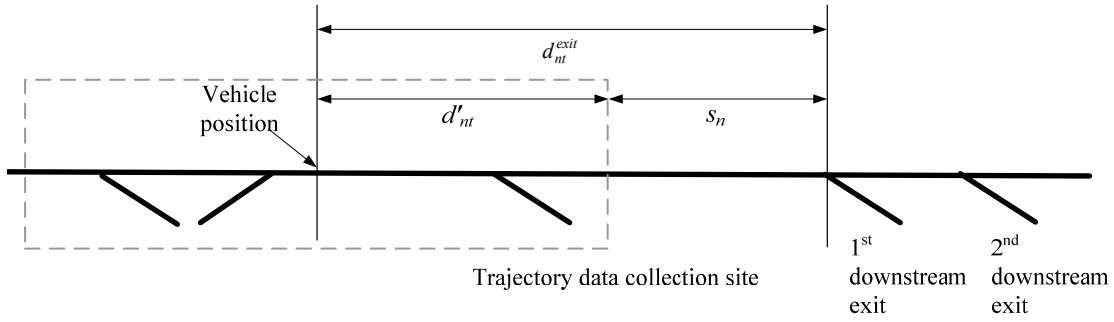
Figure 4.10: Definition of path-plan variables

The probability mass function of the distance beyond the downstream end of the section to the off-ramps used by drivers is given by the following expression:

$$p(s_n) = \begin{cases} \pi_1 & \text{for } s_n = s^1 \\ \pi_2 & \text{for } s_n = s^2 \\ 1 - \pi_1 - \pi_2 & \text{for } s_n = s^3 \end{cases} \tag{4.9}$$

Where,

s_n = remaining distance to the exit point of driver n

s^1, s^2, s^3 = distance beyond the downstream end of the section to the first, second and subsequent exits, respectively

π_1, π_2 = parameters to be estimated

The first and second exit distances (s^1 and s^2) were extracted from maps and an infinite distance was used for the subsequent exits ($s^3 = \infty$). This corresponds to an

assumption that on the section being studied, drivers that use these subsequent exits have path-plans that are not constraining.

The joint probability of a combination of target lane (l) and lane action (j) observed for driver n at time t, conditional on the distance to the exit point (s_n) and the individual-specific characteristic (υ_n) is given by:

$$P_n(l_t, j_t \mid s_n, \upsilon_n) = P_n(l_t \mid s_n, \upsilon_n) P_n(j_t \mid l_t, \upsilon_n)$$
(4.10)

Where, $P_n(l_t \mid .)$ and $P_n(j_t \mid .)$ are given by Equations 4.4, 4.6 and 4.8 respectively.

Only the lane changing actions are observed. The marginal probability of the lane-changing action is therefore given by:

$$P_n\left(j_t \mid s_n, \upsilon_n\right) = \sum_{l \in L_n} P_n\left(l_t, j_t \mid s_n, \upsilon_n\right)$$
(4.11)

The behavior of driver n is observed over a sequence of T_n consecutive time intervals. Assuming that, conditional on s_n and υ_n, these observations are independent, the joint probability of the sequence of observations is given by:

$$P_n\left(j_1, j_2, ..., j_T \mid s, \upsilon\right) = \prod_{t=1}^{T_n} P_n\left(j_t \mid s_n, \upsilon_n\right)$$
(4.12)

The unconditional individual likelihood function ($\mathcal{L}_n$) is obtained by integrating (summing for the discrete variable s_n) over the distributions of the individual-specific variables:

$$\mathcal{L}_n = P_n\left(j_1, j_2, ..., j_T\right) = \int_\upsilon \sum_s P_n\left(j_1, j_2, ..., j_T \mid s, \upsilon\right) p(s) f(\upsilon) d\upsilon$$
(4.13)

Assuming that the observations from different drivers are independent, the log-likelihood function for all N individuals observed is given by:

$$\mathcal{L} = \sum_{n=1}^{N} \ln(\mathcal{L}_n)$$
(4.14)

The maximum likelihood estimates of the model parameters are found by maximizing this function.

In this study, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm implemented in the statistical estimation software GAUSS (Aptech Systems 2003) has

been used. BFGS is a quasi-Newton method, which maintains and updates an approximation of the Hessian matrix based on first-order derivative information (see, for example, Bertsekas 1999). GAUSS implements a variant of BFGS due to Gill and Murray (1972), which updates the Cholesky decomposition of the Hessian (Aptech Systems 1995). The integrals in the likelihood function were calculated numerically using the Gauss-Legendre quadrature method (Aptech Systems 2003). The likelihood function is not globally concave. For example, if the signs of all the coefficients of the individual-specific error term are reversed, the solution is unchanged due to its symmetric distribution function. To avoid obtaining a local solution, different starting points have been used in the optimization procedure. It may be noted that the estimation approach does not involve the use of any traffic simulator, and so the estimated models are simulator independent.

4.3.3 Estimation Results

All components of the model were estimated jointly using a maximum likelihood estimation procedure as described in the previous section. However, in order to simplify the presentation, estimation results for the target lane choice and gap acceptance levels are presented and discussed separately.

The summary of estimation results of the proposed lane changing model is presented in Table 4.6.

Table 4.6: Estimation results of the target lane changing model

Final log-likelihood	-875.81
Initial log-likelihood	-1434.76
Number of drivers	442
Number of observations	15632
Number of parameters	31
Adjusted rho-bar square	0.37

To demonstrate the need to include the latent plans in the freeway lane selection model by means of target lanes, the estimation results were compared against a reduced form model with restricted latent targets (Toledo *et al.* 2003). In the reduced form model (referred as the lane shift model in the subsequent discussion), only the adjacent lanes are

considered for the lane shift. The model framework is illustrated in (Figure 4.11) and detailed in Appendix C.1.
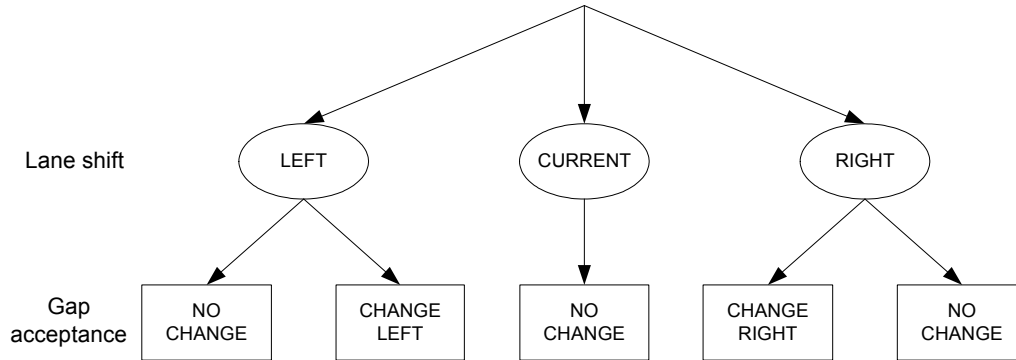
Figure 4.11: Structure of the lane-shift model (Toledo et al. 2003)

The myopic lane shift model cannot be viewed as nested within the model with explicit target lane choice, and therefore classic statistical tests cannot be applied to select between the two. For comparing the goodness-of-fit of non-nested models, the Adjusted Rho-bar square ($\bar{\rho}^2$) and the Akaike Information Criteria (AIC) have been used.

Adjusted Rho-bar square ($\bar{\rho}^2$) measures the fraction of an initial log-likelihood value explained by the model taking into account the model complexity. The measure is defined as follows:

$$\bar{\rho}^2 = 1 - \frac{L(\beta^*) - k}{L(0)} \tag{4.15}$$

Where, $L(\beta^*)$ is the maximum log-likelihood value, $L(0)$ is the maximum log-likelihood value, k is the number of estimated parameters.

Akaike (1973, 1974) developed the Akaike information criterion (AIC) as a tool for selecting between competing model specifications. The AIC penalizes the maximum likelihood value of each model to account for model complexity:

$$AIC = L(\beta^*) - k \tag{4.16}$$

In model selection, $\bar{\rho}^2$ and AIC are computed for all candidate models and the model with the larger AIC is selected (see Ben-Akiva and Lerman 1985 and Gourieroux and Monfort 1995 for details).

The test statistics are presented in Table 4.7.

Table 4.7: Model comparison

Statistic	Lane Shift (R)	Target Lane (U)
Likelihood value	-888.78	-875.81
Number of parameters (k)	26	31
Akaike information criteria (AIC)	-914.78	-906.81
Adjusted rho-bar square ($\bar{\rho}^2$)	0.362	0.368

For both statistics, the model with explicit target lane choice has larger values, which indicates that it has a better goodness-of-fit even after discounting for the increased number of parameters. The detailed estimation results are presented in the following sections.

Choice of Plan: The Target Lane Model

The driver selects the lane that he/she perceives to be the best as the target lane. A linear utility function is associated with each lane. The choice set of the driver includes all available lanes in the freeway stretch. The utility of lane target lane l of individual n at time t can be expressed as follows:

$$U_{lnt} = \beta^T X_{lnt} + \alpha^l \upsilon_n + \varepsilon_{lnt} \tag{4.17}$$

Where,

 X_{lnt} = *explanatory variables that affect the utility of lane l*
 β = *corresponding vector of parameters*
 υ_n = *individual-specific random effect (e.g. aggressiveness): $\upsilon_n \sim N(0,1)$*
 α^l = *parameter corresponding to individual specific random effect for lane l*

As discussed in Section 4.3.2, the target lane choices are affected by the attributes of the alternative lanes, the variables related to the path-plan and the neighboring vehicles as well as driver-specific characteristics. However, not all of the candidate variables mentioned in Section 4.3.2 were found to be statistically significant and/or have intuitive signs. For example, the percentage of heavy vehicles in the lane and type of the

neighboring vehicles were not found to be significant. In some cases, interactions of multiple variables have been used to better capture a particular effect. These interaction variables have been included only if there was an improvement in the goodness-of-fit. For example, in case of path-plan effect, interaction of the remaining longitudinal distance and lateral distance were found to yield an improvement in the likelihood and led to the proposed functional form. The estimation results are presented in Table 4.8.

Table 4.8: Estimation results of the target lane selection model

	Variable	Parameter	t-stat
Lane Attributes	Lane 2 constant	0.0590	1.16
	Lane 3 constant	-0.571	-1.68
	Lane 4 constant (right most lane)	-1.69	-3.03
	Lane density, vehicle/km	-0.0131	-1.21
	Average speed in lane, m/sec	0.176	1.59
Neighborhood Variables	Front vehicle spacing, m	0.0240	3.86
	Relative front vehicle speed, m/sec	0.115	1.46
	Tailgate dummy	-4.94	-1.96
Inertia Variables	Current lane (CL) dummy	2.69	1.55
	1 lane change from the CL	-0.845	-1.15
	Each additional lane change from the CL	-3.34	-1.91
Path-plan	Path-plan impact, 1 lane change required	-2.55	-4.57
	Path-plan impact, 2 lane changes required	-4.95	-2.19
	Path-plan impact, 3 lane changes required	-6.96	-1.65
	Next exit dummy, lane change(s) required	-0.872	-1.35
	Exponent of remaining distance, θ^{MLC}	-0.417	-2.48
	Probability of taking 1st exit, π_1	0.00102	0.68
	Probability of taking 2nd exit, π_2	0.0860	1.38
Heterogeneity	Coefficient of aggressiveness: Lane 1, α^{lane1}	-1.41	-2.29
	Coefficient of aggressiveness: Lane 2, α^{lane2}	-1.07	-0.50
	Coefficient of aggressiveness: Lane 3, α^{lane3}	-0.0710	-3.61
	Coefficient of aggressiveness: Lane 4, α^{lane4}	-0.0891	-1.56

The estimated values of the lane-specific constants imply that, everything else being equal, the right-most lane is the least desirable. This may be the result of drivers' preference to avoid the merging and weaving activities that take place in that lane. In general lanes that are to the left are more desirable. However, lanes 3 and 4 have non-negative constants, which may indicate that the advantage of being away from the slower right lanes is balanced by the disadvantage associated with being in lanes that are further away from the off-ramp, and by the increased interaction with vehicles traveling at higher

speeds. The results indicate that drivers are more likely to choose lanes with higher average speeds and lower densities, which is intuitive.

Some of the lane-specific variables are dependent on the current lane of the driver. For example, the required maneuver to reach a specific lane is a function of the distance of the lane from the current lane of the driver. The values of the coefficients of the number of lane changes required from the current lane to the target lane denote the disutility associated with choosing target lanes that require lane changing maneuvers. This has been modeled as a step function and the results indicate that the disutility associated with each additional lane change is much higher when more than one lane changing maneuver is associated. The positive coefficient of the current lane dummy captures the inertia preference to stay in the current lane. As expected, the sign of this coefficient is positive. As apparent in Figure 4.12, the lane-specific part of the utility thus changes depending upon the current position of the vehicle, being the highest for the current lane and diminishing with the distance from the current lane.
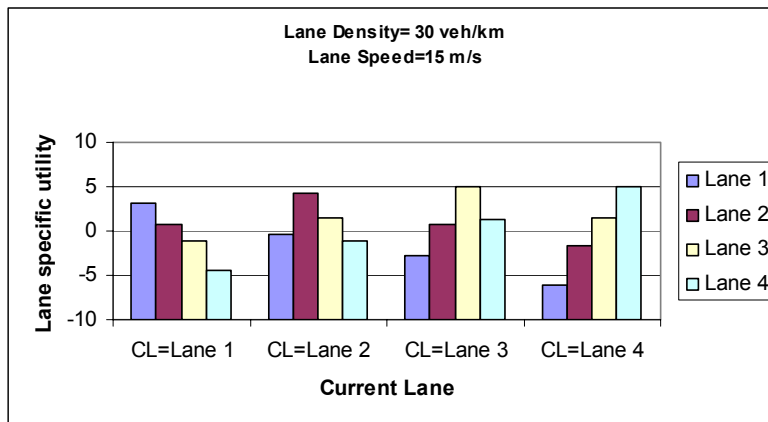
Figure 4.12: Variation of lane utilities depending on the current lane of the driver

The interactions between the subject vehicle and the vehicles in front of it in the current and adjacent lanes, also affect the target lane choice. Results show that lane utilities increase with the relative front speed and the spacing between the vehicles. The tailgating dummy variable captures drivers' tendency to move out of their current lane if they are being tailgated. Tailgating is not directly observable in the data but tailgating behavior is assumed if a vehicle is close behind the subject vehicle when traffic

conditions permit a longer headway (i.e. free-flow conditions apply). Mathematically, the tailgate dummy variable is defined by:

$$\delta_{nt}^{tailgate} = \begin{cases} 1 & gap\ behind \le 10m\ and\ level\ of\ service\ is\ A,\ B\ or\ C \\ 0 & otherwise \end{cases} \qquad (4.18)$$

Levels of service definitions are based on densities (HCM 2000). The estimated coefficient of the tailgate dummy is negative and its magnitude is large relative to the coefficients of other variables. It implies a strong preference to avoid these situations. This result is comparable with those of Ahmed (1999) and Toledo *et al.* (2002), who also found tailgating to be an important explanatory variable.

The path-plan impact variables indicate that the utility of a lane decreases with the number of lane changes the driver needs to perform in order to maintain the desired path. This effect is magnified as the distance to the off-ramp d_{nt}^{exit} decreases. This has been captured by the negative power of the distance to the off-ramp ($\theta^{MLC} = -0.417$) that guarantees that at the limits, the path-plan impact approaches 0 when $d_{nt}^{exit} \to +\infty$ and approaches $-\infty$ when $d_{nt}^{exit} \to +0$. The disutility associated with being in a wrong lane is larger when the driver needs to take the next exit. Figure 4.13 shows the impact of path-plan lane changes on the utility of a lane as a function of the distance from the off-ramp.
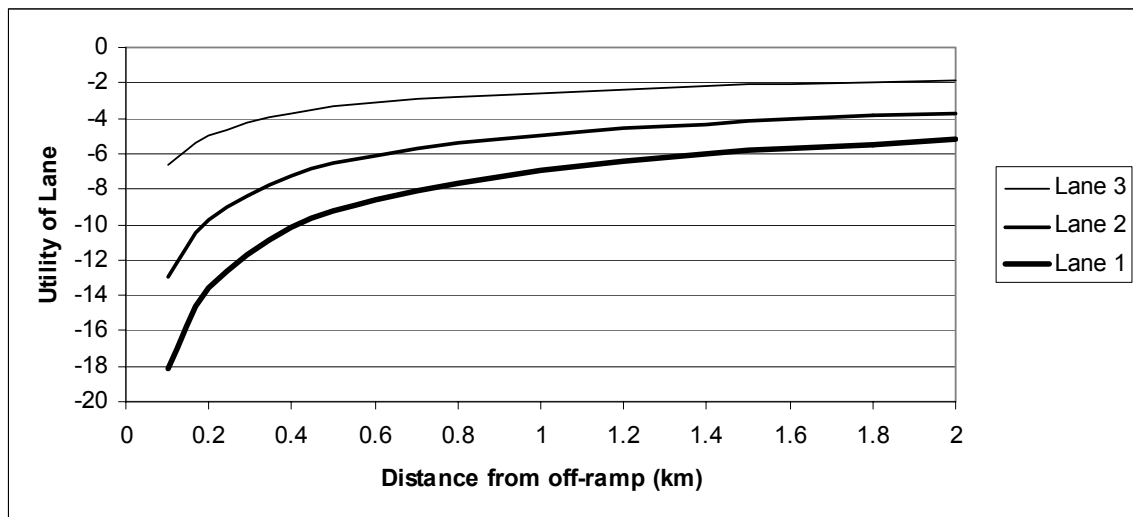
Figure 4.13: Impact of path-plan lane changes on the utility of a lane

The combined effect of path-plan and lane-specific attributes is shown in Figure 4.14.
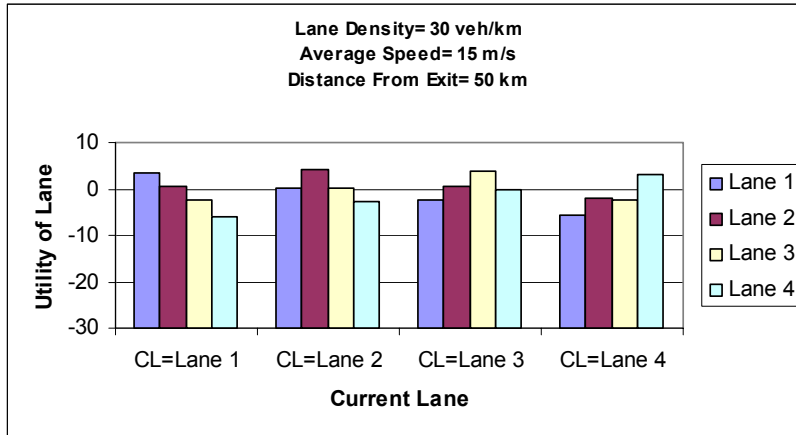
a. Distance from exit=50 km
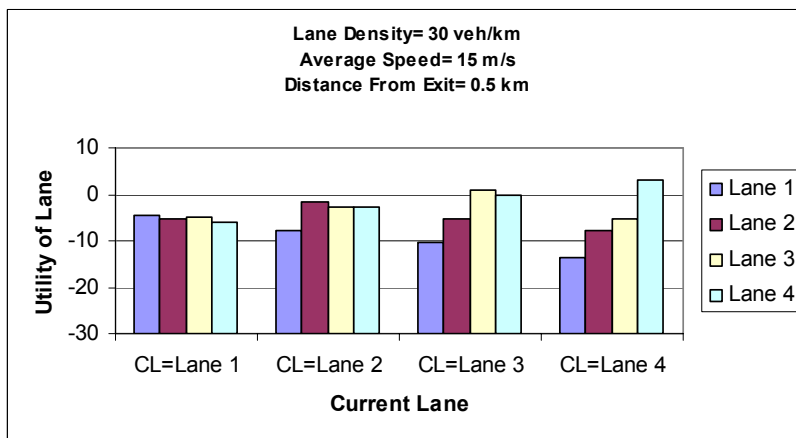
b. Distance from exit=0.5 km
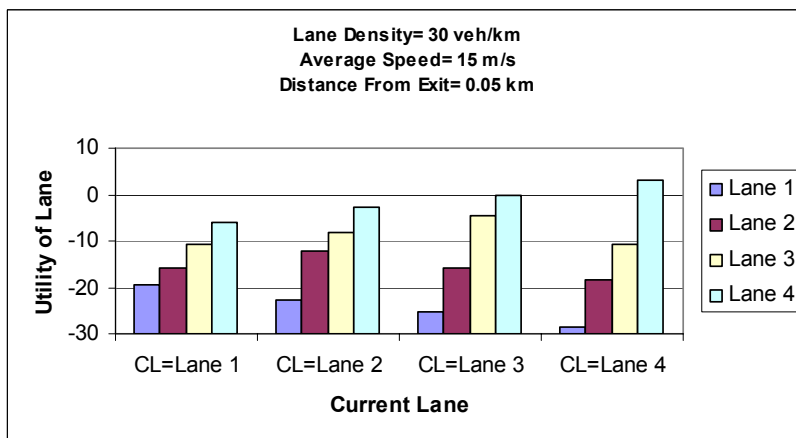
c. Distance from exit=0.05 km

Figure 4.14: Combined effects of path-plan and lane-specific attributes

In this example, the exits are on the right, Lane 4 is closest to the exit and Lane 1 is the farthest. It is interesting to note the tradeoff between path-plan and inertia of the driver. When the driver is very far from the desired exit, the lane utilities are affected primarily by the position of the lane with respect to the current lane of the driver (Figure 4.14a). As the distance to exit decreases, the disutility of being in a lane far from Lane 4 becomes more and more pronounced and the relative utility of the lanes in the direction of the exit (right in the example) gradually increase. When the driver is very close to the exit the path-plan effect clearly dominates (Figure 4.14c) and the lanes far from Lane 4 have a very high disutility.

The heterogeneity coefficients, α^{lane1}, α^{lane2}, α^{lane3} and α^{lane4} capture the effects of the individual-specific error term υ_n on the target lane choice. α^{lane1} and α^{lane2} are more negative compared to α^{lane3} and α^{lane4}. Hence, υ_n can be interpreted as correlated with aggressiveness implying aggressive drivers are less likely to choose the right lanes over the left ones compared to more timid drivers.

In summary, the target lane utility can be given by:

$$
\begin{aligned}
U_{lnt} = {} & \beta^l - 0.0131 D_{lnt} + 0.176 V_{lnt}^{avg} + 0.0240 \Delta X_{lnt}^{front} \delta_{lnt}^{CL} + 0.115 \Delta V_{lnt}^{front} \delta_{lnt}^{adj/CL} \\
& - 4.94 \delta_{nt}^{tailgate} \delta_{lnt}^{CL} + 2.69 \delta_{lnt}^{CL} - 0.845 \delta_{lnt}^{\Delta CL=1} - 3.34 (\Delta CL_{lnt} - 1) \delta_{lnt}^{\Delta CL>1} \\
& + \left[d_{nt}^{exit} \right]^{-0.417} (-2.55 \delta_{ln}^1 - 4.95 \delta_{ln}^2 - 6.96 \delta_{ln}^3) - 0.872 \delta_{nt}^{next\ exit} \Delta Exit_{ln} \\
& - \alpha^l \upsilon_n + \varepsilon_{lnt}
\end{aligned}
\tag{4.19}
$$

Where,

β^l $\quad$ =constant for lane l

D_{lnt} $\quad$ =density of lane l,vehicle/km

V_{lnt}^{avg} $\quad$ = average speed in lane l, m/s

ΔX_{lnt}^{front} =spacing of the front vehicle in lane l, m

ΔV_{lnt}^{front} = relative speed of the front vehicle in lane l, m/s

δ_{lnt}^{CL} $\quad$ = current lane dummy, 1 if lane l is the current lane, 0 otherwise

$\delta_{lnt}^{adj/CL}$ = current /adjacent lane dummy, 1 if lane l is the
$\quad$ current/adjacent lane, 0 otherwise

$\delta_{nt}^{tailgate}$ = tailgate dummy, 1 if vehicle n is being tailgated at time t, 0 otherwise

$\delta_{lnt}^{\Delta CL=1}$ = required change dummy, 1 if lane l involves one lane change from the current lane,0 otherwise

$\delta_{lnt}^{\Delta CL>1}$ = required change dummy, 1 if lane l involves more than one lane changes from the current lane,0 otherwise

ΔCL_{lnt} = number of lane changes required to get from the current lane to lane l

d_{nt}^{exit} = distance to the exit driver n intends to take

δ_{ln}^{lk} = indicator with value 1 if lane l is k (k=0,1,2,3) lanes away from the desired exit of individual n, 0 otherwise

$\delta_{nt}^{next\ exit}$ = indicator with value 1 if lane i is 1 lane away from the desired exit of driver n, 0 otherwise

$\Delta Exit_{ln}$ =number of lane changes required to get from lane l to the exit lane of driver n

α^l = heterogeneity coefficient of lane l

Choice of Action: The Gap Acceptance Model

The direction of the target lane indicates the direction of immediate lane change and the driver is assumed to evaluate the available adjacent gap in the target lane and decide whether or not to change lanes immediately. In order for the gap to be acceptable both the lead and lag gaps, must be acceptable. That is, the available lead and lag gaps must be larger than the corresponding critical gaps. As presented in Equation 4.5, in order to ensure that the critical gaps are always positive, they are assumed to follow lognormal distributions:

$$\ln(G_{lnt}^{lead\ cr}) = \beta^T X_{lnt}^{lead} + \alpha^{lead} \upsilon_n + \varepsilon_{lnt}^{lead}$$
$$\ln(G_{lnt}^{lag\ cr}) = \beta^T X_{lnt}^{lag} + \alpha^{lag} \upsilon_n + \varepsilon_{lnt}^{lag}$$

(4.20)

Where,

$G_{lnt}^{lead\ cr}, G_{lnt}^{lag\ cr}$ = lead and lag critical gap in the direction of target lane l, measured in distance units (e.g. meters)

$X_{lnt}^{lead}, X_{lnt}^{lag}$ = explanatory variables that affect the lead and lag critical gaps respectively in the direction of target lane l

$\alpha^{lead}, \alpha^{lag}$ = coefficients of individual-specific latent variable υ_n for lead and lag gap acceptance

$\varepsilon_{lnt}^{lead}, \varepsilon_{lnt}^{lag}$ = random terms: $\varepsilon_{lnt}^{lead} \sim N\left(0, \sigma_{lead}^2\right)$, $\varepsilon_{lnt}^{lag} \sim N\left(0, \sigma_{lag}^2\right)$

Similar to the target lane choice model, not all candidate variables were supported by the data. For example, the remaining distance to the desired exit did not have any significant effect on critical gaps. The estimation results of the gap acceptance model are presented in Table 4.9.

Table 4.9: Estimation results of the gap acceptance model

Lead Critical Gap		
Variable	Parameter	t-stat
Constant	1.54	5.59
Relative lead speed positive, $Max(\Delta V_{nt}^{lead},0)$,m/sec.	-6.21	-3.60
Relative lead speed negative, $Min(\Delta V_{nt}^{lead},0)$,m/sec.	-0.130	-2.09
Heterogeneity coefficient of lead gap, α^{lead}	-0.00801	-3.17
Standard deviation of lead gap, σ_{lead}	0.854	1.29
Lag Critical Gap		
Constant	1.43	5.35
Relative lag speed positive, $Max(\Delta V_{nt}^{lag},0)$, m/sec.	0.640	3.36
Heterogeneity coefficient of lag gap, α^{lag}	-0.205	-0.48
Standard deviation of lag gap, σ_{lag}	0.954	4.80

The lead critical gap decreases with the relative lead speed, i.e. it is larger when the subject vehicle is faster relative to the lead vehicle. The effect of the relative speed is strongest when the lead vehicle is faster than the subject. In this case, the lead critical gap quickly diminishes as a function of the speed difference. This result suggests that drivers perceive very little risk from the lead vehicle when it is getting away from them.

In the gap acceptance model, the lag critical gap increases with the relative lag speed: the faster the lag vehicle is relative to the subject, the larger the lag critical gap. In contrast to the lead critical gap, the lag gap does not diminish when the subject is faster. A possible explanation is that drivers may maintain a minimum critical lag gap as a safety buffer since their perception of the lag gap is not as reliable as it is for the lead gap due to the use of mirrors.

Median lead and lag critical gap variations, as a function of the relative speeds are presented in Figure 4.15.
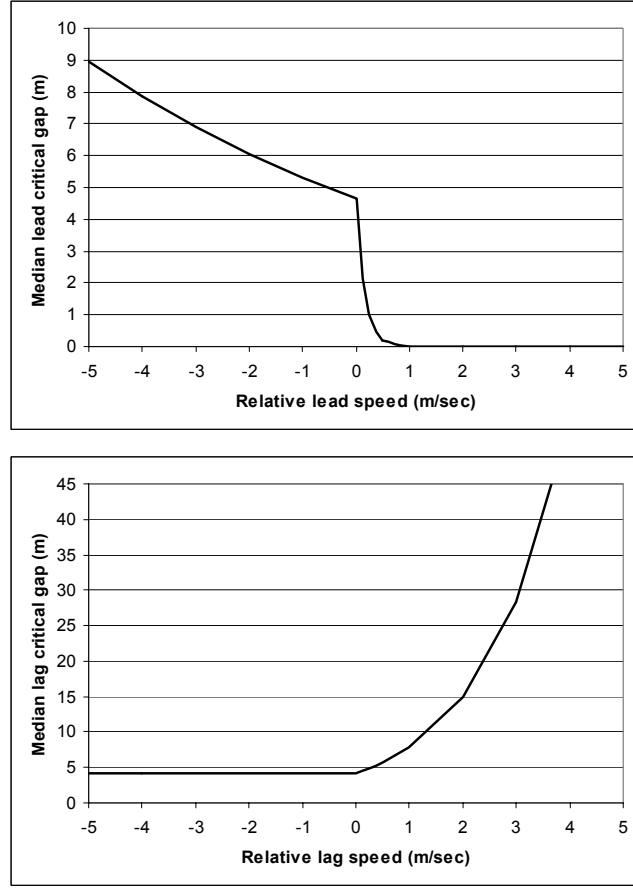
Figure 4.15: Median lead and lag critical gaps as a function of relative speed

Estimated coefficients of the unobserved driver characteristics variable, υ_n, are negative for both lead and lag critical gaps. This is consistent with the interpretation of υ_n as being negatively correlated with aggressive drivers, who require smaller gaps for lane changing compared to timid drivers.

In summary the estimated lead and lag gaps are given by:

$$G_{lnt}^{lead\ cr} = \exp\left(\begin{array}{l} 1.541 - 6.210 Max\left(0, \Delta V_{nt}^{lead}\right) - \\ -0.130 Min\left(0, \Delta V_{nt}^{lead}\right) - 0.008\upsilon_n + \varepsilon_{nt}^{lead} \end{array}\right)$$

$$G_{lnt}^{lag\ cr} = \exp\left(1.426 + 0.640 Max\left(0, \Delta V_{lnt}^{lag}\right) - 0.240\upsilon_n + \varepsilon_{lnt}^{lag}\right) \tag{4.21}$$

$$\varepsilon_{lnt}^{lead} \sim N\left(0, 0.854^2\right) \text{ and } \varepsilon_{lnt}^{lag} \sim N\left(0, 0.954^2\right)$$

4.4 Aggregate Calibration and Validation in MITSIMLab

The aggregate validation demonstrates the benefits that can be derived from using the modified models in traffic simulators. For this the estimated model (which is simulator independent) is implemented within the microscopic traffic simulator MITSIMLab (Yang and Koutsopoulos 1996) and calibrated and validated using data collected from a different site. The details of the aggregate validation are presented below.

4.4.1 Data

The data for aggregate calibration and validation consists of sensor data and aggregate trajectory data collected from a highly congested 1.5 miles section of I-80, in Emeryville and Berkeley, California. This freeway serves approximately 275,000 vehicles daily, and is one of the most vital transportation links in the San Francisco Bay Area. South of the study area, I-80 connects to the Bay Bridge and downtown San Francisco, as well as freeway interchanges to I-880 and downtown Oakland, and I-580 East. To the north of the study area are residential East Bay neighborhoods and I-580 West, leading to U.S. 101 and Marin County. Most of the drivers traveling in this area are local commuters. The left-most lane is an HOV lane that can be accessed to and exited from at any point in the section. The presence of this unlimited access HOV lane results in high difference in the level of service among different lanes and is therefore useful to test the target lane changing model.
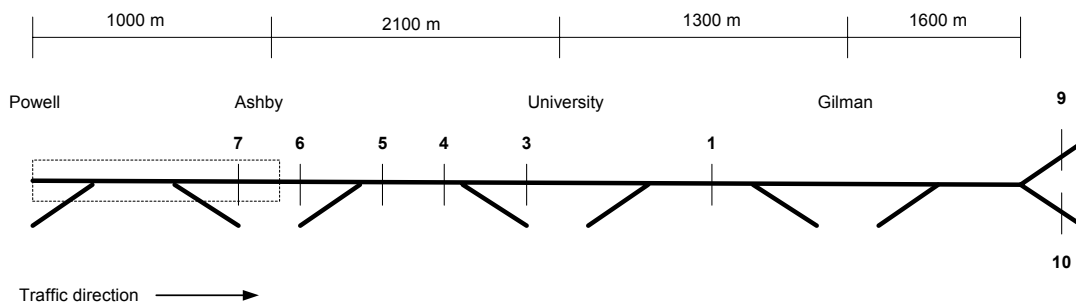
Figure 4.16: Schematic diagram of the I-80 data collection section
(not to scale)

The selected segment includes four on-ramps and three off-ramps (shown schematically in Figure 4.16). The downstream end of the network extends beyond the I-80/I-580 split, which is the major bottleneck in this area. These boundaries have been selected in order to ensure that possible queues forming at this bottleneck could be captured and explained in the model. The geometry of this section is particularly useful for validation of the lane changing model for several reasons:

- The presence of the unlimited access HOV lane and the high level of service differential associated with it.

- The section includes weaving sections that are required to test the lane changing model.

- The multiple ramps that exist in this section provide the ability to verify the path-plan based lane pre-positioning ('look ahead') effects that are incorporated in the model.

- The selected network is a corridor and therefore complex route choice situations do not arise. This is a desirable property for this study since it eliminates route choice as a source of modeling error, and so results should be more indicative of the effect of the driving behavior.

The data for aggregate calibration and validation consist of sensor data at the locations shown schematically in Figure 4.16. The data is available for two weeks (10 weekdays) at 30-second intervals and includes lane-specific traffic counts, occupancies and speeds. In addition trajectory data from the part of the corridor between the Powell Street and Ashby Street interchanges (showed by a dotted rectangular in Figure 4.16) is also available for one day between 2.35PM and 3.05PM. Aggregate statistics derived from these trajectories, which provide richer information compared to the sensor data, are also used in the validation (referred as trajectory data in the subsequent sections). The traffic counts and speeds from 2:35PM to 3:05PM are used for calibration and validation. The available sensor data has been split into two data sets with one week of data in each. The first week of data is used for aggregate calibration of the MITSIMLab model (the calibration methodology is detailed in Appendix B). The second week of data is used only for validation of the calibrated model and therefore allows independent validation.

4.4.2 Aggregate Calibration

MITSIMLab (Yang and Koutsopoulos 1996) is used to estimate the OD flows and calibrate the sensitive behavioral parameters. To identify the sensitive parameters, each candidate parameter was allowed to change over 10 iterations, while keeping all other parameters fixed and the parameters that results the larger improvements in the objective function were identified. The estimation data collection site (I-395) did not have any HOV lane and the effect of the HOV lane has also been captured during the aggregate calibration process by introducing an HOV dummy. This parameter was calibrated simultaneously with other behavioral parameters. The parameters chosen for calibration and their initial and calibrated values are shown in Table 4.10:

Table 4.10: Initial and calibrated values of the parameters of the target-lane model

Parameter		Initial Value	Calibrated Value
Car following	Acceleration Constant	0.040	0.042
	Deceleration Constant	-0.042	-0.084
Desired Speed	Mean	0.100	0.175
	Variance	0.150	0.254
Lane changing	Rightmost Lane Constant	-1.696	-1.052
	Current Lane Dummy	2.686	2.800
	HOV Dummy	0.000	1.521

As seen in Table 4.10, many of the parameters changed significantly during calibration. This is expected since the estimation dataset (collected from I-395, VA) and the aggregate calibration and validation dataset (collected from I-80, CA) had substantial differences in geometry and level of service as well as driver characteristics.

The model fit after calibration is presented in Figure 4.17. The lane shift model was also calibrated with the same aggregate data in a similar manner.

Calibrated Model:
Observed vs. Simulated Counts
(2:30 pm to 6:00 pm at 15 min intervals)

Calibrated Model:
Observed vs. Simulated Speeds
(2:30 pm to 6:00 pm at 15 min intervals)
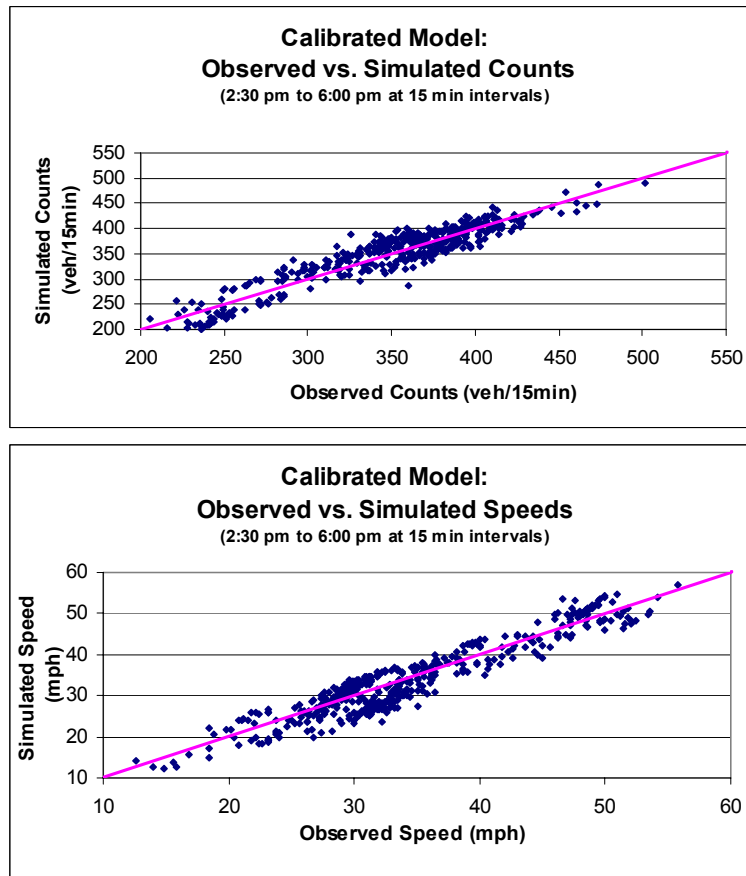
Figure 4.17: Calibration results for the target lane model

4.4.3 Aggregate Validation

The purpose of aggregate validation is to determine the extent to which the simulation model replicates the real system. At this step, the behavior parameters obtained in the aggregate calibration step are fixed and the model predictions are compared against the second set of traffic measurements, which have not been used for calibration. A separate OD matrix is estimated for the validation measurements.

The following measures of performance (MOPs) are selected based on their relevance to the evaluation of the lane changing model:

- End lane distribution of vehicles with respect to the starting lane
- Lane-specific sensor speeds
- Number of lane changes by vehicles
- Lane changes 'From' and 'To' lanes

A number of goodness of fit measures were used to evaluate the overall performance of a simulation model. Root Mean Square Error (*RMSE*), Root Mean Square Percent Error (*RMSPE*), Mean Error (*ME*) and Mean Percent Error (*MPE*). These measures are defined below:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(Y_n^{sim} - Y_n^{obs}\right)^2} \tag{4.22}$$

$$RMSPE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{Y_n^{sim} - Y_n^{obs}}{Y_n^{obs}}\right)^2} \tag{4.23}$$

$$ME = \frac{1}{N}\sum_{n=1}^{N}\left(Y_n^{sim} - Y_n^{obs}\right) \tag{4.24}$$

$$MPE = \frac{1}{N}\sum_{n=1}^{N}\frac{Y_n^{sim} - Y_n^{obs}}{Y_n^{obs}} \tag{4.25}$$

Where, Y_n^{obs} and Y_n^{sim} are the averages of observed and simulated measurements at space-time point *n*, calculated from all available data (i.e. several days of observations and/or multiple simulation replications).

RMSE and *RMSPE* penalize large errors at a higher rate relative to small errors. *ME* and *MPE* indicate systematic under-prediction or over-prediction in the simulated measurements.

End Lane Distribution

The distribution of vehicles across lanes at the end of the section with respect to the starting lane was extracted from the aggregate trajectory data and compared with the simulated lane distributions of both the models. Figure 4.18 presents the results of the comparison.
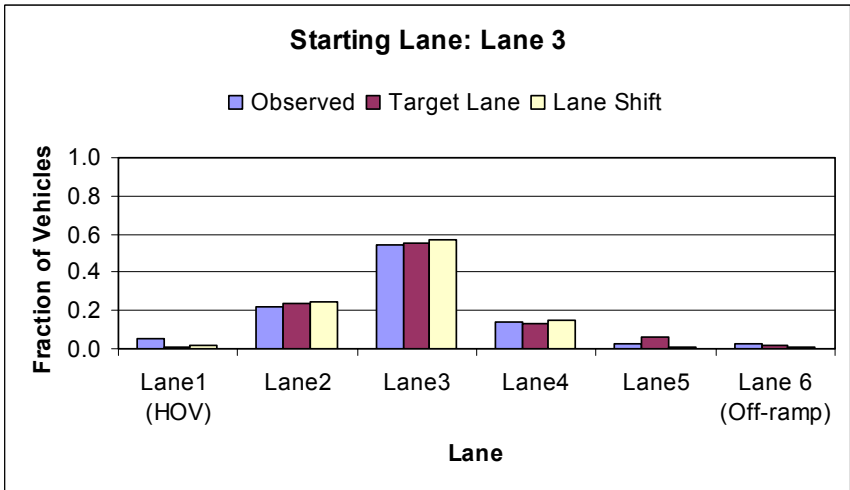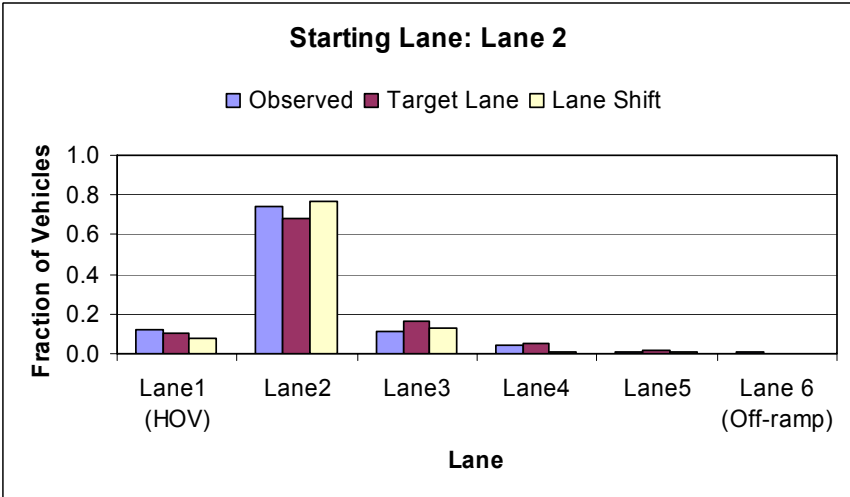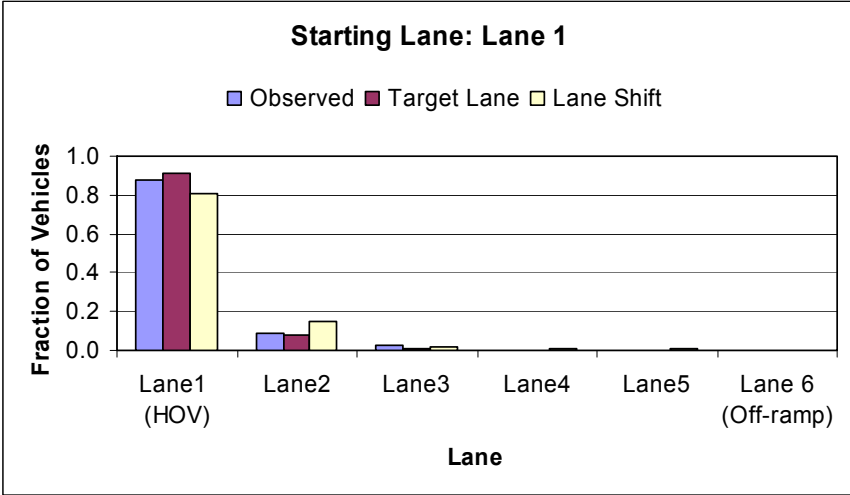
Figure 4.18: Comparison of end lane distribution of vehicles
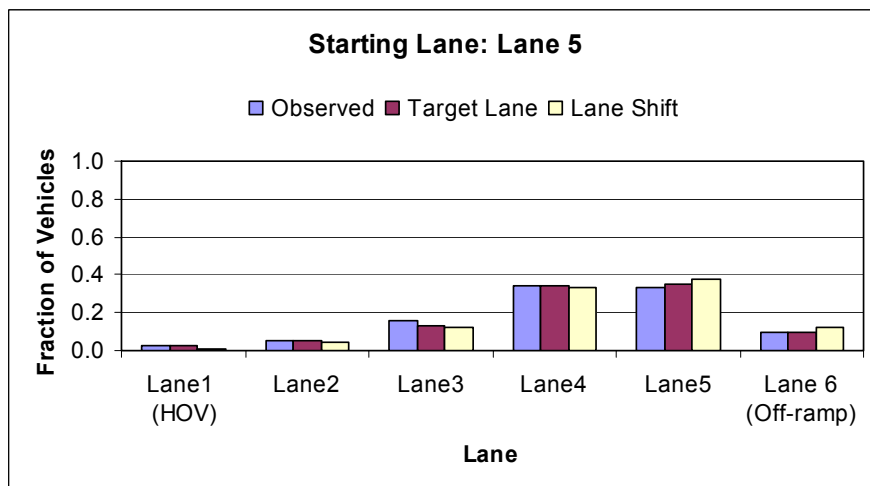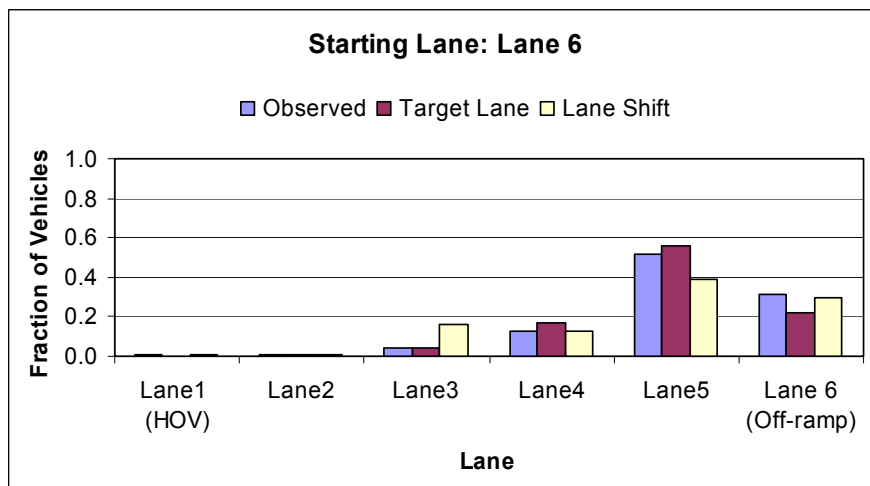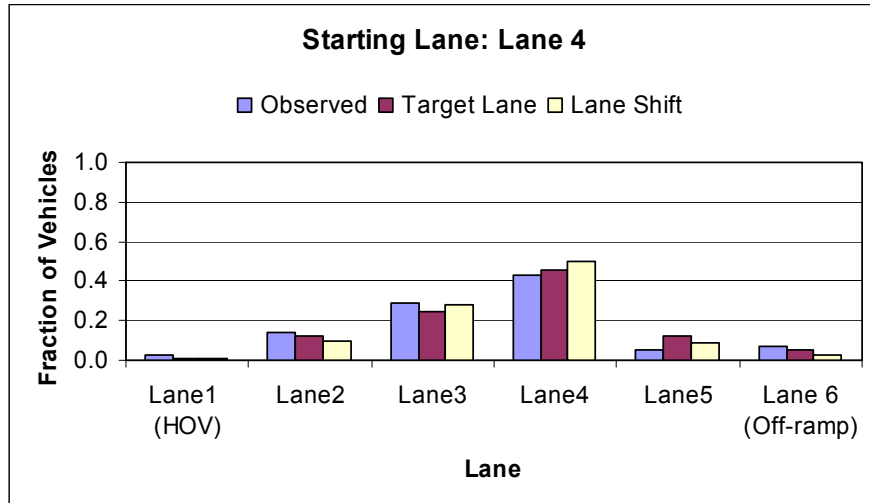
Figure 4.18: Comparison of end lane distribution of vehicles (contd.)
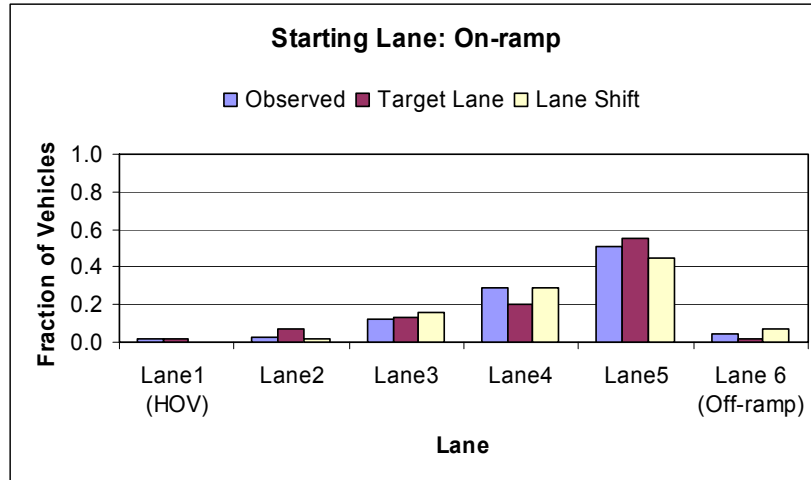
Figure 4.18: Comparison of end lane distribution of vehicles (contd.)

Overall, the model with explicit target lane matched the observations better and the RMSE was calculated to be 0.032 for the target lane model and 0.041 for the lane shift model denoting a 20.04 % improvement. A closer look at the results of the lane shift model shows a significant proportion of the error is due to incorrect representation of the HOV lane. The RMSE of the percentage of vehicles moving to the HOV from all starting lanes is 4.8% for the lane shift model and 2.1 % for the target lane model. This result indicates that the lane shift lane changing model is unable to correctly capture the attractiveness of the HOV lanes and therefore underestimated its use. These underestimations of the HOV flows is a potential source of discrepancy in the lane-specific traffic speed outputs discussed next, since the reduced flow rates on the HOV lane results in increased speeds on this lane.

Lane-specific Speeds

A separate set of lane-specific speed measurements from sensors (not used for calibration) has been used for validation purpose. Comparisons of the goodness of fit measures are presented in Table 4.11. As seen in the table, The target lane model consistently performs better particularly in terms of Mean Error and Mean Percent Error. The discrepancy in the lane distribution can be a potential source of the speed mismatch in the lane shift model since the erroneously lower calculations of the flows in the HOV lane result in increased speed outputs of the HOV lanes and reduced speed outputs in the other lanes.

Table 4.11: Goodness of fit statistics for the traffic speed comparison

Statistic	Lane Shift Model	Target Lane Model	Improvement
RMSE, m/sec	3.92	3.10	20.92 %
RMSPE (%)	14.89	12.15	18.40 %
ME (m/sec)	1.59	-0.83	47.80 %
MPE (%)	5.17	-3.33	35.59 %

Lane Changes by Vehicles

The number of lane changes by vehicles as observed in the trajectory data was compared against the simulated results of the target lane model and the lane shift model with the results presented in Figure 4.19.
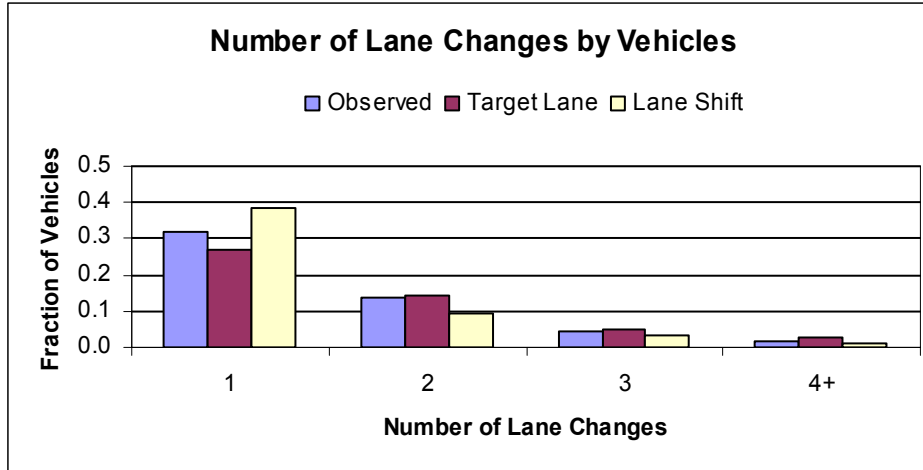
Figure 4.19: Comparison of number of lane changes by vehicles

The lane shift model under predicted the number of more-than-one lane changes probably due to the lane shift model including only adjacent lanes in the choice set of the driver and the higher level of service prevailing in lanes further away not being taken into account. The target lane model performed much better than the lane shift model particularly in terms of predicting the higher number of lane changes. The RMSE for the fraction of vehicles in the lane shift model and the target lane model were 0.040 and 0.024 respectively indicating an improvement of 38.33 %.

Lane changes 'From' and 'To' lane

The number of lane changes by 'From' (starting) and 'To' (ending) lanes was also compared. As seen in Figure 4.20, the lane shift model has a significantly small number of lane changes to the HOV Lane. The target lane model performs much better in this respect.
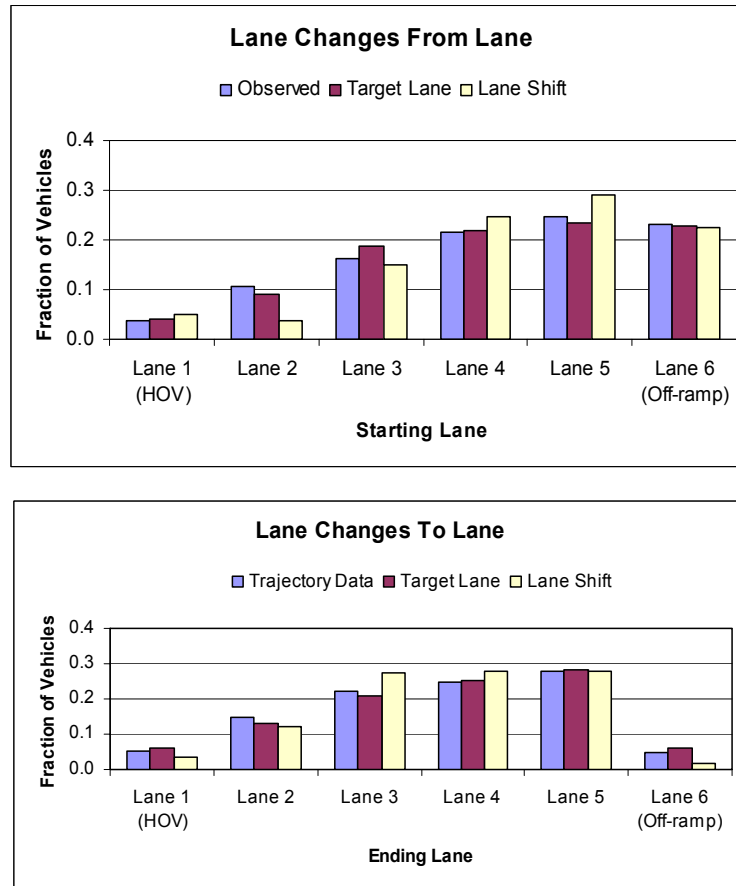
Figure 4.20: Comparison of lane changes 'From' and 'To' lanes

4.5   Model Validation in Other Simulators

As part of the NGSIM project of the FHWA, the new lane changing model was also tested independently in three commercial simulators by the model development teams. The results are summarized below. The detailed results are reported in Barceló *et al*. (2006), Speirs (2006) and Vortisch and Rössel (2006)[3].

_____

[3] The Target Lane model is referred as NGSIM Freeway Lane Selection Algorithm in these reports.

AIMSUN

The target lane model was implemented in AIMSUN (AIMSUN Target Lane) and compared against the default lane changing model of AIMSUN (AIMSUN Original). The comparison results are presented in Table 4.12 and Table 4.13. In the tables, the first rows denote the observed flows and speeds in the trajectory data and the simulated flows and speeds of the AIMSUN Target Lane and the AIMSUN Original models are presented in second and third rows respectively. The *RMSE* values for the two models compared with the observed data are presented in the last columns.

Table 4.12: Comparison of flows (vph)

	5	10	15	20	25	30	Avg.	RMSE
Observed	9804	10344	10020	9216	9648	7764	9466	
AIMSUN Target Lane	10281	10296	9972	9192	9432	7728	9484	216.28
AIMSUN Original	10320	10320	9948	9348	9204	7812	9492	285.44

Source: Commercial Validation of Freeway Lane Selection Model: Report on Testing the NGSIM Lane Selection Model with AIMSUN (Barceló *et al*. 2006).

Table 4.13: Comparison of speeds (mph)

	5	10	15	20	25	30	Avg.	RMSE
Observed	58.28	58.62	58.76	57.5	41.62	35.77	51.76	
AIMSUN Target Lane	58.53	58.32	58.45	58.56	44.42	36.9	52.53	0.31
AIMSUN Original	58.6	58.51	58.7	58.6	44	38.1	52.75	0.4

Source: Commercial Validation of Freeway Lane Selection Model: Report on Testing the NGSIM Lane Selection Model with AIMSUN (Barceló *et al*. 2006).

As seen in these tables, the target lane changing model performed better than the base model in terms of both flow and speed and has been selected to be incorporated in the commercial version.

Paramics

In the validation study in Paramics, the general finding was the algorithm works under a wide range of scenarios achieving its goal of encouraging drivers to consider lanes other than those strictly adjacent as viable or desirable to travel in. Rigorous comparisons against the default Paramics lane changing model were not conducted.

VISSIM

In the validation study conducted within the simulator VISSIM (Vortisch and Rössel 2006), the modeling of exclusive lanes was felt to be easier and more straight forward in the target lane changing model compared to the existing model in VISSIM (Figure 4.21). The *RMSE* for flow was reported to improve to 252 vph (VISSIM Target lane) from 288 vph (VISSIM Original). Although the speed was slightly worse though (6.7m/sfor VISSIM Original and 7.0 m/s for VISSIM Target Lane).
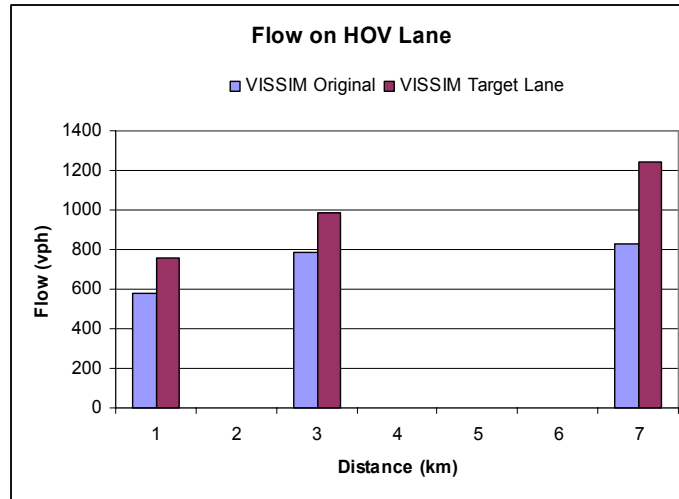
Figure 4.21: Comparison of flow on HOV lane

Source: Commercial Validation of the NGSIM Freeway Lane Selection Algorithm in VISSIM (Vortisch and Rössel 2006).

4.6   Summary

A lane changing model with explicit choice of a target lane is developed to capture the effect of latent planning in the immediate maneuvers of the driver. This approach differs from existing models that assume that drivers evaluate the current and adjacent lanes and choose a direction of change (or not to change) based on the relative utilities of these lanes. While the proposed model is applicable to any general freeway situation, it is most useful in cases where there exists a high difference in the level of service among the lanes, e.g. with an HOV lane.

The target lane model parameters have been estimated using a maximum likelihood estimator and detailed vehicle trajectory data. Comparison of goodness-of-fit test

statistics indicates significant improvement over the lane shift model that ignores the latent targets of the driver.

The improvement in the model performance was demonstrated through a detailed validation study within MITSIMLab where the simulation capability of the target lane model is compared against that of the lane shift model. Test statistics calculated in the aggregate validation stage indicates that the target lane model provides significantly better prediction for all measures of performance. The improvements in the modeling capability are further strengthened by independent validation within three commercial microscopic simulators AIMSUN, Paramics and VISSIM.

In the target lane model, the choice of target lane in subsequent instants is assumed to be independent of each other. That is, the driver is assumed to re-evaluate the situation at each time step and if required, change the latent plan by selecting a different target lane. This indirectly captures the evolution of the latent plans and dynamicity of driving behavior.

The data used for estimating the model was not from a highly congested situation and the lane changes were assumed to be through normal gap acceptance and the target gap were always the adjacent gap.

In this model, the heterogeneity in planning capability of the drivers has been ignored and it is assumed that all drivers are aware of the location of their exit from the beginning and choose lanes accordingly. This assumption has been relaxed in the arterial lane selection models discussed in Chapter 6.

Chapter 5

Freeway Merging

In this chapter, the latent plan involving the merging decision of drivers in a freeway on-ramp is presented. The similar two-stage general decision structure presented in the previous chapters, that is, choice of latent plans followed by selection of action to execute the plan, is also applicable in this scenario. However, the geometric and traffic characteristics associated with the merging situation lead to a model framework that is different from that of the freeway mainline lane selection presented in Chapter 4.

The chapter is organized as follows: the background of the research is presented in Section 5.1. In Section 5.2, the structure of the latent plan merging model is detailed in three sub-sections: the model components are presented in 5.2.1, and the descriptions of how these components lead to different plans and actions are presented in 5.2.2 and 5.2.3 respectively. The details of the model estimation are presented in Section 5.3. This section includes description of the data used to estimate the model parameters, the likelihood function and the estimation results. The comparison of the goodness-of-fit of the latent plan model against a reduced form model is also presented in this section. The chapter concludes with the validation results within the microscopic traffic simulator MITSIMLab and a summary of the findings.[4]

5.1   Background

Freeway merging involves a complex decision process. The target lane for freeway merging in right hand drive traffic is always the rightmost lane in the mainline. An on-

---

[4] The model presented in this chapter has been developed as part of NGSIM program of the FHWA. The results presented in this chapter have been reported in Choudhury *et al.* (2006, 2007a, 2007b). The results of the simplified merging models have been developed by Lee (2006) and Rao (2006).

ramp driver approaching the mainline seeks suitable gaps in the target lane for merging. The merge is executed when the gaps in the target lane are acceptable. However, in congested situations, when acceptable gaps are often not available, more complicated merging phenomena may be observed. For example, in highly congested situations, due to restricted maneuverability in the longitudinal direction, it may not be possible for a driver to preposition himself to a non-adjacent gap and he/she may decide to merge to the adjacent gap through courtesy of the lag driver in the target lane or decide to force in and compel the lag driver to slow down. In such situations, the chosen merging tactic dictates the plan/state of the driver, which in turn affects the driver's merging behavior. The execution of the plan involves acceptance of available adjacent gaps. The gap acceptance behavior models may differ depending on the merging tactic. For example, the acceptable gaps are smaller in case of courtesy merging compared to normal merging since there is less risk associated with it. However, the chosen plan/state is unobserved and only the action, that is the execution of the merge through gap acceptance, is observed.

Further, the plan/state may evolve dynamically as the immediate execution of the chosen merging plan may not be feasible. For example, a driver may begin with a plan of normal merging and then change to a plan of forced merging as the merging lane is coming to an end. The probabilities of transitions from one plan to another are affected by the perception of risk associated with the merge (anticipation), the inertia to continue the previously chosen merging plan (state-dependence) as well as the latent characteristics of the driver like impatience, urgency and aggressiveness.

Existing microscopic traffic simulators, such as AIMSUN (TSS 2004), Paramics (Quadstone 2004) and VISSIM (PTV 2004), use basic or modified versions of their lane changing models to model freeway merging behavior. These models consider gaps created by adjacent vehicles, and in some cases model reduced gap acceptance thresholds under congested conditions, but they do not explicitly consider all three merging tactics in a single framework. Thus the existing models often fail to capture these phenomena in the merging vicinity and represent congestion incorrectly.

The literature review (in Chapter 2) shows that several disjoint models have been developed specifically to model the cooperative lane changing and forced merging behaviors (Ahmed 1999, Hidas 2002, Wang *et al.* 2005), but none of these models

integrate the three merging mechanisms into a single framework. Hidas (2005) developed a merging model that includes both cooperative and forced merge components but the cooperative lane change part only consists of modeling the decision of the lag driver (whether or not to provide courtesy to the merging driver) and not the decision of the merging driver (whether or not to initiate or execute the courtesy lane change based on the behavior of the lag driver). The unified decision framework of the merging driver is thus not addressed in any of these models. Therefore, these models fail to capture drivers' transition from normal to cooperative or forced merge. The limitations of the existing models and the need for improving them is also reflected in the findings of the NGSIM study on Identification and Prioritization of Core Algorithm Categories, where development of freeway merging and weaving model was ranked fourth in importance by both model developers and users (Alexiadis *et al.* 2004).

5.2   Model Structure

The discussion in the previous section demonstrates the need to introduce the choice of merging tactic in the decision framework of the driver. The merging driver may merge through normal gap acceptance, merge through courtesy of another driver or decide to force in. In the case of a courtesy merge, the lag driver decelerates voluntarily whereas in the case of a forced merge, the lag driver is forced to decelerate. The execution of the merge involves acceptance of available gaps. The plan and the decision process of the driver are latent and only the end action of the driver (change to the lane in the mainline) is observed. The framework of the proposed combined merging model is summarized in Figure 5.1. Latent decisions are shown in ovals and observed actions are shown in rectangles.

The choice of merging tactic is hierarchical. The model hypothesizes four levels of decision-making: normal gap acceptance, decision to initiate courtesy merging, decision to initiate forced merging and gap acceptance for courtesy or forced merging. The merging driver first compares the available lead and lag gaps in the mainline to the corresponding minimum acceptable gaps (critical gaps) for normal gap acceptance. Critical gaps are functions of explanatory variables related to the subject driver and

his/her neighboring conditions. If both the lead and the lag gaps are greater than the critical gaps, the merge can be executed.
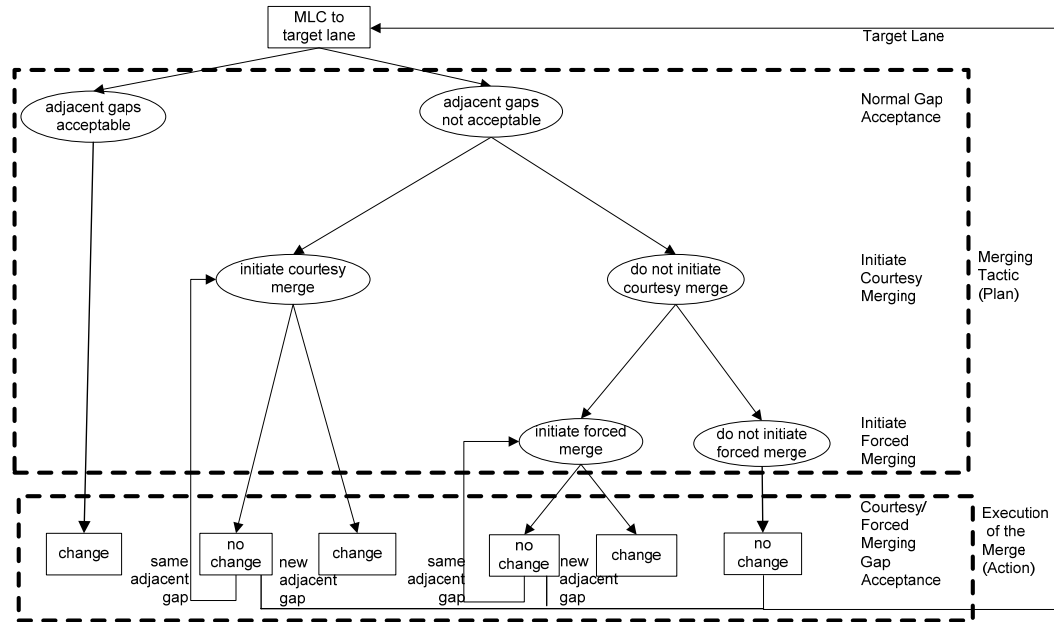
Figure 5.1: Structure of the combined merging model

If the gaps are not acceptable, the merging driver evaluates the speed, acceleration and relative position of the through vehicles and tries to evaluate whether or not the lag driver is providing courtesy. The courtesy or discourtesy of the lag driver is reflected in the anticipated gap. If the lag driver has decided to provide courtesy to a merging vehicle and has started to decelerate, the anticipated gap increases. The anticipated gap of a particular driver also depends on the length of the time horizon over which it is estimated. Differences in perception and planning abilities among drivers are captured by the distribution of the length of the time horizon. If the anticipated gap is acceptable, the merging driver perceives that he/she is receiving courtesy from the lag driver and initiates a courtesy merge. The immediate completion of the initiated courtesy merge however may not be possible due to unacceptable adjacent gaps.

If the anticipated gap is unacceptable, the driver decides whether to force his/her way to the mainline compelling the lag driver to slow down or not. This decision can depend on the urgency of the merge, driver characteristics (e.g. risk averseness) and traffic conditions. Similar to courtesy merge, the immediate completion of the initiated forced merge may not be possible due to unacceptable adjacent gaps.

If the driver does not initiate a courtesy or forced merge, the entire decision process is repeated in the next instant. However, if the driver has initiated a courtesy or forced merge, and is adjacent to the same gap, the subsequent decisions only involve evaluation of the adjacent gaps for completion of the initiated merge. After deciding to initiate a courtesy or forced merge, the choice of merging tactic is not reevaluated unless there is a significant change in neighborhood conditions e.g. the lead and/or lag in the mainline changes and the driver is adjacent to a new gap.

The decision tree of the driver thus differs depending on the previously chosen plan and action. The driver may be in one of the following three states at any instant:

- Normal merging (l_t =M),
- Courtesy merging (l_t =C) and
- Forced merging (l_t =F).

If the driver has not initiated a courtesy or forced merge previously, the state is normal. If the driver initiates a courtesy merge but the adjacent gaps are not immediately acceptable for executing the merge, there is a transition to the courtesy merging state. Similarly, if the driver initiates a forced merge but the adjacent gaps are not acceptable for immediate execution of the merge, there is a transition to forced state.
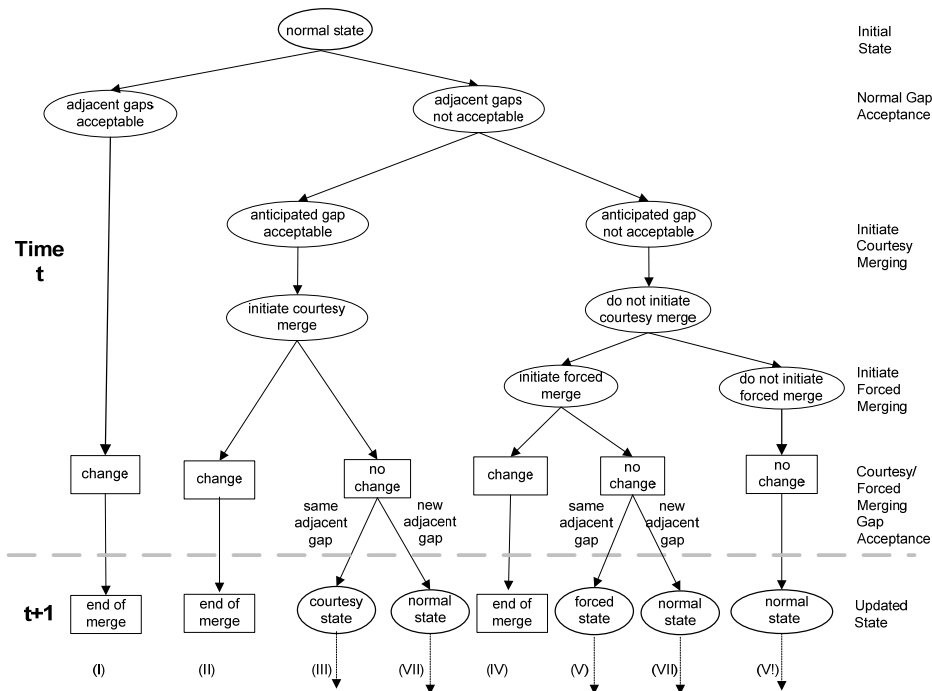
Figure 5.2: Decision tree for normal initial state

109

If the driver is in the normal state at an instant, the full decision tree is in effect. That is, the decision process starts from the top of the tree presented in Figure 5.1 and normal, courtesy and forced merging plans are evaluated sequentially. The intermediate decisions are detailed in Figure 5.2. As shown in the figure, while being in the normal state, the driver may perform any of the following:

I. Change lanes and complete the merge through normal gap acceptance,

II. Change lanes by initiating a courtesy merge and immediately complete it,

III/VII. Initiate a courtesy merge but do not complete it immediately,

IV. Change lanes by initiating a forced merge and immediately complete it,

V/VII. Initiate a forced merge but do not complete it immediately, or

VI. Do not initiate a courtesy or forced merge.

A lane change (I, II or IV) denotes the end of the merging process. If there are no lane changes, the decision process continues but there can be a transition to courtesy (III) or forced (V) merging state or the state can remain the same (VI). Further, if a courtesy or forced merge has been initiated but not completed from the normal state and the driver is adjacent to a new gap, the state of the driver is reset to normal (VII).

In the courtesy lane changing state, if the driver is adjacent to the same gap, the merging plan is not reevaluated and the full decision tree is not active. Rather, the decisions only involve evaluation of the adjacent gaps to complete the courtesy merge (Figure 5.3).
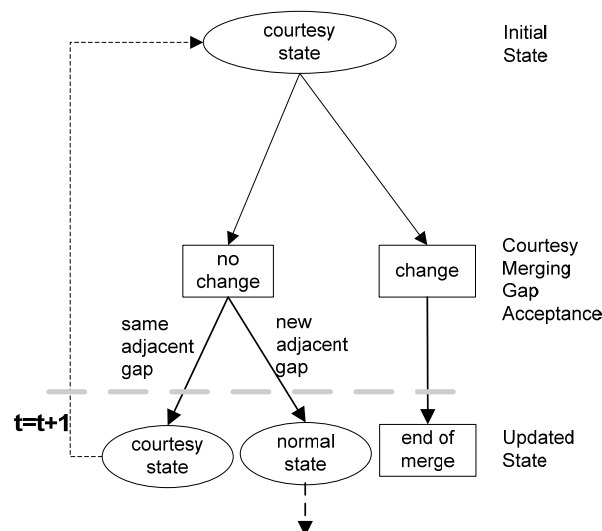
Figure 5.3: Decision tree for courtesy initial state

Thus once a transition is made from the normal to courtesy merging state, the state cannot change to forced merging or normal merging unless the driver is adjacent to a new gap. If the driver is adjacent to a new gap, the state is however reset to normal.

Similar to the courtesy merging state, in the forced merging state, if the driver is adjacent to the same gap, the entire decision process is not repeated. Rather, the decisions only involve evaluation of the adjacent gaps to complete the forced merge (Figure 5.4). Thus, once a transition is made from the normal to forced merge state, the state cannot go back to normal and it cannot change to the courtesy merge state unless the driver is adjacent to a new gap. The state of the driver is however reset to normal if the driver is adjacent to a new gap.
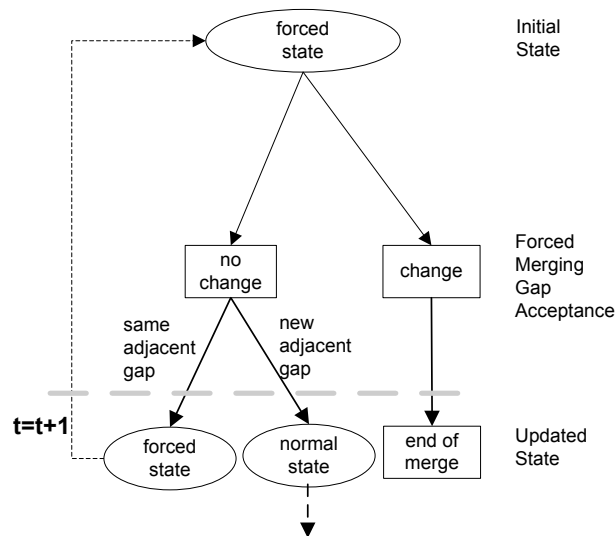
Figure 5.4: Decision tree for forced initial state

Thus when the driver is adjacent to the same gap in two subsequent time instants, the following state transitions are possible:

- Normal to Normal ($l_{t+1} = M \,|\, l_t = M$)
- Normal to Courtesy ($l_{t+1} = C \,|\, l_t = M$)
- Normal to Forced ($l_{t+1} = F \,|\, l_t = M$)
- Courtesy to Courtesy ($l_{t+1} = C \,|\, l_t = C$)
- Forced to Forced ($l_{t+1} = F \,|\, l_t = F$)

When the driver is adjacent to a new gap, the following transitions are possible.

- Normal to Normal ($l_{t+1} = M \mid l_t = M$)

- Courtesy to Normal ($l_{t+1} = M \mid l_t = C$)

- Forced to Normal ($l_{t+1} = M \mid l_t = F$)

The decision components affecting the state transitions and subsequent actions are discussed in Section 5.2.1. These include normal gap acceptance, decision to initiate a courtesy merge (anticipated gap acceptance), gap acceptance for completion of the courtesy merge, decision to initiate a forced merge and gap acceptance for completion of the forced merge. Descriptions of how these components lead to different plans and actions are presented in Sections 5.2.2 and 5.2.3 respectively.

5.2.1 Model Components

Normal Gap Acceptance

The normal gap acceptance model indicates whether or not a normal merge is possible using the existing gaps. In the model developed for this dataset, lead or lag vehicles are defined as the closest vehicles in the corresponding adjacent lanes within the current section of the subject vehicle (Figure 5.5). The lead gap is the clear spacing between the rear of the lead vehicle and the front of the subject vehicle. Similarly, the lag gap is the clear spacing between the rear of the subject vehicle and the front of the lag vehicle. One, or both, of these gaps may be negative if the vehicles overlap.
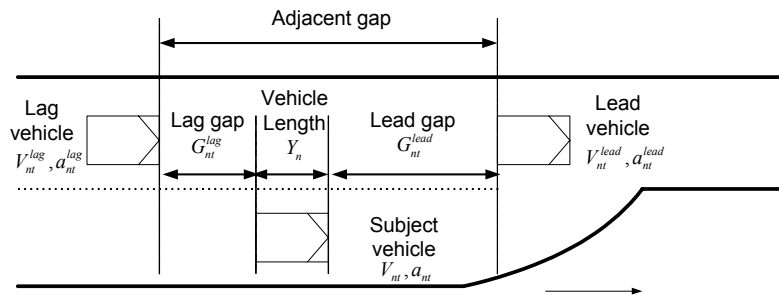
Figure 5.5: Vehicle relationships in a merging situation

An available gap is acceptable if it is greater than the critical gap. Similar to the critical gaps for gap acceptance for freeway lane changing, the critical gaps for normal

merging are assumed to follow lognormal distributions, the mean gap being a function of explanatory variables. This can be expressed as follows:

$$\ln\left(G_{nt}^{Mg}\right) = G\left(X_{nt}, \upsilon_n, \beta^{Mg}, \alpha^{Mg}\right) + \varepsilon_{nt}^{Mg} \quad g \in \{lead, lag\} \tag{5.1}$$

Where,

G_{nt}^{Mg} = critical normal gap, $g \in \{lead, lag\}$

X_{nt} = vector of explanatory variables

υ_n = individual-specific random effect : $\upsilon_n \sim N(0,1)$

β^{Mg}, α^{Mg} = parameters for normal gap acceptance

ε_{nt}^{Mg} = random term for normal gap acceptance: $\varepsilon_{nt}^{Mg} \sim N\left(0, \sigma_{Mg}^2\right)$

Gap acceptance can be affected by the interaction between the subject vehicle and the lead and lag vehicles in the adjacent lane. It can be also affected by the urgency of the merge that can be captured through the variable remaining distance or time to the mandatory lane changing (MLC) point. Candidate variables affecting normal gap acceptance include speed and acceleration of the subject, lead and lag vehicles, distance remaining to the MLC point on the ramp, type of vehicles etc.

The normal gap acceptance model assumes that the driver must accept both the lead gap and the lag gap to change lanes. Probability of driver *n* in normal state (M) making a lane change through normal gap acceptance at time *t* can be expressed as follows:

$$P_n\left(accept\ lead\ gap | l_t = M, \upsilon_n\right) P_n\left(accept\ lag\ gap | l_t = M, \upsilon_n\right)$$
$$= P_n\left(G_{nt}^{lead} > G_{nt}^{M\,lead} | l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{M\,lag} | l_t = M, \upsilon_n\right)$$
$$= \Phi\left[\frac{\ln\left(G_{nt}^{lead}\right) - G_{nt}^{M\,lead}}{\sigma_{M\,lead}}\right] \Phi\left[\frac{\ln\left(G_{nt}^{lag}\right) - G_{nt}^{M\,lag}}{\sigma_{M\,lag}}\right] \tag{5.2}$$

Decision to Initiate Courtesy Merge (Anticipated Gap Acceptance)

If the adjacent gaps are not acceptable, the merging driver evaluates the speed, acceleration and relative position of the through vehicles and anticipates the gap that will be available after τ_n seconds. Because of the difference in perception among individuals, the anticipation time τ_n may vary among individuals. The anticipated gap for individual *n* at time *t* is given by:

113

$$\overline{G}_{nt}(\tau_n) = G_{nt}^{lead} + G_{nt}^{lag} + Y_n + \tau_n \left(V_{nt}^{lead} - V_{nt}^{lag} \right) + \frac{1}{2}\tau_n^2 \left(a_{nt}^{lead} - a_{nt}^{lag} \right) \tag{5.3}$$

Where, as shown in Figure 5.5 for individual n at time t,

$\overline{G}_{nt}$ = *anticipated gap*

Y_n = *length of the subject vehicle*

$G_{nt}^{lead}, G_{nt}^{lag}$ = *available lead and lag spacings respectively*

$V_{nt}^{lead}, V_{nt}^{lag}$ = *lead and lag speeds respectively*

$a_{nt}^{lead}, a_{nt}^{lag}$ = *lead and lag accelerations respectively*

The anticipated gap is thus calculated based on the assumption that other drivers maintain their current accelerations. Therefore, if the lag driver of the merging driver is decelerating to provide courtesy, the anticipated gap is likely to increase.

If this anticipated gap is acceptable, the driver initiates a courtesy merge. The anticipated gap is acceptable if it is larger than the corresponding critical anticipated gap. Critical anticipated gaps are non-negative and assumed to be log-normally distributed[5]. The mean of the distribution is a function of explanatory variables.

$$\ln\left(G_{nt}^{A} \right) = G\left(X_{nt}, \upsilon_n, \beta^{A}, \alpha^{A} \right) + \varepsilon_{nt}^{A} \tag{5.4}$$

Where,

G_{nt}^{A} = *critical gap of individual n at time t for anticipated gap acceptance*

X_{nt} = *explanatory variables*

υ_n = *individual-specific random effect:* $\upsilon_n \sim N(0,1)$

ε_{nt}^{A} = *random term for anticipated gap acceptance:* $\varepsilon_{nt}^{A} \sim N\left(0,\sigma_A^2\right)$

β^{A}, α^{A} = *parameters for anticipated gap acceptance*

Candidate variables affecting the decision to initiate a courtesy merge include:

- Status of the lag vehicle in the mainline: speed and acceleration of the lag vehicle, type of the lag vehicle (heavy vehicle or not) etc.
- Traffic conditions: level of congestion in the mainline etc.

Probability of individual n initiating a courtesy merge at time t can be expressed as follows:

---

[5] Other non-negative distributions (truncated normal, truncated lognormal etc.) were also tested and the log-normal distribution had better fit than other distributions.

$$P_n(initiate\ courtesy\ merge|\ l_t = M, \upsilon_n)$$

$$= P_n\left(\overline{G}_{nt}(\tau_n) > G_{nt}^A \mid l_t = M, \upsilon_n\right) \tag{5.5}$$

$$= \Phi\left[\frac{\ln\left(\overline{G}_{nt}(\tau_n)\right) - \left(G_{nt}^A\right)}{\sigma_A}\right]$$

Decision to Complete the Courtesy Merge

After a driver initiates a courtesy merge, the completion of the merge depends on the acceptance of the immediate adjacent gaps. An available gap is acceptable for courtesy merge if it is greater than the corresponding critical gap (assumed to follow a lognormal distribution).

$$\ln\left(G_{nt}^{Cg}\right) = G\left(X_{nt}, \upsilon_n, \beta^{Cg}, \alpha^{Cg}\right) + \varepsilon_{nt}^{Cg} \quad g \in \{lead, lag\} \tag{5.6}$$

Where,

$G_{nt}^{Cg} = critical\ courtesy\ gap,\ g \in \{lead, lag\}$

$X_{nt} = vector\ of\ explanatory\ variables$

$\upsilon_n = individual\text{-}specific\ random\ effect:\ \upsilon_n \sim N(0,1)$

$\beta^{Cg}, \alpha^{Cg} = parameters\ for\ courtesy\ gap\ acceptance$

$\varepsilon_{nt}^{Cg} = random\ term\ for\ courtesy\ gap\ acceptance:\ \varepsilon_{nt}^{Cg} \sim N\left(0, \sigma_{Cg}^2\right)$

Though the critical gaps for completion of courtesy merge have the same general functional form as normal merge, the variables and the associated parameters can be different. Also, the critical gaps are assumed to be independent of the initial state that is the critical courtesy gap is assumed to be the same if the driver was in courtesy merging state at the beginning of the decision step (l_t =C) or was at normal merging state in the beginning and have just initiated the courtesy merge (l_t =M). In other words, it is assumed that the time elapsed after the driver has initiated a courtesy merge does not affect the critical gap for execution of the courtesy merge.

The gap acceptance model assumes that the driver must accept both the lead gap and the lag gap to change lanes. Probability of individual *n* executing a courtesy merge at time *t* given initial state *i* can be expressed as follows:

$$P_n \left(accept\ lead\ gap | l_{t=i},\ \upsilon_n \right) P_n \left(accept\ lag\ gap | l_{t=i},\ \upsilon_n \right)$$

$$= P_n \left(G_{nt}^{lead} > G_{nt}^{C\,lead} \mid l_t = i,\ \upsilon_n \right) P_n \left(G_{nt}^{lag} > G_{nt}^{C\,lag} \mid l_t = i,\ \upsilon_n \right)$$

$$= \Phi \left[\frac{\ln\left(G_{nt}^{lead} \right) - G_{nt}^{C\,lead}}{\sigma_{C\,lead}} \right] \Phi \left[\frac{\ln\left(G_{nt}^{lag} \right) - G_{nt}^{C\,lag}}{\sigma_{C\,lag}} \right] \tag{5.7}$$

$$\forall i \in M, C$$

Decision to Initiate a Forced Merge

If the current gaps are not acceptable and the driver perceives that a courtesy merge is also infeasible (anticipated gap is not acceptable), the driver evaluates whether or not to initiate a forced merge. By initiating a forced merge, the merging driver imposes a deceleration on the lag vehicle in the mainline. The utility of initiating a forced merge can be expressed as follows:

$$U_{nt}^F = U \left(X_{nt}, \upsilon_n, \beta^F, \alpha^F, \varepsilon_{nt}^F \right) \tag{5.8}$$

Where,

$U_{nt}^F =$ is the utility of initiating a forced merge by individual n at time t

$\beta^F, \alpha^F =$ parameters associated with initiating a forced merge

$\varepsilon_{nt}^F =$ random term for initiating forced merge

Candidate variables affecting the decision to initiate a forced merge include:

- Status of the merging driver: distance to the MLC point, delay (time elapsed since the driver is in MLC condition, as a proxy for impatience), speed, type of vehicle (heavy vehicle or not) etc.

- Status of the lag vehicle in the mainline: speed and acceleration of the lag vehicle, type of the lag vehicle (heavy vehicle or not) etc.

- Traffic conditions: level of congestion in the mainline, queue behind (merging vehicles waiting behind the subject vehicle) etc.

By assuming that the relationship between the influencing variables are linear and that the random error terms ε_{nt}^F are independently and identically extreme value distributed, the probability of initiating a forced merge can be modeled as a logit model and can be expressed as follows:

$$P_n \left(initiate\ forced\ merge \middle| l_t = M, \upsilon_n \right) = \frac{1}{1 + \exp\left(-\beta^F X_{nt} - \alpha^F \upsilon_n \right)} \tag{5.9}$$

It may be noted that the individual-specific term υ_n is assumed to have a linear effect in the utility in this case. It can have other non-linear forms as well (e.g. interaction with other variables in the utility).

Decision to Complete a Forced Merge

After a driver decides to initiate a forced merge, the actual merge is executed only when the available gaps are acceptable in comparison with the critical gaps for the forced merge. Similar to normal and courtesy merge the critical gap for forced merge is assumed to be log-normally distributed, the parameters being different from the other types of merge.

$$\ln \left(G_{nt}^{Fg} \right) = G \left(X_{nt}, \upsilon_n, \beta^{Fg}, \alpha^{Fg} \right) + \varepsilon_{nt}^{Fg} \quad g \in \{lead, lag\} \tag{5.10}$$

Where,

$G_{nt}^{Fg} = critical\ forced\ gap,\ g \in \{lead, lag\}$

$X_{nt} = vector\ of\ explanatory\ variables$

$\upsilon_n = individual\text{-}specific\ random\ effect:\ \upsilon_n \sim N(0,1)$

$\beta_{nt}^{Fg}, \alpha_{nt}^{Fg} = parameters\ for\ forced\ gap\ acceptance$

$\varepsilon_{nt}^{Fg} = random\ term\ for\ forced\ gap\ acceptance:\ \varepsilon_{nt}^{Fg} \sim N\left(0, \sigma_{Fg}^2\right)$

Further, similar to courtesy merging gap acceptance, the forced merging critical gap is assumed to be independent of the time the driver has been in forced merging state and all else being equal the probability of forced gap acceptance is the same if the initial state at the beginning of the time period was forced (l_t =F) or normal (l_t =M).

Probability of individual n executing a forced merge at time t given initial state i can be expressed as follows:

$$P_n \left(accept\ lead\ gap \middle| l_t = i, \upsilon_n \right) P_n \left(accept\ lag\ gap \mid l_t = i, \upsilon_n \right)$$
$$= P_n \left(G_{nt}^{lead} > G_{nt}^{F\ lead} \middle| l_t = i, \upsilon_n \right) P_n \left(G_{nt}^{lag} > G_{nt}^{F\ lag} \mid l_t = i, \upsilon_n \right)$$
$$= \Phi \left[\frac{\ln \left(G_{nt}^{lead} \right) - G_{nt}^{F\ lead}}{\sigma_{F\ lead}} \right] \Phi \left[\frac{\ln \left(G_{nt}^{lag} \right) - G_{nt}^{F\ lag}}{\sigma_{F\ lag}} \right] \tag{5.11}$$
$$i \in M, F$$

5.2.2 Choice of Plan: Selecting the Merging Tactic

The driver first evaluates whether or not a lane change is possible using the existing adjacent gaps without a courtesy or forced merge. So, the initial state and plan is always normal. The driver then evaluates the courtesy merge and forced merge plans sequentially. The transition probabilities from one plan/state to another are discussed next.

Courtesy Merge

If the adjacent gaps are not acceptable under normal gap acceptance, the merging driver evaluates the speed, acceleration and relative position of the through vehicles and decides whether or not to initiate a courtesy merge (Equation 5.5). If the driver initiates a courtesy merge but is unable to complete it immediately, there is a transition from the normal merging to courtesy merging state. If the adjacent gap is the same, the probability of a transition from normal to courtesy merge is therefore the combined probabilities of not accepting the normal gap, initiating a courtesy merge and not completing the courtesy merge and can be expressed as follows:

$$
\begin{aligned}
&P_n\left(l_{t+1} = C\,|\,l_t = M, \upsilon_n, \tau_n\right)\\
&= \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{M\,lead}\,|\,l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{M\,lag}\,|\,l_t = M, \upsilon_n\right)\right]\\
&\quad \left[P_n\left(\overline{G}_{nt} > G_{A\,nt}\,|\,l_t = M, \upsilon_n, \tau_n\right)\right]\\
&\quad \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{C\,lead}\,|\,l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{C\,lag}\,|\,l_t = M, \upsilon_n\right)\right]
\end{aligned}
\tag{5.12}
$$

Where, the probabilities of the components can be calculated using Equations 5.2, 5.5 and 5.7.

Once the driver has initiated a courtesy merge, as long as he/she is adjacent to the same gap, the probability of being in the courtesy merge state is 1. On the other hand, if the driver has already initiated a forced merge, and is adjacent to the same gap, the probability of initiating a courtesy merge is 0. However, if a courtesy merge has been initiated, but not completed, and the vehicle is adjacent to a new gap (i.e. the lead and/or lag vehicle has changed), the state of the driver is reset to normal. The transition probabilities to courtesy merge state or plan are summarized in the following equations:

$$P_n\left(l_{t+1} = C \mid l_t = M, \upsilon_n, \tau_n\right)$$

$$= \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{M\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{M\,lag} \mid l_t = M, \upsilon_n\right)\right]$$

$$\left[P_n\left(\overline{G}_{nt} > G_{Ant} \mid l_t = M, \upsilon_n, \tau_n\right)\right]$$

$$\left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{C\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{C\,lag} \mid l_t = M, \upsilon_n\right)\right]\delta_{nt}$$

$$P_n\left(l_{t+1} = C \mid l_t = C, \upsilon_n, \tau_n\right)$$

$$= \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{C\,lead} \mid l_t = C, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{C\,lag} \mid l_t = C, \upsilon_n\right)\right]\delta_{nt}$$

$$P_n\left(l_{t+1} = C \mid l_t = F, \upsilon_n, \tau_n\right)$$

$$= 0$$

(5.13)

Where,

$\delta_{nt} = 1$ *if the driver is adjacent to the same gap at time t and t+1, 0 otherwise.*

Forced Merge

If the current gaps are not acceptable and the driver perceives that a courtesy merge is also infeasible (anticipated gap is not acceptable), the driver chooses whether or not to initiate a forced merge (Equation 5.9). However, if the driver initiates a forced merge, is unable to complete it immediately, and is adjacent to the same gap, the driver remains in the forced merging state. In case of the same adjacent gap, the probability of a transition from normal to forced merge is therefore the combined probability of not accepting the normal gap, not initiating the courtesy merge, initiating a forced merge and not completing the forced merge and can be expressed as follows:

$$P_n\left(l_{t+1} = F \mid l_t = M, \upsilon_n, \tau_n\right)$$

$$= \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{M\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{M\,lag} \mid l_t = M, \upsilon_n\right)\right]$$

$$\left[1 - P_n\left(\overline{G}_{nt} > G_{Ant} \mid l_t = M, \upsilon_n, \tau_n\right)\right]\left[\frac{1}{1 + \exp\left(-\beta^F X_{nt} - \alpha^F \upsilon_n\right)}\right]$$

$$\left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{F\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{F\,lag} \mid l_t = M, \upsilon_n\right)\right]$$

(5.14)

Where, the probabilities of the components can be calculated using Equations 5.2, 5.5, 5.9 and 5.11.

Similar to the courtesy merge, the probability that a driver is in the force merge state depends on the previous state: the probability is 1 if the driver had already initiated a forced merge to the same gap and 0 if the driver had already initiated a courtesy merge to

119

the same gap. However, if the driver cannot complete a forced merge that has been initiated while he/she is adjacent to the same gap, the state is reset to the normal state. The probability of being in forced merging state can therefore be expressed as follows:

$$P_n\left(l_{t+1} = F \mid l_t = M, \upsilon_n, \tau_n\right)$$
$$= \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{M\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{M\,lag} \mid l_t = M, \upsilon_n\right)\right]$$
$$\left[1 - P_n\left(\overline{G}_{nt} > G_{A\,nt} \mid l_t = M, \upsilon_n, \tau_n\right)\right]\left[\frac{1}{1 + \exp\left(-\beta^F X_{nt} - \alpha^F \upsilon_n\right)}\right]$$
$$\left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{F\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{F\,lag} \mid l_t = M, \upsilon_n\right)\right]\delta_{nt}$$

$$P_n\left(l_{t+1} = F \mid l_t = C, \upsilon_n, \tau_n\right) \tag{5.15}$$
$$= 0$$

$$P_n\left(l_{t+1} = F \mid l_t = F, \upsilon_n, \tau_n\right)$$
$$= \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{F\,lead} \mid l_t = F, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{F\,lag} \mid l_t = F, \upsilon_n\right)\right]\delta_{nt}$$

Normal Merge

If the driver does not initiate a courtesy or forced merge, the state remains normal. The probability of a transition from the normal to normal state is therefore the combined probability of not accepting the normal gap, not initiating courtesy and not initiating a forced merge and can be expressed as follows:

$$P_n\left(l_{t+1} = M \mid l_t = M, \upsilon_n, \tau_n\right)$$
$$= \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{M\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{M\,lag} \mid l_t = M, \upsilon_n\right)\right] \tag{5.16}$$
$$\left[1 - P_n\left(\overline{G}_{nt} > G_{A\,nt} \mid l_t = M, \upsilon_n, \tau_n\right)\right]\left[1 - \frac{1}{1 + \exp\left(-\beta^F X_{nt} - \alpha^F \upsilon_n\right)}\right]$$

Where, the probabilities of the components can be calculated using Equations 5.2, 5.5 and 5.9.

Further, whenever the driver is adjacent to a new gap, the state is reset to normal. The probability of being in normal merging state can therefore be expressed as follows:

$$P_n\left(l_{t+1} = M \mid l_t = M, \upsilon_n, \tau_n\right)$$

$$= \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{M\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{M\,lag} \mid l_t = M, \upsilon_n\right)\right]$$

$$\left[1 - P_n\left(\overline{G}_{nt} > G_{Ant} \mid l_t = M, \upsilon_n, \tau_n\right)\right]\left[1 - \frac{1}{1 + \exp\left(-\beta^F X_{nt} - \alpha^F \upsilon_n\right)}\right]\delta_{nt} + (1 - \delta_{nt}) \qquad (5.17)$$

$$P_n\left(l_{t+1} = M \mid l_t = C, \upsilon_n, \tau_n\right) = 1 - \delta_{nt}$$

$$P_n\left(l_{t+1} = M \mid l_t = F, \upsilon_n, \tau_n\right) = 1 - \delta_{nt}$$

5.2.3 Choice of Action: Execution of the Merge

The initial state of the driver can be normal, courtesy or forced. The observed action involves execution of the plan. The decision tree of the driver and the critical gaps vary with the initial state.

Normal State

If the driver is in the normal state at an instant, the driver can execute a lane change in three ways (Figure 5.2):

1. Change lanes through normal gap acceptance,
2. Change lanes by initiating a courtesy merge and immediately completing it,
3. Change lanes by initiating a forced merge and immediately completing it.

The probability of observing a lane change conditional on the initial state being normal can thus have the following three components:

Change lanes through normal gap acceptance:

A lane change through normal gap acceptance is possible if both lead and lag gaps are acceptable for a normal merge and can be expressed as follows:

$$Component\ 1 = P_n\left(G_{nt}^{lead} > G_{nt}^{M\,lead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{M\,lag} \mid l_t = M, \upsilon_n\right) \qquad (5.18)$$

These probabilities can be calculated using Equations 5.1 and 5.2.

Change lanes by initiating a courtesy merge and immediately completing it:

These lane changes occur when the adjacent gap is not acceptable for normal merge but the driver perceives that the lag driver in the mainline is providing courtesy to him, initiate a courtesy merge and complete the courtesy merge in the same time step. The probability of such a lane change is therefore the combined probability of not accepting

the normal gap, initiating courtesy and accepting the adjacent gap for courtesy gap acceptance, all at the same time step. This can be expressed as follows:

$$Component\ 2 = \left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{Mlead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{Mlag} \mid l_t = M, \upsilon_n\right)\right]$$

$$\left[P_n\left(\overline{G}_{nt} > G_{nt}^{A} \mid l_t = M, \upsilon_n, \tau_n\right)\right]\left[P_n\left(G_{nt}^{lead} > G_{nt}^{Clead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{Clag} \mid l_t = M, \upsilon_n\right)\right]$$

(5.19)

These probabilities can be calculated using Equations 5.2, 5.5 and 5.7.

Change lanes by initiating a forced merge and immediately completing it:

This category of lane change occurs when the adjacent gap is not acceptable for normal merge, and the anticipated gap is not acceptable for initiating courtesy but the driver decides to initiate a forced merge and the adjacent gap is acceptable for immediate execution of the forced merge.

The probability of this type of lane change thus refers to the joint probability of not accepting the normal gaps, not accepting the anticipated gap, deciding to initiate a forced merge and accepting the lead and lag gaps through forced gap acceptance. This can be expressed as follows:

$$Component\ 3 =$$

$$\left[1 - P_n\left(G_{nt}^{lead} > G_{nt}^{Mlead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{M\ nt}^{lag} \mid l_t = M, \upsilon_n\right)\right]$$

$$\left[1 - P_n\left(\overline{G}_{nt} > G_{nt}^{A} \mid l_t = M, \upsilon_n, \tau_n\right)\right]\left[\frac{1}{1 + \exp\left(-\beta^F X_{nt} - \alpha^F \upsilon_n\right)}\right]$$

(5.20)

$$\left[P_n\left(G_{nt}^{lead} > G_{nt}^{Flead} \mid l_t = M, \upsilon_n\right) P_n\left(G_{nt}^{lag} > G_{nt}^{Flag} \mid l_t = M, \upsilon_n\right)\right]$$

This can be calculated using Equations 5.2, 5.5, 5.9 and 5.11.

Probability of making a lane change given the initial state is normal is the sum of the above mentioned components.

$$P_n\left(j_t = 1 \mid l_t = M, \upsilon_n, \tau_n\right)$$
$$= Component\ 1 + Component\ 2 + Component\ 3$$

(5.21)

Where, the three components are given by Equations 5.18, 5.19 and 5.20 respectively.

Probability of no lane change conditional that the initial state is normal merge can be expressed as follows:

$$P_n\left(j_t = 0 \mid l_t = M, \upsilon_n\right) = 1 - P_n\left(j_t = 1 \mid l_t = M, \upsilon_n\right)$$

(5.22)

Courtesy Merging State

In the courtesy merge state, if the driver is adjacent to the same gap, the entire decision process is not repeated. Rather, the decisions only involve evaluation of the adjacent gaps to complete the courtesy merge (Figure 5.3). Probability of a lane change conditional on the initial state is courtesy merge can therefore be expressed as follows:

$$P_n \left(j_t = 1 \middle| l_t = C, \upsilon_n \right)$$
$$= P_n \left(G_{nt}^{lead} > G_{nt}^{Clead} \middle| l_t = C, \upsilon_n \right) P_n \left(G_{nt}^{lag} > G_{nt}^{Clag} \mid l_t = C, \upsilon_n \right) \tag{5.23}$$

This can be calculated using Equation 5.7.

Probability of no lane change conditional on the initial state is courtesy merge can be expressed as follows:

$$P_n \left(j_t = 0 \middle| l_t = C, \upsilon_n \right) = 1 - P_n \left(j_t = 1 \middle| l_t = C, \upsilon_n \right) \tag{5.24}$$

Forced Merging State

Similar to the courtesy merge state, in the forced merge state, if the driver is adjacent to the same gap, the entire decision process is not repeated. Rather, the decisions only involve evaluation of the adjacent gaps to complete the forced merge (Figure 5.4).

Probability of a lane change conditional on the initial state is forced merge can therefore be expressed as follows:

$$P_n \left(j_t = 1 \middle| l_t = F, \upsilon_n \right)$$
$$= P_n \left(G_{nt}^{lead} > G_{nt}^{Flead} \middle| l_t = F, \upsilon_n \right) P_n \left(G_{nt}^{lag} > G_{nt}^{Flag} \mid l_t = F, \upsilon_n \right) \tag{5.25}$$

This can be calculated using Equation 5.11.

Probability of no lane change conditional on the initial state is forced merge can be expressed as follows:

$$P_n \left(j_t = 0 \middle| l_t = F, \upsilon_n \right) = 1 - P_n \left(j_t = 1 \middle| l_t = F, \upsilon_n \right) \tag{5.26}$$

Depending on the chosen plan and decision state, the lane action can thus have different probabilities.

5.3 Model Estimation

5.3.1 Data

Study Area

The data used in the estimation of the driving behavior model represents travel on a 502.9 meters northbound section of Interstate 80 (I-80) in Emeryville, California (Figure 5.6).

Figure 5.6: Estimation data collection site

The data was collected and processed as part of the FHWA's NGSIM program. The data was collected using video cameras mounted on a 30-story building adjacent to I-80. The University of California at Berkeley maintains traffic surveillance capabilities at the building and the segment is known as the Berkeley Highway Laboratory (BHL) site.
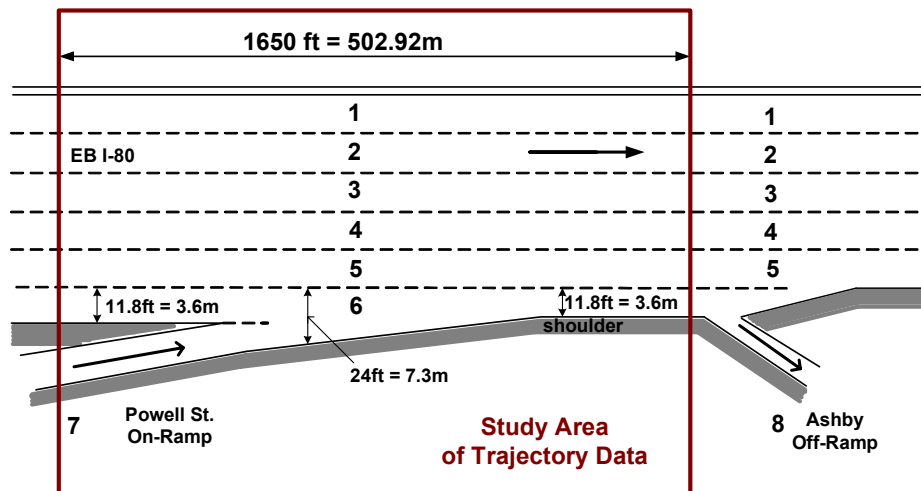
Figure 5.7: Schematic of the estimation data collection site

(not in scale)

Complete vehicle trajectories were recorded at a resolution of 10 frames per second. 45 minutes of data were collected on April 13, 2005 at a resolution of 0.1 second during the time intervals 4:00 to 4:15 p.m., 5:00 to 5:15 p.m., and 5:15 to 5:30 p.m. The 4:00 to 4:15 p.m. period is representative of a transitional traffic period in the build up to congested conditions, and the 5:00 to 5:30 p.m. period is representative of congested conditions.

For data handling tractability, the combined dataset was sampled at the rate of 1 in 10 observations, meaning the locations of vehicles were known at one-second time steps. The resulting dataset had 540 merging vehicles with 17,352 observations.

Characteristics of the Estimation Dataset

As shown in the schematic representation of the study area in Figure 5.7, there are no physical lane markings separating the on-ramp vehicles from the mainline vehicles. The absence of a physical lane demarcation over a long stretch made it difficult to specify when a lane change has occurred, and necessitated the definition of an imaginary lane boundary.

The mandatory lane changing (MLC) point, as shown in Figure 5.8, is defined as the point where the width of the rightmost lane assumes the single lane width (3.6 meter). The definition of this point is important as it defines whether or not a merge has occurred.

A merge is classified as completed when the center point of the vehicle has crossed this imaginary line/lane-mark (X in Figure 5.8).
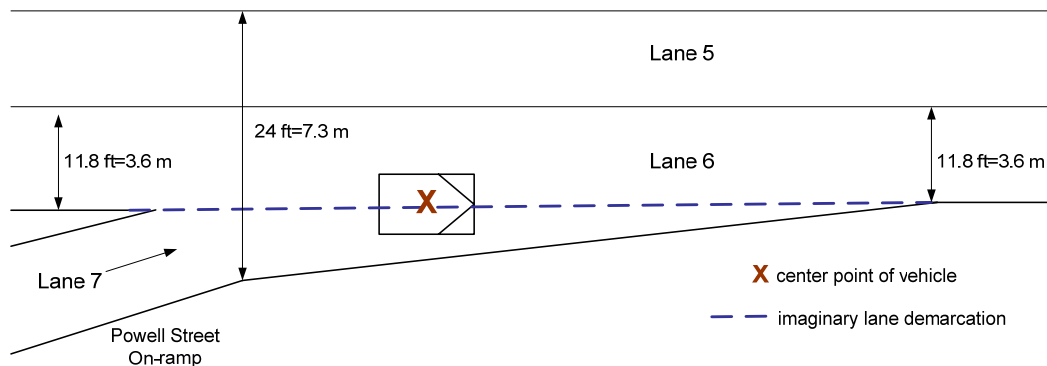
Figure 5.8: Definition of merge point

The vehicle trajectory data containing the coordinates of the merging and mainline vehicles in the section were used to derive the required variables for estimation of speed, acceleration, average density, etc.

Speeds in the merging section (the on-ramp and part of lane 6 as defined in Figure 5.8) vary from 0 m/sec to a maximum of 20.7 m/sec with a mean of 4.2 m/sec. There are many stop-and-go situations present in the dataset. Densities calculated 150 meters downstream of the merging vehicles in lane 6 range from 0 veh/km/lane to 126.7 veh/km/lane with an average of 61.9 veh/km/lane. 1.4 percent of the merging vehicles in the dataset are heavy vehicles (trucks in this case).

The distributions of speed, acceleration, and density in lane 6 and distance to the MLC point in the entire dataset are shown in Figure 5.9.
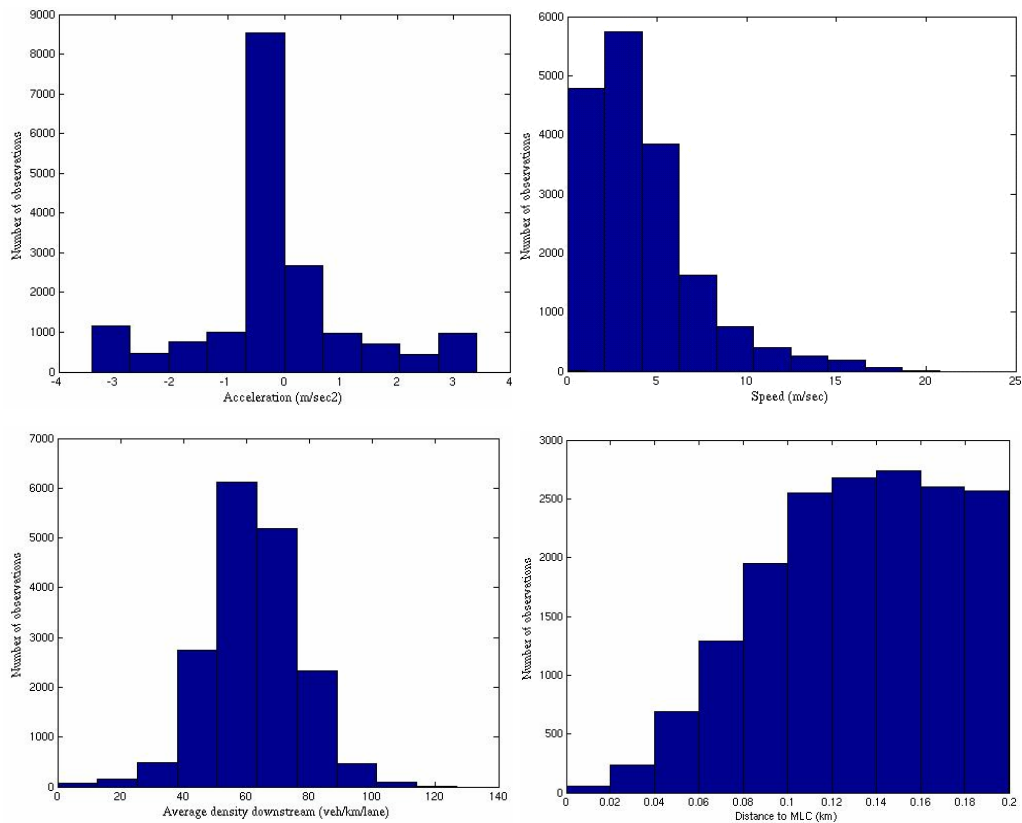
Figure 5.9: Distributions of speed, acceleration, density, and distance to MLC point

As defined earlier in this chapter, the lead gap is the distance between the front of the subject vehicle to the rear of the lead vehicle in the target lane, and the lag gap is the distance between the rear of the subject vehicle and the front of the lag vehicle in the

target lane (Figure 5.2). Negative gaps imply overlap between the subject and the lead/lag vehicle. The statistics relating to the subject vehicle are shown in Table 5.1.

Table 5.1: Statistics of variables related to the subject vehicle

Variable	Mean	Std Dev	Median	Minimum	Maximum
Speed (m/sec)	4.2	3.11	3.34	0	20.7
Average Density d/s (veh/km/lane)	61.9	15.3	60.0	0	126.7
Distance to MLC (km)	0.13	0.04	0.13	0	0.20
Acceleration (m/sec^2)	0.61	1.03	0	0	3.41
Deceleration (m/sec^2)	-0.65	1.07	-0.006	-3.41	0

Table 5.2 presents the descriptive statistics for the lead and lag vehicle relative to the subject vehicle. Relative speeds are defined as the speed of the lead (lag) vehicle less the speed of the subject vehicle. The table summarizes statistics of the lead and lag gaps (i.e. the gaps vehicle changed lanes into) both for the accepted gaps and for the entire dataset (both accepted and rejected gaps). Accepted lead gaps vary from 0.13 meters to 102.9 meters, with a mean of 9.92 meters. Accepted lag gaps vary from 0.48 meters to 172.9 meters. Statistics for the entire dataset are presented in parentheses.

Table 5.2: Statistics for the lead and lag vehicles of merging vehicles

Variable	Mean	Std Dev	Median	Minimum	Maximum
Lead Relative Speed (m/sec)	0.24 (-0.29)	1.26 (2.15)	0.24 (0.01)	-6.21 (-16.80)	5.60 (8.13)
Lead Gap (m)	9.92 (4.83)	9.01 (8.83)	7.57 (2.94)	0.13 (-19.43)	102.9 (160.6)
Lag Relative Speed (m/sec)	-0.55 (-0.41)	1.56 (2.15)	-0.51 (-0.15)	-10.98 (-14.25)	5.38 (18.09)
Lag Gap (m)	11.35 (5.25)	11.58 (8.85)	8.43 (3.39)	0.48 (-19.9)	172.9 (178.25)

As expected, the mean accepted gaps are larger than the mean gaps in the traffic stream for both the lead and lag gaps. Similarly, mean lead relative speeds in the accepted gaps are higher than those in the entire dataset and mean lag relative speeds in the accepted gaps are lower than those in the entire dataset. This implies that when a gap is accepted, the subject vehicle is traveling slower than the lead vehicle and faster than

the lag vehicle. The distributions of the speeds and spacing with respect to the lead and lag vehicles for the entire dataset are shown in Figures 5.10 and 5.11, respectively, and those for the accepted gaps are shown in Figures 5.12 and 5.13, respectively.
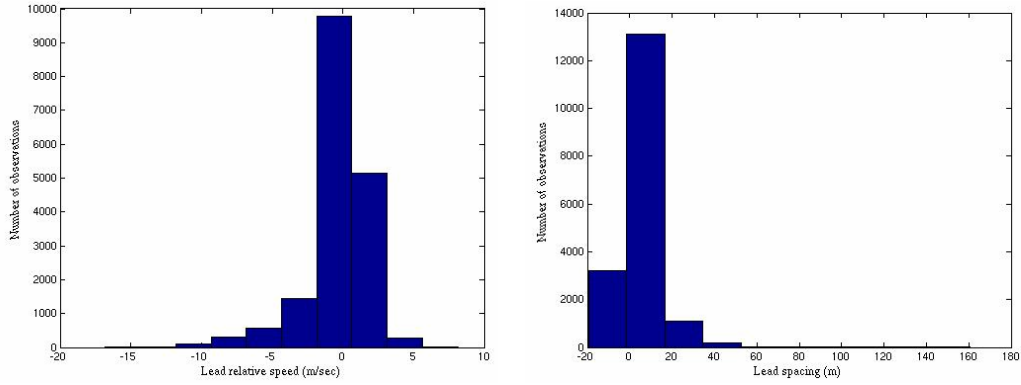
Figure 5.10: Distributions of lead relative speed and spacing in the full dataset
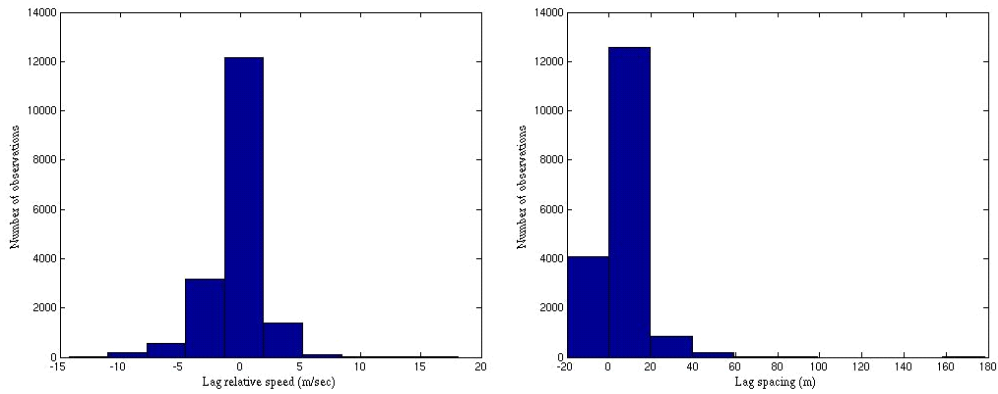
Figure 5.11: Distributions of lag relative speed and spacing in the full dataset
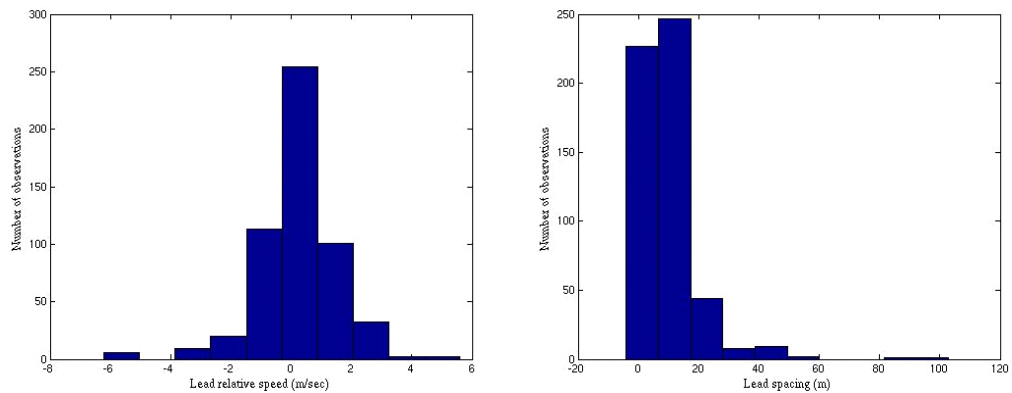
Figure 5.12: Distributions of lead relative speed and spacing for the accepted gaps
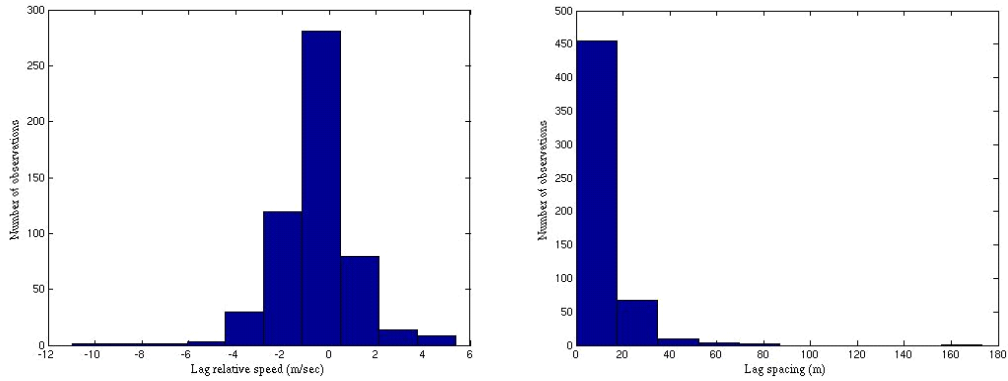
Figure 5.13: Distributions of lag relative speed and spacing for the accepted gaps

From the dataset, it was observed that more than 80 percent of the merges occur when the distance to the mandatory lane changing point, as defined by the imaginary lane boundary, is less than 100 m. Figure 5.14 shows the distribution of the number of merges with distance to the mandatory lane changing point in the section.
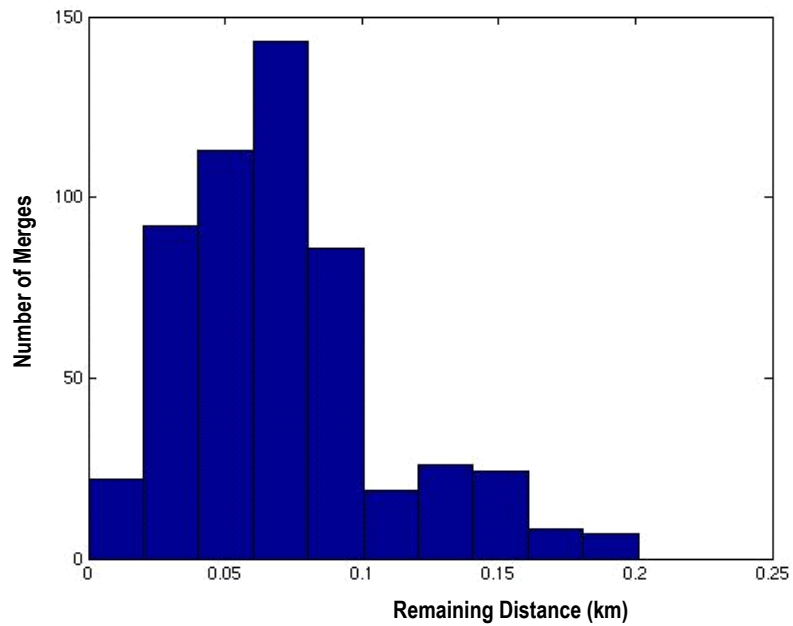
Figure 5.14: Distribution of number of merges with distance to MLC point

5.3.2 Likelihood of the Trajectory

All model parameters were estimated jointly using a maximum likelihood technique. The likelihood function that was maximized is presented in this section.

At any time t, an individual can be in one of the following states:

- Courtesy merging ($l_t = C$),

- Forced merging ($l_t = F$), or

- Normal lane changing ($l_t = M$).

The lane changing decisions of the driver depends on the state. The state of the driver at any instant depends on his/her previous state(s).

According to the first-order Markov assumption:

- The state at a given time period t depends only on the state at time (t-1) and action of all previous time periods (1: t-1).

- The lane action at a given time period t depends only on the state at time period t.

Further, in the merging data, the observation of a driver ends when he makes a lane change. That is, there is always a sequence of 'no changes' followed by a lane change in the last time step. Therefore, the fact that the driver is in state l_t at time t conditional that the previous state was l_{t-1} indicates the following:

- The lane action in the previous state l_{t-1} was 'no change' ($j_{t-1} = 0$) and

- There has been a transition to state l_t from state l_{t-1} at $(t-1)^{th}$ time step, where, $l_t, l_{t-1} \in M, C, F$.

The probability of being in state l_t is therefore the product of the probability of being in state l_{t-1} at time (t-1), and the joint probability of no change at the previous time period ($j_{t-1} = 0$) and probability of a transition from state l_{t-1} to state l_t at time (t-1).

The lane actions at time t (j_t) are conditional on the state at time t (l_t). As discussed in Section 5.2.3, many decision state sequences can lead to the same state at time t. At time t for driver n, the probability of observing a particular lane action j is the sum of

probabilities that he/she is observed to execute lane action j given that the selected merging plan is l, over all sequence of plans that could have led to plan l_t.

$$P_n(j_t \mid j_{1:t-1}, \upsilon_n, \tau_n) = \sum_{(l_1,\dots,l_t)} P_n(j_t \mid l_t, \upsilon_n) P_n(l_t, j_{t-1} \mid l_{t-1}, j_{1:t-2}, \upsilon_n, \tau_n)$$

$$= \sum_{(l_1,\dots,l_t)} P_n(j_t \mid l_t, \upsilon_n) P_n(l_t \mid l_{t-1}, j_{1:t-1}, \upsilon_n, \tau_n) P_n(j_{t-1} \mid l_{t-1}, \upsilon_n)$$

(5.27)

The probability of observing the entire trajectory of driver n can be calculated recursively and is given by the following equation:

$$P_n\left(j_1,\dots,j_T \mid \upsilon, \tau\right)$$

$$= \sum_{(l_1,\dots,l_{T_n})} P_n\left(j_{T_n} \mid l_{T_n}, \upsilon_n\right) P_n\left(l_{T_n}, j_{T_n-1} \mid l_{T_n-1}, j_{1:T_n-2}, \upsilon_n, \tau_n\right) \cdots P_n\left(l_3, j_2 \mid l_2, j_1, \upsilon_n, \tau_n\right) P_n\left(l_2, j_1 \mid l_1, \upsilon_n, \tau_n\right)$$

$$= \sum_{l_{T_n}} P_n\left(j_{T_n} \mid l_{T_n}, \upsilon_n\right) \sum_{l_{T_n-1}} P_n\left(l_{T_n}, j_{T_n-1} \mid l_{T_n-1}, j_{1:T_n-2}, \upsilon_n, \tau_n\right) \cdots \sum_{l_2} P_n\left(l_3, j_2 \mid l_2, j_1, \upsilon_n, \tau_n\right) P_n\left(l_2, j_1 \mid l_1, \upsilon_n, \tau_n\right)$$

(5.28)

$$= \sum_{l_{T_n}} P_n\left(j_{T_n} \mid l_{T_n}, \upsilon_n\right) \sum_{l_{T_n-1}} P_n\left(l_{T_n} \mid l_{T_n-1}, j_{1:T_n-1}, \upsilon_n, \tau_n\right) P_n\left(j_{T_n-1} \mid l_{T_n-1}, \upsilon_n\right)$$

$$\cdots \sum_{l_2} P_n\left(l_3 \mid l_2, j_2, \upsilon_n, \tau_n\right) P_n\left(j_2 \mid l_2, \upsilon_n\right) P_n\left(l_2 \mid l_1, j_1, \upsilon_n, \tau_n\right) P_n\left(j_1 \mid l_1, \upsilon_n\right)$$

$$l_1 = M; l_{1:t} \in M, C, F; \; j_{T_n} = 1; \; j_{1:T_n-1} = 0$$

Where, the state transition probabilities are given by Equations 5.13, 5.15 and 5.17 and the lane action probabilities are given by Equations 5.21 through 5.26.

The unconditional individual likelihood is given by:

$$\mathcal{L}_n = \iint_{\upsilon \; \tau} P_n(j_1,\dots,j_T \mid \upsilon, \tau) f(\upsilon) f(\tau) \, d\upsilon \, d\tau$$

(5.29)

Where,

    $f(\upsilon) = $ *standard normal probability density function*

    $f(\tau) = $ *probability density function of a doubly truncated normal distribution*

        with mean μ_τ and variance σ_τ^2

Assuming that the observations from different drivers are independent, the log-likelihood function for all N individuals observed is given by:

$$\mathcal{L} = \sum_{n=1}^{N} \ln(\mathcal{L}_n)$$

(5.30)

The maximum likelihood estimates of the model parameters are found by maximizing this function.

5.3.3 Estimation Results

All model parameters: the parameters of the gap acceptance models, the plan/state transition models and the agent effect are estimated simultaneously with detailed vehicle trajectory data using maximum likelihood estimation technique as described in the previous section. However, in order to simplify the presentation, estimation results for the various components of the model are presented and discussed separately. The presentation order follows the hierarchy of the hypothesized decision-making process: the normal gap acceptance model is presented first, followed by the initiation and execution of courtesy merge models, and initiation and execution of forced merge models.

The summary of estimation results is presented in Table 5.3. Table 5.4 presents the parameter estimates of the normal merge model, Tables 5.5 and 5.6 present the results of the courtesy merge model and Tables 5.7 and 5.8 present the results of the forced merge model.

Table 5.3: Estimation results of the merging model

Final log-likelihood	-1609.65
Initial log-likelihood	-13763.75
Number of cases	540
Number of observations	17352
Number of parameters	42
Adjusted rho-bar square	0.88

The state-dependent merging model is compared with a reduced form model with no latent mechanism (Lee 2006). The instantaneous model aims at capturing the normal, forced and courtesy behavior of drivers through a single gap acceptance level by including variables relevant to all three types of merges in a single critical gap function. The model structure is shown in Figure 5.15. The model is estimated with the same trajectory data.

The latent plan model is an extension of the single level model. The summary statistics of the estimation results for the two models, presented in Table 5.4, show an improvement in the fit of the model, even when accounting for the larger number of parameters in the latent model.
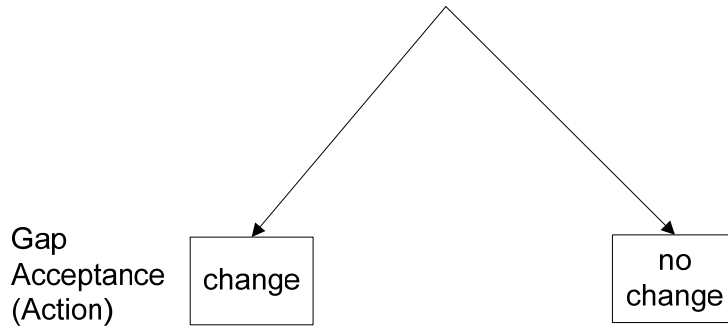
Figure 5.15: Framework of the single level merging model (Lee 2006)

Table 5.4: Model comparison

Statistic	Single Level (R)	Combined Merging (U)
Likelihood value	-1639.69	-1609.65
Number of parameters (k)	17	42
Akaike information criteria (AIC)	-1622.69	-1567.65
Adjusted rho-bar square ($\bar{\rho}^2$)	0.87	0.88

The model with explicit target lane choice has larger values in terms of both AIC and $\bar{\rho}^2$ (detailed in Chapter 4). This indicates that the inclusion of the latent plans in the decision framework results in an improved goodness-of-fit even after discounting for the increase in the number of parameters.

The detailed estimation results of the model components are presented below:

Execution of Normal Merge

In the hypothesized decision making process, the driver first evaluates the adjacent lead and lag gaps to decide whether or not to merge through normal gap acceptance. In order for the gap to be acceptable both the lead and lag gaps, must be acceptable.

The critical lead and lag gaps are functions of the relative speeds and accelerations of the adjacent vehicles and the remaining distance to the mandatory lane changing point. The estimated coefficients are presented in Table 5.5.

Table 5.5: Estimation results of the normal gap merging model

Normal Lead Gap		Parameter	t-stat
Variable		Parameter	t-stat
Normal lead constant		-0.230	-0.33
* Relative average speed (positive) (m/sec)		0.521	0.81
* Relative lead speed (negative) (m/sec)		-0.505	-3.13
* Remaining distance function	Distance to MLC point (10 m)	1.32	3.64
	Constant	0.420	0.89
	Heterogeneity coefficient, $\alpha^{RemDistLead}$	0.355	1.68
Standard deviation for normal lead gap, $\sigma_{M\,lead}$		3.42	9.67
Heterogeneity coefficient for normal lead gap, $\alpha^{M\,lead}$		-0.819	-3.12
Normal Lag Gap			
Normal lag constant		0.198	2.87
* Relative lag speed (positive) (m/sec)		0.208	1.78
* Relative lag speed (negative) (m/sec)		0.184	1.63
* Remaining distance function	Distance to MLC point (10 m)	0.239	5.09
	Constant	0.0242	0.03
	Heterogeneity coefficient, $\alpha^{RemDistLag}$	0.0180	0.03
* Lag acceleration (positive) (m/sec^2)		0.0545	0.61
Standard deviation for normal lag gap, $\sigma_{M\,lag}$		0.840	3.03
Heterogeneity coefficient for normal lag gap, $\alpha^{M\,lag}$		-0.0076	-0.01

* same coefficients in normal, courtesy and forced gap acceptance levels

The lead critical gap is a function of the average speed in the mainline relative to the subject vehicle's speed, the relative speed of the lead with respect to the subject and the remaining distance to the mandatory lane changing point and can be expressed as follows:

$$G_{nt}^{M\,lead} = \exp\left(\begin{array}{c} -0.230 + 0.521 V_{nt}' - 0.505 Min\left(0, \Delta V_{nt}^{lead}\right) + \dfrac{1.32}{1+\exp(0.420 + 0.355 \upsilon_n)} d_{nt} \\ -0.819 \upsilon_n + \varepsilon_{nt}^{M\,lead} \end{array} \right) \qquad (5.30)$$

Where,

$G_{nt}^{M\,lead}$ = critical lead gap for the normal gap acceptance level (m)

$V_{nt}^{'}$ =relative average speed factor (m/sec)

ΔV_{nt}^{lead} =relative speed of the lead vehicle with respect to the subject (m/sec)

d_{nt} =remaining distance to the mandatory lane changing point (10 m)

υ_n = individual-specific random effect

$\varepsilon_{nt}^{M\,lead}$ = random error term associated with normal lead gap : $\varepsilon_{nt}^{M\,lead}\ \sim N\left(0,3.83^2\right)$

The lag critical gap is a function of the subject vehicle speed relative to the lag vehicle, the remaining distance to the mandatory lane changing point and the acceleration of the lag vehicle. This can be expressed as follows:

$$G_{nt}^{M\,lag}\ =\exp\left(\begin{array}{c}0.198+0.208Max\left(0,\Delta V_{nt}^{lag}\right)+0.184Min\left(0,\Delta V_{nt}^{lag}\right)\\+\dfrac{0.239}{1+\exp(0.0242+0.018\upsilon_n)}d_{nt}+0.0545Max\left(0,a_{nt}^{lag}\right)-0.0076\upsilon_n+\varepsilon_{nt}^{M\,lag}\end{array}\right) \qquad (5.31)$$

Where,

$G_{nt}^{M\,lag}$ = critical lag gap for the normal gap acceptance level (m)

ΔV_{nt}^{lag} =relative speed of the lag vehicle with respect to the subject (m/sec)

d_{nt} =remaining distance to the mandatory lane changing point (10 m)

a_{nt}^{lag} =acceleration of the lag vehicle (m/sec^2)

υ_n = individual-specific random effect

$\varepsilon_{nt}^{M\,lag}$ = random error term associated with normal lead gap: $\varepsilon_{nt}^{M\,lag}\sim N\left(0,0.532^2\right)$

The lead critical gap increases with the average speed of the mainline. As the mainline average speed increases, the driver needs larger critical gaps to adjust the speed to the speed of the mainstream. However, critical gap does not increase linearly with increasing average speeds in the mainline (Figure 5.16), rather it increases as a diminishing function $\beta^{\,avg}V_{nt}^{'}$, where, $V_{nt}^{'}=\left(\dfrac{1}{1+\exp\left(-Max\left(0,\Delta V_{nt}^{avg}\right)\right)}\right)$, ΔV_{nt}^{avg} being the relative speed between the average mainline and the subject vehicle (m/sec).
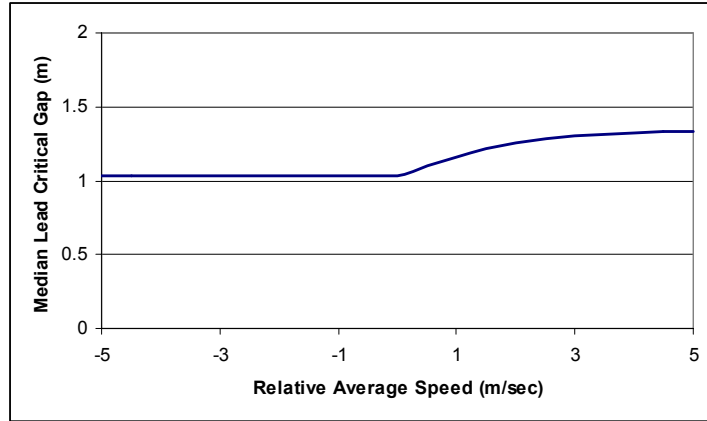
Figure 5.16: Lead critical gap as a function of relative average speed in the mainline

The lead critical gap is larger when the lead vehicle is moving slower than the subject since the driver perceives an increased risk when the lead is slowing down and he/she is getting closer to the lead vehicle (Figure 5.17).
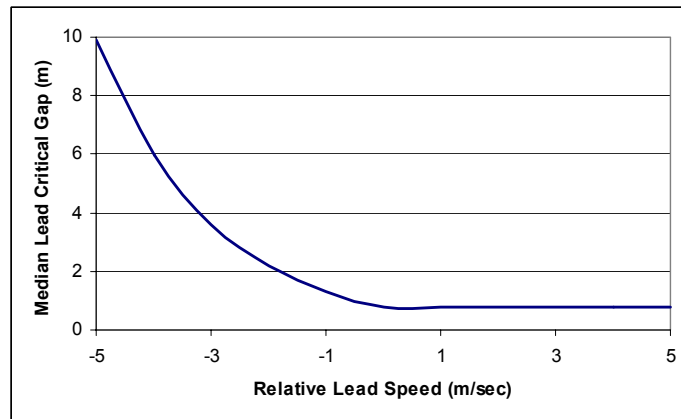
Figure 5.17: Lead critical gap as a function of relative lead speed

The lag critical gap increases with the relative lag speed: the faster the lag vehicle is relative to the subject, the larger the critical gap (Figure 5.18).

The lag critical gap increases as the acceleration of the lag vehicle increases (Figure 5.19), due to the higher perceived risk of merging into the mainstream when the lag vehicle is accelerating.
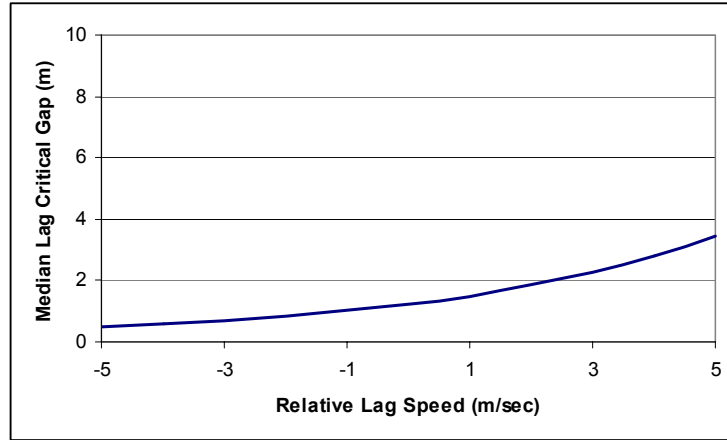
Figure 5.18: Lag critical gap as a function of relative lag speed
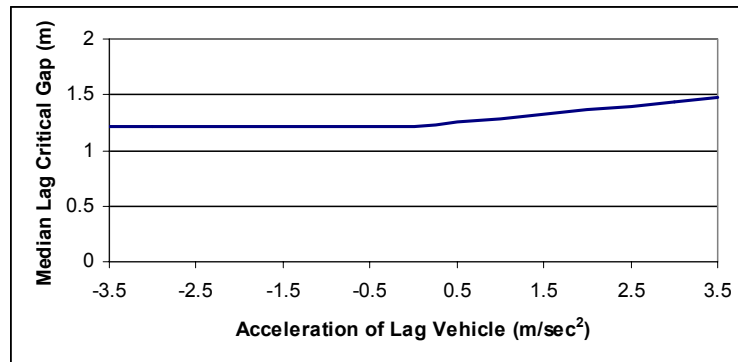
Figure 5.19: Lag critical gap as a function of lag vehicle acceleration

Both the lead and lag critical gaps decrease as the distance remaining to the mandatory lane changing point decreases. This is because as the driver approaches the point where the ramp ends, the urgency to make the merge increases and he/she is willing to accept lower gaps to merge. To capture drivers' heterogeneity, an individual-specific random term has been introduced in the coefficient of the remaining distance. Aggressive and timid drivers can thus have different critical gaps, the remaining distance being equal. The aggressiveness/timidity of the driver captures the heterogeneity among the driver population and is assumed to have a continuous distribution (truncated normal in this case) rather than discrete having a discrete class membership. All other variables having no effect, the lead and lag critical gaps as a function of remaining distance for aggressive drivers are much smaller than the gaps for timid drivers. Thus, aggressive drivers can find lead and lag gaps to be acceptable even when they are far from the MLC

point. On the other hand, timid drivers have large critical gaps till they reach the end of the ramp. The sensitivity of the lead and lag critical gaps as a function of the remaining distance according to the individual characteristics of the driver is shown in Figure 5.20 and Figure 5.21 respectively. As seen in Figure 5.20, the timid drivers have an unusually large critical lead gap till they are closer to the MLC point, implying that they do not consider lane changes at the beginning of the on-ramp. It may be noted that the sign of the unobserved driver characteristics is consistent for both gaps as well as other choice dimensions. The t-statistics for the linear part of the coefficient of remaining distance is found to be very significant both for lead and lag gaps.
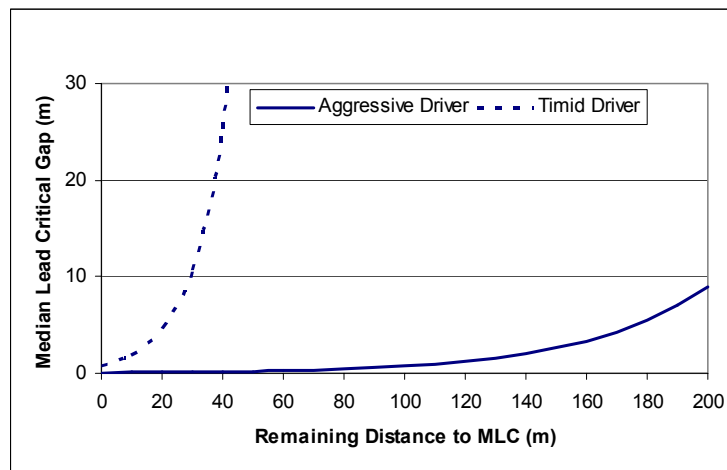
Figure 5.20: Lead critical gap as a function of remaining distance to MLC point
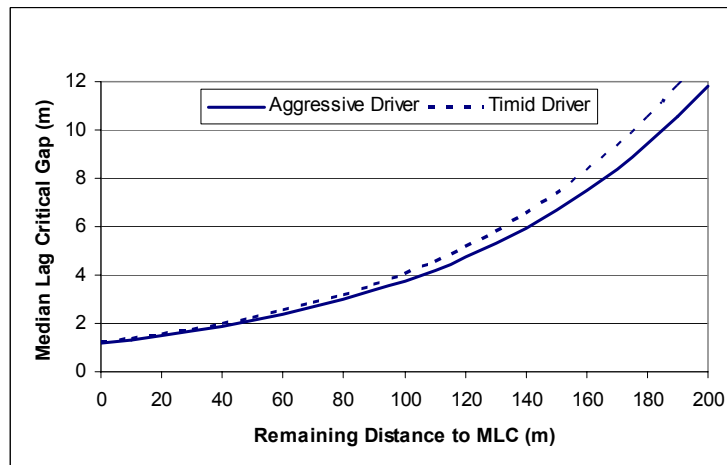
Figure 5.21: Lag critical gap as a function of remaining distance to MLC point

Estimated coefficients of the unobserved driver characteristics (υ_n) are negative for both the lead and lag critical gaps. This implies that an aggressive driver requires smaller gaps for lane changing compared with a timid driver.

Initiation and Execution of Courtesy Merge

If the available lead and lag gaps are not acceptable for normal merge, the merging driver evaluates the speed, acceleration and relative position of the through vehicles and tries to evaluate whether or not the lag driver is providing courtesy to him/her. The courtesy or discourtesy of the lag driver is reflected in the anticipated gap which is defined as the total gap after time τ_n (anticipation time):

$$\overline{G}_{nt}(\tau_n) = G_{nt}^{lead} + G_{nt}^{lag} + Y_n + \tau_n(V_{nt}^{lead} - V_{nt}^{lag}) + \frac{1}{2}\tau_n^2(a_{nt}^{lead} - a_{nt}^{lag}) \tag{5.32}$$

Where, for individual n at time t,

$\overline{G}_{nt} = anticipated\ gap,\ m$

$Y_n = length\ of\ the\ subject\ vehicle,\ m$

$G_{nt}^{lead}, G_{nt}^{lag} = available\ lead\ and\ lag\ spacing\ respectively,\ m$

$V_{nt}^{lead}, V_{nt}^{lag} = lead\ and\ lag\ speeds\ respectively,\ m/sec$

$a_{nt}^{lead}, a_{nt}^{lag} = lead\ and\ lag\ accelerations\ respectively,\ m/sec^2$

The anticipated gap is compared against the critical anticipated gap and if deemed acceptable, the merging driver perceives that he/she is receiving courtesy from the lag driver and initiates a courtesy merge. The anticipated gap is acceptable if it is larger than the corresponding critical anticipated gap. Critical gaps are assumed to be log-normally distributed (a better fit than other non-negative distributions). The mean of the distribution is a function of explanatory variables: the relative lag speed, remaining distance, and density of the traffic stream. The estimated parameters are presented in Table 5.6.

Table 5.6: Estimation results of the initiate courtesy model

Initiate Courtesy Merge		Parameter	t-stat
Variable		Parameter	t-stat
Anticipated gap constant		1.82	1.00
Relative average speed (positive) (m/sec)		1.82	2.13
Relative lead speed (m/sec)		-0.153	-0.97
Remaining distance function	Distance to MLC point (10 m)	0.244	1.50
	Constant	0.449	0.49
	Heterogeneity coefficient, $\alpha^{RemDistA}$	0.360	0.18
Standard deviation for anticipated gap (σ_A)		0.0106	0.07
Heterogeneity coefficient for anticipated gap (α^A)		-0.231	-1.90
Mean of anticipation time (μ_τ)		1.87	9.51
Standard deviation of anticipation time (σ_τ)		1.44	17.71

The estimated functional form of the critical anticipated gap is given by:

$$G_{nt}^A = \exp\left(\begin{array}{l} 1.82 + 1.82 Max\left(0, \Delta V_{nt}^{lag}\right) - 0.153\rho_{nt} + \dfrac{0.244}{1+\exp(0.449 + 0.360\upsilon_n)}d_{nt} \\ -0.231\upsilon_n + \varepsilon_{nt}^A \end{array} \right)$$

(5.33)

Where,

G_{nt}^A = critical anticipated gap for the initiating courtesy merge (m);

ΔV_{nt}^{lag} = relative speed of the lag vehicle with respect to the subject (m/sec);

d_{nt} = remaining distance to the mandatory lane changing point (10 m);

ρ_{nt} = density in the rightmost lane of the mainline (veh/10 m); and

υ_n = unobserved driver characteristics.

ε_{nt}^A = random error terms $\varepsilon_{nt}^A \sim N\left(0, 0.0106^2\right)$

Similar to normal critical gaps, the critical anticipated gap is higher at higher lag speeds. It decreases as the remaining distance decreases and it is smaller for aggressive drivers than timid drivers. Courtesy yielding/merging more commonly occurs in dense traffic conditions and hence the probability of merging through courtesy increases with the density of mainline traffic. The critical anticipated gap therefore decreases with

density of traffic in the rightmost mainline lane. Median critical anticipated gap as a function of density is presented in Figure 5.22.
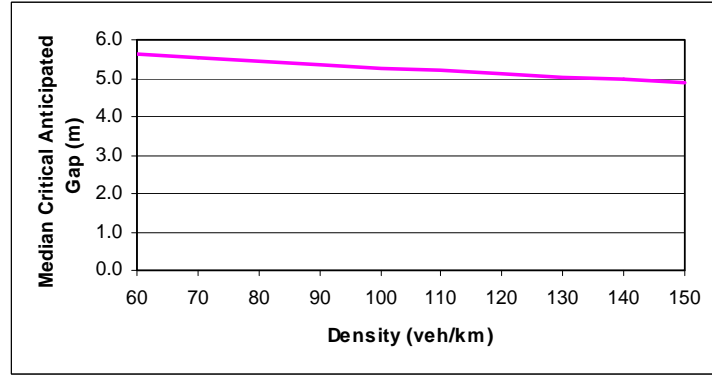
Figure 5.22: Median critical anticipated gap as a function of density in target lane

On initiating a courtesy merge, the driver decides whether to complete the merge by accepting or not the available gap based on the respective lead and lag critical gaps. For identification purposes, except for the constant and the unobserved driver characteristics, the coefficients of variables in these levels are restricted to be the same as for the normal gap acceptance level (Table 5.7).

The estimated functional form of the lead and lag critical gaps for courtesy can be expressed by the following equations:

$$
G_{nt}^{Clead} = \exp\left(\begin{array}{l} -0.582 + 0.521V_{nt}^{'} - 0.505Min\left(0, \Delta V_{nt}^{lead}\right) + \dfrac{1.32}{1+\exp(0.420+0.355\upsilon_n)}d_{nt} \\ -0.054\upsilon_n + \varepsilon_{nt}^{Clead} \end{array}\right) \quad (5.34)
$$

$$
\varepsilon_{nt}^{Clead} \sim N\left(0, 0.0109^2\right)
$$

$$
G_{nt}^{Clag} = \exp\left(\begin{array}{l} -1.23 + 0.208Max\left(0, \Delta V_{nt}^{lag}\right) + 0.184Min\left(0, \Delta V_{nt}^{lag}\right) \\ + \dfrac{0.439}{1+\exp(0.0242+0.00018\upsilon_n)}d_{nt} + 0.0545Max\left(0, a_{nt}^{lag}\right) - 0.554\upsilon_n + \varepsilon_{nt}^{Clag} \end{array}\right) \quad (5.35)
$$

$$
\varepsilon_{nt}^{Clag} \sim N\left(0, 0.554^2\right)
$$

Where,

$G_{nt}^{Clead}, G_{nt}^{C\,lag} =$ *lead and lag critical gaps for the courtesy gap acceptance level respectively*

$\varepsilon_{nt}^{Clead}, \varepsilon_{nt}^{Clag} =$ *random error terms*

Table 5.7: Estimation results of the courtesy gap acceptance model

Courtesy Lead Gap			
Variable		Parameter	t-stat
Courtesy lead constant		-0.582	-0.20
*Relative average speed (positive) (m/sec)		0.521	0.81
*Relative lead speed (negative) (m/sec)		-0.505	-3.13
*Remaining distance function	Distance to MLC point (10 m)	1.32	3.64
	Constant	0.420	0.89
	Heterogeneity coefficient, $\alpha^{RemDistLead}$	0.355	1.68
Standard deviation for courtesy lead gap, σ_{Clead}		0.0109	0.08
Heterogeneity term for courtesy lead gap, α^{Clead}		-0.0540	-0.03
Courtesy Lag Gap			
Courtesy lag constant		-1.23	-0.07
*Relative lag speed (positive) (m/sec)		0.208	1.78
*Relative lag speed (negative) (m/sec)		0.184	1.63
*Remaining distance function	Distance to MLC point (10 m)	0.439	5.09
	Constant	0.0242	0.03
	Heterogeneity coefficient, $\alpha^{RemDistLag}$	0.000180	0.03
*Lag acceleration (positive) (m/sec^2)		0.0545	0.61
Standard deviation for courtesy lag gap, σ_{Clag}		0.554	0.05
Heterogeneity term for courtesy lag gap, α^{Clag}		-0.0226	-0.04

* same coefficients in normal, courtesy and forced gap acceptance levels

The estimation results show that all other things held constant, a driver is more willing to accept smaller lead and lag gaps when he/she is in the courtesy merging state than in normal or forced merging states. This is intuitive since in case of courtesy merging, the lag vehicle is slowing down and therefore, a smaller buffer space is sufficient.

Initiation and Execution of Forced Merge

If the driver perceives that a normal lane change is not possible and there is no courtesy yielding of the lag driver (anticipated gap is not acceptable), the driver chooses whether or not to initiate a forced merge. As described in Section 5.2.1, this is modeled as a binary logit model.

Table 5.8: Estimation results of the initiate forced merge model

Initiate Forced Merge		
Variable	Parameter	t-stat
Initiate force constant	-6.41	-4.63
Heavy lag vehicle dummy	-1.25	-0.63
Heterogeneity term for initiated forced merge(α^F)	5.43	3.26

The decision to initiate a forced merge was found to be dependent on the aggressiveness of the driver and whether the lag vehicle in the mainline is a heavy vehicle or not. In particular, the coefficient of aggressiveness has a significant impact on the decision to initiate a forced merge. If the lag is a heavy vehicle, the probability of initiating a forced merge decreases, as the driver perceives a higher risk in undertaking such a maneuver. The variable remaining distance (urgency of the merge) and delay (impatience) of the driver were assumed to impact forced merge, but the estimated coefficients of these two variables did not have the expected signs. This may be due to the fact that in the estimation dataset, many of the forced merges actually occurred in the beginning of the section as opposed to the end.

The probability of initiating a forced merge is given by the following equation:

$$P_{nt}^F = \frac{1}{1+\exp\left(6.41+1.25\delta_{nt}^{hv}-5.43\upsilon_n\right)}$$

(5.36)

Where,

δ_{nt}^{hv} = *heavy lag vehicle dummy, 1 if the lag vehicle is a heavy vehicle, 0 otherwise*

Similar to courtesy merging, on initiating a forced merge, the driver decides whether to complete the merge by accepting the available gap or not based on the respective lead and lag critical gaps. For identification purposes, except for the constant and the

unobserved driver characteristics, the coefficients of variables in these levels are restricted to be the same as for the normal gap acceptance level (Table 5.9).

Table 5.9: Estimation results of the forced merging model

Forced Lead Gap		Parameter	t-stat
Variable		Parameter	t-stat
Forced lead constant		3.11	2.11
*Relative average speed (positive) (m/sec)		0.521	0.81
*Relative lead speed (m/sec)		-0.505	-3.13
*Remaining distance function	Distance to MLC point (10 m)	1.32	3.64
	Constant	0.420	0.89
	Heterogeneity coefficient, $\alpha^{RemDistLead}$	0.355	1.68
Standard deviation for forced lead gap, σ_{Flead}		7.95	5.82
Heterogeneity term for forced lead gap, α^{Flead}		-0.0401	-0.07
Forced Lag Gap			
Forced lag constant		-2.53	-3.42
*Relative lag speed (positive) (m/sec)		0.208	1.78
*Relative lag speed (negative) (m/sec)		0.184	1.63
*Remaining distance function	Distance to MLC point (10 m)	0.439	5.09
	Constant	0.0242	0.03
	Heterogeneity coefficient, $\alpha^{RemDistLag}$	0.000180	0.03
*Lag acceleration (positive) (m/sec^2)		0.0545	0.61
Standard deviation for forced lag gap, σ_{Flag}		0.465	2.49
Heterogeneity term for forced lag gap, α^{Flag}		-0.0239	-0.19

* same coefficients in normal, courtesy and forced gap acceptance levels

The estimated functional form of the lead and lag critical gaps for courtesy can be expressed by the following equations:

$$G_{nt}^{Flead} = \exp\left(\begin{array}{l} 3.11 + 0.521V_{nt}' - 0.505 Min\left(0, \Delta V_{nt}^{lead}\right) + \dfrac{1.32}{1+\exp(0.420+0.355\upsilon_n)}d_{nt} \\ -0.0401\upsilon_n + \varepsilon_{nt}^{Flead} \end{array} \right) \quad (5.37)$$

$$\varepsilon_{nt}^{Flead} \sim N\left(0, 7.95^2\right)$$

$$G_{nt}^{Flag} = \exp\left(\begin{array}{c} -2.53 + 0.208 Max\left(0, \Delta V_{nt}^{lag}\right) + 0.184 Min\left(0, \Delta V_{nt}^{lag}\right) \\ + \dfrac{0.439}{1+\exp(0.0242+0.00018\upsilon_n)} d_{nt} + 0.0545 Max\left(0, a_{nt}^{lag}\right) - 0.0239\upsilon_n + \varepsilon_{nt}^{Flag} \end{array} \right) \qquad (5.38)$$

$$\varepsilon_{nt}^{Flag} \sim N\left(0, 0.465^2\right)$$

Where,

$G_{nt}^{Flead}, G_{nt}^{Flag}$ = *lead and lag critical gaps for the forced gap acceptance level respectively*

ε_{nt}^{Flead} *and* ε_{nt}^{Flag} = *random error terms*

The constant term for the lag critical gap for forced merging is smaller than for the normal and courtesy merges. However, the lead critical gap for the forced merging case is found to be larger than the case of the normal merge. This reflects the fact that once the driver has initiated a forced merge (pushed the front bumper establishing the right of way), the lead gap plays a dominant role in the completion of the merge. Once initiated, the forced merge is completed only when the lead gap is sufficiently large since the maneuver involves significantly higher risk than for normal gap acceptance.

Distribution of Anticipation Time

The anticipation time is assumed to follow a doubly truncated normal distribution. Estimation results indicated that it is normally distributed within 0 to 4 sec.[6] The estimated distribution of anticipation time is

$$f(\tau_n) = \begin{cases} \dfrac{1}{0.833}\phi\left(\dfrac{\tau_n - 1.87}{1.44}\right) & \text{if } 0 \leq \tau_n \leq 4 \\ 0 & \text{otherwise} \end{cases} \qquad (5.39)$$

---

[6] Different values between 0 to 6 sec were tested as the upper limit of anticipation time and the selected value (4 sec) provided the best goodness-of-fit.
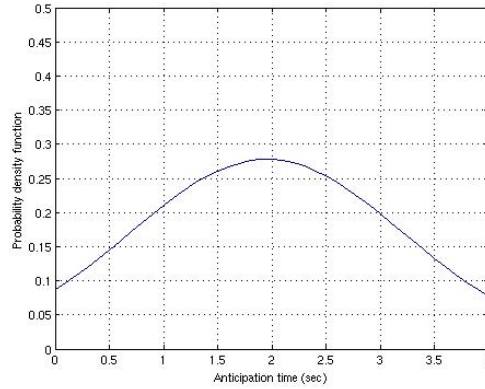
Figure 5.23: Distribution of anticipation time

5.4 Model Validation

Both the latent plan model and the single level model were implemented in the microscopic traffic simulator MITSIMLab (Yang and Koutsopoulos 1996) for aggregate validation. In the validation process, part of the aggregate data was first used to calibrate the overall behavioral parameters of MITSIMLab. The calibrated MITSIMLab outputs were then compared with the remaining data.

5.4.1 Data

U.S. 101 dataset was collected on a 2100 feet (640 meter) southbound section of U.S. Highway 101, in Los Angeles (California) with five mainline and one auxiliary lane connecting to the Ventura on-ramp and the Cahuenga off-ramp (Figure 5.24).
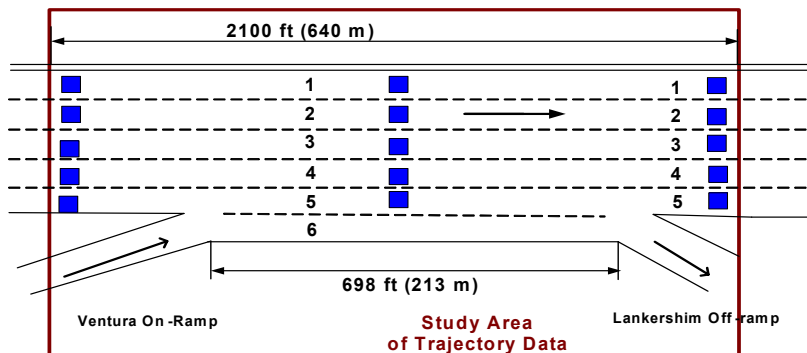
Figure 5.24: Validation data collection site

This site has an auxiliary lane after the onramp, which was not the case for the I-80 site. 45 minutes of data (7:50 a.m.-8:35 a.m.) were available. Based on the trajectory

data, 'synthetic' sensor data was created in three locations (Figure 5.24).   This sensor data replicated counts and speeds (aggregated over every five minutes) that would have been recorded by sensors located in these locations.

5.4.2 Aggregate Calibration

The aggregate calibration problem can be formulated as an optimization problem which seeks to minimize a function of the deviation of the simulated traffic measurements from the observed measurements (Toledo *et al.* 2004). The number of behavioral parameters in the simulation model is very large and it is not feasible to calibrate all of them. Based on previous experience and sensitivity test results, the following parameters of the combined model were selected for calibration:

- Acceleration and deceleration constants
- Desired speed mean and sigma
- Intercepts (constants) and standard deviations (sigma's) of normal critical gap
- Anticipation time mean and sigma
- Probability of yielding of the mainline vehicle
- Constant for forced merging
- Individual-specific random errors in the remaining distance terms

The fit of the combined model to the calibration data are presented in Table 5.10.

Table 5.10: Calibration results of the combined model

Lane-specific Counts			
	Before Calibration	After Calibration	Improvement
RMSE (vehicles/15 mins)	11.05	7.18	35.02%
RMSPE (%)	10.47	5.11	51.19%
Lane-specific Speeds			
	Before Calibration	After Calibration	Improvement
RMSE (m/s)	8.22	5.59	32.00%
RMSPE (%)	32.34	20.09	37.88%

5.4.3 Aggregate Validation

The validation process involved data that was not used for calibration. A comparison of the following simulated and observed statistics was conducted:

- Lane-specific point speeds for the remaining 15 mins (8:20-8:35 am)
- Lane-specific flows for the remaining 15 mins (8:20-8:35 am)
- Distribution of location of merges (summarized from aggregate trajectory data)

The measures of performance of the combined model were compared against the performance of the 'reduced form' single level model. The OD flows used for this step were calculated directly from the trajectory data.

Lane-specific Sensor Speeds

A separate set of lane-specific speed measurements from sensors (not used for calibration) was used for validation purpose. The comparisons of the goodness-of-fit measures are presented in Table 5.12. As is evident from the RMSE and RMSPE, the performance of the models improved with complexity of the model: the latent plan merging model performed better than the single level model.

Table 5.11: Comparison of lane-specific speeds

	Single Level Model	Combined Merging Model	Improvement
RMSE (m/s)	9.16	8.82	3.71 %
RMSPE (%)	24.27	22.26	8.28 %

Lane-specific Sensor Counts

The simulated lane-specific sensor counts of the latent plan merging model were compared against the actual observations and the simulated counts of the single level model. As observed in Table 5.13, the combined merging model had a significantly better match with the actual observations.

Table 5.12: Comparison of lane-specific counts

	Single Level Model	Combined Merging Model	Improvement
RMSE (vehicles/5 mins)	19.18	13.22	31.07 %
RMSPE (%)	12.18	7.52	38.26 %

Location of Merge

The simulated locations of merges were compared against the observed locations. The latent plan model had a significantly better prediction of the location of merges than the single level model (see Figure 5.17). In particular, the simpler model tends to over predict merges occurring toward the end of the auxiliary lane since courtesy and forced merging plans are not considered explicitly in this model.
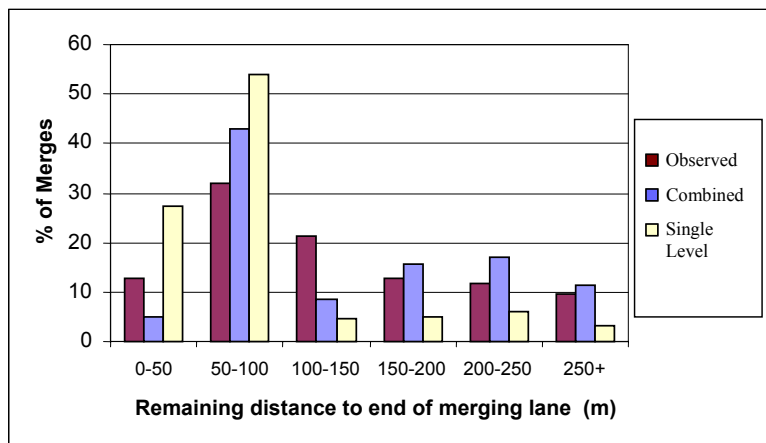
Figure 5.25: Comparison of merge locations

5.5   Summary

The detailed structure, estimation results and validation results of a latent plan based combined merging model has been presented in this chapter. The model integrates normal, cooperative and forced merging types into a single framework. Parameters of the models are estimated with detailed vehicle trajectory data collected from I-80, in California. The effect of unobserved driver/vehicle characteristics on the lane changing process was captured by driver-specific random terms included in different model components.

Important explanatory variables were found to be the urgency of the driver (e.g. distance remaining to the end of the merging section), the relation of the merging vehicle with neighboring vehicles (e.g. lead and lag speed and position etc.), the traffic conditions (e.g. average speed and density) and driver heterogeneity. Statistical comparisons of estimation results indicate that the estimated combined model has significantly better goodness-of-fit compared to a reduced form simpler model that does not explicitly consider courtesy and forced merging. This was supported by a validation case study where the performance of the two models was compared in a different network setting. The combined model performed significantly better than the simpler models across all measures.

In the current model, the latent plans were assumed to include only lateral decisions involved with the merging decision. The extent of the improvements obtained with the enhancements in the merging models presented in this application indicates that further advances in merging models may lead to improvements in their ability to replicate observed vehicle trajectories. In particular, including target gap choice and acceleration in the model is a possible future direction of research. As observed in the validation results, the combined model was better at replicating counts than speeds. Inclusion of target gap choice and speed adjustment to reach a targeted gap in the decision framework of the merging driver may improve the match of speed as well as ensuring a better prediction of merge locations.

Chapter 6

Lane Selection on Urban Arterials

The latent plans involving the lane selections of drivers on urban arterials are investigated in this chapter. The specific models discussed here are the lane choice model for urban intersections and the lane changing model for arterial mainline sections. The general structure of both models is the same as presented in the previous chapters: latent plan followed by observed actions. However, because of differences in geometric and operational characteristics, the latent plans involving lane selection of drivers on urban mainline and side streets are likely to be quite different than those of freeway mainline (discussed in Chapter 4) and on-ramps (discussed in Chapter 5).

The chapter is organized as follows: the background of the research is presented in Section 6.1. Description of the estimation data and the details of the model estimations are presented in Section 6.2: the intersection lane choice model in Section 6.2.2 and the mainline lane changing model in Section 6.2.3. Each section includes the model structures, the likelihood formulations and the estimation results of each of the two models. The chapter concludes with the aggregate calibration and validation results within the microscopic traffic simulator MITSIMLab and a summary of the findings.[7]

6.1   Background

Travelers on arterial networks face special challenges regarding lane positioning strategies. Arterial corridors have a set of varied driving activities that differ by lane and location. These activities encompass trip destination activities (parking, entering transit

---

[7] The model presented in this chapter has been developed as part of the NGSIM program of FHWA. The results presented in this chapter have been reported in Choudhury *et al.* (2007). A simplified version of the mainline lane changing model has been developed by Ramanujam (2007).

stops, right turns, left turns etc.), trip origination activities (exiting a parking spot, exiting transit stops etc.), and complex routing behaviors (permissive left turns, pedestrian-impeded right turns etc.). Drivers familiar with the network may be aware of these activities and be mindful about how they vary by lane and location. These drivers often make appropriate tactical lane positioning decisions to minimize their travel times and driving efforts on these complex facilities. The familiarity and planning ability of the drivers that is: how far they 'look-ahead' or 'plan-ahead' affect their tactical plans and thus impact their driving decisions.

Due to situational constraints, immediate execution of the tactical lane selection plan may not be possible. For example, at a particular instant, conflicts with other vehicles can delay movement to the target lane. Further, changes in circumstances may lead to changes in the tactical plan: a long queue build-up in the chosen target lane for example can lead to amendment to the original target.   The chosen target lanes are thus unobserved and only the immediate choice of lanes is observed.

The lane positioning decisions generally manifest themselves inside lane changing models in existing simulation systems. Lane changing models are often generalized between freeway and arterial facilities. On arterial networks, existing models rely on standard lane changing logic to determine vehicle positioning behavior. Some models address the pre-positioning of drivers for path-plan considerations using rule based lane changing models (Jin *et al.* 1999, Wei *et al.* 2000). But the complexity of the tactical plans behind the immediate decisions of the drivers, as well as the heterogeneity in their planning behaviors, is ignored in the existing models. Moreover, none of the arterial lane selection models involve rigorous statistical estimation using detailed traffic data. These weaknesses of the existing models often lead to unrealistic spillbacks and uneven queue distributions across lanes. This has been also reflected in the findings of the NGSIM study on Identification and Prioritization of Core Algorithm Categories, where development of arterial lane selection model has been identified as the most important research area by both model developers and users (Alexiadis *et al.* 2004).

Lane selections in urban arterials include lane changes within the arterial mainline sections as well as lane choices at intersections. The lane changes in the arterial mainline sections involve repeated decisions of drivers while intersection lane choices are

intermittent decisions that are evaluated only when the driver is turning at intersections (since changing lanes within the intersection are not permissible). Further, the nature of conflict with other vehicles while turning at an intersection is distinctly different from that of lane changing in a mainline section. These lead to differences in the detailed frameworks of the intersection lane choice and mainline lane changing models in this study.

A driver turning at a signalized intersection is likely to choose the lane that he/she perceives to be the best and plans to move to that lane. However, because of conflicts with other vehicles having the right of way, it may not be possible to execute the plan immediately. The immediate lane choice of the driver, that is the lane where he/she is observed just after turning, thus may not be the same as the originally targeted lane. The plan of the driver is thus unobserved and only the immediate lane selection is observed. The intersection lane selection is therefore a two level decision:

- Choice of target lane (plan)
- Choice of immediate lane (action)

In an unsignalized intersection, the choice may involve additional levels like gap acceptance, target gap selection and/or decision whether or not to move towards the target lane using alternate gap acceptance tactics (e.g. by courtesy of another driver or by forcing in).

Once the driver gets on the main arterial, he/she is likely to have a latent plan based lane changing decision structure similar as in the freeway mainline: target lane selection followed by gap acceptance to reach the target lane. However, in urban arterials where average speeds and headways are significantly lower than the freeway, duration of the lane changing maneuver may be longer than that of a freeway. Drivers trying to reach their target lane therefore are not instantaneously observed to complete the lane change in the direction of the target lane even if an acceptable adjacent gap is available. Rather, a lane change is observed in the direction of the chosen target lane in presence of an acceptable adjacent gap when the execution of the lane change has been completed (center point of the vehicle has passed the lane boundary).

The simplest lane changing maneuver of drivers in the arterial mainline can thus be defined as a three stage decision:

- Choice of target lane (plan)
- Decision to accept available gaps (plan)
- Execution of the lane change (action)

The components of the chosen plan (target lane selection and gap acceptance) are latent or unobserved and only the completion of the execution of the lane change is observed. It may be noted that in congested urban arterials, the plan may include additional levels like target gap selection and choice of lane changing tactics in the decision framework.

As mentioned, in an urban arterial with closely spaced turns, the plan-ahead distance of the driver generally has a strong influence on the lane selection. The tactical plans of the driver are affected by influencing factors within the plan-ahead distance of the driver. For a driver familiar with the network and turning in a subsequent section, this implies that he/she does not consider the path-plan in lane selection until the desired turn is within the plan-ahead distance. It may also imply that the driver does not look beyond the plan-ahead distance while considering lane-specific variables such as average speeds, density and queue lengths. The plan-ahead distance of the driver is expected to vary among the driver population and can depend on different factors such as personal traits, network familiarity, congestion level etc. While the continuous plan-ahead distance is more appropriate for lane changing scenarios where the decisions are evaluated continuously, a discrete approach (where plan-ahead distances are multiples of section lengths) is more relevant in case of intersection lane choice since the decisions are taken intermittently at distinct points in the network.

6.2 Model Estimation

6.2.1 Estimation Data

Study Area

The two arterial models in discussion: the intersection lane choice and the within section lane change model are estimated from data collected from Lankershim Boulevard

in Los Angeles, California. Vehicle trajectory data was collected in 2005 as part of the FHWA's NGSIM project on a segment of the arterial located near the intersection with US highway 101 (Hollywood Freeway) (Figure 6.1).

Figure 6.1: Lankershim Boulevard arterial section

The study site is approximately 1600 feet (488 m) in length. It consists of four signalized intersections, and three to four through lanes in each direction in each section. Five video cameras were used to collect the trajectory data for a 22 minute period (8:28 am to 9:00 am). These cameras were mounted on top of a 36-story building, 10 University Plaza, located adjacent to the US 101 and Lankershim Boulevard interchange.
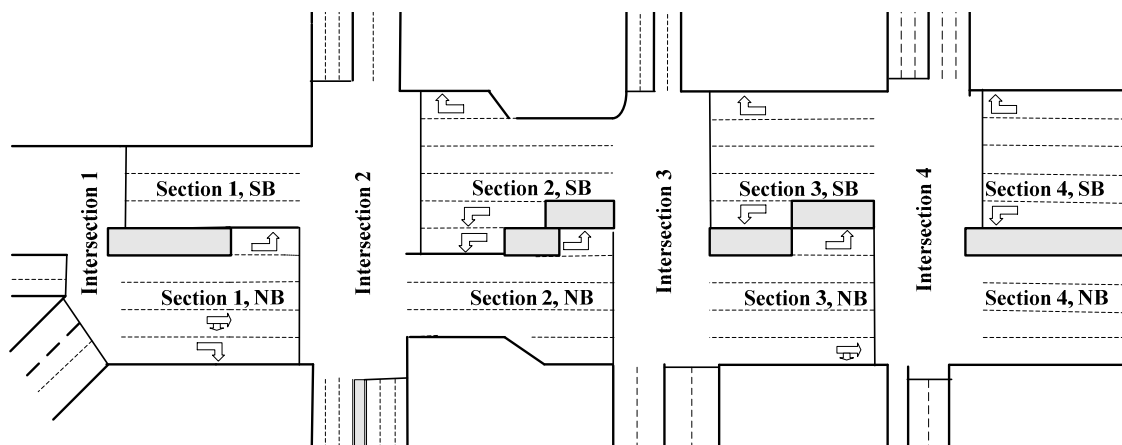
Figure 6.2: A schematic representation of the arterial stretch

(not in scale)

Figure 6.2 shows a schematic of the arterial segment constituting the study area. It also provides details regarding the reference indices used for demarcating the

intersections and sections. Lane numbering is assigned starting from the left most lane (i.e. in a four lane section, the rightmost lane index is 4). Almost every section has exclusive turning bays in the approach leading to the intersection.

The trajectory data have been split into two parts: one part containing observations of vehicles between each set of intersections (2016 vehicles), and the other part containing observations of vehicles in the vicinity of the intersections including the side streets. The lane changing model for the mainline was estimated using the first part of the data. The intersection lane selection model was based on observations of side street vehicles entering the main arterial at the four intersections (703 vehicles). Intersection lane choices of vehicles turning to side streets from the main arterial were excluded because of lack of information of the downstream conditions in the side streets.

For estimation purposes, the dataset with the mainline section observations was sampled randomly at the rate of 1 per every 5 vehicles, with the objective of establishing a computationally tractable representative dataset for arterial lane changing behavior. Sampling was also applied in the time dimension, at a rate of 1 per every 10 time instances of each sampled vehicle. As the original dataset had a 0.1 second time resolution, this sampling step converted the information to a 1 second resolution. Non-conforming and erroneous observations (e.g. through vehicles positioned in the turning bays, vehicles making turns from the wrong lanes etc.) were not used for estimation.

For the intersection lane choice, the variables were generated using the last observations of each vehicle in the side street before entering the intersections (the decision point for intersection lane choice). The first observations of these vehicles after entering the main arterial section indicated their observed choices.

Characteristics of Estimation Dataset

The vehicle trajectory data of the various drivers in the study area and the speeds and accelerations derived from these trajectories were used to generate the required variables.

The dataset used for estimating the intersection lane choice model included 703 observations (1 observation per vehicle): 629 northbound and 74 southbound. The vehicles are mostly passenger cars with only a small percentage (3.5%) of trucks and buses present. Of these vehicles 269 (38.1%) turn into the closest receiving lanes 435

(61.9%) later change to different lanes within the section. The majority of the entering vehicles are observed for more than one sections (80.2%) with more than half (55.9%) observed for more than two sections. However, most of the vehicles observed for more than two sections do not make any turn within the observed data collection area. The distribution of directions is presented in Figure 6.3.

Figure 6.3: Distribution of directions

The dataset of mainline vehicles within section consists of a total of 400 vehicles, of which 160 are northbound and 240 are southbound. The average vehicle observation duration is 51.3 seconds, with the maximum duration of observation being 170 seconds. Out of the 400 vehicles in the sampled dataset, 150 vehicles (around 37.5%) exited from the arterial within the study area, i.e. their destination is a side street at one of the four intersections within the study area.

The arterial sections are mostly 3-lane and 4-lane roadways, with exclusive turning bays widening the section at the approach to every downstream intersection. For data analysis and estimation purposes, lanes have been categorized on the basis of permitted vehicular movements. Statistics on the relevant aggregate lane-specific variables are presented in Table 6.1. It may be noted that there were no shared through and left turn lane in the data collection site.

Table 6.1: Aggregate lane-specific statistics

	Through lane	Shared through and right turn lane	Right turn bay	Left turn bay	Extra turn bay
Average Speed (m/s)	10.32	8.67	18.43	13.93	6.50
Average Queue Length (m)	1.07	1.93	0.18	1.44	2.08
Max Queue Length (m)	15.0	12.0	7.0	18.0	11.0

The presence of turning vehicles, and the conflicts arising due to their movements in conjunction with through vehicles, provides a reasonable explanation for the low average speeds observed in both through and turning lanes. The maximum queue length values are observed during red intervals at traffic signals. The presence of the exclusive turn bays and associated restrictions on turning vehicles has been considered while generating from trajectory data the explanatory variables for the lane changing model.

In the sampled dataset, there are a total of 249 lane changes observed. Of these, 104 (41.8%) are made by turning vehicles, i.e., those exiting the arterial within the observed stretch. A portion of these turning vehicles (19% out of the 104 turning vehicles) change lanes before reaching the last section. It may be noted that the data collection site had a mix of local commuters (familiar drivers) and tourists (unfamiliar drivers).

For vehicles going through the entire arterial stretch (also termed the through vehicles in the subsequent discussion), the distribution of lane changes over different sections is given in Table 6.2:

Table 6.2: Distribution of locations of lane change points for through vehicles

	Northbound	Southbound	Total
Section 2	13	13	26
Section 3	56	16	72
Section 4	26	21	47
Total	95	50	145

In the model developed for this dataset, lead or lag vehicles are defined as the closest vehicles in the corresponding adjacent lanes within the current section of the subject vehicle (Figure 6.4). The lead gap is the clear spacing between the rear of the lead vehicle and the front of the subject vehicle. Similarly, the lag gap is the clear spacing between the

rear of the subject vehicle and the front of the lag vehicle. One, or both, of these gaps may be negative if the vehicles overlap.
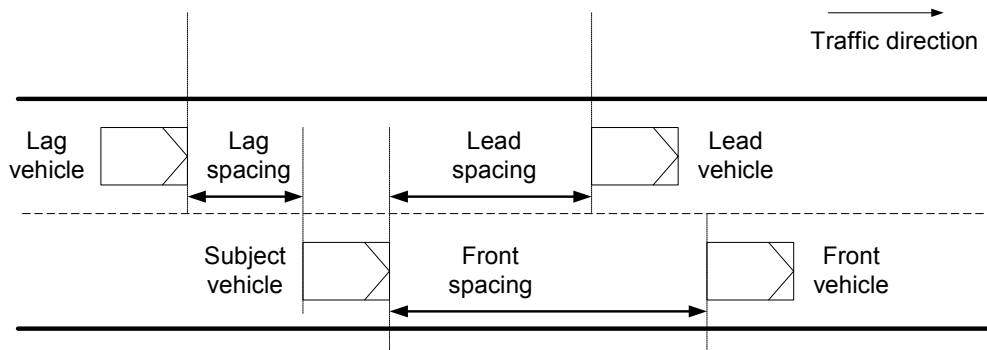
Figure 6.4: Definitions of the lead and lag vehicles and the gaps they define

A notable aspect of the Lankershim dataset is a significant percentage of observations in which a lead or lag vehicle is absent in adjacent lanes during the lane changes. The effect of signals is a major factor behind this phenomenon. This aspect of the dataset is summarized in Table 6.3.

Table 6.3: Vehicle observations without lead/lag vehicle in adjacent lane

	All Observations (total 16696)		Lane Changing Observations (total 249)	
	No.	%	No.	%
Lead Vehicle Absent	3749	22.45	155	62.24
Lag Vehicle Absent	3811	22.83	151	60.64

To accommodate observations without lead and/or lag gap, it is hypothesized that the lead and lag gap lengths considered by the drivers in such instances are the distances from the nearest intersection boundaries lying within either gap. This approach represents a reasonable assumption given that traffic regulations do not allow vehicles to make a lane change within an intersection. The definition of gaps in such situations is illustrated in Figure 6.5.
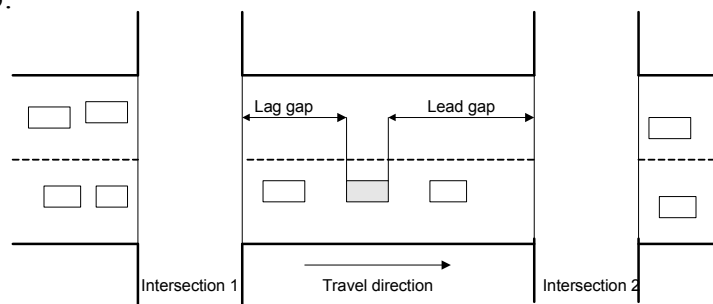
Figure 6.5: Definitions of the lead and lag gaps in absence of lead and/or lag vehicles

Table 6.4 presents the descriptive statistics for the lead and lag vehicle relative to the subject vehicle. Relative speeds are defined as the speed of the lead (lag) vehicle less the speed of the subject vehicle. The table also summarizes statistics for the accepted lead and lag gaps. Accepted lead gaps vary from 0.22 meter to 118.73 meters, with a mean of 23.57 meters. Accepted lag gaps vary from 0.75 meter to 128.52 meters. Statistics for all gaps in the dataset are shown in parenthesis in the same table (Table 6.4). In these statistics, negative spacing values indicate that the subject and the lead vehicle partly overlap (this is possible because they are in different lanes). As expected, the mean accepted gaps are larger than the mean gaps in the traffic stream. Similarly, lead relative speeds in accepted gaps are larger than in the mean of the dataset and lag relative speeds are smaller in the entire dataset (i.e. on average, in accepted gaps the subject vehicle is slower relative to the lead vehicle and faster relative to the lag vehicle compared to the gaps in the entire dataset).

Table 6.4: Statistics describing the lead and lag vehicles

Variable	Mean	Std Dev	Max	Min
Lead Relative Speed	-2.05	3.87	3.50	-14.58
(m/sec)	(0.35)	(3.55)	(15.92)	(-15.73)
Lead Gap	23.57	19.24	118.73	0.22
(m)	(11.60)	(18.74)	(155.80)	(0.00)
Lag Relative Speed	-0.93	3.90	7.30	-15.25
(m/sec)	(0.35)	(3.65)	(15.62)	(-15.73)
Lag Gap	9.18	23.47	128.52	0.75
(m)	(3.51)	(20.24)	(152.28)	(0.00001)

More detailed information on the Lankershim Blvd. dataset is available in the NGSIM data analysis report (Cambridge Systematics, Inc. 2006) and Ramanujam (2007).

As noted previously, arterial lane selection behavior is captured by two sub-models, the intersection lane choice model and the lane changing model for the mainline section. The details of the intersection lane choice model are presented first followed by the details of the lane changing model for the mainline section.

6.2.2 Lane Choice at Intersection

Model Structure

The intersection lane choice model involves the lane selection of drivers entering the arterial from a side street (Figure 6.6). These drivers are likely to target the lane that they perceive to be the best in the subsequent section of their path and plan to move to that lane. The choice set of target lane is likely to include all lanes in the subsequent section regardless of their availability. Due to situational constraints, immediate execution of the tactical lane selection plan may not be possible. For example, at a particular instant, conflicts with other vehicles can delay movement to the target lane. The immediate lane choice of the driver, that is the lane where he/she is observed just after turning, thus may not be the same as the originally targeted lane. Further, changes in circumstances may lead to changes in the target lane: a long queue build-up in the chosen target lane for example can lead to amendment to the original target. The chosen target lanes are thus unobserved and only the immediate choice of lanes is observed.
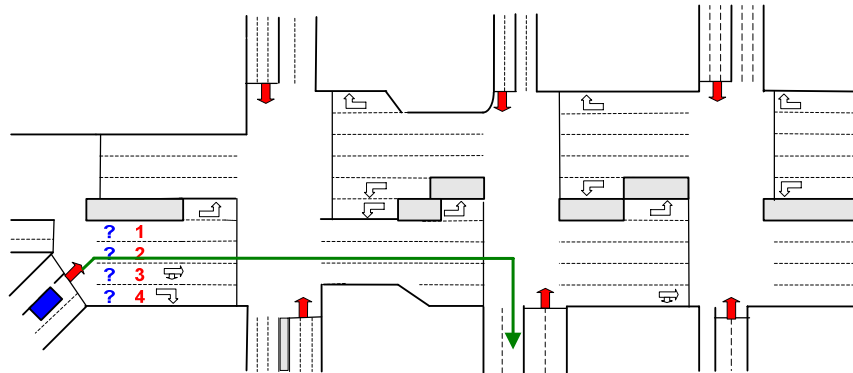
Figure 6.6: Intersection lane selection

The intersection lane selection can therefore be modeled as a two level decision:

- Choice of target lane
- Choice of immediate lane

The choice of target lane is a tactical decision of the driver whereas the choice of immediate lane is governed by maneuverability considerations. The framework of the model is shown in Figure 6.7. Latent choices are shown as ovals, observed ones are

shown as rectangles. It should be noted that once the driver enters the arterial, then the mainline lane changing model is applicable.
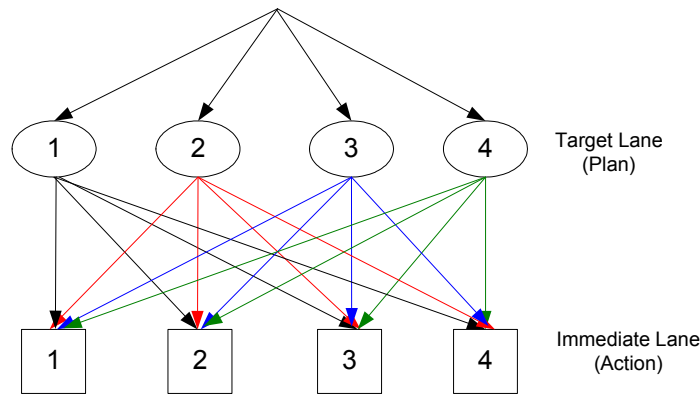
Figure 6.7: Structure of the intersection lane selection model

The two levels of decision are detailed in the following sections.

Choice of Plan: The Target Lane Choice

At the first level, the driver chooses the most desirable lane as the target lane. The target lane choice set constitutes of all the available lanes the driver is eligible to move to. The target lane utilities are affected by a wide range of factors. These include variables related to the path-plan of the driver, such as the distance to a point where the driver needs to be in specific lanes and the number of lane changes required from the target lane to the correct lanes. However, the effects of path-plan in the target lane choice also depend on the planning capability of the driver and his/her familiarity with the network. Drivers who are familiar with the network and 'plan-ahead' are likely to pre-position themselves in the correct lanes well-ahead of the section prior to the turn. These drivers may also be aware of the lane-specific obstructions in downstream sections and take into account the anticipated delays associated with staying in a lane while making their decisions. On the other hand, drivers who are not familiar with the network and/or do not plan-ahead are not likely to be affected by path-plan considerations or anticipated delay beyond their immediate sections.
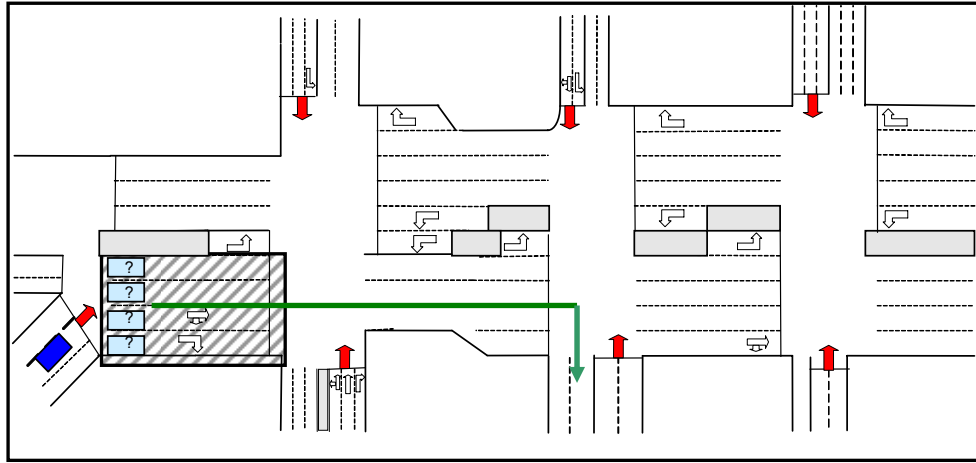
Figure 6.8: Perspective of myopic drivers
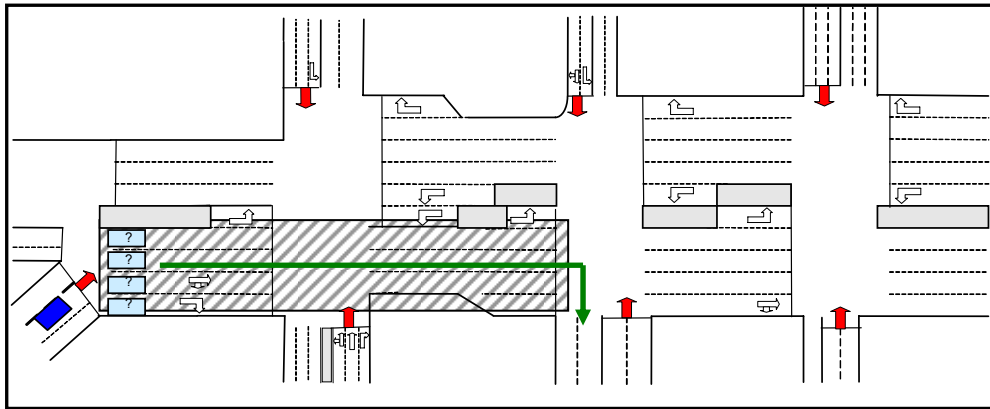
Figure 6.9: Perspective of drivers who plan-ahead

Depending on the familiarity and planning capability, the drivers can thus belong to either of the two classes:

- Class 1: Myopic drivers. Drivers who consider the path-plan and anticipated delay only in their immediate subsequent section while making the lane selections belong to this class.

- Class 2: Drivers who plan-ahead. These drivers consider path-plan and anticipated delay beyond their immediate subsequent section while making the lane selections.

Parameters associated with the target lane of the driver may be class-specific, indicating significant difference in sensitivity to influencing variables among driver classes. The perspectives of each class of driver are presented in Figures 6.8 and 6.9.

The utilities of the various target lanes can be expressed as follows:

$$U_{ln} = V_{ln} + \varepsilon_{ln}, \quad l \in L_n \tag{6.1}$$

Where,

V_{ln} = systematic part of the utility of target lane l of driver n

ε_{ln} = random error term associated with target lane choice

L_n = choice set of target lane of driver n

Variables likely to influence the target lane choice of the driver include the following:

- Path-plan variables: Number of lane changes the driver needs to make in order to be in the correct lane to follow his/her path and the distance to the point by which he/she needs to be in the correct lane;

- Lane attributes: Queue lengths, average speeds, and queue discharge rates;

- Driving style and capabilities: Individual driver/vehicle characteristics, such as the plan-ahead distance and aggressiveness of the driver; and

- Expected maximum utility from the immediate lanes: The expected maximum utility (EMU) that can be derived from choosing the immediate lanes given a particular target lane can also affect the target lane choice. The choice of target lane is thus indirectly influenced by variables that influence the immediate lane choices given the target lane choice.

The systematic utility consists of interaction of these sets of variables and can be expressed as follows:

$$V_{ln} = V\left(X_{ln}, \beta, \alpha^l, \lambda_n, \upsilon_n, EMU_{ln} \right) \tag{6.2}$$

Where,

X_{ln} = attributes of lane l for driver n, can be function of λ_n

λ_n = individual-specific look-ahead/plan-ahead distance

β = cofficients

υ_n = individual-specific random effect (e.g. aggressiveness): $\upsilon_n \sim N(0,1)$

α^l = coefficient of individual-specific random effect for lane l for target lane choice

EMU_{ln} = expected maximum utility from choosing lane l as target lane

I

t may be noted that since the path-plan and other influencing variables are effective only when these are within the plan-ahead distance of a driver and the plan-ahead distance varies among the driver population with network familiarity and planning capability, the

variables related to the lane attributes can differ among drivers at the same intersection with the same path-plan and surrounding conditions.

The driver chooses as the target lane the lane with the highest utility. Different choice models are obtained depending on the assumption made about the distributions of the random terms ε_{ln}. Assuming that they are independently and identically extreme value distributed, target lane choice probabilities, conditional on individual-specific characteristics (planning-ahead and aggressiveness), can be expressed as follows:

$$P_n\left(l \mid \upsilon_n, \lambda_n\right) = \frac{\exp\left(V_{ln}\right)}{\displaystyle\sum_{l' \in L_n} \exp\left(V_{l'n}\right)} \quad \forall l, l' \in L_n \tag{6.3}$$

The Immediate Lane Choice Model

Given the choice of the target lane the driver selects the immediate lane. The immediate lane selection depends on the choice of target lane but is also influenced by maneuverability considerations. For example, a lane may be unavailable as an immediate lane if it is already full. To make the model more flexible, the choice set for the immediate lane is assumed to include all available lanes in the roadway irrespective of the target lane and the current position of the driver. The structure can thus accommodate cases when the target lane and lanes in the direction of the target lane are blocked by other vehicles and the driver has no option but to move to a different connecting lane.
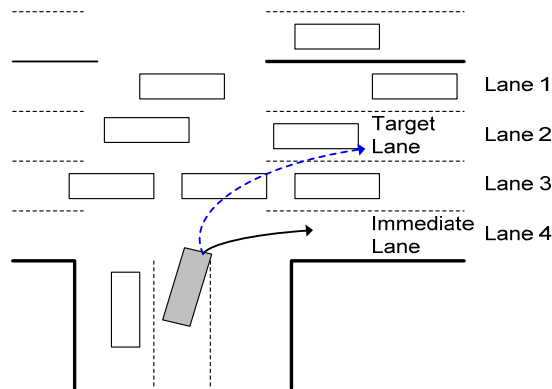
Figure 6.10: Example of a situation when the target lane is blocked

This extreme situation is illustrated in Figure 6.10 with a hypothetical example where the target lane of the driver is Lane 2 (the path to which is blocked) and the driver

chooses Lane 4 as the immediate lane. The other option for the driver is to wait till the vehicles in Lane 3 move forward and maneuver to Lane 2 when possible.

The immediate lane choice is thus affected by maneuverability considerations and the driving effort needed to reach a particular lane, and is conditional on the choice of target lane. The utility of choosing a lane as immediate lane can be expressed as follows:

$$U_{jln} = V_{jln} + \varepsilon_{jln}, \quad j \in J_n \tag{6.4}$$

Where,

V_{jln} = *systematic part of utility of immediate lane j of driver n given target lane l*

ε_{jln} = *random error term associated with immediate lane choice*

J_n =*choice set of immediate lane of driver n*

Variables likely to influence the immediate lane choice of the driver include the following:

- Current position of the driver: Proximity of a given lane to the receiving lane closest to the driver;

- Neighborhood variables: Presence of other vehicles and their actions, relative position and speed of the subject vehicle with respect to neighboring vehicles, geometric elements of the roadway, signals and signs;

- Choice of target lane: Proximity of the immediate lane to the chosen target lane; and

- Driving style and capabilities: Individual driver/vehicle characteristics, such as the aggressiveness of the driver and performance capabilities of the vehicle (e.g., required turning radius).

The systematic utility consists of interaction of these factors and can be expressed as follows:

$$V_{jln} = V\left(X_{jln}, \beta, \alpha^j, \upsilon_n\right) \tag{6.5}$$

Where,

X_{jln} = *attributes of lane j for driver n given target lane l*

α^j = *coefficient of individual-specific random effect for lane j*

J_n = *choice set of immediate lanes of driver n*

Assuming the random error terms ε_{jln} are independently and identically extreme value distributed, immediate lane choice probabilities, conditional on target lane l and individual-specific characteristics can be expressed as follows:

$$P_n\left(j\,|\,l,\upsilon_n\right) = \frac{\exp\left(V_{jln}\right)}{\displaystyle\sum_{j'\in J_n}\exp\left(V_{j'ln}\right)} \quad \forall j, j' \in J_n \tag{6.6}$$

Likelihood Function

The probability that driver n selects lane j is the joint probability of selecting lane j given target lane l and the probability of choosing target lane l and can be expressed as:

$$P_n\left(j\,|\,\upsilon_n,\lambda_n\right) = \sum_{l\in L_n} P_n\left(j\,|\,l,\upsilon_n\right) P_n\left(l\,|\,\upsilon_n,\lambda_n\right) \tag{6.7}$$

The unconditional probability of driver n selecting lane j at a given time can be expressed as:

$$P_n\left(j\right) = \int_\upsilon \sum_\tau P_n\left(j\,|\,\upsilon_n,\lambda_n\right) p(\lambda) f(\upsilon)d\upsilon$$

$$p(\lambda) = \begin{cases} \pi_1 & \textit{Class 1 (Myopic driver)} \\ 1-\pi_1 & \textit{Class 2 (Driver who plans-ahead)} \end{cases} \tag{6.8}$$

Where, the probability that the driver belongs to Class 1 (π_1) or Class 2 $(1-\pi_1)$ is estimated from the data along with other parameters.

Assuming that the observations from different drivers are independent, the log-likelihood function for all N individuals observed is given by:

$$\mathcal{L} = \sum_{n=1}^{N} \ln(P_n(j)) \tag{6.9}$$

The parameters of the model are estimated by maximizing this function.

Estimation Result

All components of the model were estimated jointly using a maximum likelihood estimation procedure as described in the previous section. The summary of estimation results of the proposed model is presented in Table 6.5.

Table 6.5: Estimation results of the target lane changing model

Final Log-likelihood	-2115.8
Initial Log-likelihood	-2797.9
Number of drivers	703
Number of observations	703
Number of parameters	20
Adjusted rho-bar square	0.237

The goodness-of-fit of the new model was compared with a reduced form model estimated with the same data (Table 6.7). The reduced form model is a single level lane choice model with no latent targets (Figure 6.11). The estimation results of this model are presented in Appendix C.3.
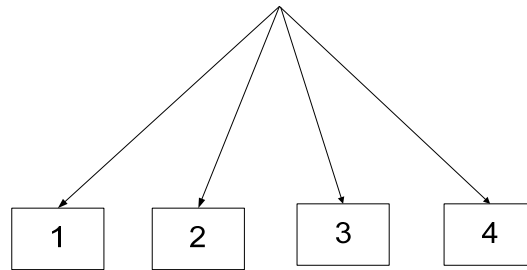
Figure 6.11: Simple model structure

Table 6.6: Model comparison

Statistic	Single Level (R)	Target Lane (U)
Likelihood value	-2120.4	-2115.8
Number of parameters (k)	19	20
Akaike information criteria (AIC)	-2139.3	-2135.8
Adjusted rho-bar square ($\bar{\rho}^2$)	0.235	0.237

The model with explicit target lane choice has larger values in terms of both AIC and $\bar{\rho}^2$ (detailed in Chapter 4). This indicates that the inclusion of the latent plans in the decision framework results in an improved goodness-of-fit even after discounting for the increase in the number of parameters..

The detailed estimation results are presented next.

Choice of Plan: Target Lane Choice

The target lane choice model describes drivers' choice of lane they would want to travel in. The choice set of the driver includes all lanes in the mainline. The target lane choice of the driver is affected by the path-plan, the lane attributes and driver characteristics. Path-plan variables include number of lanes a driver has to cross (if any) in order to take a turn or exit while following the path. Lane attributes include queue length, queue discharge rate, average speed etc. of each lane. In this model, the queue length and queue discharge rates are combined in a single variable anticipated delay. This variable represents the delay associated with the queue (in time unit) and is calculated by dividing the current queue length by the average queue discharge rate. The variables affecting the immediate lane choice also have indirect effect son target lane choice. This effects have been captured through Expected Maximum Utility (EMU) variables.

Different functional forms of variables as well as interactions between multiple variables have been tested during estimation and the functions resulting improvements in goodness-of-fits are selected. For example, the path-plan variables were found to improve the goodness-of-fit when interacted with the individual specific random term (aggressiveness). Also, some parameters were excluded because of non-intuitive signs and/or statistical insignificance. For example, the effect of average speed on target lane choice was tested but had a non-intuitive sign. The high correlation between the queue length and the average speed variables may have caused this and so the latter has been excluded from the model.

The estimation results of the target lane selection are presented in Table 6.7. The magnitudes of the lane-specific constants indicate that all else being equal, the drivers prefer lanes on the right (the rightmost lane being the most preferred lane). It should be noted that though the model has been developed with data where the receiving mainline section had 3 or 4 lanes, the model structure is flexible for application in other scenarios with a different number of available lanes. For this, the lane constants in particular need to be re-calibrated.

Table 6.7: Intersection lane choice: Target lane model

Variable		Parameter	t-stat
Target Lane			
Lane 2 constant		-0.837	-3.64
Lane 3 constant		1.30	7.62
Lane 4 constant		3.25	8.16
Anticipated delay (second)		-0.477	-0.56
Lanes away from turning lane (myopic)	coefficient-myopic drivers	-0.0240	-0.63
	constant-myopic drivers	1.43	0.83
	heterogeneity coefficient -myopic drivers	1.53	0.75
Lanes away from turning lane (with plan-ahead)	coefficient-drivers who plan-ahead	-4.08	-1.98
	constant-drivers who plan-ahead	2.05	3.01
	heterogeneity coefficient-drivers who plan-ahead	0.466	0.74
Expected maximum utility from immediate lane		0.915	7.22
Driver Class			
Driver population with >1 section plan-ahead (%)		18.3	2.07

As described in the earlier section, a latent class formulation has been used for the model to capture the heterogeneity in planning capability of drivers. The probability of the driver being a myopic driver (Class 1) or a driver who plans-ahead (Class 2) is calculated along with the other model parameters. The estimated probability that the driver is of Class 2 was found to be 18.3 %.

The influencing variables differ depending upon the plan-ahead distance of the driver. For example, familiar drivers may consider the anticipated delay in subsequent sections while making their target lane choices. Therefore, an anticipated delay value was calculated for each class of driver based on what segments they are considering while making their lane choices. The functional form of the anticipated delay variable can be expressed as follows:

$$\tilde{q}_{nt}^{kl} = \frac{1}{1 + exp\left(-q_{nt}^{kl}\right)} \quad k=1,2 \tag{6.10}$$

$$q_{nt}^{1l} = d_{nt}^{1l}/r_n^{1l}, q_{nt}^{2l} = q_{nt}^{1l} + d_{nt}^{2l}/r_n^{2l}$$

Where,

q_{nt}^{kl} = anticipated delay in lane l considering k sections ahead

d_{nt}^{ki} = queue length in lane i in section k at time t (vehicles)

r_{nt}^{ki} = average queue discharge rate of lane i in section k (vehicles/sec)

The anticipated delay has a diminishing effect on the utility of target lane as illustrated in Figure 6.12. The sensitivity to anticipated delay was however not found to be significantly different for the two classes of drivers.
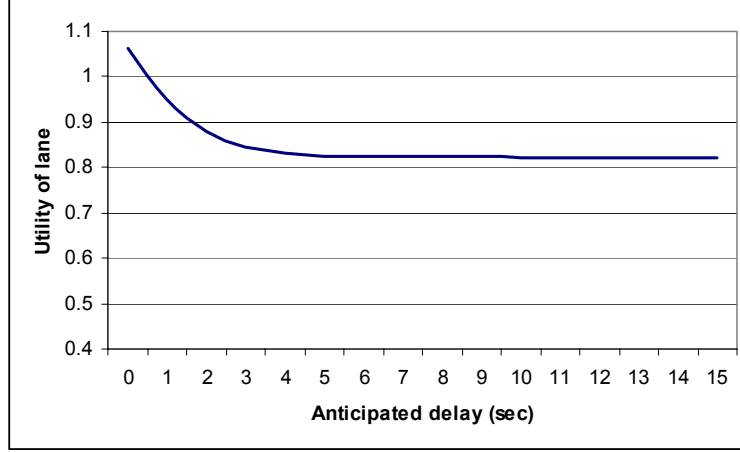
Figure 6.12: Effect of anticipated delay

The path-plan of the driver has an important role in the target lane selection. The two classes of drivers are found to have different sensitivities to path-plan considerations, which in this case has been modeled as an interaction between the number of lanes away from the correct lane and the aggressiveness of the driver. The functional form best fitting the data is found to be as follows:

$$\left[\frac{\theta_1}{\psi_1 + \alpha_1 \upsilon_n} \left(e_n^{1l} \right) \left(1 - \delta_n \right) + \frac{\theta_2}{\psi_2 + \alpha_2 \upsilon_n} \left(e_n^{2l} \right) \delta_n \right] \tag{6.11}$$

Where,

$\delta_n = 1$ *if the driver plans-ahead beyond the immediate section*

e_n^{1l} = *lanes away from turning lane for myopic drivers*

e_n^{2l} = *lanes away from turning lane for drivers who plan-ahead*

$\theta_i, \psi_i, \alpha_i$ = *coefficients of vehicle class i*

As seen from the estimates, for both classes of drivers, utility of lanes reduce if they are away from the lane that the driver needs to be in to follow his path. This disutility is, however, less for aggressive drivers, since they are more prone to make aggressive lane changes later if needed. The disutility was found to be larger and more significant for

drivers who plan-ahead (Class 2). The effect of path-plan for each driver class is explained in Figure 6.13.
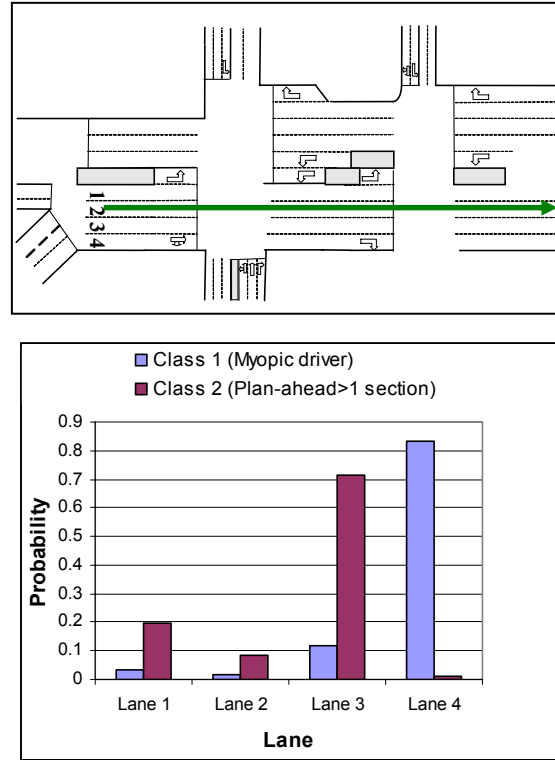
Figure 6.13: Effect of path-plan and driver class

In this example, Lane 4 becomes a right turn only lane in the second section. Therefore, drivers who are going straight have to change later if they choose Lane 4 in the first section. Drivers who plan-ahead beyond the immediate section are less likely to choose this lane as target compared to drivers who do not plan-ahead.

The expected maximum utility term captures the maximum utility that can be derived from selecting a particular lane as the immediate lane. It has a significant effect on the target lane choice. The expected maximum utility (EMU) can be calculated as the logsum of the immediate lanes given the target lane (see Ben-Akiva 1973, Ben-Akiva and Lerman 1985). Mathematically, this refers to the following:

$$
\begin{aligned}
EMU_{ln} &= E\left(\max\left(U_{1\,ln}, U_{2\,ln}, ..., U_{j\,ln}, ..., U_{J_l\,ln}\right)\right) \\
&= \ln\left(\exp\left(V_{1\,ln}\right) + \exp\left(V_{2\,ln}\right) + ... + \exp\left(V_{j\,ln}\right) + ... + \exp\left(V_{J_l\,ln}\right)\right)
\end{aligned}
\tag{6.12}
$$

Where,

EMU_{ln} = expected maximum utility derived from lane l

U_{jln} = utility of immediate lane j for driver n given target lane l

The estimated utility of the target lane can thus be expressed as follows:

$$U_{ln} = \beta_l - 0.477 \left[(\tilde{q}_{ln}^{1})(1-\delta_n) + (\tilde{q}_{ln}^{2})(\delta_n) \right] - \frac{0.024}{1.43 + 1.53\upsilon_n}(e_{ln}^{1})(1-\delta_n)$$

$$- \frac{4.08}{2.05 + 0.466\upsilon_n}(e_{ln}^{2})\delta_n + 0.915(EMU_{ln})$$

(6.13)

Where,

β_l = constant for lane l

$\tilde{q}_{ln}^{1}$ = anticipated delay function in lane l for myopic drivers

$\tilde{q}_{ln}^{2}$ = anticipated delay function in lane l for drivers who plan-ahead

e_{ln}^{1} = lanes away from correct lane for myopic drivers

e_{ln}^{2} = lanes away from correct lane for drivers who plan-ahead
(consider path-plan beyond current section)

EMU_{ln} = expected maximum utility derived by driver n from selecting
lane l as target lane (consider delay beyond current section)

δ_n = 1 if the driver plans-ahead beyond immediate section

Choice of Action: Immediate Lane Choice

Immediate lane choices were found to be influenced by maneuverability considerations and inertia to continue to the naturally connecting lane. The estimation results are summarized in Table 6.8.

Table 6.8: Intersection lane choice: Immediate lane model

Variable		Parameter	t-stat
Lanes away from connecting lane	coefficient	-1.01	-1.19
	constant	0.691	1.94
	heterogeneity coefficient	1.96	3.48
Target lane dummy		3.16	4.54
Lanes away from target lane	coefficient	-4.42	-3.00
	constant	2.12	2.14
	heterogeneity coefficient	0.0904	0.36
Conflict dummy		-1.76	-9.63

Inertia effects are captured by variables like current lane inertia and number of lanes away from the connecting lane. The inertia effect was greater for aggressive drivers. Aggressive drivers tend to stay in their current lane as long as possible and then make aggressive changes if a lane change is warranted by the path-plan. Drivers were also found to have a strong preference to reach their target lane and lanes closer to their target lanes. The combined effect of inertia and preference for moving to lanes nearer to target lanes for aggressive and normal drivers are illustrated in Figure 6.14.
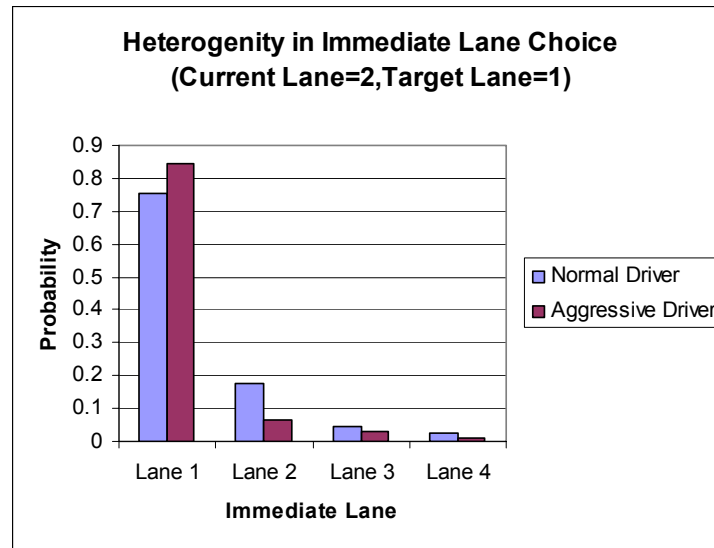
Figure 6.14: Heterogeneity in immediate lane choice

Maneuver to a given lane may not be possible due to conflicts with neighboring vehicles. In the case of such obstructions or conflicts, the driver can choose an immediately available lane, or can wait until the neighboring vehicle moves and there are no obstructions to maneuver to the intended target lane. As a result, if there are conflicting vehicles in the direction of a lane, the driver was found to have a lower preference for that lane.

The utility of immediate lane j is summarized in the following equation:

$$U_{jn} = -\frac{1.01}{0.691+1.96\upsilon_n}(c_{jn})+3.16(l_{jn}=0)-\frac{4.42}{2.12+.0904\upsilon_n}(l_{jn})-1.76\gamma_{jn} \qquad (6.14)$$

Where,

$c_{jn} = lanes\ away\ from\ connecting\ lane$

$l_{jn} = lanes\ away\ from\ target\ lane\ l,\ l \in L_n$

$\gamma_{jn} = 1\ if\ maneuver\ to\ lane\ j\ is\ obstructed\ by\ adjacent\ vehicle$

6.2.3 Mainline Lane Changing

Model Structure

As discussed in Section 6.1, the within section lane changing maneuver of drivers can be modeled as a three stage process:

- Choice of target lane (plan)
- Decision to accept available gaps (plan)
- Execution of the lane change (action)

The structure of the model is shown in Figure 6.15 with a hypothetical scenario of a four lane road with the driver in Lane 3.
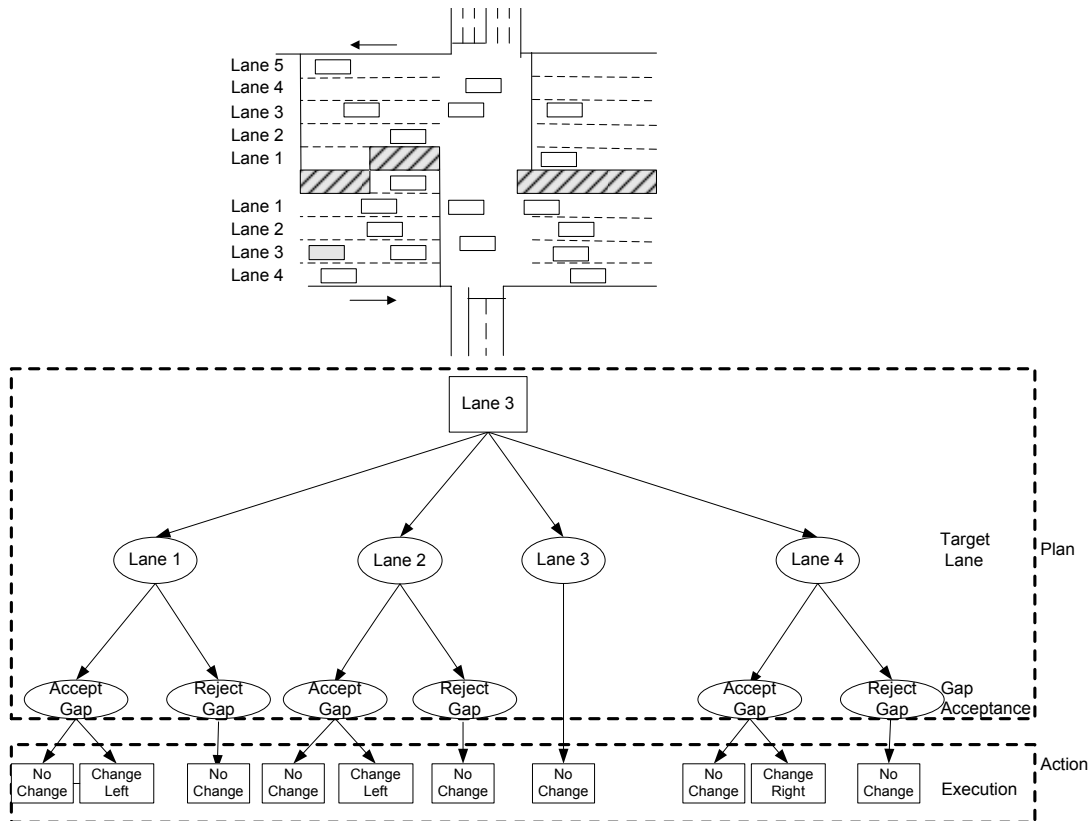
Figure 6.15: Framework for within section lane changing model

First, the driver selects the target lane - the lane the driver perceives as best to be in, depending upon the prevalent driving conditions and the path-plan. The choice set for the target lane selection includes all allowable lanes in the current section in the direction of travel. The choice of the target lane indicates the direction of immediate lane shift of the driver. In the hypothetical scenario, if the driver is in Lane 3, a choice of Lane 3 as the target lane means that the driver has decided not to pursue a lane change and to continue in the current lane. If the driver perceives that moving to another lane would improve the condition, he/she chooses that lane as the target lane. In the above example, the immediate direction of lane shift for the driver is the left lane if either Lane 1 or 2 is chosen as the target lane, while it is the right if Lane 4 is chosen as the target lane. If the target lane is different than the current lane, the driver evaluates the gaps in the corresponding adjacent lane in the direction of the target lane.

The available lead and lag gaps in the adjacent lane are compared with respective critical gaps (unobserved) and the driver decides to accept or reject the gap. For the gap to be acceptable, both the lead and the lag gaps have to be acceptable, i.e. greater than the respective critical gap values.

Even if the driver perceives that the gap is acceptable, the execution of the lane change may take some time. In urban situations, lane change durations are found to range from 3.4 to 13.6 seconds with a mean duration of 6.0 seconds (Hetrick 1997). In the trajectory data, the completion of the lane change is manifested as an execution decision of the driver. In presence of acceptable gaps, the driver can either decide to complete the lane change in a given instant and move to the adjacent lane (Change Right or Change Left), or not to execute the lane change in that instant and stay in the current lane (No Change). In instances when the driver chooses the current lane as the target lane or chooses a different lane as the target lane but does not find the gaps acceptable in the immediate lane in the direction of the target lane, he/she does not consider the execution step (probability of execution is zero), and the driver is observed in the current lane (No Change). This decision process is repeated at every time step.

The first two steps in the decision process, target lane choice and gap acceptance, are latent. Only the driver's final actions, constituting no lane change or lane change

execution to the left or right, are observed. Latent choices are shown as ovals, and observed ones are shown as rectangles.

Choice of Plan

In the target lane model, the driver chooses the target lane and plans to move to that lane through gap acceptance. The choice of plan is modeled as a two stage decision, target lane selection and gap acceptance.

The Target Lane Selection

At the highest level of lane changing, the driver chooses the lane with the highest utility as the target lane. The target lane choice set constitutes all available lanes in the roadway. Due to closely spaced turns and lane use restrictions (e.g. turn-only and no-turn lanes) in urban arterials, path-plan variables are likely to have a significant impact on target lane choice. The network familiarity of the driver as well as his/her planning capabilities is likely to affect the path-plan variables. For example, a driver who is not familiar with the network and/or does not plan-ahead may not take into account the path-plan considerations before getting very close to the turn. In contrast to the intersection lane choice case where the heterogeneity in planning ability of the driver is modeled using a latent class methodology, in the mainline model, where a series of lane changing decisions of the same driver are modeled over a section, a continuous distribution of look-ahead or plan-ahead distance is more intuitive. Estimation results also support this hypothesis and therefore a continuous plan-ahead distance is integrated in the model.

The total utility of lane *l* as a target lane for driver *n* at time *t* can be expressed by:

$$U_{lnt} = \beta^T X_{lnt}(\lambda_n) + \alpha^l \upsilon_n + \varepsilon_{lnt} \qquad \forall l \in L_n \tag{6.15}$$

Where,

υ_n = *individual-specific random effect (e.g. aggressiveness):* $\upsilon_n \sim N(0,1)$

α^l = *coefficient of individual-specific random effect for lane l*

ε_{lnt} = *random term associated with the target lane utilities*

L_n = *choice set of target lane of driver n*

λ_n = *look-ahead / plan-ahead distance of the driver*

X_{lnt} = *explanatory variables that affect the utility of lane l, function of* λ_n

β = *corresponding vector of parameters*

The target lane utilities of a driver are likely to be affected by the following variables:

- Path-plan variables: Distance to the point when the driver needs to be in a specific lane to follow the path, and the number of lane changes required to be in the correct lane;

- Lane attributes: Queue lengths, average speeds, and queue discharge rates;

- Current position of the driver: Inertia to stay in the current lane, proximity of a lane to the current lane of the driver;

- Neighborhood variables: Presence of other vehicles and their actions, relative position and speed of the subject vehicle with respect to neighboring vehicles, geometric elements of the roadway, signals and signs , and available capacity of the lane; and

- Driving style and capabilities: Individual driver/vehicle characteristics, such as the plan-ahead distance and aggressiveness of the driver;

Different choice models are obtained depending on the assumption made about the distribution of the random term ε_{lnt}. Assuming that these random terms are independently and identically extreme value distributed, choice probabilities for target lane l, conditional on the individual-specific error term (υ_n) are given by a logit model and expressed as follows:

$$P_n(l_t \mid \upsilon_n, \lambda_n) = \frac{\exp(\beta^T X_{lnt}(\lambda_n) + \alpha^l \upsilon_n)}{\displaystyle\sum_{l' \in L_n} \exp(\beta^T X_{l'nt}(\lambda_n) + \alpha^{l'} \upsilon_n)} \qquad \forall l, l' \in L_n \qquad (6.16)$$

The Gap Acceptance Model

In the gap acceptance stage, the driver evaluates the adjacent gaps in the direction of the target lane and decides whether or not a lane change in the chosen direction can be undertaken. The adjacent gap in the target lane is defined by the lead and lag vehicles in that lane as described in Section 6.2.1 and shown in Figure 6.16.

The driver compares the available lead and lag gaps to the corresponding critical gaps, which are the minimum acceptable gaps. An available gap is acceptable if it is greater than the critical gap. Critical gaps are modeled as random variables with their

means being functions of explanatory variables. The individual-specific term in this mean function captures correlations between the critical gaps of the same driver over time.
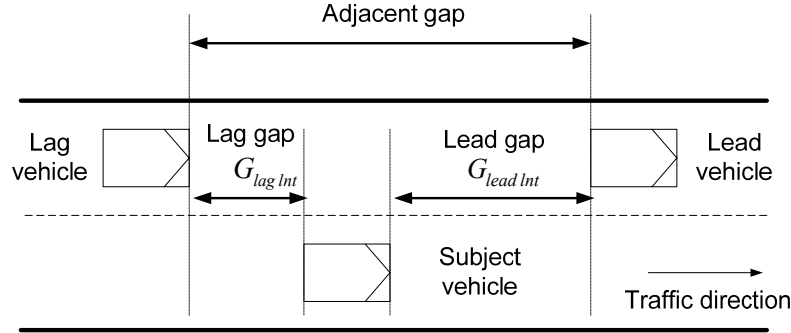
Figure 6.16: Definitions of the lead and lag vehicles and the gaps they define

Critical gaps are assumed to follow lognormal distributions to ensure that they are always non-negative and can be expressed as follows:

$$G_{lnt}^{g\,cr} = \exp(\beta^T X_{lnt}^g + \alpha^g \upsilon_n + \varepsilon_{lnt}^g) \quad g \in \{lead, lag\} \tag{6.17}$$

Where,

$G_{lnt}^{g\,cr}$ = critical gap g in the direction of target lane l, measured in
   distance units (e.g. meters)

X_{lnt}^g = explanatory variables that affect the critical gap g in the
   direction of target lane l

β = coefficients of explanatory variables

α^g = coefficients of individual-specific latent variable υ_n for gap acceptance

ε_{lnt}^g = random term: $\varepsilon_{lnt}^g \sim N\left(0, \sigma_g^2\right)$

Gap acceptance is affected by the interaction between the subject vehicle and the lead and lag vehicles in the adjacent lane. This may be captured by variables such as the subject relative speed with respect to the lead and lag vehicles, type of lag vehicle etc. It can be also affected by the urgency of the lane change. Remaining distance to reach the mandatory lane changing point can be used as an indicator of the urgency of the lane change.

The gap acceptance model assumes that the driver must accept both the lead gap and the lag gap to consider lane change execution. The probability of accepting the adjacent

gap, conditional on the individual-specific term υ_n and the chosen target lane is therefore given as follows:

$$P_n(i_t = 1 \mid l_t, \upsilon_n) = P_n(accept\ lead \mid l_t, \upsilon_n) P_n(accept\ lag \mid l_t, \upsilon_n)$$
$$= P_n(G_{lnt}^{lead} \geq G_{lnt}^{lead\ cr} \mid l_t, \upsilon_n) P_n(G_{lnt}^{lag} \geq G_{lnt}^{lag\ cr} \mid l_t, \upsilon_n) \qquad (6.18)$$

Based on the assumption that critical gaps follow lognormal distributions (ε_{glnt} is normally distributed), the conditional probability that gap $g \in \{lead, lag\}$ is acceptable is given by:

$$P_n\left[G_{lnt}^g > G_{lnt}^{g\ cr}(\upsilon_n) \right]$$
$$= P_n\left[\ln\left(G_{lnt}^g\right) > \ln\left(G_{lnt}^{g\ cr}(\upsilon_n)\right) \right] = \Phi\left[\frac{\ln\left(G_{lnt}^g\right) - \left(\beta^T X_{lnt}^g + \alpha^g \upsilon_n\right)}{\sigma_g} \right] \qquad (6.19)$$
$$\varepsilon_{lnt}^g \sim N(0, \sigma_g^2)$$

$\Phi[\cdot]$ denotes the cumulative standard normal distribution.

The Lane Change Execution Model

The driver considers the lane change execution decision step if he/she chooses a target lane that is different from the current lane and finds the adjacent gaps in the immediate lane in the direction of the chosen target lane acceptable. Given the above latent decisions, the decision to execute the lane change in the current time step can be modeled as a binary logit model. The choice set in this decision step contains two alternatives, to execute the lane change (completely move to the adjacent lane in the direction of target lane) or not.

The probability of executing the lane change in the current time instant t is given by

$$P_n(j_t \mid i_t, l_t, \upsilon_n) = \begin{cases} \dfrac{1}{1 + \exp(-(\beta^T X_{jlnt} + \alpha^j \upsilon_n))} & if\ i_t = 1\ (accept\ gap) \\ 0 & otherwise \end{cases} \qquad (6.20)$$

Where,

    j_t = *lane changing action*

    X_{jlnt} = *explanatory variables that affect the driver's execution decision*

    α^j = *parameter of individual-specific latent variable υ_n for execution level*

The essence of modeling the execution decision is to capture indirectly the time required to complete the lane change after initiating one. Analysis of trajectory data by Toledo and Zohar (2007) shows that lane changing durations are affected by traffic density, direction of change (left or right), relative speed and spacing of front, lead and lag vehicles, speed of the subject vehicle, type of vehicle (heavy or not) etc.

Further, the inclusion of the execution of the lane change level is more relevant in this particular study, where there are a number of successive observations where the driver faces large and unchanging adjacent gaps (including those instances where there are no lead or lag vehicles). It is observed that the driver changes lane in one of these instances. The execution step in a way models the way the driver differentially evaluates these seemingly similar scenarios (see Ramanujam 2007 for details).

Another key aspect that affects this instantaneous decision is the time resolution of the subsequent observations of the same driver. The probability of executing a lane change is reduced as the time resolution decreases. In the current dataset, the time resolution is constant and hence this effect is likely to be embedded in the constant terms of the execution level. But in datasets where the time steps of observation vary across the drivers, it is possible to estimate the influence of this attribute in the final decisions of the driver.

Likelihood Function

In this section, the likelihood function of lane changing actions observed in the data is presented. The joint probability density of a combination of target lane (l), gap acceptance (i) and lane action (j) observed for driver n at time t, conditional on the individual-specific characteristics, λ_n and υ_n are given by:

$$P_n(l_t, i_t, j_t \mid \upsilon_n, \lambda_n) = P_n(l_t \mid \upsilon_n, \lambda_n) P_n(i_t \mid l_t, \upsilon_n) P_n(j_t \mid i_t, l_t, \upsilon_n) \qquad (6.21)$$

Where, $P_n(l_t \mid .)$, $P_n(i_t \mid .)$ and $P_n(j_t \mid .)$ are given by Equations 6.16, 6.18 and 6.21, respectively.

Only the lane changing actions are observed. The marginal probability of the lane changing action is therefore given by:

$$P_n\left(j_t \mid \upsilon_n, \lambda_n\right) = \sum_{l \in L_n} \sum_{i \in I_n} P_n\left(l_t, i_t, j_t \mid \upsilon_n, \lambda_n\right) \qquad l \in L_n, \ i \in I_n = 1,0 \tag{6.22}$$

The behavior of driver n is observed over a sequence of T_n consecutive time intervals. Assuming that, conditional on υ_n and λ_n, the observations are independent, the joint probability of the sequence of observations, is given by:

$$P_n\left(j_1, j_2, ..., j_T \mid \upsilon, \lambda\right) = \prod_{t=1}^{T_n} P_n\left(j_t \mid \upsilon_n, \lambda_n\right) \tag{6.23}$$

The unconditional individual likelihood function ($\mathcal{L}_n$) is obtained by integrating over the distributions of the individual-specific variables:

$$\mathcal{L}_n = P_n\left(j_1, j_2, ..., j_T\right) = \int_\upsilon \int_\lambda P_n\left(j_1, j_2, ..., j_T \mid \upsilon, \lambda\right) f(\upsilon) d\upsilon f(\lambda) d\lambda \tag{6.24}$$

Where, $f(\upsilon)$ and $f(\lambda)$ are assumed to have normal and doubly truncated normal distributions respectively.

Assuming that the observations from different drivers are independent, the log-likelihood function for all N individuals observed is given by:

$$\mathcal{L} = \sum_{n=1}^{N} \ln(\mathcal{L}_n) \tag{6.25}$$

The maximum likelihood estimates of the model parameters are found by maximizing this function.

Estimation Results

All components of the model have been estimated jointly using a maximum likelihood estimation procedure as described in the previous section. However, in order to simplify the presentation, estimation results for the target lane choice, gap acceptance and execution levels are presented and discussed separately.

The summary of estimation results of the proposed lane changing model is presented in Table 6.9.

Table 6.9: Estimation results of the target lane changing model

Final log-likelihood	-1003.2
Initial log-likelihood	-2094.9
Number of drivers	400
Number of observations	16696
Number of parameters	22
Adjusted rho-bar square	0.53

The improvement in the goodness-of-fit of the new model was compared with the simpler lane-shift model proposed by Toledo *et al.* (2003) and illustrated in Figure 6.17. In this model, the driver evaluates the current and the adjacent lanes and decided whether or not to make a lane change. The model was reestimated with the same Lankershim Boulevard arterial data. The detailed model structure and estimation results are presented in Appendix C.4. The statistical tests for comparing non-nested models imply that the new model has a statistically significant improvement in goodness-of-fit. The test results are presented in Table 6.10.
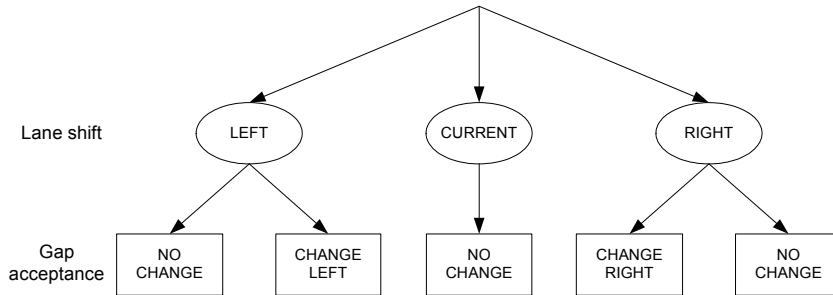
Figure 6.17: Structure of lane-shift model (Toledo et al. 2003)

The model with explicit target lane choice has larger values in terms of both statistical test criteria, which indicates that it better fits the data, and supports the inclusion of the latent planning in the model framework.

Table 6.10: Model comparison

Statistic	Lane Shift (R)	Target lane (U)
Log-likelihood value	-1186.9	-1003.6
Number of parameters (k)	17	22
Akaike information criteria (AIC)	-1203.9	-1126.1
Adjusted rho-bar square ($\bar{\rho}^2$)	0.441	0.531

Choice of Plan: The Target Lane Selection and the Gap Acceptance Models

The lane that the driver perceives to be the best is selected as the target lane. The choice set of the driver includes all available lanes in the freeway stretch. The utility of lane target lane *l* of individual *n* at time *t* can be expressed as follows:

$$U_{lnt} = \beta^T X_{lnt}(\lambda_n) + \alpha^l \upsilon_n + \varepsilon_{lnt} \qquad \forall l \in L_n \tag{6.26}$$

Where,

$\lambda_n = look\text{-}ahead \ / \ plan\text{-}ahead \ distance \ of \ the \ driver$

$X_{lnt} = explanatory \ variables \ that \ affect \ the \ utility \ of \ lane \ l, \ function \ of \ \lambda_n$

$\beta = corresponding \ vector \ of \ parameters$

$\upsilon_n = individual\text{-}specific \ random \ effect \ (e.g. \ aggressiveness): \upsilon_n \sim N(0,1)$

$\alpha^l = coefficient \ of \ individual\text{-}specific \ random \ effect \ for \ lane \ l$

$\varepsilon_{lnt} = random \ term \ associated \ with \ the \ target \ lane \ utilities$

$L_n = choice \ set \ of \ target \ lane \ of \ driver \ n$

As discussed in a previous section, the target lane choices are affected by the variables related to the path-plan and inertia, the neighborhood variables and the attributes of the alternative lanes, as well as driver-specific characteristics. However, not all of the candidate variables were found to be statistically significant and/or have intuitive signs. For example, the coefficients of neighborhood variables were not found to be significantly different than zero supporting the hypotheses that the path-plan considerations, inertia effect and lane attributes are pre-dominant factors behind arterial lane changing decisions. In case of some of the variables, interaction of multiple variables have been included in the model (based on the goodness-of-fit improvements). The estimated parameter values and the corresponding t-statistics for the target lane selection models are presented in Table 6.11.

Table 6.11: Estimation results of the target lane selection model

Target Lane Selection		Parameter	t-stat
Variable		Parameter	t-stat
Inertia effect	Current lane dummy	0.168	0.21
	Heterogeneity term for current lane, α^{CL}	-0.479	-3.12
	Lanes away from the current lane	-3.71	-4.71
Path-plan impact	Lanes away from exit/turn lane	-2.22	-8.41
	Lanes away from exit/turn lane interacted with distance from exit	-0.338	-1.05
	Exponent of distance to exit in the interaction term	-4.10	-2.51
Queue ahead (vehicles/lane)	Less than 3 vehicles	-0.408	-4.00
	More than 3 vehicles	-1.24	-4.33
Mean plan-ahead distance (10 m)		37.9	10.39
Std deviation of plan-ahead distance (10 m)		4.04	12.83

As can be seen from the estimation results, the lane utilities are affected by the path-plan related variables, the current lane inertia variables, the lane-specific attributes (queue length) and driver-specific characteristics (planning capability and aggressiveness). The heterogeneity in planning capability of drivers is captured by the mean and standard deviation of plan-ahead distance (assumed to be truncated normally distributed). The target lane utilities can be expressed as follows:

$$U_{lnt} = (0.168 - 0.479\,\upsilon_n)\,\delta_{lnt}^{CL} - 3.71\,\Delta CL_{lnt} - 2.22\,\Delta Exit_{ln}$$
$$- 0.338\,\varepsilon_{lnt}\,(d_{nt}^{exit})^{-4.10} - 0.408\,(q_{lnt})(q_{lnt} < 4) - 1.24(q_{lnt} > 3) \tag{6.27}$$

Where,

δ_{lnt}^{CL} = *current lane dummy, 1 if lane l is current lane, 0 otherwise*

q_{lnt} = *queue ahead in lane l*

ΔCL_{lnt} = *number of lane changes required from current lane to lane l*

$\Delta Exit_{ln}$ = *number of lane changes required to take the desired exit/turn of driver n from lane l*

d_{nt}^{exit} = *remaining distance to exit/turn*

The trade-off between the effects of the path-plan and inertia variables in the utility of the target lanes is illustrated in Figure 6.18, for a standard case of 4 lanes (which represents the typical case in the current dataset), with Lane 4 being the exit/turn lane. In these cases, it is assumed that all drivers are taking the path-plan into account, that is their plan-ahead distances are greater that 400 meters.

In the first case, the driver is far from the exit/turn (remaining distance 400m which is about 2 sections in the study dataset) and the current lane inertia dominates the target lane selection (Figure 6.18a). As the driver approaches the turn, the path-plan effect starts to dominate and the preference shifts to the exit/turn lane (Lane 4 in this case). Thus when the driver is 60 m from the exit/turn, there is a very high probability of choosing Lane 4 irrespective of the current lane, as illustrated by Figure 6.18b.
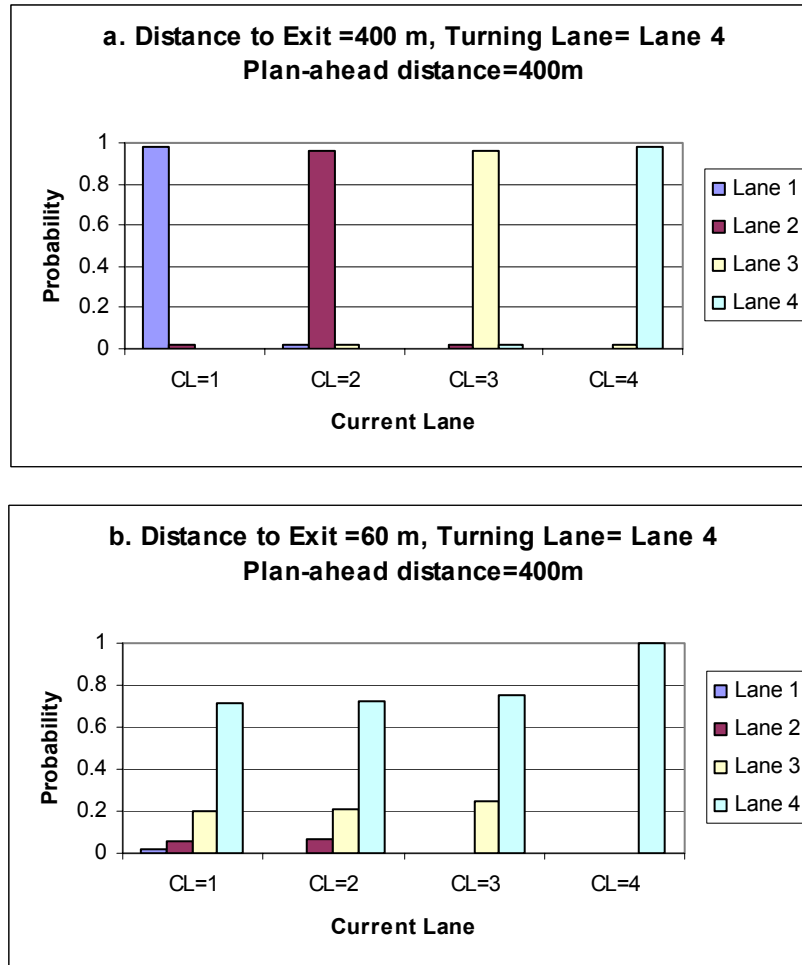
Figure 6.18: Trade-off between current lane inertia and path-plan effect

The plan-ahead distance of the driver also has a significant impact on how the path-plan considerations affect the utilities. All else being equal, the lane preferences for a driver with 50 m plan-ahead distance (as opposed to 400m as in Figure 6.18) is presented in Figure 6.19.

In this case, the path-plan is not a factor until the driver reaches the last 50m before the exit/turn. Therefore, even when he/she is 60m from the exit/turn the current lane inertia dominates the target lane selection.
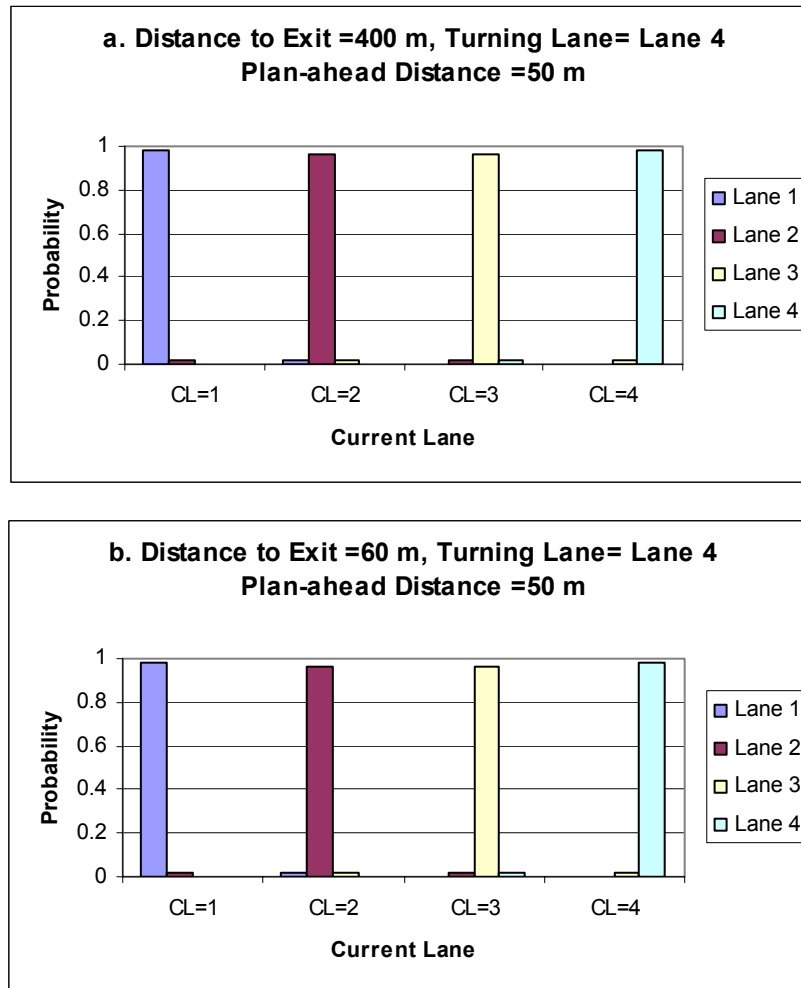
Figure 6.19: Trade-off between current lane inertia and path-plan effect

The direction of the target lane indicates the direction of immediate lane change and the driver is assumed to evaluate the adjacent gap in the immediate target lane and decide whether or not to change lanes. In order for the gap to be acceptable, both the lead and lag gaps must be acceptable. That is, the available lead and lag gaps must be larger than the corresponding critical gaps. As presented in Equation 6.17, in order to ensure that the critical gaps are always positive, they are assumed to follow lognormal distributions:

$$\ln(G_{lnt}^{lead\ cr}) = \beta^T X_{lnt}^{lead} + \alpha^{lead} \upsilon_n + \varepsilon_{lnt}^{lead}$$
$$\ln(G_{lnt}^{lag\ cr}) = \beta^T X_{lnt}^{lag} + \alpha^{lag} \upsilon_n + \varepsilon_{lnt}^{lag}$$

(6.28)

Where,

$G_{lnt}^{lead\ cr}, G_{lnt}^{lag\ cr}$ = *lead and lag critical gap in the direction of target lane l, measured in distance units (e.g. meters)*

$X_{lnt}^{lead}, X_{lnt}^{lag}$ = *explanatory variables that affect the lead and lag critical gaps respectively in the direction of target lane l*

$\alpha^{lead}, \alpha^{lag}$ = *coefficients of individual-specific latent variable υ_n for lead and lag gap acceptance*

$\varepsilon_{lnt}^{lead}, \varepsilon_{lnt}^{lag}$ = *random terms: $\varepsilon_{lnt}^{lead} \sim N\left(0, \sigma_{lead}^2\right), \varepsilon_{lnt}^{lag} \sim N\left(0, \sigma_{lag}^2\right)$*

The critical gaps are likely to be affected by the speed, position and type of the lead and lag vehicles, remaining distance to the desired turn etc. However, not all candidate variables were supported by the data. For example, the remaining distance to the desired turn did not have any significant effect on critical gaps.

The estimation results for the gap acceptance model are presented in Table 6.10.

Table 6.12: Estimation results for the gap acceptance model

Gap Acceptance		
Variable	Parameter	t-stat
Lead Critical Gap		
Lead gap constant	2.38	32.23
Relative lead speed, ΔV_{nt}^{lead} (m/s)	-0.0216	-1.42
Standard deviation of lead gap, σ_{lead}	0.00761	0.07
Heterogeneity coefficient of lead gap, α^{lead}	-1.75	-36.39
Lag Critical Gap		
Lag gap constant	1.44	25.68
Relative lag speed, ΔV_{nt}^{lag} (m/s)	0.264	14.73
Standard deviation of lag gap, σ_{lag}	0.00851	0.21
Heterogeneity coefficient of lag gap, α^{lag}	-1.86	-40.38

The critical gaps are affected by relative speeds of the associated lead and lag vehicles and the aggressiveness of the driver of the subject vehicle.

The critical gap of the driver can be expressed by the following equation:

$$G_{nt}^{lead\ cr} = \exp(2.38 - 0.0216\Delta V_{lead\ nt} - 1.75\upsilon_n + \varepsilon_{lead\ nt})$$

$$\varepsilon_{nt}^{lead} \sim N(0, 0.00761^2)$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (6.29)$$

$$G_{nt}^{lag\ cr} = \exp(1.44 + 0.264\Delta V_{lag\ nt} - 1.86\upsilon_n + \varepsilon_{lag\ nt})$$

$$\varepsilon_{nt}^{lag} \sim N(0, 0.00851^2)$$

Where,

$$\Delta V_{nt}^g = V_{nt}^g - V_{nt}$$

$V_{nt} =$ *speed of subject vehicle n at time t*

$V_{nt}^g =$ *speed of vehicle associated with gap g of subject n at time t*

The critical gap decreases with the relative lead speed, i.e. it is larger when the subject vehicle is faster relative to the lead vehicle. The lag critical gap increases with the relative lag speed: the faster the lag vehicle is relative to the subject, the larger the lag critical gap. The influence of the included explanatory variables on the critical gap lengths are summarized in Figure 6.20 and Figure 6.21.
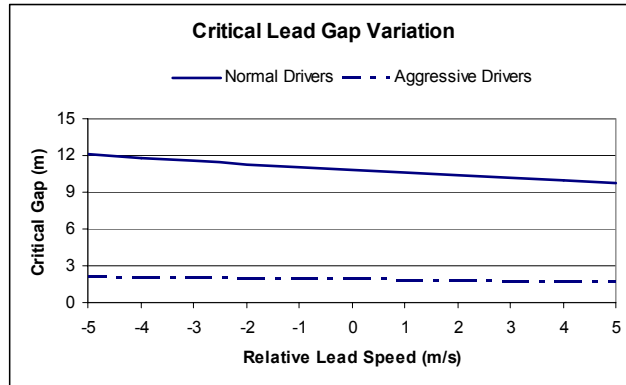
Figure 6.20: Variation of lead critical gap with lead speed and aggressiveness
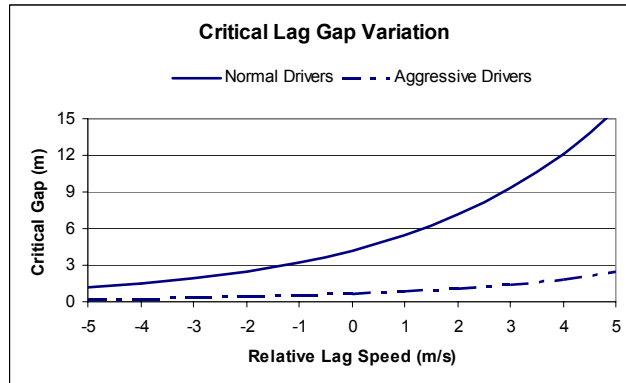
Figure 6.21: Variation of lag critical gap with relative lead speed and aggressiveness

As seen in the figures, the sensitivity to the relative speed is higher for the lag gap than the lead gap. For both cases, the critical gaps are smaller for aggressive drivers than for normal drivers.

Choice of Action: Execution of Change

Even if both lead and lag gaps are acceptable, it may take time to complete the lane change. The duration of the lane change is reflected by the execution decisions of the lane change. The probability of executing the lane change is expressed as follows:

$$P_n(j_t|i_t,l_t,\upsilon_n) = \begin{cases} \dfrac{1}{1+\exp\left(-\left(\beta^T X_{jlnt}+\alpha^j\upsilon_n\right)\right)} & \text{if } i_t=1 \ (accept\ gap) \\ 0 & otherwise \end{cases}$$

(6.30)

Where,

j_t = lane changing action

X_{jlnt} =explanatory variables that affect the driver's execution decision

α^j = parameter of individual-specific latent variable υ_n for execution level

In the execution model, several variables were tested: the density and average speed of the traffic stream, relative lead and lag speeds, rate of change in gap size etc. But none of these variables gave the correct signs in estimation. In the final model, only the speed, the indicator of change in gap reduction and aggressiveness of the subject driver were included along with an intercept. The estimation results are presented in Table 6.11.

Table 6.13: Estimation results of the execution model

Execution Decision		
Variable	Parameter	t-stat
Intercept	-3.26	-6.41
Speed of subject vehicle, V_{nt}	0.627	6.44
Gap reduction dummy, ξ_{nt}^i	0.593	0.68
Heterogeneity coefficient of execution α_j	0.266	0.42

The executions of the lane change of the driver, given the adjacent gaps in the direction of the target lane are acceptable, can be expressed by the following equation:

$$P_n\left(j_t\!=\!1|i_t\!=\!1,l_t \neq CL\right)=\cfrac{1}{1+\exp\left(-\left(-3.26+0.627V_{nt}+0.598\xi_{nt}^i+0.266\upsilon_n\right)\right)} \tag{6.31}$$

Where,

$\xi_{nt}^i = 1$ *if the accepted adjacent gap* (i_t) *is reducing that is* $((V_{lag\,nt}-V_{lead\,nt})>0)$

The negative intercept indicates that all else being equal, there is delay associated with the lane change. The execution of the lane change is faster if the speed of the subject vehicle is high. If the corresponding adjacent gap is reducing (the lag vehicle is faster than the lead vehicle) the execution of the lane change becomes more urgent and the lane change is faster. The coefficient of aggressiveness of the driver was found to be positive which agrees with the earlier hypothesis that aggressive drivers have less inertia to stay in their current lanes and require less time to execute the lane change.

It is expected that the probability of executing the lane change will be lower if the time resolution of the data is smaller. However, the effect of the time step could not be tested in this study since the entire data had the same time resolution of 1 sec. The effect of time step is thus embedded in the intercept term and needs to be calibrated if the model is implemented in a simulator with a time step different than 1 sec.

Distribution of Plan-ahead Distance

The plan-ahead distance of drivers is assumed to follow a normal distribution truncated on both sides. The distribution is given by:

$$f(\lambda_n)=\begin{cases}\cfrac{\dfrac{1}{\sigma_\tau}\phi\left(\dfrac{\lambda_n-\mu_\tau}{\sigma_\tau}\right)}{\Phi\left(\dfrac{\lambda_{max}-\mu_\tau}{\sigma_\tau}\right)-\Phi\left(\dfrac{\lambda_{min}-\mu_\tau}{\sigma_\tau}\right)} & \text{if } \lambda_{min}\leq\lambda_n\leq\lambda_{max}\\[2em]0 & \text{otherwise}\end{cases} \tag{6.32}$$

Estimation results indicated that it is normally distributed within 50m to 500m.[8] The estimated distribution of anticipation time can be expressed as follows:

---

[8] Different ranges between 30 m to 600 m were tested as the upper limit of anticipation time and the selected range provided the best goodness-of-fit.

$$f(\tau_n) = \begin{cases} \dfrac{1}{40.35}\phi\left(\dfrac{\tau_n - 379}{40.44}\right) & \text{if } 50 \leq \tau_n \leq 500 \\ 0 & \text{otherwise} \end{cases} \tag{6.33}$$

6.3   Aggregate Validation

This section describes the aggregate validation process including the dataset used, the details of the calibration, the validation process that was applied, and the results obtained.

The calibration process involves adjusting the values of the parameters of the behavioral models and estimating travel demand, in the form of OD flows, on the network being studied in order to obtain a better fit of the model output with the actual traffic flow. In this study, the trajectory data collected from the same site was used for calibration and validation in the absence of other suitable data. Exact vehicle OD flows were available from the trajectory data and no route choice was involved. The calibration process therefore only involved calibration of the driving behavior parameters.

6.3.1   Data

The total dataset was available for a 32 minute period (8:28 a.m. to 9:00 a.m.). The first 22 minutes of data in the north bound direction was used for calibration and the remaining 10 minutes was used for validation.

The trajectory data was aggregated to generate synthetic sensor counts and speeds that are used for calibration. The locations of these sensors are illustrated in Figure 6.22.
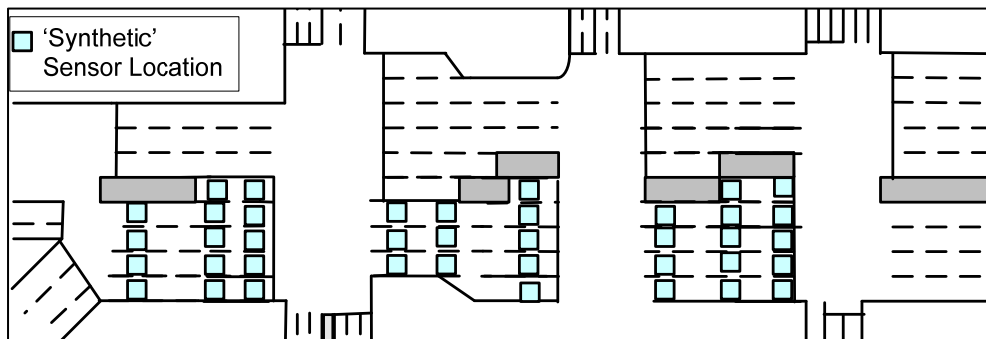
Figure 6.22: Locations of synthetic sensors

6.3.2 Aggregate Calibration

Based on previous experience and sensitivity test results (described in Section 4.4.2), the following parameters of the combined model were selected for calibration:

- Acceleration and deceleration constants;
- Desired speed mean and sigma;
- Intercepts (constants) and variance (sigmas) of critical gap;
- Constant in the execution level;
- Intercept (constant) of lane 3 in the intersection lane choice model.

The calibration parameters and their before and after values are listed in Table 6.14.

Table 6.14: Calibration parameters

Model/Variable	Calibrated Parameter	Parameter Value	
		Initial	Calibrated
Car following*	Acceleration constant	0.0400	0.042
	Deceleration constant	-0.0420	-0.029
Desired Speed*	Mean	0.100	0.056
	Variance	0.150	0.540
Gap Acceptance	Lead gap constant	2.38	1.56
	Lead gap sigma	.00751	0.0406
	Lag gap constant	1.44	-0.0612
	Lag gap sigma	.00845	0.0517
Intersection Lane Choice	Lane 3 constant	1.31	1.10
	Target lane dummy	3.16	2.13
Within Section	Away from exit lane	-1.27	-0.0101
	Current lane dummy	3.02	1.57
	Execution constant	-3.38	-1.37

*Note: General parameters of MITSIMLab. These variables are described in Ahmed (1999).

The driver-specific variables (desired speed, plan-ahead distance and the coefficients of aggressiveness) are expected to vary since the model was validated at a different site from the estimation data collection site. Among these parameters, the desired speed parameter, the current lane dummy and the execution constant made the most significant contributions to improving the performance of the model. When these parameters were unconstrained, the model performed better (the objective function for calibration improved significantly) compared to the case when these parameters were fixed at the originally estimated values. The improvements after the calibration are presented in Table 6.15.

Table 6.15: Calibration results

Lane-specific Counts			
	Before Calibration	After Calibration	Improvement
RMSE (vehicles/20 mins)	18.80	15.70	16.49%
RMSPE	0.83	0.73	12.05%
Lane-specific Speeds			
	Before Calibration	After Calibration	Improvement
RMSE (mph)	24.05	12.65	47.40%
RMSPE	1.32	0.64	51.52%

The original MITSIMLab lane changing model reestimated with arterial data (referred to as base model) parameters were also calibrated in a similar manner.

6.3.3 Aggregate Validation

The purpose of system validation is to determine the extent to which the simulation model replicates the real system. At this step, the behavior parameters obtained in the system calibration step are fixed, and the model predictions are compared against the second set of traffic measurements, which were not used for calibration.

The validation process is comparative. In this study, the goodness-of-fit statistics of the new model are compared with those of the base MITSIMLab model that includes a lane-shift model (Toledo *et al.* 2003). The details of the model are shown in Appendix C.4.

In this study, the sensor measurements used for the validation are the 'synthetic' sensor counts and speeds generated using the last 10 minutes of available trajectory data (8:50 a.m. to 9:00 a.m.).

The validity of the calibrated model was tested using the several measures of effectiveness (MOEs) that were obtained from the synthetic sensor data and from the summaries of the trajectory data. These included measures related to the mainline traffic conditions as well as measures related to the merging lane (auxiliary lane) traffic conditions:

- Lane-specific flows
- Lane-specific speeds
- Lane distributions by location

Lane-specific Flows

Lane-specific flow (vehicle/unit time) was compared among the observed data, new arterial models and default MITSIMLab models. As seen in Table 6.14, the new model performs better than the base model for all measures.

Table 6.16: Comparison of lane-specific counts

	Base	New	Improvement
RMSE (vehicles/ 5 mins)	13.71	12.26	10.58%
RMSPE	0.59	0.49	16.95%
ME (vehicles/5 mins)	4.91	-0.34	93.08%
MPE	0.25	-0.004	98.40%

Lane-specific Speeds

Speed distribution in lanes was compared among the observed data, new arterial models and default MITSIMLab models. As seen in Table 6.15, the new model performs better in terms of *RMSE* and *RMSPE*. But the significant improvements are in *ME* and *MPE*.

RMSE and *RMSPE* tend to penalize large errors. The results suggest that there are large discrepancies in some of the observations of the new model that contribute to the large *RMSE* and *RMSPE* values. But the average errors are not high. As discussed before the limitations associated with absence of accurate signal inputs could have resulted in such errors.

Table 6.17: Comparison of lane- specific speeds

	Base	New	Improvement
RMSE (mph)	12.61	11.84	6.11%
RMSPE	2.51	1.13	54.98%
ME (mph)	9.56	4.51	52.82%
MPE	0.76	0.48	36.84%

Lane Distributions by Location

Distribution of vehicles in lanes was compared among the observed data, base model and the new models. In each section distributions are calculated at three locations:

Station 1: At the beginning of the section

Station 2: In the middle of the section

Station 3: At the end of the section.

The locations of the sensors are presented in Figure 6.23.
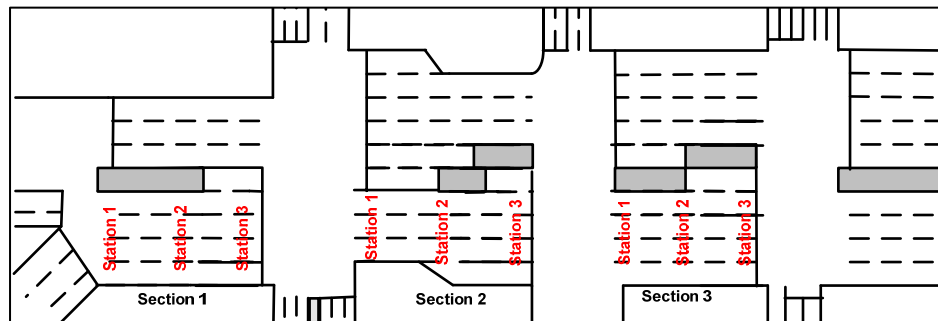
Figure 6.23: Locations of the sensors

The results for the north bound sections are presented in Figure 6.24Figures 6.24, 6.25 and 6.26. As seen in the diagrams, the lane distributions of the new model have a better fit to the observed data than the lane shift model. Specially, the lane shift models over predict the through lane occupancies. The latent plan models better capture the vehicle positioning.
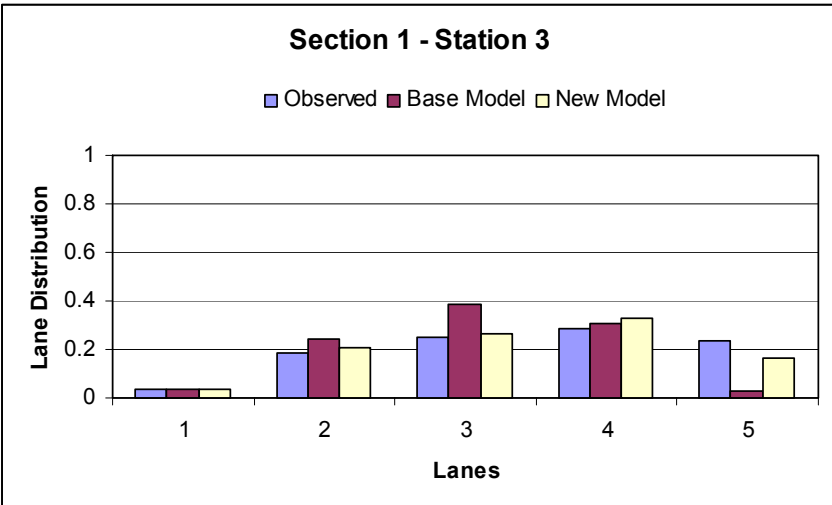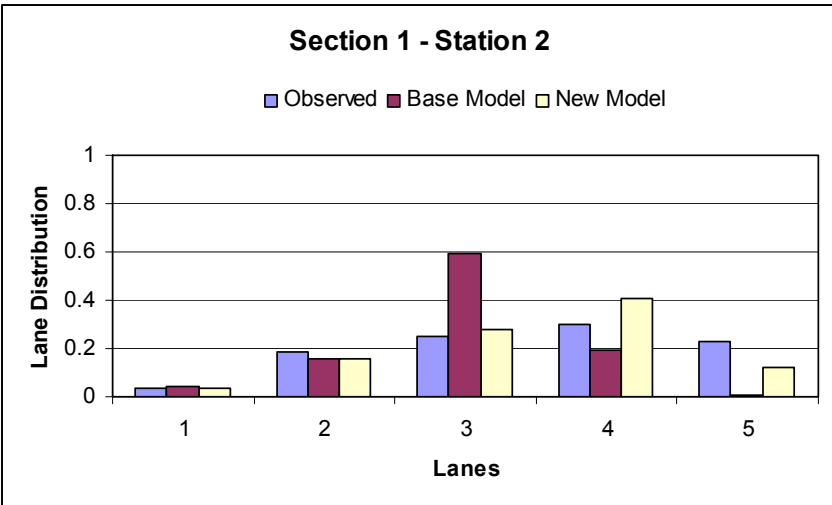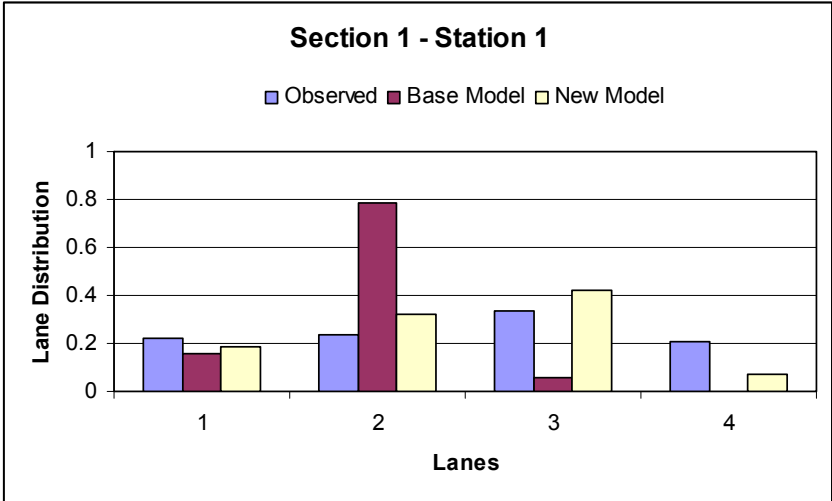
Figure 6.24: Comparison of lane distributions (Section 1)
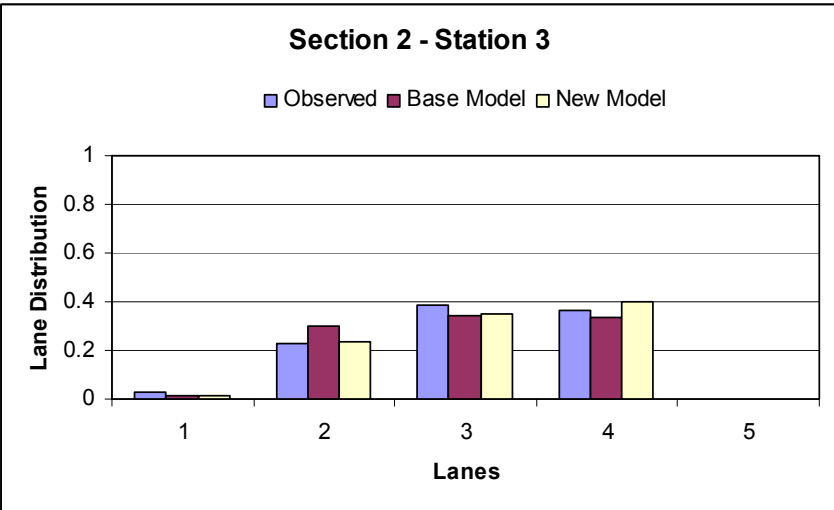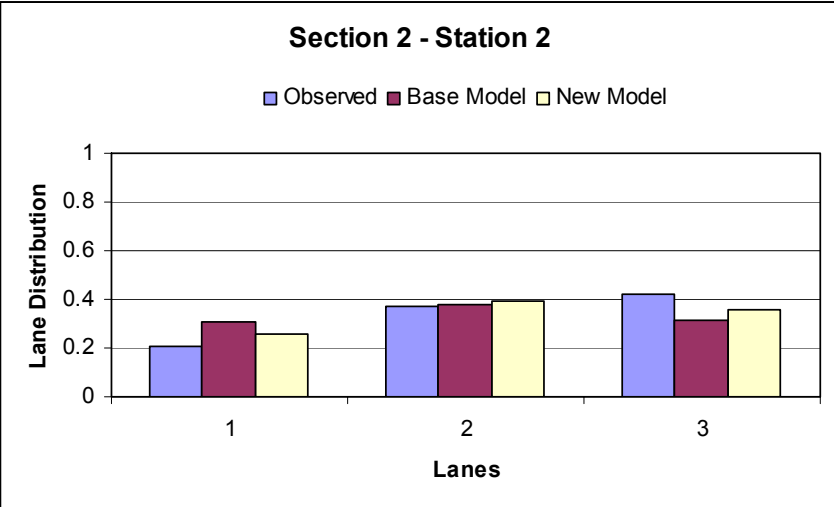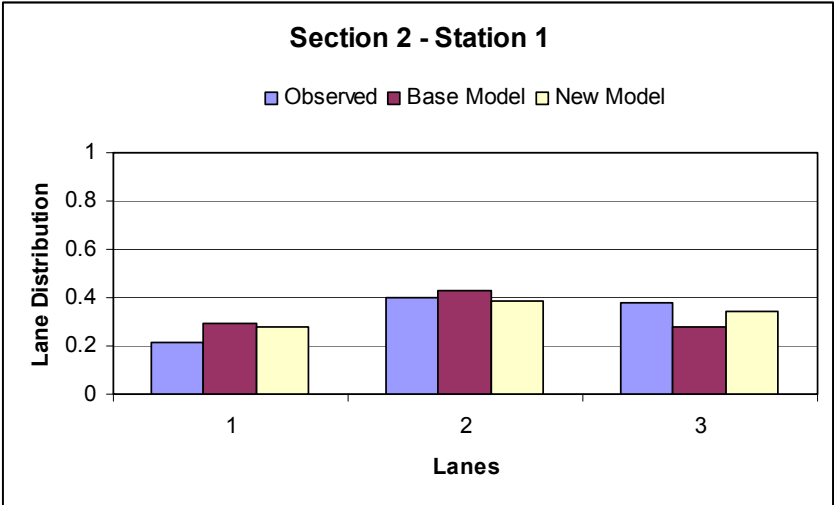
Figure 6.25: Comparison of lane distributions (Section 2)
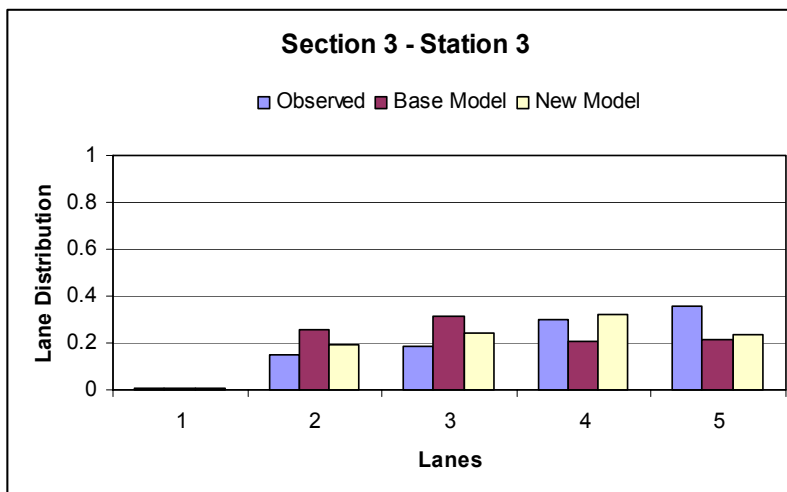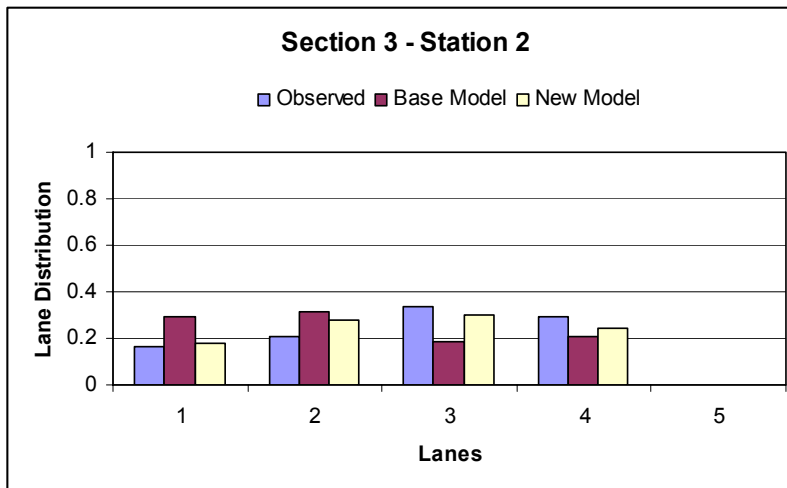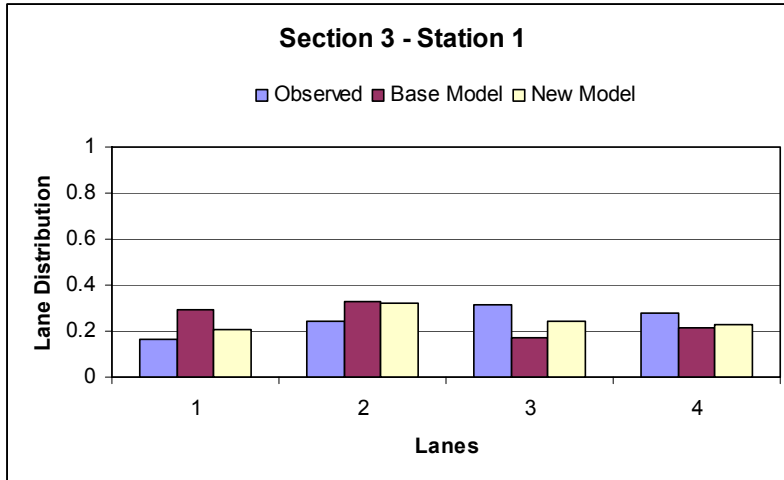
Figure 6.26: Comparison of lane distributions (Section 3)

6.4 Summary

A latent plan based intersection lane choice and a mainline lane changing model were presented in this chapter. Both models were estimated with detailed trajectory data collected from Lankershim Boulevard, in Los Angeles, California.

The intersection lane choice model involves the lane choice of drivers entering the arterial from a side street. The choice is modeled as a two step process: target lane choice (plan) and immediate lane selection based on the target lane selection (action). The choice of target lane was unobserved and only the final maneuvers of the driver to immediate lanes were observed. The choice of target lane is influenced more by path-plan variables and lane-specific attributes whereas immediate lane choices are governed by maneuverability considerations. The heterogeneity of the driver population was explicitly taken into account in the model formulation. In particular, the planning capability of the driver was allowed to vary among drivers using a latent class model formulation.

The lane changing model for the mainline involves target lane choice, gap acceptance decisions to make a lane change towards the direction of the target lane, and execution of the lane change to the accepted gap. The choice of target lane is unobserved and only the final lane actions are observed. The choice of target lane was found to be influenced by neighboring vehicle speeds and positions; lane-specific attributes such as queue length; and factors such as the path-plan of the driver and driver characteristics (planning capability and aggressiveness). The planning capability of the driver was allowed to vary among drivers using a latent 'plan-ahead' distance of the driver. Gap acceptance was influenced by relative speeds of lead and lag vehicles. The execution or completion of the lane change was found to be a function of the speed of the driver.

The goodness-of-fit of the latent plan models were compared against simpler models estimated with the same data in each case. Statistical tests on estimation results showed significant improvement in the goodness-of-fit of the latent plan models.

The new models were validated against simpler reduced form models that do not address the latent targets behind the actions. These models consisted of a rule based intersection lane choice model and a simpler within section lane shift model re-estimated with the arterial data. The measures of validation included comparison of the lane-

specific flows and speeds, lane distributions in different locations, number of lane changes per vehicle and number of incomplete trips. The validation results support improvement in the simulation capabilities of the new models.

One issue of interest within the domain of driver behavior models is the ability to capture the effect of the time resolution of data on the estimation results. The introduction of a third level (execution level) in the lane changing model structure to model the lane change execution decision facilitates the explicit consideration of the impact of the time resolution on the driving decisions. This is an aspect of research related to this work that can be explored in the future.

A critical feature that has not yet been incorporated within the model structure for lane changing and acceleration decisions of a mainline driver has been that of state-dependence among the successive decisions of an individual in traffic. For example, in this research, if the gaps are acceptable but the lane change is not completed, the target lane and target gap decisions are reevaluated at the next time step. This approach assumes that the state-dependence is captured by the explanatory variables (if a gap is acceptable now it will continue to be acceptable till the lane change is completed). However, the validity of this assumption needs to be tested by comparison with models with explicit consideration for state-dependence. This would form a very important extension to the current work.

Chapter 7

Conclusion

This chapter summarizes the research presented in the thesis and highlights the major contributions. Directions for future research are suggested at the end.

7.1   Summary

In many situations, drivers first select a plan by choosing a target or tactic. The plan is then manifested through subsequent maneuvers. The plan however is unobserved or latent and only the end actions are observed. The effect of planning is more evident in certain geometric and traffic conditions, for example in freeway/arterial with closely spaced exits/turns, in congested or incident affected situations etc. Ignoring the plans behind the actions can lead to unrealistic traffic flow characteristics and incorrect representation of congestion particularly in the above mentioned scenarios. This was reflected in the findings of the NGSIM study on Identification and Prioritization of Core Algorithm Categories (Alexiadis *et al.* 2004), where the urban arterial lane selection, oversaturated freeway behavior, freeway lane changing and weaving section behaviors were identified as weak points of the state-of-the-art traffic simulation tools.

The focus of this thesis is on latent plan based driving behavior models that capture the unobserved plans behind the observed driving decisions. This new framework includes the effects of past decisions as well as anticipated future conditions in the current decisions of drivers.

In this thesis, a general methodology for modeling behaviors with unobserved or latent plans was presented in two phases. The methodology was presented first for a basic

case with only serial correlation and no state-dependence and was then extended for a case with state-dependence.

For a case without any state-dependence, the plans and actions of the driver conditional on individual-specific driving characteristics were assumed to be independent over time. The interdependencies and causal relationships between the choice of plan and choice of action of the same driver were captured through individual-specific latent variable of aggressiveness. The aggressiveness of the driver remains unchanged across all choice dimensions and all observations of the same driver. In addition, individual-specific planning capability of the drivers was introduced in the latent plan model framework. Two approaches were used for this: a continuous latent 'plan-ahead' distance approach and a latent class approach. In the continuous latent plan-ahead distance approach, the plan-ahead distance was assumed to have a truncated normal distribution and the parameters of the distribution were estimated along with other parameters. In the latent class approach, the drivers were assigned to discrete classes based on their plan-ahead distances.

The methodology was then extended to capture the effects of state-dependence between subsequent plans and actions of the driver. This was done with a Hidden Markov Model (HMM), which was employed to make the state-dependent model computationally tractable. According to the HMM assumptions, the current plan depends only on the plan of the previous time step (i.e. not on all previous plans) and all previous actions. The current action depends only on the current plan. These assumptions enabled calculating the probability of the plans recursively.

The latent plan modeling approach was demonstrated by lane changing models of drivers in freeway and urban traffic scenarios. These include the following lane selection scenarios:

- Freeway mainline lane changing
- Freeway on-ramp merges
- Urban arterial intersection lane choice and
- Urban arterial mainline lane changing.

The general framework was the same in all these cases, latent plans followed by observed actions. As described in Section 1.2, in the general case, the latent plans can

include multiple levels: target lane selection, target gap selection, choice of lane changing tactic and gap acceptance for execution of the merge for example (Figure 7.1). However, depending on the geometric and traffic attributes, one or more of these levels may be redundant in particular scenarios. For example, in a moderately congested freeway situation, if acceptable gaps are readily available, the target gap is always the adjacent gap and the lane change is always through normal gap acceptance. The target gap selection and choice of lane changing tactic are therefore redundant in such situations. Similarly, in a freeway merging situation the target lane is by default the rightmost lane of the freeway. The detailed structures of the estimated models were thus guided by characteristics of the trajectory data used for the model estimation.
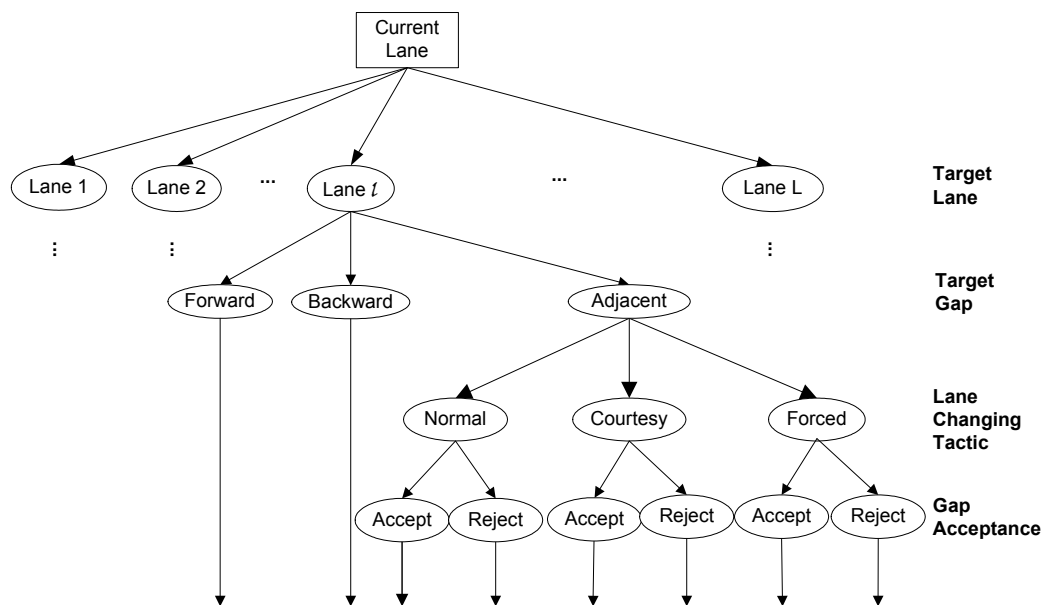
Figure 7.1: Framework of choice of plan

The frameworks of the estimated freeway and arterial models are presented in Figures 7.2 and 7.3 respectively (the lane indices start from the left, i.e. Lane L indicates the rightmost lane).

In the freeway lane selection models (Figure 7.2), the lane selection plan was represented by choice of target lane and lane changing tactics. However, the behavior of the mainline and on-ramp drivers have been modeled separately with different data. For the mainline driver, because of lower level of congestion in the estimation data, the lane changes were through normal gap acceptance and choice of lane changing tactic

component was redundant. The on-ramp merging model on the other hand was estimated with data from a highly congested situation and supported the choice of lane changing/merging tactic component. In the merging model, the initial target lane was observed (Lane L, the rightmost lane in the mainline) and the lane selection model was redundant.
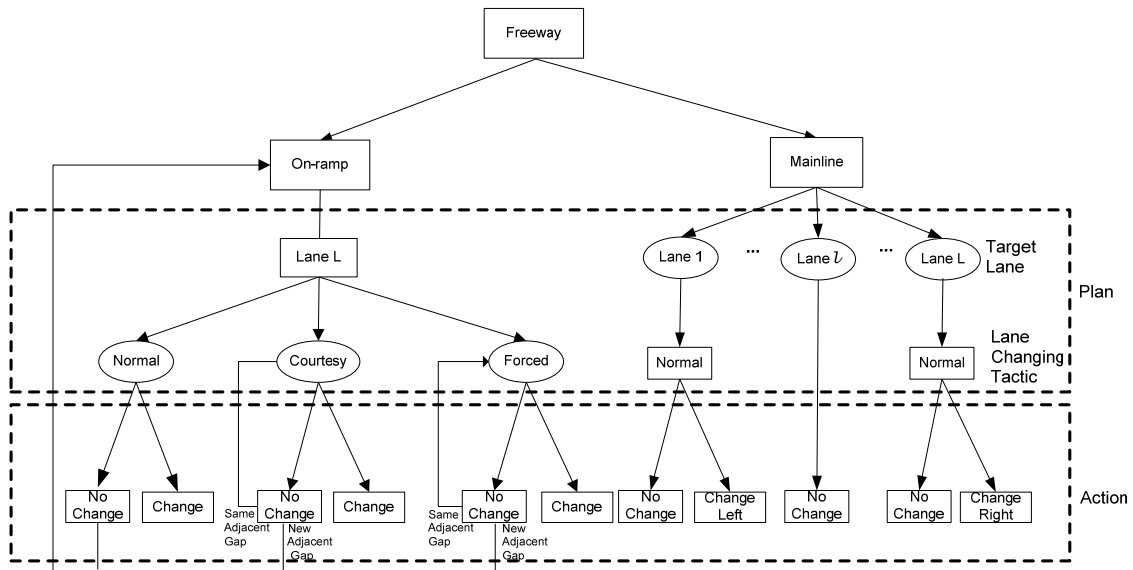
Figure 7.2: Estimated model framework for freeway lane selection model
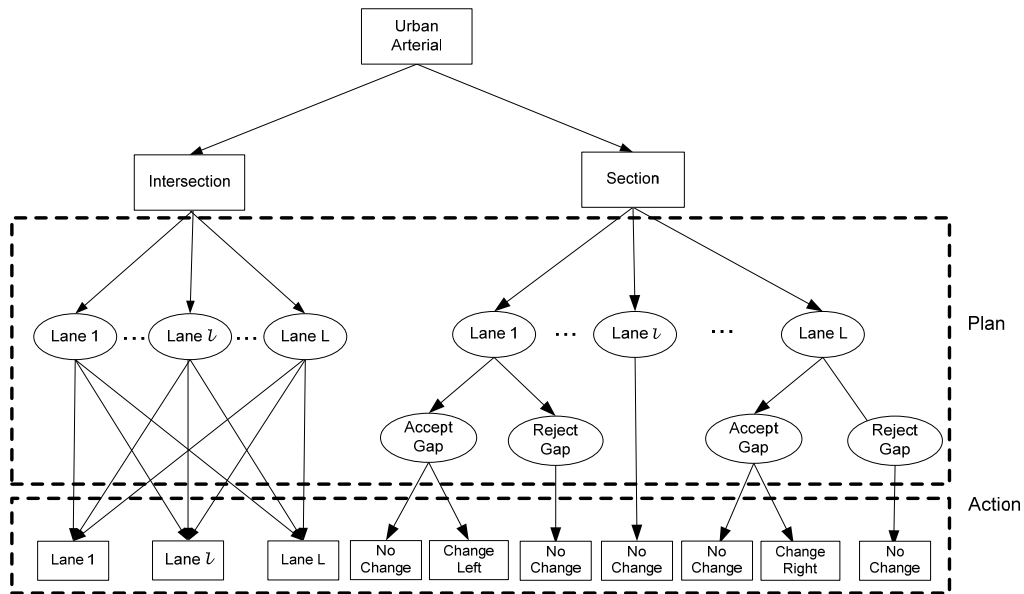
Figure 7.3: Estimated model framework for urban arterial lane selection model

In the arterial lane selection models (Figure 7.3), the intersection lane choice and mainline lane changing models were estimated separately using data from the same site. In the intersection lane choice case, because of the intermittent nature of decisions (rather than continuous as in the lane changing case), a different plan and action framework was adapted where the plan includes the choice of the target lane and the action involves choice of the immediate lane based on the target lane. In the mainline arterial model, the duration of the lane changes was found to have a significant impact and was modeled as a separate decision level in addition to the target lane and gap acceptance levels as in the freeway mainline.

General Findings

The new set of models was evaluated using comparison of goodness-of-fits of estimation results and aggregate validation results within MITSIMLab. The goodness-of-fits of the new models were compared with simpler reduced form models for each case. The reduced form models were estimated with the same data but they do not model the latent plans of the driver. Statistical tests on estimation results indicated significant improvement in the goodness-of-fit of each of the four latent plan models.

For evaluation of the enhancements in the simulation capabilities of the latent plan models, the models were implemented in MITSIMLab and validated using aggregate data. Part of the available aggregate data was used first to calibrate the overall simulation system. The remaining aggregate data (not used for calibration) were then compared with the corresponding outputs of the calibrated MITSIMLab. Validation results also supported the improved performance of the latent plan models as compared to the myopic models. The specific findings for each of the models are presented below:

Freeway Lane Changing

In freeway lane changing, an explicit choice of target lane was introduced to capture the effect of latent plans in the immediate maneuvers of the driver. In this approach, the driver selects as the target the lane he/she perceives to be the best among all available lanes. The driver then looks for gaps in the adjacent lane in the direction of the target lane. A lane change is executed in the direction of target lane when both lead and lag gaps in the adjacent lane are acceptable. This approach differs from existing models that

assume that drivers evaluate the current and adjacent lanes and choose a direction of change (if any) based on the relative utilities of these lanes. The choice of target lane was found to be affected by lane-specific attributes, path-plan considerations and speed of neighboring vehicles as well as individual-specific characteristics like aggressiveness (latent) of the driver. The gap acceptance decisions were found to be affected by relative speed and spacing of lead and lag vehicles in the adjacent lane (in the direction of the target lane) and the aggressiveness of the driver.

While the proposed latent plan model is applicable in any general freeway situation, it is most useful in cases with large differences in the level of service among the lanes (e.g. in the presence of an HOV lane) where traditional modeling approaches tend to fail. The target lane model, which has the ability to capture immediate lane changes to a worse adjacent lane in order to execute the chosen latent plan to reach a better target lane in that direction, performed substantially better compared to state-of-the-art driving behavior models in such scenario.

Freeway Merging

In the freeway merging model, the target lane is observed and the latent plan manifests as a choice of merging tactic. Traditional merging models are based on the concept of an 'acceptable gap' emerging and the driver merging into this gap. However, in congested situations, acceptable gaps are often not available and more complex merging phenomena are observed. For example, drivers may merge through courtesy of the lag driver in the target lane or become impatient and decide to force in, compelling the lag driver to slow down. The plan of the driver, thus, involves selection of the merging tactic, which in turn affects the driver's merging behavior. However, the chosen plan is unobserved and only the action, that is the execution of the merge through gap acceptance, is observed. The acceptable gaps for completion of the merge at any instant depend on the plan at that time. For example, the acceptable gaps are smaller in case of courtesy merging compared to normal merging since there is less risk associated with it. Further, the plan may evolve dynamically as the immediate execution of the chosen merging plan may not be feasible. For example, there can be a switch to the forced merging plan if the driver is unable to merge through normal gap acceptance for a

considerable duration. The probabilities of transitions from one plan to another were found to be affected by the risk associated with the merge and the characteristics of the driver such as impatience, urgency and aggressiveness as well as a strong inertia to continue the previously chosen merging tactic (state-dependence). These effects were captured by variables such as relative speed and acceleration of the mainline vehicles, delay associated with the merge, density of traffic, remaining distance to the end of merging lane, etc. To capture the effect of predicted behavior of other drivers in the decision making of the subject driver, changes in position of the other drivers within the anticipation time of the subject driver was explicitly taken into account. In the trajectory data, only the final execution of the merge is observed but the tactics used for the merge and the sequence of plans that led to the chosen merging tactics are unobserved. A HMM formulation was used to formulate the likelihood of the observations.

The new merging model thus integrated, for the first time, all merging tactics of the driver (normal, courtesy and forced) in a combined decision framework. The combined decision framework gives the flexibility to model the transition between the merging tactics that are beyond the scope of disjoint merging models. Also, for the first time, the state-dependence among decisions that had been ignored in the previous state-of-the-art models was captured.

Urban Arterial Intersection Lane Choice

The intersection lane choice model involves the lane choice of drivers entering the arterial from a side street. Their latent plan was captured through the choice of target lane. The immediate lane selections observed in the data are based on the target lane selections. The choice of target lane was found to be influenced by path-plan variables and lane-specific attributes whereas immediate lane choices was found to be governed by maneuverability considerations. The heterogeneity in planning capability of the driver was allowed to vary among drivers using a latent class model formulation. The latent plan based lane selection model thus had the flexibility to capture the pre-positioning of some drivers based on path-plan before they reach their terminal section. The urban intersection lane choice model replaced the traditional rule-based assignment technique of vehicles in their subsequent lanes used in state-of-the-art traffic simulators.

Urban Arterial Mainline Lane Changing

Mainline lane changing models for urban arterials were estimated rigorously for the first time using disaggregate data collected from urban arterials. Similar to the freeway lane changing model, the latent plan in the mainline lane changing model within urban arterial sections involves target lane choice. However, the duration of the lane change that is the time elapsed between the initiation and completion of the lane change was found to be significant. The gap acceptance decisions (indicting the maneuverability considerations for making the change in the direction of the target lane) were therefore followed by an additional execution level to mark the completion of the lane change. The gap acceptance decisions were however unobserved like the choice of target lane and only the final lane actions were observed. The choice of target lane was found to be influenced by neighborhood vehicle speeds and positions, lane-specific attributes like queue length, path-plan of the driver and driver characteristics (planning capability and aggressiveness). The planning capability of the driver was allowed to vary among drivers using a continuous latent variable: the 'plan-ahead' distance of the driver. Gap acceptance was found to be influenced by relative speeds of lead and lag vehicles. The execution or completion of the lane change was found to be a function of the speed of the driver and the trend in change in gap size.

7.2 Contributions

The thesis advances the state-of-the-art driving behavior models through explicit inclusion of the effects of latent plans in the decision framework of the drivers. The new modeling approach gives a better representation of the decision mechanism by capturing the causal relationships between plans and actions of the driver and results in more realistic traffic simulation.

The above contributions were demonstrated through four lane selection scenarios that were identified as weak points of traffic simulation by model users and developers in the NGSIM study on Identification and Prioritization of Core Algorithm Categories (Alexiadis 2004). In each scenario, the inclusion of the latent plans was justified by comparison of goodness-of-fit of estimation and aggregate validation results. The comparison of goodness-of-fit of estimation results exhibited the improvements in model

estimates as compared to the reduced form models that do not have any latent mechanism. The aggregate validation results demonstrated this through improvements in the simulation capability in comparison to the state-of-the-art models that use instantaneous decisions of drivers based on myopic considerations.

Though the benefits of the latent plan model are likely to be more in extreme traffic conditions, for instance, when there are substantial differences in level of service among lanes or there is severe congestion leading to increased cooperation among drivers, the improvements were also observed for general situations.

7.3 Directions for Future Research

In this thesis, a general framework for latent plan models supported by four applications of the framework in modeling driving decisions was presented. The concept of latent plan and the proposed framework has enormous potential both in modeling driving decisions and modeling decisions in other scenarios. Some of the directions in which further research is needed are presented below:

- Additional dimensions: In this thesis the latent plan methodology was applied to model the lane changing decisions in different freeway and urban scenarios. These models however do not capture the effect of acceleration behavior to facilitate lane changing. For example, drivers planning to make a lane change may target a gap and adapt their acceleration in order to better position themselves to maneuver to the gap chosen for lane changing. These additional dimensions of the plan of the driver, target gap selection and acceleration behavior for example, need to be integrated into the latent plan decision framework.

- State-dependence: In this thesis the effect of state-dependence was modeled only in the case of freeway merging. Similar state-dependence among observations may also prevail in other lane changing scenarios. For example, in case of normal lane changing, it was assumed in both freeway and arterial models that the instantaneous choices of target lanes capture the dynamicity of the lane changing decisions at every instant. However it ignores the possibility that drivers may have preference to follow their initial plan even if situational

constraints reduce the attractiveness of the plan. The hypothesis regarding such inertia and persistence behavior of the driver was not tested in this thesis due to computational limitations related to initial conditions. Application of HMM methodology to simplify the computation of the state-dependent target lane model is an interesting direction to extend this research in the future.

- Future expectations of the driver: In this research, the utilities of future options of the driver were not considered in the current utility. But in reality, the expectations/payoffs of future decisions can influence the current decisions of the drivers. For example, utility of executing a merge now can be influenced by expected utilities of merging in later time steps. Dynamic programming can be an effective approach to capture the future consequences of the current action. This approach will involve inclusion of the temporally discounted future utilities in the expected utility of the current instant. But this was not pursued in this research due to computational burden and need to be explored in future.

- Driver heterogeneity: In the disaggregate estimation data used in this research, no driver specific information was available. The driver heterogeneity was hence captured through statistical distributions using a latent variable estimation methodology. Combination of the trajectory data with socio-economic data of the drivers can be used in future to further enrich these models.

- Additional applications: The simulation tools enhanced with latent plan models have demonstrated better performance in the current research. The improved tools have the potential to be used to investigate aggregate traffic dynamics. For example, they can be used to investigate traffic shockwave propagations.

The latent plan models developed in this thesis focus on the latent plans involved with the driving decisions. The same methodology can be applied in many other cases, both in driving behavior models and other discrete choice models where the decisions of individuals involve unobserved planning. In particular, the proposed method to integrate Discrete Choice and Discrete Hidden Markov methods could be

effectively used in explicitly capturing the dynamics of plan and action in different scenarios. Examples include route choice models (Ben-Akiva *et al.* 1984, Bierlaire *et al.* 2006), shopping destination choice (Ben-Akiva and Lerman 1985), activity participation and travel behavior models, and many other choice situations involving 'hidden' decision layers and latent alternatives.

Appendix A

Microscopic Traffic Simulation Laboratory (MITSIMLab)

MITSIMLab (Yang *et al.* 1996) is a simulation-based laboratory that was developed for evaluating alternative traffic management system designs at the operational level. MITSIM, which represents the 'real-world' with detailed traffic and network elements and mimics the behavior of individual drivers, provides as an ideal tool for testing the performance of different driving behavior models.

The various components of MITSIMLab are organized in three modules:

1. Microscopic Traffic Simulator (MITSIM)
2. Traffic Management Simulator (TMS)
3. Graphical User Interface (GUI)

The main elements of MITSIM are network components, travel demand and driving behavior. The road network is represented with nodes, links, segments and lanes. The vehicle movements and the traffic control and surveillance devices are represented at the microscopic level.

If disaggregate travel demand data is available (vehicles starting at a given time interval from each origin to each destination), exact time-dependent origin-destination (OD) trip tables can be directly provided as an input to MITSIM. If such detailed data is not available, simulated travel demand based on estimated OD flows (detailed in Appendix B) are given as the input. The aggregate OD flows are translated into individual vehicles wishing to enter the network at a specific time. A probabilistic route

choice model is used to capture drivers' route choice decisions, which may be based on historical or real-time travel time information.

Each vehicle/driver combination is assigned behavior parameters (e.g. desired speed, aggressiveness, anticipation time, plan-ahead distance etc.) and vehicle characteristics (type, old vs. new etc.). The vehicles move through the network according to acceleration and lane changing models. The acceleration model captures drivers' response to conditions ahead as a function of relative speed, headway, and other traffic measures. The lane changing models are replaced by the latent plan models to be tested. The default driving behavior models implemented in MITSIMLab were estimated and validated by Ahmed (1999) and Toledo (2002).

TMS mimics the traffic control system in the network under consideration. A wide range of traffic control and route guidance systems can be simulated. These include intersection controls, ramp control, freeway mainline control, lane control signs, variable speed limit signs, portal signals, variable message signs, and in-vehicle route guidance. TMS can represent different designs of such systems with logic at varying levels of sophistication (pre-timed, actuated, or adaptive). An extensive graphical user interface is used for both debugging purposes and demonstration of traffic impacts through vehicle animation. A detailed description of MITSIMLab appears in Yang and Koutsopoulos (1996) and Yang *et al.* (2000).

The proposed freeway models: the target lane changing model and the combined merging model and the proposed arterial lane selection models have been implemented in MITSIM for the validation study. The acceleration model proposed by Ahmed (1999) has been used to simulate the longitudinal movement in both cases.

Appendix B

Calibration Methodology

B.1 Calibration Framework

The process of calibration of the simulation system aims to set the various parameters so that observed traffic conditions are accurately replicated. The overall calibration framework is summarized in Figure B.1.

βo   =Originally estimated parameters
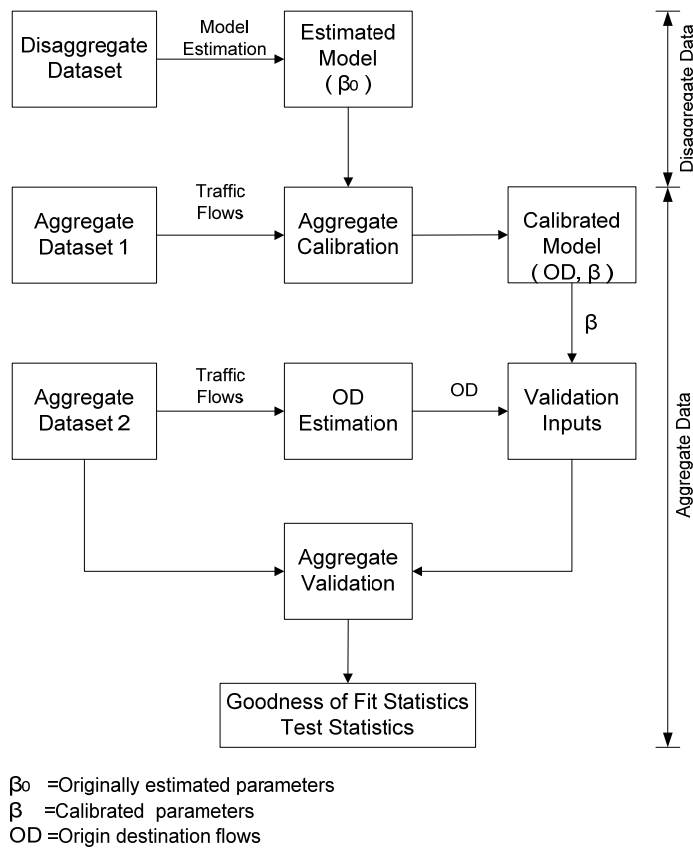β   =Calibrated parameters
OD =Origin destination flows

Figure B. 1: Calibration and validation framework

The calibration process consists of two steps: initially, the individual models of the simulation are estimated using disaggregate data. Disaggregate data includes detailed driver behavior information such as vehicle trajectories. The required explanatory variables including speeds and relations between the subject vehicle and other vehicles can be generated from the trajectory data. The disaggregate analysis is performed within statistical software and does not involve the use of a simulation system.

In the second step, the simulation model as a whole is calibrated using aggregate data like flows, speeds, occupancies, time headways, travel times, queue lengths etc. The process of aggregate calibration of the simulation system aims to adjust the various parameters so that observed traffic conditions are accurately replicated. These parameters consist of the parameters of the behavior model (initially estimated parameters β_0 adjusted to β) and the travel demand (expressed in terms of origin - destination or OD flows). Also, in special cases, due to limitations of the available disaggregate dataset, it may not be possible to estimate all the parameters of the model in the first step. For example, if the estimation dataset does not have a high occupancy vehicle (HOV) lane, it will not be possible to capture the effects of the HOV lane-specific variables during the estimation step. In such cases, the values of these omitted parameters can be captured during the aggregate calibration.

Once the calibration is complete, the values of the full set of behavioral parameters are fixed (β) and a second set of data is used for validation. Application of the simulation to replicate this dataset also requires OD flows as input. However, these may be different from the ones obtained in the calibration phase and so the OD estimation component of the calibration must be re-done for this dataset. These new OD flows and the calibrated parameter values are used as inputs to the simulation system.

B.2 Problem Formulation

Aggregate calibration can be formulated as an optimization problem, which seeks to minimize a function of the deviation of the simulated traffic measurements from the observed measurements and of the deviation of calibrated values from the a-priori estimates of the OD flows and the estimated behavior parameters. The formulation presented here assumes that the observations are drawn during a period in which steady

state traffic conditions prevail. That is, while OD flows and model parameters may vary for various observation days, these differences are due to random effects and do not represent a change in the underlying distributions of these variables. Furthermore, driving behavior parameters are assumed to be stable over the period of observation. It is important to note that the steady state assumption concerns the variability between observation days, and not within each observation day.

The formulation is shown below. The first and second terms in the objective function are a measure of deviation between observed and simulated measurements and between a priori OD flows and the estimated OD flows respectively. The first constraint shows the dependence of simulated measurements on the driving behavior parameters, OD flows and the network conditions. The second constraint is a non-negativity constraint for the OD flows.

$$\min_{\beta,OD} \sum_{i=1}^{N} \left(M^{sim} - M_i^{obs} \right)^T W^{-1} \left(M^{sim} - M_i^{obs} \right) + \left(OD - OD^o \right)^T V^{-1} \left(OD - OD^o \right)$$

$$s.t. \quad M^{sim} = S\left(\beta, OD \right) \tag{B.1}$$

$$OD \geq 0$$

Where,

β =driving behavior parameters

OD =OD flows

OD^o =a priori OD flows

N =number of days for which sensor data is available

M^{sim} =simulated measurements

M_i^{obs} =observed measurements for day i

S =the simulation model function, which generates simulated traffic measurements

W= variance-covariance matrix of the sensor measurements

V = variance-covariance matrix of the OD flows

The sensor measurements in this case constitute of the traffic flows and speeds measurements at all sensor stations and all time intervals.

The formulation presented above is difficult to solve because of the absence of analytical formulations that relate the affect of behavior parameters to the sensor measurements and relatively large number of parameters to calibrate. An iterative

solution approach is therefore adopted. In each iteration, first the driving behavior parameters are kept fixed and the OD flows are estimated. Then the OD flows are kept fixed and the driving behavior parameters are estimated.

The number of behavior parameters in the simulation model is very large. It is not feasible to calibrate all of them. A sensitivity analysis is often done to identify the parameters that contribute most in improvement of the objective function. In sensitivity analysis, the impact of an individual factor on the overall predictive quality of the simulator is measured while keeping all other parameters at their original values.

The details of the calibration methodology are presented by Ben-Akiva *et al.* (2003).

Appendix C

Reduced Form Models

C.1 Lane Shift Model (Toledo et al. 2003)

This section describes the structure and the parameter estimates of the lane shift model (Toledo et al. 2003), which is the reduced form model for the freeway target lane model. The model integrates the mandatory and discretionary lane changing considerations of the driver in a single framework. The lane changing process consists of two steps: 1) choice of a lane shift and 2) gap acceptance decisions. The choice set for lane shift consists of current and adjacent lanes (restricted targets) and does not include the full set of latent targets as in the latent plan freeway lane selection model. The choice of lane shift is however unobservable; only the driver's lane changing actions are observed. The structure of the model is shown in Figure C.1. Latent choices variables are shown as ovals, and observed ones are shown as rectangles.
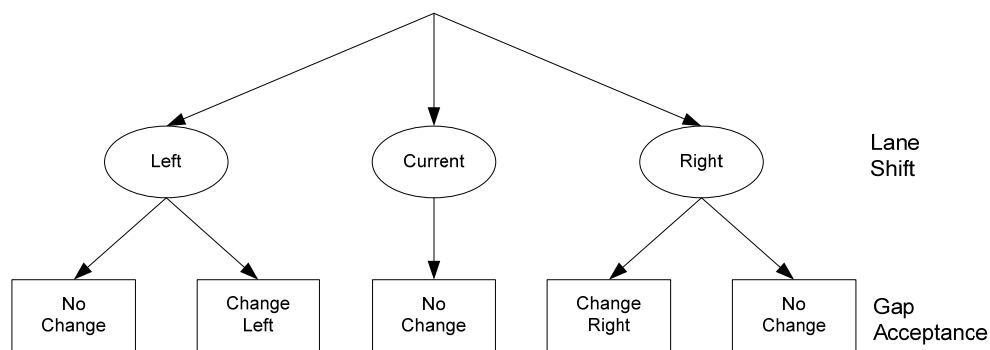

Figure C. 1: Structure of the lane shift model

The lane shift is the direction of change (or decision not to change) that the driver perceives as best to undertake. The Current branch corresponds to a situation in which the driver decides not to pursue a lane change. In the Right and Left branches, the driver

perceives that moving in these directions, respectively, would improve his/her condition. In these cases, the driver evaluates the adjacent gap in the lane in the chosen direction and decides whether the lane change can be executed or not. Only if the driver perceives that the gap is acceptable the lane change is executed (Change Right or Change Left); otherwise, the driver does not execute the lane change (No Change).

Estimation results of the lane shift model are presented in Table C.1.

Table C. 1: Estimation results of lane-shift model (Toledo et al. 2003)

Variable	Parameter	t-stat
Lane Shift		
CL constant	2.490	3.74
RL constant	-0.173	-0.51
Right-most lane dummy	-1.230	-3.89
Subject speed, m/sec.	0.062	1.59
Relative front vehicle speed, m/sec.	0.163	3.02
Relative Lag speed, m/sec.	-0.074	-1.30
Front vehicle spacing, m.	0.019	3.42
Tailgate dummy	-3.162	-1.68
Path-plan impact, 1 lane change required	-2.573	-4.86
Path-plan impact, 2 lane changes required	-5.358	-5.94
Path-plan impact, 3 lane changes required	-8.372	-5.70
Next exit dummy, lane change(s) required	-1.473	-2.30
θ^{MLC}	-0.378	-2.29
π_1	0.004	0.46
π_2	0.009	0.77
α^{CL}	0.734	4.66
α^{RL}	2.010	2.73
Lead Critical Gap		
Constant	1.353	2.48
$Max\left(\Delta V_n^{lead}(t),0\right)$, m/sec.	-2.700	-2.25
$Min\left(\Delta V_n^{lead}(t),0\right)$, m/sec.	-0.231	-2.42
α^{lead}	1.270	2.86
σ^{lead}	1.112	2.23
Lag Critical Gap		
Constant	1.429	6.72
$Max\left(\Delta V_n^{lag}(t),0\right)$, m/sec.	0.471	3.89
α^{lag}	0.131	0.64
σ^{lag}	0.742	3.68

C.2 Single Level Gap Acceptance Model (Lee 2006)

This section describes the structure and the parameter estimates of the single level gap acceptance model (Lee 2006), which is the reduced form model for the freeway merging model. The single level model aims at capturing the normal, forced and courtesy behavior of drivers through one gap acceptance level. In this model the merging driver evaluates the adjacent lead and lag gaps for merging and compares them with the corresponding critical gaps. An adjacent gap is acceptable if both lead and lag gaps are acceptable. The model structure is shown in Figure C.2.
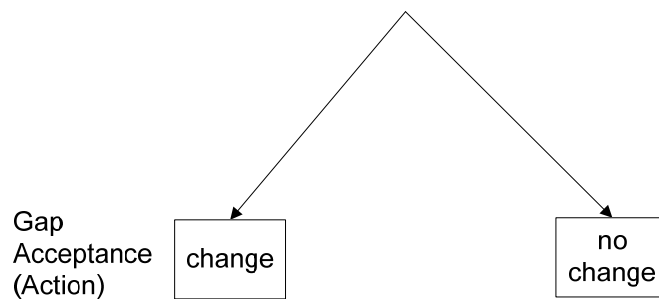
Figure C. 2: Framework of single level gap acceptance model (Lee 2006)

Critical gaps are modeled to have log-normal distribution their means being function of explanatory variables. Variables capturing courtesy and forced merging are included by means of relevant variables (acceleration of lag vehicle as an indicator of courtesy, remaining distance to the merge as an indicator of forced merge). The estimation results of the single level gap acceptance model are presented in Table C.2.

Table C. 2: Estimation results of the single level gap acceptance model

Variable	Parameter	t-stat
Lead Gap		
Constant	0.181	0.203
Max(0,average speed - subject speed)(m/sec)	1.45	4.59
Min(0,lead speed - subject speed) (m/sec)	-0.571	-3.53
remaining distance to MLC point (10 meters)	1.029	4.29
Remaining distance constant	-0.492	-0.81
$\alpha^{\mathrm{Re}\,mDisLead}$	0.798	2.66
$\sigma^{M\,lead}$	4.27	5.86
$\alpha^{M\,lead}$	-0.00016	-0.0033
Lag Gap		
Constant	0.379	0.89
Max(0,lag speed - subject speed)(m/sec)	0.178	1.36
Min(0,lag speed - subject speed)(m/sec)	0.0909	0.707
remaining distance to MLC point (10 meters)	0.178	1.74
Remaining distance constant	-2.21	-0.55
$\alpha^{\mathrm{Re}\,mDisLag}$	2.88	0.73
Max(0,acceleration of lag vehicle)(m/sec^2)	0.0766	0.81
$\sigma^{M\,lag}$	0.914	5.63
$\alpha^{M\,lag}$	-0.00012	-0.0025

C.3 Single Level Intersection Lane Choice Model

The reduced form model for the intersection lane choice model is a single level multinomial logit model. The structure of the model is illustrated in Figure C.3. In this model the driver evaluates the utilities of the available lanes and selects the lane with the highest utility. The model was estimated with same data as the latent plan intersection lane choice model.
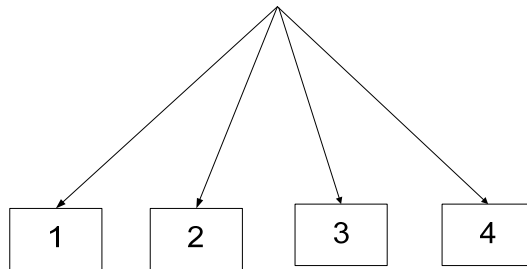

Figure C. 3: Framework of single level intersection lane choice model

The estimation results are presented in Table C.3.

Table C. 3: Estimation results of single level intersection lane choice model

Variable		Parameter	t-stat
Lane Selection			
Lane 2 constant		-0.0181	0.12
Lane 3 constant		1.10	6.91
Lane 4 constant		3.29	7.43
Anticipated delay-myopic drivers (second)		-0.331	-0.87
Anticipated delay-drivers who plan-ahead (second)		-0.477	-0.49
Lanes away from turning lane (myopic)	coefficient-myopic drivers	-1.11	-0.49
	constant-myopic drivers	0.968	0.79
	heterogeneity coefficient -myopic drivers	0.427	1.03
Lanes away from turning lane (with plan-ahead)	coefficient-drivers who plan-ahead	-4.24	-6.36
	constant-drivers who plan-ahead	.415	0.84
	heterogeneity coefficient -drivers who plan-ahead	1.51	2.10
Lanes away from connecting lane	coefficient	-1.01	-2.92
	constant	0.427	1.03
	heterogeneity coefficient	0.648	4.46
Conflict dummy		-1.95	-11.01
Driver Class			
Driver population with >1 section plan-ahead (%)		3.31	0.33

C.4 Lane Shift Model for Urban Arterials

This section describes the reduced form model for the mainline lane changing model for the urban arterial. The structure of the lane shift model for urban arterials is same as the lane shift model for freeway developed by Toledo *et al.* (summarized in C.1). But the model has been re-estimated using the same trajectory data that has been used to estimate the latent plan model for the urban arterial mainline presented in Section 6.2.3. Estimation results of the model are presented in Table C4.

Table C. 4: Estimation results of the re-estimated lane changing model

Variable	Parameter	t-stat
Lane Shift		
Current Lane Dummy	4.36	0.34
Path-plan impact : No. of lane changes to exit lane	-1.12	-3.64
Path-plan impact : No. of lane changes to exit interacted with distance from exit	-0.71	-2.53
Exponent of dist. to exit in no. of lanes to exit- dist. to exit interaction	0.344	1.20
Queue length ahead in lane	-0.087	-1.22
Front vehicle relative speed	0.0186	3.35
α^{CL}	1.15	5.20
Lead Critical Gap		
Lead gap constant	0.798	20.18
$\Delta V_{nt}^{lead,TL}$ (m/s)	0.757	59.21
σ^{lead}	0.000141	0.08
α^{lead}	-0.94	-31.31
Lag Critical Gap		
Lag gap constant	-1.205	-10.51
$\Delta V_{nt}^{lag,TL}$ (m/s)	0.257	9.60
α^{lag}	0.000157	0.002
σ^{lag}	-1.78	-116.94

Appendix D

Related Publications

1. Choudhury C., Ben-Akiva M., Toledo T., Rao A. and Lee G. (2007), Modeling State-dependence in Lane-changing Behavior, paper presented at the 17th International Symposium on Transportation and Traffic Theory, London, UK.

2. Choudhury C., Ben-Akiva M. and Toledo, T. (2007), Modeling Lane-changing Behavior in Presence of Exclusive Lanes, paper presented at the 11th World Conference on Transport Research, Berkley, USA.

3. Ben-Akiva M., Abou Zeid M., Choudhury C. (2007), Latent Choices, paper presented at the 9e Journée Transport, Paris, France.

4. Ben-Akiva M., Abou Zeid M., Choudhury C. (2007), Hybrid Choice Models: from Static to Dynamic, paper presented at the Triennial Symposium on Transportation Analysis VI, Phuket, Thailand.

5. Choudhury C., Ben-Akiva M., Ramanujam, V., Toledo, T. and Rathi, V. (2007), Arterial Lane-selection Model, Federal Highway Administration, Technical Report No. FHWA-HOP-07-079.

6. Choudhury C., Ben-Akiva M., Toledo, T., Lee G. and Rao A. (2007), Modeling Cooperative Lane-changing and Forced Merging Behavior, paper presented at the 86th Transportation Research Board Annual Meeting, Washington DC, USA.

7. Ben-Akiva M., Choudhury C. and Toledo, T., Lane-changing Models (2006), paper presented at the 2nd International Symposium of Transport Simulation, Lausanne, Switzerland.

8. Ben-Akiva M., Choudhury C. and Toledo, T. (2006), Modeling Latent Choices: Application to Driving Behavior, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.

9. Choudhury C., Ben-Akiva M. ,Toledo, T., Rao, A. and Lee, G. (2006), Cooperative Lane-changing and Forced Merging Model, Federal Highway Administration, Technical Report No. FHWA-HOP-07-078.

10. Choudhury C., Ben-Akiva M. and Toledo, T. (2005), Lane-selection Model, Federal Highway Administration, Technical Report No. FHWA-HOP-06-012.

11. Toledo, T., Choudhury C. and Ben-Akiva M. (2005), A Lane-changing Model with Explicit Target Lane Choice, Transportation Research Record, Journal of the Transportation Research Board, No. 1934, pp. 157-165.

Bibliography

Abdulhai B., Sheu J.B. and Recker W. (1999) Simulation of ITS on the Irvine FOT area using Paramics 1.5 scalable microscopic traffic simulator, Phase I: model calibration and validation, California PATH Research Report UCB-ITS-PRR-99-12.

Addison, P. S. and Low, D. J. (1998) A novel nonlinear car following model, Chaos: An Interdisciplinary Journal of Nonlinear Science, 8, pp. 791–799.

Adebisi, O. and Sama, G. N. (1989) Influence of stopped delay on driver gap acceptance behaviour, ASCE Journal of Transportation Engineers, 115, pp. 305–315.

Ahmed K.I. (1999) Modeling Drivers' Acceleration and Lane changing Behavior, PhD Dissertation, Department of Civil and Environmental Engineering, MIT.

Ahmed K.I., Ben-Akiva M., Koutsopoulos H.N. and Mishalani R.G. (1996) Models of Freeway Lane changing and Gap Acceptance Behavior, in Lesort J. B. (ed.) Proceedings of the 13th International Symposium on the Theory of Traffic Flow and Transportation, pp. 501-515.

Akaike H. (1973) Information theory and an extension of the maximum likelihood principle, in Petrov B.N. and Csake F. (ed.), Second International Symposium on Information Theory, pp. 267-281.

Akaike H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19, pp. 716-723.

Alexiadis V., Gettman D. and Hranac R. (2004) NGSIM Task D: Identification and Prioritization of Core Algorithm Categories, Federal Highway Administration, FHWA-HOP-06-008.

Aptech System (1995) GAUSS applications: maximum likelihood estimation, Maple Valley, WA.

Aptech System (2003) GAUSS manual, Volume I and II, Maple Valley, WA.

Astrom K.J. (1965) Optimal control of Markov decision processes with incomplete state estimation, Journal of Mathematical Analysis and Applications, 10, pp. 174-205.

Baker J.K. (1975) The dragon system-An overview, IEEE Trans. Acoustic, Speech and Signal processing, ASSP-23-1, pp. 24-29.

Barto A., Bradtke S., and Singh S. (1995) Learning to act using real-time dynamic programming. Artificial Intelligence, Special Volume on Computational Research on Interaction and Agency, 72(1), pp. 81-138.

Baum L. E. and Petrie T. P. (1966) Statistical inference for probabilistic functions of finite state markov chains, Annals of Mathematical Statistics, 37, pp. 1554-1563.

Baum, L.E. (1972) An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. Inequalities, 3, pp. 1-8.

Behbahanizadeh K. and Hidas P. (1999) Microscopic Simulation of Lane changing under Incident Conditions, in Ceder A. (ed.) Proceedings of the 14th International Symposium on the Theory of Traffic Flow and Transportation, pp. 53-69.

Bekey, G. A., Burnham, G. O. and Seo, J. (1977) Control theoretic models of human drivers in car following, Human Factors, 19, pp. 399–413.

Bellman, R. E. (1957) A Markovian Decision Process, Journal of Mathematics and Mechanics, 6, pp. 679-684.

Bellman, R. E. (1957), Dynamic Programming, Princeton, NJ: Princeton University Press.

Ben-Akiva M. (1973) Structure of passenger travel demand models, PhD Dissertation, Department of Civil and Environmental Engineering, MIT.

Ben-Akiva, M. (1974) Structure of passenger travel demand models, Transportation Research Record, 526, pp. 26-42.

Ben-Akiva M., Bergman M., Daly A. and Ramaswamy R. (1984) Modeling inter-urban route choice behavior, in Vollmuller J. and Hamerslag R. (ed.) Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Science Press, Utrecht, Netherlands, pp. 299-330.

Ben-Akiva M., Darda D., Jha M., Toledo T. and Koutsopoulos H. (2003) Calibration of microscopic traffic simulation models with aggregate data, Transportation Research Board, 82nd Annual Meeting.

Ben-Akiva M., Lerman S. (1985) Discrete Choice Analysis, Cambridge, MA: The MIT Press.

Ben-Akiva, M. and Bierlaire, M. (2003) Discrete choice models with applications to departure time and route choice, in Hall R. (ed.) Handbook of Transportation Science, Kluwer.

Ben-Akiva M., Choudhury C. and Toledo, T., Lane-changing Models (2006), paper presented at the 2nd International Symposium of Transport Simulation, Lausanne, Switzerland.

Ben-Akiva M., Choudhury C. and Toledo, T. (2006), Modeling Latent Choices: Application to Driving Behavior, paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.

Ben-Akiva M., Abou Zeid M., Choudhury C. (2007), Latent Choices, paper presented at the 9e Journée Transport, Paris, France.

Ben-Akiva M., Abou Zeid M., Choudhury C. (2007), Hybrid Choice Models: from Static to Dynamic, paper presented at the Triennial Symposium on Transportation Analysis VI, Phuket, Thailand.

Benekohal, R. and Treiterer, J. (1988) Carsim: car-following model for simulation of traffic in normal and stop-and-go conditions, Transportation Research Record, 1194, pp. 99–111.

Bertsekas D.P. (1999) Nonlinear programming, Belmont MA: Athena Scientific.

Bexelius, S. (1968) An extended model for car following, Transportation Research, 2, pp. 13–21.

Bierlaire M. (2001) A theoretical analysis of the cross-nested logit model, Report RO-011218, http://roso.epfl.ch/mbi/papers/crossnested-report.pdf [accessed 05.26.07]

Bierlaire M., Frejinger E., Stojanovic J. (2006) A Latent Route Choice Model in Switzerland, European Transport Conference, Strasbourg, France.

Bilmes J. (2002) What HMMs can do, University of Washington Technical Report, https://www.ee.washington.edu/techsite/papers/documents/UWEETR-2002-0003.pdf [accessed 03.01.07]

Bhat C. R. (1998) Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice, Transportation Research Part B: Methodological, 32-7, pp. 455-466.

Blythe J. (1999) An Overview of Planning Under Uncertainty, AI Magazine, 20(2), Summer 1999, pp. 37–54.

Boutilier, C., Dean T. and Hanks S. (1999) Decision-Theoretic Planning: Structural Assumptions and Computational Leverage, Journal of Artificial Intelligence Research 11, pp. 1-94.

Boehm T. P., Herzog H. W. and Schlottmann A. L. (1991) Intra-Urban Mobility, Migration, and Tenure Choice, The Review of Economics and Statistics, 73-1, pp. 59-68.

Boutilier, C., Dearden, R., & Goldszmidt, M. (1995) Exploiting structure in policy construction. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, Canada, pp. 1104-1111.

Brilon, W. (Ed.) (1988) Intersections Without Traffic Signals I, New York, NY: Springer.

Brilon, W. (Ed.) (1991) Intersections Without Traffic Signals II, New York, NY: Springer.

Cambridge Systematics, Inc. (2004) NGSIM: High-Level Verification and Validation Plan.

Cambridge Systematics, Inc. (2004) NGSIM: I-80 Data Analysis. http://www.ngsim.fhwa.dot.gov/

Cambridge Systematics, Inc. (2005) NGSIM: U.S. 101 Data Analysis. http://www.ngsim.fhwa.dot.gov

Cambridge Systematics, Inc. (2006) NGSIM: Lankershim Data Analysis. http://www.ngsim.fhwa.dot.gov

Cascetta E. and Nguyen S. (1988) A Unified Framework for Estimating or Updating Origin Destination Matrices from Traffic Counts. Transportation Research Part B, 22(6), pp. 437–455.

Cascetta E. and Postorino M.N. (2001) Fixed Point Approaches to the Estimation of O/D Matrices Using Traffic Counts on Congested Networks, Transportation Science, 35, pp. 134 –147.

Chandler, R., Herman, R. and Montroll, E. W. (1958) Traffic dynamics: studies in car following, Operations Research, 6, pp. 165–184.

Choudhury C.F. (2005) Modeling Lane changing Behavior in Presence of Exclusive Lanes, MST thesis, Department of Civil and Environmental Engineering, MIT.

Choudhury C., Ben-Akiva M., Toledo, T., Lee G. and Rao A. (2007), Modeling Cooperative Lane-changing and Forced Merging Behavior, paper presented at the 86th Transportation Research Board Annual Meeting, Washington DC, USA.

Choudhury C., Ben-Akiva M., Toledo T., Rao A. and Lee G. (2007), Modeling State-dependence in Lane-changing Behavior, paper presented at the 17th International Symposium on Transportation and Traffic Theory, London, UK.

Choudhury C., Ben-Akiva M. and Toledo, T. (2007), Modeling Lane-changing Behavior in Presence of Exclusive Lanes, paper presented at the 11th World Conference on Transport Research, Berkley, USA.

Cleveland W.S. (1979) Robust locally weighted regression and smoothing scatter plots, Journal of the American Statistical Association 74, pp. 829-836.

Cleveland W.S. and Devlin S.J. (1988) Locally weighted regression: an approach to regression analysis by local fitting, Journal of the American Statistical Association 83, pp. 596-610.

Cleveland W.S., Devlin S.J. and Grosse E. (1988) Regression by local fitting: methods, properties and computational algorithms, Journal of Econometrics 37, pp. 87-114.

Daly A.J. (1987) Estimating Tree Logit Models, Transportation Research Part B 21, pp. 251-268.

Daly A.J., Zachary S. (1976) Improved multiple choice models. In: Proceedings of the Fourth PTRC Summer Annual Meeting. University of Warwick, England, 1976.

Daly A.J., Zachary S. (1978) Improved multiple choice models. In: Hensher D.A., Dalvi, M.Q. (Eds.), Determinants of Travel Choice. Saxon House, Sussex.

Daganzo C.F. (1981) Estimation of gap acceptance parameters within and across the population from direct roadside observation, Transportation Research 15B, pp. 1-15.

Darroch, J. N. and Rothery, R. W. (1972) Car following and spectral analysis, in: G. F. Newell (Ed.) Proceedings of the 5th International Symposium on the Theory of Traffic Flow and Transportation, pp. 47–56, New York, NY: Elsevier.

Dean, T., Kaelbling, L., Kirman, J., & Nicholson, A. (1995). Planning under time constraints in stochastic domains. Artificial Intelligence, 76 (1-2), pp. 3-74.

Dearden R., and Boutilier C. (1994). Integrating planning and execution in stochastic domains. In Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, Washington, DC, pp. 162-169.

Drew, D. R., LaMotte, L. R., Buhr, J. H. and Wattleworth, J. A. (1967) Gap acceptance in the freeway merging process, Highway Research Record, 208, pp. 1–36.

Feldman, J. A., and Sproull, R. F. 1977. Decision theory and artificial intelligence ii: The hungry monkey. Cognitive Science, 1, pp. 158–192.

FHWA (1985) Freeway Data Collection for Studying Vehicle Interactions, Technical Report FHWA/RD-85/108, Federal Highway Administration, US Department of Transportation, McLean, Virginia.

FHWA (1998) CORSIM User Manual (1.04 ed.), Federal Highway Administration, US Department of Transportation, McLean, Virginia.

Gabard, J. F., Henry, J. J., Tuffal, J. and David, Y. (1982) Traffic responsive or adaptive fixed time policies? A critical analysis with SITRA-B, in: Proceedings of the International Conference on Road Traffic Signalling, Institution of Electrical Engineers, pp. 89–92.

Gazis, D., Herman, R. and Potts, B. (1959) Car following theory of steady-state traffic flow, Operations Research, 7, pp. 499–505.

Gazis, D., Herman, R. and Rothery, R. (1961) Nonlinear follow-the-leader models of traffic flow, Operations Research, 9, pp. 545–567.

Gill P.E. and Murray W. (1972) Quasi-Newton methods for unconstrained optimization. Journal of the Institute of Mathematics and its Applications 9, pp. 91-108.

Gipps P.G. (1986) A Model for the Structure of Lane changing Decisions, Transportation Research, 20B, pp. 403-414.

Gipps, P. G. (1981) A behavioural car-following model for computer simulation, Transportation Research Part B, 15, pp. 101–115.

Gopinath D. (1995) Modeling Heterogeneity in Discrete Choice Processes: Application to Travel Demand, PhD thesis, Department of Civil and Environmental Engineering, MIT.

Halati A., Lieu H. and Walker S. (1997) CORSIM – Corridor Traffic Simulation Model, in Proceedings of the Traffic Congestion and Traffic Safety in the 21st Century Conference, pp. 570-576.

Hamed, M. M., Ease, S. M. and Batayneh, R. R. (1997) Disaggregate gap-acceptance model for unsignalized T-intersections, ASCE Journal of Transportation Engineering, 123, pp. 36–42.

Heckman J.J. (1981) Statistical models for discrete panel data. In C.F. Manski and D. McFadden (Ed.), Structural analysis of discrete data with econometric applications, pp. 114-178.

Helly, W. (1961) Simulation in bottlenecks in single-lane traffic flow, in: Proceedings of the Symposium on the Theory of Traffic Flow, pp. 207–238 (New York, NY: Elsevier).

Herman, R. and Rothery, R. W. (1965) Car following and steady-state flow, in: J. Almond (Ed.) Proceedings of the 2nd International Symposium on the Theory of

Traffic Flow, pp. 1–11 (Paris: Organization for Economic Co-operation and Development (OECD)).

Herman, R. and Weiss, G. H. (1961) Comments on the highway crossing problem, Operations Research 9, pp. 838–840.

Hess S. (2005), Advanced Discrete Choice Models with Applications to Transport Demand, PhD Dissertation, Center for Transport Studies, Imperial College, University of London.

Hetrick S., Examination of Driver Lane Change Behavior and the Potential Effectiveness of Warning Onset Rules for Lane Change or "Side" Crash Avoidance Systems, MSc thesis, Industrial and Systems Engineering, Virginia Tech, Blacksburg VA, 1997.

Hidas P. and Behbahanizadeh K. (1999) Microscopic Simulation of Lane Changing under Incident Conditions, in Proceedings of the 14th International Symposium on the Theory of Traffic Flow and Transportation, pp. 53-69.

Hidas, P (2002) Modeling Lane changing and Merging in Microscopic Traffic Simulation, Transportation Research Part C: Emerging Technologies, Volume 10, pp. 351-371, Elsevier Science Ltd.

Hidas, P (2005) Modeling Vehicle Interactions in Microscopic Simulation of Merging and Weaving, Transportation Research Part C: Emerging Technologies, Volume 13, pp. 37-62, Elsevier Science Ltd..

Hranac R., Gettman D., Toledo T., Kovvali V. and Alexiadis V. (2004) NGSIM: Core Algorithms Assessment, Federal Highway Administration, FHWA-HOP-06-009.

Hranac R., Margiotta R. and Alexiadis V. (2004) NGSIM: High-Level Data Plan, Federal Highway Administration, FHWA-HOP-06-011.

Hunt J.G., Lyons G.D., (1994) Modeling dual carriageway lane changing using neural networks, Transportation Research-C, Vol. 2, No.4, pp.231-245

Jelinek F. (1976) Continuous speech recognition by statistical methods, Proceedings of IEEE, 64, pp. 532-536

Jin W., Zheng Y. and Li J. (1999) Microscopic Simulation of Traffic Flow at Signalized Intersection based on Cellular Automata, Vehicle Electronics Conference, Proceedings of the IEEE International, Vol 1, pp. 106-109.

Kita H. (1993) Effect of Merging Lane Length on the Merging Behavior at Expressway On-ramps, in Proceedings of the 12th International Symposium on the Theory of Traffic Flow and Transportation, pp. 37-51.

Koshi, M., Kuwahara, M. and Akahane, H. (1992) Capacity of sags and tunnels on Japanese motorways, ITE Journal, 62, pp. 17–22.

Koski T. (2001) Hidden Markov Models for Bioinformatics, Kluwer, Dordrecht.

Koppelman, F. S. & Sethi, V. (2000), Closed form discrete choice models, in D. Hensher & K. Button, eds, 'Handbook of Transport Modeling', Pergamon Press, Oxford, pp. 211–228.

Krishnamurthi L., and Raj S. P. (1988) A Model of Brand Choice and Purchase Quantity Price Sensitivities, Marketing Science, 7-1, pp. 1-20.

Kushmerick N., Hanks S. and Weld D.S. (1994) An algorithm for probabilistic least-commitment planning, Proceedings AAAI-94, Seattle, WA, pp. 1073- 1078.

Lee G. (1966) A generalization of linear car following theory, Operations Research, 14, pp. 595–606.

Lee G. (2006) Modeling Gap Acceptance at Freeway Merges, MST thesis, Department of Civil and Environmental Engineering, MIT.

Limanond T., Niemeier D. A. and Mokhtarian P.L. (2005) Specification of a tour-based neighborhood shopping model, 32-2, pp. 105-134.

DYMO (1999) Modeling of ITS applications: test of four dynamic models, Centre of Traffic Simulation (CTR), Royal Institute of Technology (KTH) in co-operation with TRANSEK, ISSN 1104-683X, Stockholm, Sweden.

Luenberger D. G. (1973) Introduction to Linear and Nonlinear Programming. Addison-Wesley, Reading, Massachusetts.

Madanat, S. M., Cassidy, M. J. and Wang, M. H. (1993) A probabilistic model of queuing delay at stop controlled intersection approaches, ASCE Journal of Transportation Engineering, 120, pp. 21–36.

Mahmassani, H. and Sheffi, Y. (1981) Using gap sequences to estimate gap acceptance functions, Transportation Research Part B, 15, pp. 143–148.

May, A. and Keller, H. (1967) Non-integer car following models, Highway Research Record, v. 199, pp. 19–32.

McFadden D. (1978) Modeling the choice of residential location, in A. Karlquist, ed., Spatial Interaction Theory and Planning Models, North Holland, Amsterdam, pp. 75–96.

Miller, A. J. (1972) Nine estimators of gap-acceptance parameters, in: G. F. Newell (Ed.) Proceedings of the 5th International Symposium on the Theory of Traffic Flow, pp. 215–235 (New York, NY: Elsevier).

National Highway Traffic Safety Administration (1997) NHTSA 2020 Report, U.S. Department of Transportation.

Ort´uzar, J. de D. (2001), 'On the development of the nested logit model', Transportation Research Part B: Methodological 35(2), 213–216.

Ozaki, H. (1993) Reaction and anticipation in the car-following behaviour, in: C. F. Daganzo (Ed.) Proceedings of the 12th International Symposium on the Theory of Traffic Flow and Transportation, pp. 349–366 (New York, NY: Elsevier).

Peot M. A. and Smith D. E. (1992) Conditional nonlinear planning, Proceedings of the First International Conference on Artificial Intelligence Planning Systems, pp. 189–197.

Pindyck R. and Rubinfeld D. (1997) Econometric models and economic forecasts, 4th edition. Irwin McGraw-Hill, Boston MA.

Pipes, L. A. (1953) An operational analysis of traffic dynamics, Journal of Applied Physics, 24, pp. 274–281.

PTV (2004) VISSIM 4.00 User Manual, Karlsluhe, Germany.

Quadstone (2004) Paramics V5.0, Technical Notes, Version No. 1.0, October, 2004.

Rabiner, L. (1989) A tutorial on Hidden Markov Models and selected applications in speech recognition, Proceedings of the IEEE, 77-2

Ramanujam V. (2007) Lane Changing Models for Arterial Traffic, M.S. thesis, Department of Civil and Environmental Engineering, MIT.

Rao, A. (2006) Modeling Anticipatory Driving Behavior, M.S. thesis, Department of Civil and Environmental Engineering, MIT.

Reuschel, R. (1950) Fahrzeugbewegungen in der Kolonne, Osterreichisches Ingenieur Archiv, 4, pp. 193–215.

Ribeiro, T. (2004): Hidden Markov models and their applications to estimation, forecasting and policy analysis in panel data settings. Working paper.

Ribeiro, T. , Adams (2003) Healthy, wealthy and wise? Tests for direct causal path between health and socioeconomic status, Journal of Econometrics, 112(1), pp. 3-56.

Rust J. (1994) Structural Estimation of Markov Decision Processes in R. Engle and D. McFadden (eds.) Handbook of Econometrics, North Holland: Amsterdam.

Rust, J. (1994) Estimation of Dynamic Structural Models, Problems and Prospects: Discrete Decision Processes, in J. J. Laffont and C. Sims reds,), Advances in Econometrics: Sixth World Congress, Vol. II,Cambridge University Press.

Saad, F., Delhomme, P. and Van-Elslande, P. (1990) Drivers' speed regulation when negotiating intersections, in: M. Koshi (Ed.) Proceedings of the 11th International Symposium on the Theory of Traffic Flow and Transportation, pp. 193–212 (New York, NY: Elsevier).

Schiller E. (1998) The roads ahead: the economic and environmental benefits of congestion pricing, www.pacificresearch.org [accessed 02.15.07]

Schrank D. and Lomax T. (2005) Urban Mobility Report, Texas Transportation Institute, The Texas A&M University System.

Simon H.A. (1955) A behavioral model of rational choice. Quarterly Journal of Economics, 69, pp. 99-118.

Simmons, R. and Koenig, S. (1995) Probabilistic robot navigation in partially observable environments. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, Canada, pp. 1080-1087.

Smallwood R. D. and Sondik E. J. (1973) The Optimal Control of Partially Observable Markov Processes Over a Finite Horizon, Operations Research, Vol. 21, No. 5. (Sep. - Oct., 1973), pp. 1071-1088.

Sparmann, U. (1978) Spurwechselvorgänge auf Zweispurigen BAB-Richtungsfahrbahnen. Forschung Straßenbau und Straßenverkehrstechnik, Heft 263.

State Environmental Research Center, Fact Pack on Congestion Relief http://www.serconline.org/trafficcongestionrelief/fact.html [accessed 3rd Nov 2004]

Tash, J. and Russell, S. (1994) Control strategies for a stochastic planner. In Proceedings of the Twelfth National Conference on Artificial Intelligence, Seattle, WA., pp. 1079-1085.

Thorton J. (2000) Physician choice of medical specialty: do economic incentives matter?, Applied Economics, 32, pp. 1419 – 1428.

Toledo T. (2002) Integrated Driving Behavior Modeling, PhD Dissertation, Department of Civil and Environmental Engineering, MIT.

Toledo T. (2007) Driving Behaviour: Models and Challenges, Transport Reviews, Vol. 27, 1, pp. 65–84.

Toledo T. and Koutsopoulos H. (2004) Statistical Validation of Traffic Simulation Models, Transportation Research Record No. 1876, pp. 142-150.

Toledo T. and Zohar D. (2007) Modeling the duration of lane changes, Transportation Research Board, 86th Annual Meeting.

Toledo T., Koutsopoulos H. and Ben-Akiva M. (2003) Modeling Integrated Lane changing Behavior, Transportation Research Record 1857, pp 30-38.

Transportation Research Board (1997) Highway Capacity Manual, Washington, DC.

Transportation Research Board (2000) Highway Capacity Manual, Washington, DC.

TSS-Transport Simulation Systems (2004) AIMSUN 5.0 Microsimulator User's Manual 5.0.5.

Victoria Transport Policy Institute (2004) Online TDM Encyclopedia http://www.vtpi.org/tdm/tdm96.htm [23rd Sep 2004]

Vogel S., Ney H. and Tilmann C. (1996) HMM-based word alignment in statistical translation, Proceedings of the 16th conference on Computational Linguistics, 2, pp.836-841.

Waddell P. (1996) Accessibility and residential location: the interaction of workplace, residential mobility, tenure, Lincoln Land Institute TRED Conference.

Walker J.L. (2001) Extended discrete choice models: integrated framework, flexible error Structures and latent variables. PhD thesis, Department of Civil and Environmental Engineering, MIT.

Wang J., Liu, R., Montgomery F. (2005) A simulation Laboratory for Motorway Merging Behavior. In: Mahmassani, H (ed.) Transportation and Traffic Theory: Flows, Dynamics and Human Interaction, Elsevier, 5(3), pp.127-140.

Webstar N., Suzuki T., Chung E., and Kuwahara M. (2007) Tactical Driver Lane Change Model Using Forward Search, Transportation Research Board, 86th Annual Meeting.

Wei H., Lee J., Li Q. and Li C.J. (2000) Observation-based Lane-vehicle-assignment Hierarchy for Microscopic Simulation on an Urban Street Network, Transportation Research Board, 79th Annual Meeting.

Wen, C.-H. & Koppelman, F. S. (2001) The Generalized Nested Logit Model, Transportation Research Part B: Methodological 35(7), 627–641.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit, Environment and Planning A, 9, pp. 285-344.

Williams H.C.W.L., Senior, M.L. (1977) Model based transport policy assessment II: removing fundamental inconsistencies from the models. Traffic Engineering and Control, 18, pp. 464-469.

Wooldridge J. M. (2005) Simple solutions to the initial conditions problem in dynamic non-linear panel data models with unobserved heterogeneity, Journal of Applied Econometrics, 20, pp. 39-54.

Yan, J., K. Small, and E. Sullivan (2002) Choice Models of Route, Occupancy, and Time-of-Day with Value Priced Tolls, Transportation Research Record, 1812, pp. 69-77.

Yang Q. and Koutsopoulos H.N. (1996) A Microscopic Traffic Simulator for Evaluation of Dynamic Traffic Management Systems, Transportation Research, 4C, pp. 113-129.

Yang Q., Koutsopoulos H.N. and Ben-Akiva M. (2000) A simulation laboratory for evaluating dynamic traffic management systems. Transportation Research Board, 79th Annual Meeting.

Zhang Y., Owen L.E. and Clark J.E. (1998) A Multi-regime Approach for Microscopic Traffic Simulation, Transportation Research Board, 77th Annual Meeting.