MODELLING LANE CHANGING BEHAVIOUR IN APPROACHES TO ROADWORKS: CONTRASTING AND COMBINING DRIVING SIMULATOR DATA WITH STATED CHOICE DATA

Stephane Hess
Choice Modelling Centre
Institute for Transport Studies
University of Leeds
34-40 University Road, LS2 9JT, Leeds, United Kingdom
Email: S.Hess@its.leeds.ac.uk

Charisma F. Choudhury
Choice Modelling Centre
Institute for Transport Studies
University of Leeds
34-40 University Road, LS2 9JT, Leeds, United Kingdom
Email: C.F.Choudhury@leeds.ac.uk

Michiel C.J. Bliemer
The University of Sydney Business School
The University of Sydney
378 Abercrombie Street, Darlington, NSW 2006, Australia
Email: michiel.bliemer@sydney.edu.au

Daryl Hibberd
Institute for Transport Studies
University of Leeds
34-40 University Road, LS2 9JT, Leeds, United Kingdom
Email: D.L.Hibberd@leeds.ac.uk

Word count:
5,407 words text + 4 tables/figures = 6,407 words
ABSTRACT
Drivers approaching lane closures due to roadworks tend to choose a target lane (plan) and seek suitable gaps to execute the plan (action). The plan is however latent or unobserved as the driver may or may not be able to move to the target lane due to the constraints imposed by the surrounding traffic. Hence, only the actions of the driver (as manifested by their final lane occupancies) are observed in the trajectory data. This paper analyses such behaviour in detail with data from a controlled driving simulator experiment and a simple stated preference survey with the same group of participants. While in the former drivers face similar constraints in implementing the plans as in the real world, in the simple stated choice survey the same drivers elicit their preferred target lanes without a need to put the plan into action. We contrast the findings from the two sources and also show correlations between the latent plan and stated target components in a latent class model. The results provide useful insights for improving the driving behaviour models and potentially reducing the costs of data collection for the model development.

Keywords: Driving behaviour, lane changing, driving simulator, stated choice
BACKGROUND

Road works, which are inevitable for maintenance of the aging infrastructure and/or capacity expansion, very often create traffic bottlenecks [1]. They also have significant safety impacts. In 2010, in United States alone, 87,606 crashes occurred in work zones (1.6% of the total number of roadway crashes in the US that year) out of which 0.6% were fatal crashes, 30% were injury crashes, and 69% were property damage only crashes [2]. Over the years, this has prompted researchers to identify the factors that affect traffic flow and safety in road work zones. These have revealed that the safety and capacity of work zones are significantly affected by the roadway geometry (e.g. grade, ramp configuration), traffic features (e.g. traffic composition and temporal variation), weather conditions and the features of the road work activities (e.g. intensity and duration of construction activities, lighting levels, etc.) [e.g. 3,4]. These studies have however been mostly based on macroscopic analyses using field data which is difficult and expensive to collect and which is often characterised by insufficient variability to develop detailed models.

On the other hand, driving simulator data has been used in a number of studies for analysing driver performance on roads approaching work zones in a simulated environment. These studies have however primarily focused on validation of the simulator for work zone scenarios [e.g. 6, 7, 8], comparison of performance of alternate traffic control measures [e.g. 9] evaluating effects of work-load and distraction in the presence of road works [e.g. 10]. A review of the literature reveals that there has been very limited research on developing mathematical models to quantify how different traffic factors and driver factors influence driving decisions in work zones [11]. Consequently, in a study conducted by the Next Generation Simulation (NGSIM) program of US Federal Highways, driving behaviour models in work zones have been identified as weak points of the traffic simulation tools [12]. This is primarily attributed to the fact that building detailed mathematical models of driving behaviour requires detailed data which are expensive to collect from the field as well as driving simulators. It may be noted that though driving simulators allow researchers to manipulate the surrounding and run various hypothetical scenarios, there has been scepticism regarding the simulator fidelity (physical and behavioural) of how well the driver’s behaviour in a simulator matches with his/her behaviour in real roads [6]. Recent research has however shown that the car-following models developed using driving simulator data have reasonable (though not perfect) transferability between the driving simulator and real world traffic with the transferability score being better when both samples are collected from drivers of the same region (Papadimitriou and Choudhury 2017).

A further complication that arises in understanding driving behaviour in these contexts is that in order to understand driver preferences, we need to filter out the impact of the constraints they face. Let us consider a simple example. A driver who finds himself in an approach to roadworks may have an underlying preference for changing into the appropriate lanes early on so as to avoid late (and potentially risky) changes. However, his ability to do so is affected by the surrounding traffic. For an analyst, it is then difficult to understand whether the driver changes lanes only at a late stage out of preference or out of necessity. The development of joint models for target lane choice and gap acceptance [15,14] is an important step towards gaining such an understanding but these models have been primarily developed using data extracted from video recordings of real traffic with complex confounding of influencing factors and less control on the external factors as well as absence of driver characteristics. In the present paper, we take this further by at the same time analysing data from a stated choice (SC) survey where drivers are asked to indicated their preferred target lanes (without needing to put the plan into action). We find differences between the two approaches, but also similarities. We believe that this work is a first step towards combining driving simulator data with SC data.
The remainder of the paper is organised as follows. We first discuss the two data efforts before looking at model specification and results. The findings are summarized in the end along with directions of future research.

**DATA**

This research aims to address this research gap by investigating drivers’ lane changing behaviour leading up to roadworks by augmenting driving simulator data with stated preference survey data. The stated preference data, which is more economic compared to driving simulator or field studies, provides the initial lane preferences of the drivers at specific critical points approaching the road works. The driving simulator data, collected using the University of Leeds Driving Simulator (UoLDS), provides the dynamic lane changing decisions of the same drivers as well as detailed information about the speeds and positions of the surrounding drivers. The two data sources used for the study are presented in this section. In each case, the experimental settings are presented first, followed by the characteristics of the datasets.

**Driving simulator data**

The study was conducted between October 2013 and April 2014 on a second-generation, motion-base, high fidelity driving simulator. The simulator vehicle is an adapted 2005 Jaguar S-type vehicle cab with fully-functional internal controls and dashboard instrumentation. The simulator vehicle is enclosed in a spherical projection dome (4m diameter) to reduce interference from external visual and auditory stimuli. Participants operate the simulator vehicle as they would any automatic transmission vehicle in the real-world. The simulator incorporates an eight degree of freedom motion system, immersive visual environment, surround sound, and driver feedback (through steering torque and brake pedal sensation). Images are generated and rendered at 60 frames per second and presented to give a forward field of view of 250°, rear field of view of 40°, with a vertical field of view of 45°. The simulator system collects data relating to driver behaviour (vehicle control), the vehicle (position, speed, accelerations, etc.) and other autonomous vehicles in the scene (e.g. identity, position and speed) at a rate of 60Hz.

A total of 40 drivers were recruited for a study of traffic management signage strategies involving up to 3 hours of driving. Participants were split evenly into four groups (male high and low experience; female high and low experience). Two levels of surrounding traffic were scripted; a low flow of 100 vehicles per lane per hour and a high flow of 1200 vehicles per lane per hour. These traffic flow rates were presented in separate drives, with the presentation order counterbalanced across the participant sample. The virtual environment presented was a four lane motorway with no hard shoulder and a concrete central reservation. Emergency refuge areas (ERAs) were visible at regular intervals.
In the data used here, we focus on an approach to roadworks where the two outside lanes are closed. Fixed plates signs displayed on both sides of the carriageway included an advance warning of upcoming roadworks, followed by four signs showing an upcoming double lane closure (lanes 1 and 2). The total length of this road section was 1750m and the speed limit remained at 70mph for the duration.

Driver lane choice was recorded once every second. 5,409 observations were collected in total, with average scenario transit time of 67.6 seconds. Data has been analysed from the approach to the lane closure only due to interruptions in driver behaviour from congestion during the roadworks section of the drive.

Stated choice data
The SC survey was conducted for 35 of the 40 respondents who had participated in the simulator experiment. There were three runs per respondent, with different traffic levels. Respondents were faced with static screenshots from the simulator (i.e. using the same graphics) but without a rear view mirror. These were taken at distances of 800, 600, 400 and 200 yards from the roadworks, and respondents were each time asked which lane they would prefer to find themselves in after the next 200 yard section. This is different from making a choice as respondents can indicate a preference without a need to make an actual change, and do so even if the gap between vehicles is very small. The environment is also static, rather than dynamic, with no indication of speed of other vehicles and no control of speed of the vehicle.

Initial analysis
As a first step, we provide an initial analysis of which lanes respondents find themselves in at different distances from the roadworks, in both the simulator and SC data. The results of this are presented in TABLE 1. For the simulator data, we have observations from before the closure signs and then at various segments along the road, as well as after the closure, noting again that this data was not included in our models. For the SC data, respondents are asked in which lane they would like to be over the next 200 yards and the first observation thus relates to the 600 to 400 yards segment, in response to the choice screen faced at 800 yards. The simulator data clearly shows a pattern of respondents moving away from the lanes that will be closed, a process that proceeds at a constant trend, showing that some drivers change lanes early while others do so later on. The same happens in the SC data, and there is a very close correspondence in the share of drivers that filter themselves into lanes 3 and 4 at different distances, with 71.38% and 68.57% at 600 to 400 yards in the simulator and SC data, respectively, where this increases to 83.64% and 83.81% at 400 to 200 yards, and 93.21% and 95.24% for the final 200 yards.

### TABLE 1: initial data analysis – location of drivers at different distances from closure

<table>
<thead>
<tr>
<th></th>
<th>Simulator data</th>
<th>Stated choice data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lane 1</td>
<td>Lane 2</td>
</tr>
<tr>
<td>before closure sign</td>
<td>53.83%</td>
<td>38.48%</td>
</tr>
<tr>
<td>between 800 and 600 yds from closure</td>
<td>18.71%</td>
<td>36.61%</td>
</tr>
<tr>
<td>between 600 and 400 yds from closure</td>
<td>7.47%</td>
<td>21.15%</td>
</tr>
<tr>
<td>between 400 and 200 yds from closure</td>
<td>2.58%</td>
<td>13.77%</td>
</tr>
<tr>
<td>final 200 yds from closure</td>
<td>0.43%</td>
<td>6.36%</td>
</tr>
<tr>
<td>at closure</td>
<td></td>
<td>67.20%</td>
</tr>
</tbody>
</table>

We also asked respondents for their preferred lane in roadworks in a general setting (i.e. outside the experimental setup) and also for how soon they would generally like to make their way into that lane in the approach to roadworks. Our original plan had been to use these variables as indicators for underlying preferences in a hybrid choice model, but the sample size and variability in the data was not sufficient for this. Nevertheless, some interesting insights can be gained. We first asked respondents the question “When approaching a lane closure such as faced here in real life, which lane would you aim to be in when lanes 1 and 2 close, traffic permitting?”. Out of the 35 respondents who completed the survey, 85.71% indicated a preference for lane 3. Comparing the response to this question and the actual lane in which respondents find themselves at the point of the closure of the outside lanes, 58.09% are in their desired lane in the SC data, where it is striking that the figure is higher at 67.14% in the simulator data, despite the question being asked in the SC survey setting. A possible reason for this is the high share of respondents indicating a preference for lane 3, which is also easier for respondents to reach than lane 4 in the simulator.

Finally, respondents were asked the question “How quickly would you aim to get into that lane?” Here, we observe that the correlation between the stated distance for moving and the actual distance in the simulator is only 0.08, which is a direct result of drivers being affected in their
ability to change lanes by traffic around them, a point already alluded to in the introduction. The
correlation is somewhat higher in the SC data, where it is 0.16 in the first run but rises to 0.30 in
the third run.

MODEL STRUCTURE
We first discuss the modelling approach used for the simulator data before turning to the stated
SC data. We finally talk about the incorporation of random heterogeneity in the latent class
model and the joint estimation on both data sources.

Specification of models for simulator data
Drivers approaching lane closures due to roadworks tend to choose a target lane (plan) and seek
suitable gaps to execute the plan (action). The driver may or may not be able to move to the
target lane due to the constraints imposed by the surrounding traffic and hence the plan is
typically unobserved/latent. An example of lane-changing structure for a subject driver in lane 2
of a 4 lane road is shown in Figure 1. The driver first selects a target lane, which is the most
preferred lane considering the traffic conditions and the constraints imposed by the lane closures.
The choice of the target lane indicates the preferred direction of lane change. For example, for
the subject driver in lane 2, lanes 1 is on the left hand side and lanes 3 and 4 are on the right hand
side. If the target lane is the same as the current lane, no lane change is required, and the
observed action is therefore no change. If the target lane is 1, the driver looks for suitable gaps
on the left. If the target lane is lane 3 or lane 4, the driver seeks suitable gaps on the right. A lane
change is observed when the driver finds an acceptable gap in the desired direction and moves to
the left (change left) or to the right (change right). Otherwise, he/she stays in the current lane.
The choice of target lane is unobserved in the trajectory data since multiple decision paths can
lead to the same decision.

Figure 1: The lane-changing framework for a driver on lane 2 of a 4-lane road

The model thus has two components, a target lane choice component and a gap acceptance
component.
**Target lane choice**

A driver is likely to prefer the lane with the highest utility. The utility function of a driver \( n \) for choosing lane \( l \) at a specific time \( t \) can be written as:

\[
U_{n,t}^l = V_{n,t}^l + \varepsilon_{n,t}^l = \delta^l + \beta^l X_{n,t}^l + \varepsilon_{n,t}^l ,
\]

where \( \delta^l \) is a constant for lane \( l \), \( X_{n,t}^l \) is a vector of attribute levels describing lane \( l \) as faced by driver \( n \) at time \( t \), with an associated vector of coefficients \( \beta^l \) which are to be estimated and which show the impact on utility of the attributes. Finally, \( \varepsilon_{n,t}^l \) is a random error term which is distributed identically and independently across choices and alternatives according to a type I extreme value distribution.

We assume that the choice set of the driver includes all lanes that are open to traffic at time \( t \). The candidate attributes affecting the choice of the target lane may include general traffic conditions (e.g. traffic density, average speed, orientation, etc. of each lane), surrounding vehicle attributes (e.g. relative speeds, types of surrounding vehicles, etc.), path-plan impact (e.g. whether or not the driver needs to take an exit or make a mandatory lane change in order to follow the path and if yes, what is the remaining distance to the exit) Further, we explored extending the utility function to include interactions with driver characteristics (e.g. age, experience, stress level, aggressiveness, etc.). In the data available for this analysis, only gender, age and driving experience were available as driver characteristics, and these did not show significant impacts on behaviour.

With the extreme value distribution, the probability of driver \( n \) choosing lane \( l \) (out of \( L \) lanes) as the target lane at time \( t \) is given by:

\[
PL_{n,t}^l (\delta^l , \beta^l ) = \frac{\theta_{n,t}^l \exp(V_{n,t}^l)}{\sum_k \theta_{n,t}^k \exp(V_{n,t}^k)} I\{1, ..., L\} ,
\]

where \( \theta_{n,t}^l \) is 1 if lane \( l \) is open to traffic at time \( t \) for respondent \( n \) and 0 otherwise, ensuring that any lanes that are closed have a probability of zero. This probability is conditional on the vectors of lane specific constants \( \delta^l \) (where one constant is to be normalised to zero) and marginal utility coefficients \( \beta^l \).

**The gap acceptance model**

Gap acceptance is the second level of lane-changing decision-making process and is a result of interaction between the subject drivers and the traffic in the adjacent lane in the direction of the target lane. The interaction can be represented by variables such as relative speed between the subject vehicle and lead and/or lag vehicle at the target lane, relative speed between the subject vehicle and the front vehicle in the current lane, types of vehicle, distance to exit etc.

The driver evaluates both lead and lag gaps against his/her acceptable gaps threshold, known as critical gaps. The lead and lag gaps are accepted if both are greater than the corresponding critical gaps. The critical gap of a driver is not constant or static; rather it can vary among drivers and for the same driver across observations depending on the surrounding conditions. In the existing models [e.g. 14, 15], critical gaps are assumed to follow Log-normal distributions (since the gaps have non-negative values) where explanatory variables represent the mean of the distribution. These can be expressed as follows:

\[
\sigma_{n,t}^{cr,j,l} = \exp(\alpha^l + \gamma^j X_{n,t}^{l,l} + \varepsilon_{n,t}^{l,l}) , j \in \{lead, lag\}
\]
where $G_{n,t}^{cr,l}$ is the critical gap $j$ in the direction of target lane $l$ of driver $n$ at time $t$, where this depends on the current lane. We again have a constant $\alpha^{l}$, along with a vector of explanatory variables $X_{n,t}^{l,1}$ associated with driver $n$ at time $t$, where the impacts of these on the critical gap $j$ are measured by the estimated vector of parameters $\gamma^{l}$. Finally, $\epsilon_{n,t}^{l,1}$ is a random error term associated with critical gap $j$ for driver $n$ at time $t$, which is assumed to follow a normal distribution $\epsilon_{n,t}^{l,1} \sim N(0, \sigma^{l})$.

A move to lane $l$ at time $t$ occurs if the driver accepts both the corresponding lead and the lag gaps. The probability of accepting available gaps in the direction of lane $l$ at time $t$ can be expressed as follows:

$$PG_{n,t}^{l}(\alpha, \gamma) = P\left(G_{n,t}^{\text{lead},l} \geq G_{n,t}^{cr,\text{lead},l}\right) \cdot P\left(G_{n,t}^{\text{lag},l} \geq G_{n,t}^{cr,\text{lag},l}\right), \quad [4]$$

where $G_{n,t}^{\text{lead},l}$ and $G_{n,t}^{\text{lag},l}$ are the available lead and lag gaps for target lane $l$ at time $t$ for driver $n$, which are of course a function of the lane in which the driver currently is.

A Log-normal distribution of the gap acceptance probability can be written as follows:

$$P\left(G_{n,t}^{l,1} \geq G_{n,t}^{cr,l,1}\right) = \Phi\left[\frac{\ln\left(G_{n,t}^{l,1}\right) - \ln\left(G_{n,t}^{cr,l,1}\right)}{\sigma^{l}}\right], \quad [5]$$

where $\Phi[.]$ is the cumulative standard normal distribution. If the available gap is “infinite” as the lane ahead is empty (for the lead gap, for example), then the gap acceptance probability becomes equal to 1.

At any given moment in time, we seek to explain the move of driver $n$ from the lane in time $t$-1 to the lane in time $t$, noting that this lane may be the same (if no change is made). We observe that the driver either moves one lane to the left (each lane change is looked at separately), one lane to the right, or makes no change at all, where this is captured by $y_{n,t}$ taking a value of $-1$, 0 or 1. We then have that the probability of the observed outcome at time $t$ is given by:

$$P_{n,t}^{l}(\delta, \beta, \alpha, \gamma) = \begin{cases} y_{n,t} = -1 \cdot P_{n,t}^{l}(l_{n,t-1} - 1)(\delta, \beta, \alpha, \gamma) \\ + (y_{n,t} = 0) \cdot P_{n,t}^{l}(l_{n,t-1})(\delta, \beta, \alpha, \gamma) \\ + (y_{n,t} = 1) \cdot P_{n,t}^{l}(l_{n,t-1} + 1)(\delta, \beta, \alpha, \gamma) \end{cases} \quad [6]$$

where $l_{n,t-1}$ is the lane for driver $n$ at time $t$-1. Of course, if $l_{n,t-1} = 1$, then $(l_{n,t-1} - 1)$ would become negative, but the first line of Equation [6] would not apply anyway as $y_{n,t}$ could not take the value $-1$. The same applies if $l_{n,t-1} = L$, in which case $y_{n,t}$ cannot take the value $+1$ and the final line of Equation [6] does not apply.

The log-likelihood function for the model estimated on the simulator data would thus be given by:

$$LL_{sim} = \sum_{n=1}^{N} \sum_{t=1}^{T_{n}} \ln\left[P_{n,t}^{l}(\delta, \beta, \alpha, \gamma)\right] \quad [7]$$

where $T_{n}$ is the number of observations for driver $n$.

**Specification of models for stated choice data**

For the SC data, the model specification is far simpler than for the simulator data in that only the target lane model is estimated, where we define $\delta_{SC}$ and $\beta_{SC}$ to be parameters specific to the SC data, giving us a probability of $PL_{n,t,SC}^{l}(\delta_{SC}, \beta_{SC})$ that corresponds to Equation [2]. The log-
likelihood for the SC data is then simply given by:

\[ \text{LL}_{SC} = \sum_{n=1}^{N} \sum_{t=1}^{T_{n,SC}} \ln [ P L_{n,t,SC}(\delta_{SC}, \beta_{SC}) ] , \]  

\[ \text{where } T_{n,SC} \text{ is the number of observations for person } n \text{ in the SC data.} \]

### Latent class models and joint estimation

Given the small number of individual respondents, the estimation of Mixed Logit models was not possible, and we instead proceeded with a Latent Class model, where all parameters are generic across classes except for the lane constants. In particular, we allow for two groups of drivers in terms of the utility for lane 3 (vs other lanes) and two groups of drivers in terms of the utility for lanes 1 and 2 (vs other lanes). The former is meant to capture an underlying preference (or otherwise) for being in lane 3 at the point of closure, while the latter targets differences across respondents in how long they remain in lanes 1 and 2 when approaching the road works.

The probability of a driver \( n \) being in class 2 for the lane 3 layer of classes (which we refer to as layer \( a \)) is given by:

\[ \pi_{n,2,a} = \frac{\exp(\mu_{2a})}{1 + \exp(\mu_{2a})} \]  

where \( \pi_{n,1,a} = 1 - \pi_{n,2,a} \). In class 1, the constant for lane 3 would be given by \( \delta_3 \), while in class 2, it would be \( \delta_3 + \Delta_{2,2a} \). A similar approach is used for the second layer of classes (layer \( b \), which affects the constants for lanes 1 and 2, where we estimate a constant in the class allocation probabilities \( \mu_{2a} \), giving us an equivalent of Equation [9] in \( \pi_{n,2,b} \), and where in class 2, we have \( \delta_i = \delta_i + \Delta_{1-2,2b} \) for \( l=1,2 \). A respondent falls probabilistically into one class in layer \( a \) and one class in layer \( b \), thus giving us a total of 4 classes. We then have that the log-likelihood function for the simulator model becomes:

\[ \text{LL}_{Sim} = \sum_{n=1}^{N} \ln \left[ \sum_{k=1}^{4} \pi_k \prod_{l=1}^{T_{n}} P L_{n,l} \left( \delta_k, \beta, \alpha, \gamma \right) \right] , \]  

\[ \text{where we define } \pi_1 = \pi_{n,1,a} \pi_{n,1,b}, \pi_2 = \pi_{n,2,a} \pi_{n,1,b}, \pi_3 = \pi_{n,1,a} \pi_{n,2,b} \text{ and } \pi_4 = \pi_{n,2,a} \pi_{n,2,b}. \]  

Each class uses a different vector of lane constants, where in class 2 and 4, we shift the base constants for lane 3 by \( \Delta_{3,2a} \), while in class 3 and 4, we shift the base constants for lane 1 and 2 by \( \Delta_{3,2b} \).

For the SC data, we use the same approach and rewrite Equation [8] as:

\[ \text{LL}_{SC} = \sum_{n=1}^{N} \ln \left[ \sum_{k=1}^{4} \pi_k \prod_{l=1}^{T_{n,SC}} P L_{n,l,SC}(\delta_{SC,k}, \beta_{SC}) \right] \]

Finally, we also estimate a joint latent class model where the same class allocation is used for both the simulator and SC data, with generic class allocation probabilities but with different shifts in baseline constants and all other parameters remaining dataset specific. This gives us a log-likelihood function of:

\[ \text{LL}_{Joint} = \sum_{n=1}^{N} \ln \left[ \sum_{k=1}^{4} \pi_k \left( \prod_{l=1}^{T_{n}} P L_{n,l}(\delta_k, \beta, \alpha, \gamma) \right) \prod_{l=1}^{T_{n,SC}} P L_{n,l,SC}(\delta_{SC,k}, \beta_{SC}) \right] \]

\[ \text{where for those 5 individuals for whom only data from the simulator experiments is available, we set } \prod_{l=1}^{T_{n,SC}} P L_{n,l,SC}(\delta_{SC,k}, \beta_{SC}) = 1. \]

### RESULTS

All models were coded and estimated in R. The estimation results are summarised in Table 2 and Table 3.
Base models
We start by looking at the results of the base model estimated on the simulator data alone. In this models, we allow for road segment specific constants for the lanes, where the constants change every time the driver receives a new warning sign about the approaching road works. We thus have constants before the first closure sign, constants between 800 and 600 yards from the closure, and so on. Each time, the constant for lane 4 is normalised to 0, thus estimating the utilities for other lanes relative to this lane. We observe that as the driver gets closer to the lane closure, the utilities for the two outside lanes (lanes 1 and 2) become more negative. A strong negative utility is also associated with any lanes other than the current lane, capturing the penalty with needing to move lanes, where this is constant independent of how many lane changes are required. The final component of the target lane model concerns the characteristics of the lanes themselves, in terms of vehicles in the lane and their speed relative to the driver’s vehicle. We note that the utility of a lane is negatively affected by each additional vehicle in front, although this effect is only weakly significant, while there is a positive change in utility for lanes where there are no vehicles visible behind the driver’s own vehicle. There is a reduced utility for lanes where the closest vehicle in front travels more slowly than the driver’s own vehicle, with a positive shift in utility if the speed of the driver is faster than that of the closest vehicle behind. We finally turn to the parameters of the gap acceptance model. In addition to two constants, which reveal that the critical gap in front is larger than the critical gap behind, we see that the critical gap increases if the driver is travelling faster than the vehicle in front (lead gap) or slower than the vehicle behind (lag gap). Finally, the critical gap reduces (for both lead and lag) as the driver gets closer to the lane closure, but this effect is not statistically significant even if it is behaviourally reasonable, implying that drivers take greater risks. For the base model estimated on the SC data, we see a very clear picture in terms of constants, showing a preference for inside lanes, with no overall significant difference between lanes 3 and 4. A richer pattern is also observed in terms of lane changes required, where each additional lane changes carries a greater disutility. With the static SC data, we are able to estimate a much stronger negative effect for the number of vehicles visible in front of the driver’s own vehicle.

Independent latent class models
For the latent class model estimated on the simulator data alone, we see only a small improvement in log-likelihood compared to the base model, where this improvement is not statistically significant given the increase in the number of parameters. We see that the split into the two classes in layer \( a \) is deterministic, with a near 100% probability of falling into class 1\( a \), and consequently no significant differences in the utility for lane 3 in class 2\( a \). On the other hand, the shift in the utility for lanes 1 and 2 in class 2\( b \) is significant, and the model assigns an overall probability of 38.17% for this class. This shows heterogeneity in the utility for lanes 1 and 2 compared to lanes 3 and 4, with a non-trivial share of respondents having a smaller dislike for these lanes than other respondents. In the presentation of the results, we refer to four overall classes, where class \( a \) as the base class, class \( b \) includes drivers with a shift in preferences for lane 3 only, class \( c \) includes drivers with a shift in preferences for lanes 1 and 2, and class \( d \) includes drivers with shifts in both lanes 1 and 2 and lane 3. The improvement offered by the latent class models is much stronger in the SC data, with a gain in log-likelihood by 10.14 units for 4 additional parameters, which is highly significant. We now see significant heterogeneity not just in the utility for lanes 1 and 2 (where the difference in the utilities is now 3.8 between the two classes) but also in the utility for lane 3.
(with a shift by 1.71 units). The split of respondents is more deterministic for the heterogeneity in lanes 1 and 2, where there is a 90.3% probability of falling into the class with the less negative utility. The probability of falling into the two classes with a higher utility for lane 3 is 36.12%.

**Joint latent class model**

Turning finally to the joint model, we observe a significant improvement in fit over a joint model not taking into account heterogeneity (which would give a log-likelihood of -935.39), at the cost of 6 additional parameters (the two class allocation constants, which are generic across the two datasets, and the dataset specific shift terms). Crucially, this model also gives us a log-likelihood which is no worse than the two separate latent class models despite assuming generic class allocation probabilities. This supports the notion that the heterogeneity retrieved by this latent class model is person specific and share across the two data environments, confirming that some respondents have inherent preferences for given lanes. We see that for the driving simulator data, the shift is again only significant in the utility for lanes 1 and 2, while for the SC data, the shift is highly significant for lane 3 and weakly significant for lanes 1 and 2. Crucially, the shifts for lanes 1 and 2 are the same sign in both data sets. We also see a more even distribution across classes in the joint models, with the lowest class probability now being 7.15% compared to 3.5% and the highest dropping from 57.68% to 53.53%. The variations of the lane constants at different distances are presented in Figure 2. As can be seen, the relative differences between the driving simulator and SC pairs of constants of each latent class of drivers are very similar across all 3 common cases (noting that lane 1 is unavailable in the 200 yards to closure section for the SC models).
<table>
<thead>
<tr>
<th></th>
<th>Base simulator model</th>
<th>Base SP model</th>
<th>Simulator LC model</th>
<th>SP LC model</th>
<th>LC model on joint data</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>-641.05</td>
<td>-294.34</td>
<td>-639.38</td>
<td>-284.20</td>
<td>-923.40</td>
</tr>
<tr>
<td>par</td>
<td>25</td>
<td>13</td>
<td>29</td>
<td>17</td>
<td>44</td>
</tr>
<tr>
<td>BIC</td>
<td>1,496.99</td>
<td>667.20</td>
<td>1,528.04</td>
<td>671.08</td>
<td>2,228.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>est</th>
<th>rob t</th>
<th>est</th>
<th>rob t</th>
<th>est</th>
<th>rob t</th>
<th>est</th>
<th>rob t</th>
<th>est</th>
<th>rob t</th>
<th>est</th>
<th>rob t</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_1 ) (before sign)</td>
<td>-1.45</td>
<td>-3.75</td>
<td>-1.74</td>
<td>-4.24</td>
<td>-1.65</td>
<td>-4.11</td>
<td>-1.65</td>
<td>-4.11</td>
<td>-1.65</td>
<td>-4.11</td>
<td>-1.65</td>
<td>-4.11</td>
</tr>
<tr>
<td>( \delta_2 ) (before sign)</td>
<td>-1.79</td>
<td>-4.63</td>
<td>-2.07</td>
<td>-4.57</td>
<td>-1.97</td>
<td>-4.86</td>
<td>-1.97</td>
<td>-4.86</td>
<td>-1.97</td>
<td>-4.86</td>
<td>-1.97</td>
<td>-4.86</td>
</tr>
<tr>
<td>( \delta_3 ) (before sign)</td>
<td>-2.94</td>
<td>-6.7</td>
<td>-2.85</td>
<td>-6.06</td>
<td>-2.86</td>
<td>-6.67</td>
<td>-2.86</td>
<td>-6.67</td>
<td>-2.86</td>
<td>-6.67</td>
<td>-2.86</td>
<td>-6.67</td>
</tr>
<tr>
<td>( \delta_4 ) (before sign)</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \delta_1 ) (800 to 600yds)</td>
<td>-4.08</td>
<td>-10.36</td>
<td>-4.68</td>
<td>11.04</td>
<td>-4.50</td>
<td>-8.96</td>
<td>-4.50</td>
<td>-8.96</td>
<td>-4.50</td>
<td>-8.96</td>
<td>-4.50</td>
<td>-8.96</td>
</tr>
<tr>
<td>( \delta_3 ) (800 to 600yds)</td>
<td>-1.70</td>
<td>-3.12</td>
<td>-1.67</td>
<td>-3.1</td>
<td>-1.62</td>
<td>-3.27</td>
<td>-1.62</td>
<td>-3.27</td>
<td>-1.62</td>
<td>-3.27</td>
<td>-1.62</td>
<td>-3.27</td>
</tr>
<tr>
<td>( \delta_4 ) (800 to 600yds)</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \delta_1 ) (600 to 400yds)</td>
<td>-3.64</td>
<td>-8.36</td>
<td>-3.76</td>
<td>-6.34</td>
<td>-4.50</td>
<td>-9.33</td>
<td>-4.25</td>
<td>-9.33</td>
<td>-4.25</td>
<td>-9.33</td>
<td>-4.25</td>
<td>-9.33</td>
</tr>
<tr>
<td>( \delta_2 ) (600 to 400yds)</td>
<td>-3.89</td>
<td>-7.32</td>
<td>-1.73</td>
<td>-3.46</td>
<td>-4.47</td>
<td>-7.48</td>
<td>1.67</td>
<td>1.42</td>
<td>-4.19</td>
<td>-6.23</td>
<td>-2.38</td>
<td>-3.44</td>
</tr>
<tr>
<td>( \delta_3 ) (600 to 400yds)</td>
<td>-1.67</td>
<td>-3.35</td>
<td>0.23</td>
<td>0.64</td>
<td>-1.65</td>
<td>-3.29</td>
<td>-0.34</td>
<td>-0.78</td>
<td>-1.58</td>
<td>-3.11</td>
<td>-0.20</td>
<td>-0.48</td>
</tr>
<tr>
<td>( \delta_4 ) (600 to 400yds)</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \delta_1 ) (400 to 200yds)</td>
<td>-3.50</td>
<td>-5.99</td>
<td>-3.44</td>
<td>-4.23</td>
<td>-4.51</td>
<td>-5.78</td>
<td>-0.18</td>
<td>-0.13</td>
<td>-4.30</td>
<td>-4.23</td>
<td>-4.21</td>
<td>-4.31</td>
</tr>
<tr>
<td>( \delta_2 ) (400 to 200yds)</td>
<td>-4.02</td>
<td>-6.98</td>
<td>-1.98</td>
<td>-2.67</td>
<td>-4.89</td>
<td>-7.09</td>
<td>1.27</td>
<td>0.97</td>
<td>-4.56</td>
<td>-5.06</td>
<td>-2.74</td>
<td>-2.85</td>
</tr>
<tr>
<td>( \delta_3 ) (400 to 200yds)</td>
<td>-0.62</td>
<td>-1.35</td>
<td>0.01</td>
<td>0.04</td>
<td>-0.61</td>
<td>-1.34</td>
<td>-0.46</td>
<td>-1.21</td>
<td>-0.53</td>
<td>-1.2</td>
<td>-0.35</td>
<td>-0.86</td>
</tr>
<tr>
<td>( \delta_4 ) (400 to 200yds)</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \delta_1 ) (final 200yds)</td>
<td>-5.13</td>
<td>-6.78</td>
<td>0</td>
<td>-</td>
<td>-6.20</td>
<td>-8.11</td>
<td>0.00</td>
<td>-</td>
<td>-6.13</td>
<td>-7</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \delta_2 ) (final 200yds)</td>
<td>-4.22</td>
<td>-9.08</td>
<td>-2.52</td>
<td>-5.01</td>
<td>-5.24</td>
<td>-6.64</td>
<td>-0.36</td>
<td>-0.35</td>
<td>-5.08</td>
<td>-5.64</td>
<td>-3.99</td>
<td>-3.47</td>
</tr>
<tr>
<td>( \delta_3 ) (final 200yds)</td>
<td>-0.89</td>
<td>-1.77</td>
<td>0.57</td>
<td>1.59</td>
<td>-0.86</td>
<td>-1.66</td>
<td>0.15</td>
<td>0.34</td>
<td>-0.79</td>
<td>-1.43</td>
<td>0.20</td>
<td>0.47</td>
</tr>
<tr>
<td>( \delta_4 ) (final 200yds)</td>
<td>0.00</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \delta_3 ) (at closure)</td>
<td>-0.23</td>
<td>-0.87</td>
<td>0.23</td>
<td>0.67</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>( \delta_4 ) (at closure)</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

\[ \Delta_{1,2a} \]  
\[ \Delta_{1-2b} \]  
\[ \mu_{2a} \]  
\[ \mu_{2b} \]  
\[ \pi_a \]  
\[ \pi_b \]  
\[ \pi_c \]  
\[ \pi_d \]
### TABLE 3: Estimation results (part 2)

<table>
<thead>
<tr>
<th></th>
<th>Base simulator model</th>
<th>Base SP model</th>
<th>Simulator LC model</th>
<th>SP LC model</th>
<th>LC model on joint data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>rob t</td>
<td>est</td>
<td>rob t</td>
<td>est</td>
</tr>
<tr>
<td>Constant for gap acceptance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front</td>
<td>2.23</td>
<td>3.67</td>
<td>2.24</td>
<td>3.41</td>
<td>2.12</td>
</tr>
<tr>
<td>Back</td>
<td>1.29</td>
<td>1.23</td>
<td>1.39</td>
<td>1.22</td>
<td>1.22</td>
</tr>
<tr>
<td>Shift in gap if slower than</td>
<td>0.10</td>
<td>1.53</td>
<td>0.10</td>
<td>1.35</td>
<td>0.11</td>
</tr>
<tr>
<td>Vehicle behind (m/s)</td>
<td>0.16</td>
<td>2.2</td>
<td>0.17</td>
<td>2.2</td>
<td>0.15</td>
</tr>
<tr>
<td>Shift in gap if faster than</td>
<td>-0.30</td>
<td>-0.32</td>
<td>-0.30</td>
<td>-0.28</td>
<td>-0.40</td>
</tr>
<tr>
<td>Vehicle front (m/s)</td>
<td>-5.80</td>
<td>-19.48</td>
<td>-5.77</td>
<td>-19.03</td>
<td>-5.77</td>
</tr>
<tr>
<td>Change in gap every km closer</td>
<td>-4.90</td>
<td>-5.9</td>
<td>-4.93</td>
<td>-5.97</td>
<td>-4.96</td>
</tr>
<tr>
<td>Change 1 lane</td>
<td>-0.07</td>
<td>-1.02</td>
<td>-0.07</td>
<td>-0.94</td>
<td>-0.07</td>
</tr>
<tr>
<td>Change 2 lanes</td>
<td>-0.07</td>
<td>-1.12</td>
<td>-0.07</td>
<td>-1.26</td>
<td>-0.07</td>
</tr>
<tr>
<td>Change 3 lanes</td>
<td>-0.06</td>
<td>-1.79</td>
<td>0.45</td>
<td>1.8</td>
<td>0.44</td>
</tr>
<tr>
<td>Vehicles visible faster than</td>
<td>0.27</td>
<td>3.5</td>
<td>0.27</td>
<td>2.75</td>
<td></td>
</tr>
<tr>
<td>Vehicle front (m/s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.28</td>
</tr>
<tr>
<td>Empty lane behind</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicles visible faster than</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle behind (m/s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The table above presents estimation results for various parameters using different models and data sets. The columns labeled `est` and `rob t` likely refer to estimated values and robust standard errors, respectively. The models include `Base simulator model`, `Base SP model`, `Simulator LC model`, `SP LC model`, and `LC model on joint data`. The table illustrates how each parameter is estimated across these models, providing insights into traffic flow and vehicle behavior.
Figure 2: Lane specific constants at different distances and for different latent classes.
CONCLUSIONS
The joint models enable us to untangle the effects of (unconstrained) target lane selection and
how the manifestation of the plan to change to the target lane is affected by the constraints
imposed by the other drivers in the driving simulator (and real traffic). This can be used in better
traffic management – in optimum placing of road closure signs for instance (i.e. in cases where
we want drivers to change target lanes sooner).

Further, the similarities and differences in the stated choice and driving simulator results
and their respective strengths (i.e. SC providing crisp data about the target/plan simulator
providing data about the implementation/action) can be used to reduce the sample sizes/duration
of the simulator experiments and hence allow economy in data collection costs without
compromising model fidelity.

Potential direction of future research can focus on more advanced modelling techniques
for joint model development - enriching the latent class membership component with driver
demographics, using the SC lane preferences as indicators in the combined model, to name a
few.

ACKNOWLEDGEMENTS
The Leeds authors acknowledge the financial support by the European Research Council
through the consolidator grant 615596-DECISIONS

REFERENCES
   caused by freeway work zones using archived work zone and ITS traffic data.
   Transportmetrica, 8(4), 307-320.
2. FHWA (2016). FHWA Work Zone Facts and Statistics
   interstate highway capacity for short-term work zone lane closures. Transportation
   Research Record: Journal of the Transportation Research Board, (1877), 85–94.
   in Highway Work Zone Bottlenecks. Project No. 046IY02. NEXTRANS Center, West
   Lafayette, IN, 2011.
   Research Record: Journal of the Transportation Research Board, (1937)
   construction work zone clarke. Transportation Research Record: Journal of the
   driving simulator using surrogate safety measures. Accident Analysis & Prevention,
   40(1), 274-288.
   design and operation. Transportation Research Record: Journal of the Transportation
   Research Board, (2248), 87-95.


15. Toledo T; Choudhury CF; Ben-Akiva ME (2005) Lane-changing model with explicit target lane choice, Transportation Research Record, pp.157-165.