

## Part B. Collections of Information Employing Statistical Methods

### 1. Describe potential respondent universe and any sampling selection method to be used.

Our target population consists of private drivers who live at the outer end of the study routes and who use these routes during morning and afternoon rush hours. The respondents may include those residents who work locally as commercial vehicle operators and who frequently drive these routes during rush hours during work. We use mail invitations for sampling and use United States Postal Service mailing lists. We sample from the area east of downtown Orlando, Florida focusing on zip codes that lie immediately northeast or southeast of the eastern most end of our East Orlando study route. We also sample from the area west of downtown Orlando Florida focusing on zip codes that lie northwest or southwest of the western most end of our West Orlando study route. We sample from the area northeast of Atlanta focusing on zip codes that lie immediately northeast and northwest of the northern most end of our North East Atlanta study route. We also sample from the area northeast and northwest of the northern most end of our North West Atlanta study route.

Mass mailings are sent to relevant zip code areas, and within these zip codes we oversample mail carrier routes with median income levels that are below the state-wide median income levels and otherwise randomly select mail carrier routes. We oversample carrier routes that have lower income levels in order to maximize the potential for including low income respondents.

Respondents who qualify for the single driver simulator experiments (because they indicate they do not easily suffer from nausea) are then randomly assigned to treatments. Treatments in the single driver simulator include variations in congestion and tolls and earnings consequences associated with the route choice in the simulator, prizes and probabilities in the risky prospects, travel times and range of earnings in the belief task, and order of certain tasks in a meeting. Respondents who do not qualify for the single driver simulator due to nausea may not participate in the Orlando regions since no multi-driver experiments are conducted there, but respondents with nausea issues in Atlanta may be reassigned to the multiple-driver experiments. The instruction that persons who easily suffer from nausea should not participate is on the web page where respondents schedule themselves. If a participant becomes too nauseous during a driving simulator session to continue in the study they will be paid the participation fees for the sessions they have

attended, and may complete other tasks in the session and be paid for those.

Respondents who qualify for multi-driver simulator experiments are also randomly assigned to treatments. Randomization across tolls is done at the session level. Randomizations based on prizes and probabilities in the risky prospects and in the belief task are done at the individual level. There is no other qualification for the multi-driver experiments than to be available at the times the meetings are scheduled.

All respondents participate in the field experiment and are randomly assigned to treatments. Sample attrition will be studied formally to assess if drop-out is correlated with any observable characteristic of the individual that affects subsequent statistical analyses and hypothesis tests. We expect about 20% attrition after the first meeting. The short duration of the study should minimize attrition.

2. Describe procedures for collecting information, including statistical methodology for stratification and sample selection, estimation procedures, degree of accuracy needed, and less than annual periodic data cycles.

### Sampling

Respondents to the mailings will attend four face to face sessions where we will collect the choice data in the tasks we present to them. The same respondents will also be provided with GPS units for collecting their field driving choices. Each respondent is assigned an anonymous participant ID that we will use to link responses across tasks. A panel data set will be built and then analyzed using Maximum Likelihood and Maximum Simulated Likelihood methods. These methods allow us to estimate non-linear, structural choice models where risk attitudes and perceptions play a key role.

In the field experiment we will vary the tolls across drivers. Toll amounts are collected individually from each participant and does not involve any toll agencies. Toll amounts are simply deducted from the payment they get each time they drive in the study. In Orlando we use 10 sets of tolls and in Atlanta 14 sets of tolls, where a set consists of a combination of positive, zero or negative tolls on each of the express and local road. A negative toll is simply a subsidy for taking one route rather than another. In addition each driver will face three of these sets. One of these will have zero tolls on both the express and local road, this is the baseline. The other two will be selected from the remaining sets. Toll amounts are randomly assigned across the participants in a uniform manner so

approximately 50 of the 1,200 participants will be assigned to each toll in the GPS recorded study drives. Participants are informed about all earnings consequences of their route choices, including what the tolls are, in the meeting preceding each driving period. Tolls do not change during a driving period.

Selections of tolls in the field experiments are random across participants. We use pseudo-random number generators in Microsoft Excel with a uniform distribution across tolls. Each participant experiences three toll levels, the first one is a baseline with a zero toll. These variations will allow us to estimate price elasticities conditional on the varying travel times and departure times, controlling for a range of driver characteristics. The variations are also necessary in order to characterize the range of risk preference and perception structures that motivate driver choices, such as the effect on responses due to variations in the attitude to travel time reliability and due to variations in perceptions of travel times.

Toll charges are always deducted from the initial payment for driving in the study. These initial payments are always larger than the toll charged so participants never pay out of pocket.

In the single driver simulator experiments tolls are also selected randomly from a set. Each toll is equally likely to be selected by any one participant. Some tasks involving a number of drives use different tolls for the same driver, but there are also tasks that use the same toll across multiple drives. The former tasks are intended to identify the risk attitude of the driver, the same way as is done in the risky prospect tasks, and the latter are intended to estimate how drivers form expectations of risk of congestion and travel time distributions and potential biases in these. Tolls vary from 50 cents to \$5.50 in 10 cent increments and are randomly selected by each participant from a uniform distribution. More information on details of the design is given in the section "Overview of experimental instruments used as determinants to estimation models". Each participant only gets a maximum of 3 tolls in any task so estimation at the individual level is not possible and data is pooled across participants. The estimated models are therefore representative agent models, although we control for differences across demographic characteristics such as income and education levels. Because of the limited number of observations on each participant in this study a much larger sample is needed than what is traditionally used in lab experiments with students. With a total sample of 840 in this experiment, we will have approximately 210 in each of our four regions. Most lab experiments that estimate 2-3 parameter models, like we do here, are based on sample sizes between 50 and 100.

In the multiple driver traffic simulator experiment every driver in the same session pays the same toll. Across tasks in a given session the toll is fixed, but the toll varies across sessions. We include only two toll levels in these experiments since the size of each cohort limits the number of cohorts we can include. Each cohort experiences both tolls but in different order. Some experience the low toll first and others the high toll. This allows us to test the reactions of drivers and the impact this has on traffic system to both increases and decreases in tolls. Three hundred and sixty subjects are expected to participate in this experiment. Each cohort of up to 40 participants make route choices in the same traffic system, so the observations across participants are not independent. This data will be used to test system wide properties of how the traffic converges to equilibrium predictions. Due to the large number of participants needed for each system there will not be enough data points to perform maximum likelihood estimations of structural models, as is done for the single driver simulator experiments and is explained in the next several sections. Instead descriptive statistics and non-parametric tests will be used for hypothesis testing. This is the approach taken in the large literature on market analysis based on experimental data so there is a long tradition using this methodology. The distribution of our recruited sample across the single and multiple driver experiments is based on trading off the value of additional observations in the core single driver experiment for increased statistical power, to the needs for having large cohort sizes in the multiple driver experiments. Both sets of experiments will be informative to the purposes of demonstrating the experimental methodology and to the methodological validity tests. In addition to the observations on route choice in the traffic system, this data can be pooled with tasks that estimate the beliefs that participants have over the travel times in the traffic simulator (as opposed from the tasks intended to estimate this for the field travel times) as well as any of the tasks intended to estimate the risk attitudes of the participants. Perceptions of travel time distributions, perceptions of congestion risk, and any biases in these, along with the risk attitudes of the drivers, are expected to explain behavior in the route choice tasks in the multiple-driver traffic simulator experiments.

### Estimation strategy

The analysis involves estimations of non-linear random utility models. Degree of accuracy needed is based on 5% significance levels in hypotheses tests. Each panel spans up to 10 weeks and we will collect 3 panels, one for summer 2011, one for fall 2011, and one for spring 2012. A respondent will participate in only one of the 3 panels. Because of the anonymity of participants, we will be unable to generate survey weights based on information about their nine-digit

ZIP code, but we will generate survey weights based on the demographic information that participants will provide to us that includes their 5-digit ZIP code; these weights will be based on corresponding weights from public-use microdata samples from the U.S. Census available at <http://www.census.gov/main/www/pums.html>. The individuals listed under this section's question 5 have previously constructed weights of this kind using comparable information in Denmark, and have extensive experience in the statistical evaluation of complex survey data.<sup>1</sup> We stratify the sample by four different regions: Orlando East, Orlando West, Atlanta North East and Atlanta North West motivated by our hypothesis that models estimated on one region can be used to predict on another region. Sample selection in terms of household income levels will be controlled for since we have household income data on all invited households from the mailing lists. The purpose of the study is to test the validity of estimated models across a limited set of populations and not to generate findings that are representative of the broader population, thus further stratification is not employed. Demographic characteristics of the respondents are observed, however, and will be used to test systematic variations in behavior.

In the following we will give an overview of our non-linear maximum likelihood methods. We start with an overview of the choice instruments, which is a necessary background to understand our models. Then we detail the estimation strategy.

### Overview of experimental instruments used as determinants to estimation models

We use a multitude of instruments to collect the behavioral data needed to perform our estimation exercises. The multitude of instruments have been designed so that we can collect relevant information that allow us to characterize each respondent by risk

<sup>1</sup> Harrison, Glenn W., Morten I. Lau, E. Elisabet Rutstrom, and Melonie B. Sullivan, "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues", *Field Experiments in Economics*, Carpenter, Jeffrey, Harrison, Glenn W., and List, John (eds), (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10), 2005, 125-218. Harrison, Glenn W., Morten I. Lau, and E. Elisabet Rutstrom, "Estimating Risk Attitudes in Denmark", *Scandinavian Journal of Economics*, 109(2), June 2007, 341-368. Harrison, Glenn W., Morten I. Lau and E. Elisabet Rutstrom, "Lost in State Space: Are Preferences Stable?", *International Economic Review*, 49(3), August 2008, 1091-1112. These papers report on field experiments conducted in Denmark where survey weights were employed to make the findings representative. For these experiments the sample was stratified by multiple geographic regions.

attitude, perception of the risk of congestion, perception of travel time distributions, demographic characteristics and driving habits. Before undertaking any task the respondent is given complete instructions on what the choice options are and how these are related to the payments they can receive. Estimation of risk attitudes requires that respondents are given full information about the probability of various payment outcomes, whether they are from the choice over risky prospect task or from the driving simulators. Estimation of changes in perception about the risk of congestion from driving simulators requires that the respondent is not given full information about the probability of various payment outcomes, but is given full information about the payments conditional on the congestion outcome. In all these cases the participant receives full information on the level of the toll, since they need to know the payment consequences of their choices. In the belief elicitation task it is again essential that the respondent knows the payment consequences conditional on certain travel time outcomes, but since the purpose is to learn what the respondent believes the travel time to be, this information cannot be given to them. Thus, the amount of information given to participants is a crucial aspect of the experimental design and reflect various information conditions that drivers have in their normal driving environment. In the field experiment, we complement our experimental observations with official data on road conditions so as to control for information that is available to them in real time.

Participants are presented with tasks that involve choices between risky prospects unrelated to traffic in order to estimate their risk attitudes. These instruments have been widely used and it is generally found that the coefficient of relative risk aversion lies in between 0.5 and 1 for both adult field respondents and students, indicating risk aversion rather than risk neutrality or risk preference. However, risk attitudes are very heterogeneous so no single measure of risk aversion can be used for the entire population. We have used many different interfaces in the past, using table presentations or visual pie chart presentations, presenting many pairwise tasks together or presenting them separately. Our judgment based on our published work is that for non-student respondents a sequential presentation of the tasks, coupled with visual presentations, make the responses more precise and robust. We therefore estimate risk attitudes both through these instruments and through tasks in the driving simulator. This allows us to compare methodologies.

One concern is that participants of low education levels may have more of a problem understanding these tasks than participants with higher education levels. Our instruction approach is based on having one-on-one research assistants available to help each

participant with tasks. These assistants walk participants through the practice opportunity before the actual tasks, giving participants additional opportunities to clarify their understanding of the tasks. Since we present participants with both risky prospect choices, i.e. the dice game discussed in part A, section 1 as well as simulator driving tasks, we will be able to see whether education levels explain differences in inferred risk attitudes across these tasks. This partly fulfills the purpose of methodological validation.

Participants also perform tasks in the driving simulator where they do not know the probability of congestion but are allowed to gain experience and adjust their choices over time. These tasks will generate data that will allow estimations of functions that capture learning and belief updating over time and that will allow a better understanding of biases in the perception of congestion risks.

Participants are also presented with tasks in which they are asked to report their beliefs about travel times on the routes used in the field experiment. These instruments rely on methods of incentive compatibility using various scoring rules. We have employed these instruments in the past on student populations. We have estimated beliefs about presidential elections, beliefs about outcomes on psychological tests, and beliefs about draws from bingo cages. We rely on these earlier findings. We have used many different interfaces and found that the visual interface with sliders that we employ here works well to produce precise and robust responses. In a study using visual simulations of forest fires the research team has found that respondents are able to make very precise predictions of the fire risks using scoring rules.<sup>2</sup>

Here we introduce these non-simulator tasks. First we introduce the pairwise choices between risky prospects, followed by the travel time belief elicitation instruments and a description of the simulator choices. For each we also present the estimation strategy.

Each participant is presented with a choice between two risky prospects, which we can call A or B. Table 1 illustrates the basic payoff matrix presented to participants. In this illustration, the first row shows that lottery A offers a 10% chance of receiving \$3 and a 90% chance of receiving \$21.60. The expected value (EV) of this lottery,  $EV^A$ , is shown in the third-last column as \$1.64, although the EV columns will not be presented to participants. Similarly, lottery B in the first row has

---

<sup>2</sup> Fiore, Stephen M., Glenn W. Harrison, Charles E. Hughes, and E. Elisabet Rutstrom, "Virtual Experiments and Environmental Policy," Journal of Environmental and Economic Management, 57(1), January 2009, 65-86.

chances of payoffs of \$3.85 and \$0.10, for an expected value of \$0.48. Thus the two lotteries have a relatively large difference in expected values, in this case \$1.17. As one proceeds down the matrix, the expected value of both lotteries increases, but the expected value of lottery B becomes greater than the expected value of prospect A. For example on row 4, the probability of getting the higher prize is 0.4, the EV of lottery A is \$1.76, the EV of lottery B is \$1.60, but on the next row the rank ordering of the EVs change so the EV of lottery A (\$1.80) is less than the EV of lottery B (\$1.98).

The participant chooses A or B in each row, and one row is later selected at random for payout for that participant. The logic behind this test for risk aversion is that only risk-loving participants would take prospect B in the first row, and only risk-averse participants would take prospect A in the second last row. Arguably, the last row is simply a test that the participant understood the instructions, and has no relevance for risk aversion at all. A risk neutral participant should switch from choosing A to B when the EV of each is about the same, so a risk-neutral participant would choose A for the first four rows and B thereafter.

These type of choices over risky prospects can be varied to capture situations when the probabilities are not known, referred to as situations of uncertainty rather than risk. Variations in stakes to captures situations of greater or lesser consequence can easily be implemented as well. The proposed project uses both pairwise choices such as these to generate decision data, but also designs choice situations in the driving simulator and in the field that are isomorphic to these risky prospects. The exact monetary amounts and probabilities used will not be identical to those used here in the illustrations, but the general logic and the implications for the estimation strategy are the same.



**Table 1: Pairwise Choices of Risky Prospects. An Illustration**

Prospect A				Prospect B				EV <sup>A</sup>	EV <sup>B</sup>	Difference
p(\$2)		p(\$1.60)	)	p(\$3.85)		p(\$0.10)	)			
0.1	\$2	0.9	\$1.60	0.1	\$3.85	0.9	\$0.10	\$1.64	\$0.48	\$1.17
0.2	\$2	0.8	\$1.60	0.2	\$3.85	0.8	\$0.10	\$1.68	\$0.85	\$0.83
0.3	\$2	0.7	\$1.60	0.3	\$3.85	0.7	\$0.10	\$1.72	\$1.23	\$0.49
0.4	\$2	0.6	\$1.60	0.4	\$3.85	0.6	\$0.10	\$1.76	\$1.60	\$0.16
0.5	\$2	0.5	\$1.60	0.5	\$3.85	0.5	\$0.10	\$1.80	\$1.98	-\$0.17
0.6	\$2	0.4	\$1.60	0.6	\$3.85	0.4	\$0.10	\$1.84	\$2.35	-\$0.51
0.7	\$2	0.3	\$1.60	0.7	\$3.85	0.3	\$0.10	\$1.88	\$2.73	-\$0.84
0.8	\$2	0.2	\$1.60	0.8	\$3.85	0.2	\$0.10	\$1.92	\$3.10	-\$1.18
0.9	\$2	0.1	\$1.60	0.9	\$3.85	0.1	\$0.10	\$1.96	\$3.48	-\$1.52
1	\$2	0	\$1.60	1	\$3.85	0	\$0.10	\$2.00	\$3.85	-\$1.85

## Estimation Models for Risky Prospect tasks

With the binary choice nature of the tasks and the variations in monetary prizes and probabilities it is possible to estimate a full structural decision model using Maximum Likelihood, instead of using methods to regress choice on explanatory variables in reduced form. This is a benefit in that it allows complete flexibility in the assumed functional forms of the underlying theoretical models.

Data from observing choices across prospects that vary in risk, as in this illustration, can then be used to estimate decision models. For example, a commonly used decision model is that of Expected Utility, and alternatives include the Rank Dependent Utility model and Prospect Theory. Here the presentation will stay with the Expected Utility model for illustrative purposes, hereafter referred to as EU. Assume for the moment that utility of income is defined by

$$U(x) = x^{(1-r)}/(1-r) \quad (1)$$

where  $x$  is the prize in the prospect and  $r \neq 1$  is a parameter to be estimated. For  $r=1$  assume  $U(x)=\ln(x)$  if needed. Thus  $r$  is the coefficient of Constant Relative Risk Aversion (CRRA):  $r=0$  corresponds to risk neutrality,  $r<0$  to risk loving, and  $r>0$  to risk aversion. Let there be  $K$  possible outcomes in a risky prospect. Under EU the probabilities for each outcome  $k$ ,  $p(k)$ , are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each prospect  $i$ :

$$EU_i = \sum_{k=1,K} [ p(k) \times U(k) ]. \quad (2)$$

The EU for each prospect pair is calculated for a candidate estimate of  $r$ , and the index

$$\nabla EU = EU_R - EU_L \quad (3)$$

calculated, where  $EU_L$  is the “left” prospect and  $EU_R$  is the “right” prospect. This latent index, based on latent preferences, is then linked to the observed choices using a standard cumulative normal distribution function  $\Phi(\nabla EU)$ . This “probit” function takes any argument between  $\pm\infty$  and transforms it into a number between 0 and 1. Thus we have the probit link function,

$$\text{prob}(\text{choose prospect R}) = \Phi(\nabla EU) \quad (4)$$

The logistic function is very similar, and leads instead to the “logit” specification.

The link function forms the critical statistical link between observed binary choices, the latent structure generating the index  $y^*$ , and the probability of that index  $y^*$  being observed. In our applications  $y^*$  refers to some function, such as (3), of the EU of two prospects; or, later, the prospective utility of two prospects. The index defined by (3) is linked to the observed choices by specifying that the R prospect is chosen when  $\Phi(\nabla EU) > \frac{1}{2}$ , which is implied by (4).

Thus the likelihood of the observed responses, conditional on the EUT and CRRA specifications being true, depends on the estimates of  $r$  given the above statistical specification and the observed choices. The “statistical specification” here includes assuming some functional form for the cumulative density function (CDF). If we ignore responses that reflect indifference for the moment the conditional log-likelihood would be

$$\ln L(r; y, \mathbf{X}) = \sum_i [ (\ln \Phi(\nabla EU) | y_i = 1) + (\ln \Phi(1-\nabla EU) | y_i = -1) ] \quad (5)$$

where  $y_i = 1(-1)$  denotes the choice of Prospect R (L) in risk aversion task  $i$ , and  $\mathbf{X}$  is a vector of individual characteristics reflecting age, sex, race, and so on.

### Extending the Estimation Model

This approach can easily be extended to alternative utility functions, and to other theories of decision under risk, including Rank Dependent Utility and Prospect Theory. For the choice tasks in driving simulators and in the field, the income argument that is used in these illustrations will be replaced by arguments that capture how their well being depends on their travel choices. The utility function to the traveler can, for example, be specified as:

$$U(d) = U \left\{ T - t(d) - \alpha \max_{\square} (0, a^* - d - t(d)) + \beta \min_{\square} (0, a^* - d - t(d)) \right\}$$

$T$  is simply the time available to spend on travel, on waiting for  $a^*$  at the destination, or on leisure time at home. Thus,  $T$  is net of time spent working. If we were to include the work time and further were to assume that the optimal choice of work time is unconstrained within the 24 hour daily cycle then the value of time would simply collapse to the (utility of the) wage rate, since the first order condition would imply that the opportunity cost of leisure time lost is equal to the marginal value of time spent at work. The researchers think it is reasonable to assume that the work time choice has some discrete constraints, such that you cannot choose to work for 6 hours and 23 minutes, for

instance. The easiest model is to assume that work time is constant and just disregard it in the model.

Since  $t$  (travel time) is increasing in  $d$  (departure time), a delay in departure will result in a decrease in utility. Baseline utility is simply  $U(T)$ , the value of leisure time at home. Based on a vast literature estimating utility functions,<sup>3</sup> we expect this to be a concave function, so that earlier departures would have an increasing marginal utility loss. The researchers simply deduct from this  $T$  the amount of time used for travel, the amount of time wasted at the destination if arriving too early, and the benefits lost from arriving too late. There is no reason to assume that the marginal disutility of waiting time at the destination, or of the travel time, or of the late penalty, should directly affect the marginal utility of the remaining leisure, other than through the direct time allocation.

$\beta$  is the per minute penalty of being late to the destination. This function may very well be nonlinear in  $t$ , but is always increasing which is what matters to the monotonicity property of  $U$  in  $t$ . We expect  $\beta$  to be weakly greater than 1 to reflect that arriving late is more costly than the opportunity cost of foregone leisure time. If it is not then going to work is always dominated by staying at home.

$\alpha$  is the marginal disutility of wasting time at the destination before  $a^*$ . The value of this parameter is crucial for the monotonicity property of  $U$  in  $t$ . If  $\alpha < 1$  then wasting time at the destination is less costly than foregoing time at home. If  $\alpha > 1$  there is an additional marginal utility loss from being at the destination too early, over and above the lost opportunity of leisure at home.

---

<sup>3</sup> Examples from this literature involving the researchers from this study include: Harrison, Glenn W., Morten I. Lau and E. Elisabet Rutstrom, "Estimating Risk Attitudes in Denmark," Scandinavian Journal of Economics, 109(2), June 2007, 341-368. Andersen, Steffen, Glenn W. Harrison, Morten I. Lau and E. Elisabet Rutstrom, "Eliciting Risk and Time Preferences," Econometrica, (76) 3, May 2008, 583-618. Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutstrom, "Preference Heterogeneity in Experiments: Comparing the Lab and the Field," (with Steffen Andersen, Glenn W. Harrison, and Morten I. Lau), Journal of Economic Behavior and Organization, 73, 2010, 209-224. Harrison, Glenn W., and E. Elisabet Rutstrom, "Risk Aversion in the Laboratory," (with Glenn W. Harrison) in J.C. Cox and G. W. Harrison (eds.), Risk Aversion in Experiments (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008). The literature is based on samples from various populations and population segments, but the general finding is consistently that very few, if any, participants are risk preferring and the estimated curvature of the utility function is clearly concave.

These type of utility functions for drivers can be employed in lieu of utility over money whether using Expected Utility, Rank Dependent Utility, or Prospect Theory.

### Other tasks and estimation models

Apart from eliciting drivers' attitudes to risk and uncertainty we also elicit their beliefs about travel times on the study routes they participate on, the most important variable in the field experiments that determine their well being from route choices. For this purpose, participants are presented with sets of travel time intervals with various possible payments conditional on what the actual travel time is.. The payment assignments are done in an incentive compatible manner so that the respondents are encouraged to report their beliefs truthfully. Several different scoring rules have been used in the literature that guarantee that an optimum response exists conditional on the underlying subjective belief of the participants.

A simple two option example of the task is when a respondent can choose between several different allocations of money earnings conditional on the actual travel time. The first possible allocation may be getting \$10 if the travel time is less than 20 minutes and getting \$0 if it is 20 minutes or more. This payment allocation would only be selected by participants that are certain, i.e. assign probability 1, to the event that the travel time is less than 20 minutes. The next option may pay \$9 or \$1 respectively if the travel time is less than or greater than 20 minutes. For each choice option the payment associated with the travel time being less than 20 minutes is decreasing and the payment for the travel time being at least 20 minutes is increasing in a nonlinear manner. Various functional forms exist for relating these payments in a way that makes it optimal for the respondent to tell the truth. Respondents will naturally be attracted to choose the option that pays them the most money, conditional on their true beliefs about travel time. From these responses it is possible to infer what the implied probability is that participants hold over the travel time intervals.

Once the participants' responses to these scoring rules have been collected they can be used to estimate the probability of an outcome in an event, such as  $p(k)$  given in equation 2. When combining responses from these belief tasks with those of choices over risky events with known probabilities, we can jointly estimate both the risk attitude and the subjective belief of the participant.

We give participants tasks where they report their subjectively held belief over travel times associated with various departure times

for the routes we are studying. For a given departure time, say 7 am, a participant is asked to report the subjective belief that the travel time falls in each of four specific travel time intervals. We randomly vary these time intervals across participants, and they also vary with the field driving route for natural reasons.

In the single driver driving simulator respondents are first told what fixed payment for a simulator drive that they can receive. From this fixed payment a cost depending on the travel time or other aspects of congestion is subtracted and if they choose to take the tolled road, the toll is also subtracted. The driver is always presented with an option of two routes: one is tolled and never congested, the other is never tolled but is congested with some probability. In some tasks the probability of congestion is known to the participant and in others it is not, although in the latter case they get to view a sample of drives before starting to make their choices. The former tasks are used to estimate risk attitudes in a driving task that is isomorphic to the risky prospect task. In the latter case the participant repeats the task multiple times and learns the probability of congestion through experience. They first select a toll by drawing a card from a deck of cards with the full range of tolls before they proceed to make their route choice. This range is 50 cents to \$5.50 in 10 cent increments so the actual distribution will be fairly continuous. We observe the variation in route choice based on varying tolls, and varying probabilities of congestion as well as varying information about these probabilities.

Using the same maximum likelihood estimation of expected utility models as was described for analysing the data from the risky prospects, with extensions to other decision theories, this route choice data allows estimation of risk attitudes as well as implied beliefs about congestion.

In the multiple driver traffic simulator respondents again receive a fixed payment for each drive, from which tolls and travel time costs are subtracted. A group of 20-40 participants make route choices independently but simultaneously. There are two routes to choose from: one is a tolled express way and the other is a local road with traffic lights and no toll. Based on the route choices made congestion is endogenously determined. Tolls are fixed throughout a series of drives, but congestion can of course vary with the choices. These data will be used to investigate the extent to which the convergence of the traffic system to its equilibrium is a function of the risk attitudes and perceptions of the drivers.

Survey questions on demographic characteristics such as sex, income and education levels, as well as survey questions on travel habits and experiences with congestion, are used as controls in the estimations.

### Hypotheses testing

Hypotheses tests are performed directly on the estimated model parameters. Our main hypothesis is that driver characteristics such as risk attitudes and precision in perceptions, coupled with various demographic characteristics, have explanatory power across the four regions and are able to explain a significant part of the variation in route choices. The hypotheses are related to the following research questions:

QUESTION: Do risk attitudes and perceptions explain a large part of the variation in route choices of our participants during congestion conditions such as morning and afternoon commutes as road pricing varies?

We identify risk attitudes through two tasks: a pairwise prospect choice task and an isomorphic driving simulator route choice task. The former task is the most popular way to identify risk attitudes in a reliable and consistent way. We include the second task in order to test if the framing as a real time driving task affects these attitudes. If they do not then the much simpler and cheaper prospect instruments can be used in policy studies to understand the risk attitudes of the driving population.

We identify perceptions of travel times and congestion in two ways. For the field routes we use a scoring rule to reward respondents for guessing travel times under various conditions in such a way that they have incentives to tell the truth. For the simulator routes we can infer perceptions of the risk of congestion from their choices and test whether biases and variations in travel time perceptions are correlated with observable demographic characteristics.

Based on the observations of route choices we collect using the GPS units in the respondents' cars we can then test whether risk attitudes and perceptions affect these route choices. Apart from the observations on congestion conditions that we get from the GPS units we also collect information on incidents, construction, weather and other major traffic events that are available to the drivers and that may affect their choices.

In summary we collect field route choices using GPS recorders, we identify risk attitudes through risky prospect choices and simulator drives, we observe perceptions of field travel times using paid guessing task, and we identify variations in accuracy of perceptions of congestion using driving simulators. Together these observations will allow us to approach the question stated.

QUESTION: Does behavior observed among drivers in one region transfer to drivers in other regions?

If a large part of the variation in driving choices is explained by risk attitudes and perceptions, and if these are either similar across regions or directly observable, transferability is expected. If this is the case then it may be possible to observe drivers' reactions to congestion pricing in one region and use this information to make informative predictions of drivers' reactions to congestion pricing in another region.

Even if risk attitudes and congestion perceptions do not in themselves explain the major part of route choice variations it may be that driver reactions to congestion and congestion pricing is sufficiently correlated with observable demographics that observing such correlations in one region may allow us to predict responses in another regions conditional on observing the same demographics in the latter. We study drivers in four different regions in order to approach this question and assess the extent to which route choices reflect primarily risk attitudes and perceptions. Two of these regions are from Orlando where we invite respondents who reside both on the east and the west side of downtown Orlando. We expect drivers on the east and west side of Orlando to share many driving habits since they live and drive in the same driving culture and on the same traffic networks. Since they live on different sides of downtown Orlando they do not, however, use the exact same commuter routes and we will observe them on different parts of the Orlando traffic network. Using drivers from regions that are very similar gives the possibility of transferability of findings across regions its best shot. If we do not find transferability in this case we are unlikely to find it when regions are less similar. We also, for the same reasons and also to provide a robustness test on the Orlando findings, include two regions in Atlanta: the northeast and the northwest. Finally, we test transferability across less similar regions by comparing behavior across Orlando and Atlanta. One difference between these two regions that may be important is the prior experience with road tolls and the significantly lower congestion levels that are present on the tolled roads that drivers in Orlando have. Atlanta drivers do not have this experience - all routes in and out of



the downtown area are heavily congested during peak hours. Thus, by including respondents from four regions we can approach the question of transferability. If we find support for this phenomenon across the four regions then extending these tests to other regions will be warranted as a way of finding the boundaries for transferability.

QUESTION: Are results from using driving simulator experiments comparable to on-the-road choices?

Observations on respondents in driving simulators are combined with field driving observations on the same respondents. Transferability here will lend support to a decreased need for relying on large and costly field tests of new ways of solving congestion, since at least some understanding of the driving population can be reached through the much less costly simulator observations. In the simulators we will also be able to observe more detailed behavioral phenomena than can be observed in the field, such as aggressiveness in acceleration and deceleration, and correlate this with both risk attitudes and route choices. The simulator also gives us the possibility of varying not just tolls but also congestion since the researcher has control of all other traffic in the simulator.

QUESTION: Does the distribution of risk attitudes, and perceptions of travel times of drivers explain important endogenous properties of the traffic system they are in?

In order to approach this question we use traffic simulators where many drivers can make independent route choice decisions at the same time. Thus we will move from studying the reactions of the individual driver to studying the reactions of a traffic system with multiple, independent drivers. The present use of traffic simulations to predict how traffic systems respond to various manipulations are all based on assumed behavioral assumptions among the drivers. We will contribute to this important methodology by providing empirical estimates of such behaviors.

QUESTION: Are there significant differences in reactions to congestion pricing across observable demographic segments of the population such as gender, age and household income?

We collect demographic information on all participants as well as data on their travel habits to correlate with their choices in the tasks.

This approach will estimate  $r$ , the curvature of the utility function, as well as  $\alpha$  and  $\beta$  that measure how important the arrival time is, and also  $p$ , the probability function over travel times on the routes. These models will be estimated on each regional subsample, and the estimated models used to predict on the remaining subsets to test how behaviour transfers across regions. Similarly, the models will be estimated on simulation data and used to predict on the field data to test the extent to which simulations can capture field behaviour.

In general our approach is to undertake maximum likelihood estimation of structural models of decision-making under risk, and to conduct hypothesis tests directly on those estimated structural parameters. We address participant heterogeneity by means of controls for observable characteristics that participants will provide, as well as more sophisticated “random coefficients” approaches that allow for unobserved individual heterogeneity.

Here are some further examples of the researchers’ publications that illustrate all of the methods to be used.

**Andersen, Steffen; Glenn W. Harrison; Arne Risa Hole; Morten I. Lau; E. Elisabet Rutstrom**, “Non-Linear Mixed Logit”, Theory and Decision, forthcoming 2011. This paper illustrates the use of Maximum Simulated Likelihood and random coefficient estimates to capture preference heterogeneity in a population for choice tasks that are designed to elicit participant’s beliefs.

**Andersen, Steffen; Glenn W. Harrison; Morten I. Lau; E. Elisabet Rutstrom**, “Behavioral Econometrics for Psychologists,” Journal of Economic Psychology, 31, 2010, 553-756. This paper reviews our basic approach of estimating full structural models using Maximum Likelihood. It is written for a psychologist research audience where these models have not been commonly used, primarily since they are not as known in that field.

**Andersen, Steffen; Glenn W. Harrison; Morten I. Lau; E. Elisabet Rutstrom**, “Eliciting Risk and Time Preferences,” Econometrica, (76) 3, May 2008, 583-618. This paper illustrates how multiple tasks and joint estimation of preference parameters can change the conclusions one draws in significant ways, compared to methods based on single tasks and independent estimation.

**Harrison, Glenn W.; E. Elisabet Rutstrom**, “Expected Utility Theory and Prospect Theory: One Wedding and a Decent Funeral,” Experimental Economics, 12(2), May 2009, 133-158. This paper

illustrates the use of Mixture Models for estimating the explanatory contribution of different theories to a set of data.

### 3. Describe methods to maximize response rate.

Participants are compensated for their participation which will lower the cost to them of participating. We have doubled the number of invitation letters sent out compared to our past research where we had a 5% response rate, expecting that the longer study period will result in lower response rates. Because our respondents are participating in a panel, we have taken great care to train research assistants and supervisors to be courteous and professional when respondents provide data. In our experience in similar panel surveys in Denmark, apart from being desirable in general, such practices pay off in terms of willingness of respondents to participate in several sessions. Invitations are sent to mail carrier routes that are adjacent and immediately outside the routes we are studying, making participation relatively convenient.

### 4. Describe tests of procedures or methods.

The discussion above shows how we have used the decision models and estimation strategies in previous studies. The validity of this general approach is therefore well documented in the publications cited. The novel aspect in this study is the combined use of driving simulators and GPS recording of field data, so no previous publication can point to the ease or difficulty of implementing these in behavioral experiments. However, the study team has employed visual, interactive computer simulations in previous projects. These are in many respects similar to the driving simulators. In a large project involving simulated forest fires the researchers found that respondents can handle the interface and give accurate responses if assisted.<sup>4</sup> The researchers have employed driving simulators in one previous (unpublished) study and found that nausea can be a serious problem, which is why the instructions and scripts designed for this study repeatedly warn invitees about this possibility and why the proposed procedures include ways of dealing with such nausea, either through a sequencing of tasks with many breaks from the simulator or by dismissing respondents from simulator tasks.

---

<sup>4</sup> Fiore, Stephen M., Glenn W. Harrison, Charles E. Hughes and E. Elisabet Rutstrom, "Virtual Experiments and Environmental Policy," Journal of Environmental and Economic Management, 57(1), January 2009, 65-86.

The researchers have performed informal tests of our single driver and multi driver simulator tasks both on a few volunteers and as part of the process of training our 40 research assistants. This has guided the researchers both in terms of the phrasing of the scripts and instructions, and in how we space out tasks to minimize nausea. The researchers have also tested the time requirements throughout these tests and consequently rearranged or eliminated tasks as needed. The researchers have also tested the GPS units using a few volunteers and their reliability in picking up signals, which has guided us in how we frame the field driving task. Some understanding of the issues that arise as research participants are observed while driving in the field was generated through a master's thesis study on toll subsidies to student drivers that was supervised by one of the researchers on this project, Dr. Glenn Harrison.<sup>5</sup>

5. Provide name and telephone number of individuals who were consulted on statistical aspects of the IC and who will actually collect and/or analyze the information.

Dr. Glenn W Harrison, Center for the Economic Analysis of Risk, Georgia State University, phone (404) 413-7456

Dr. Morten Lau, Department of Finance, Durham University, UK, phone +44 (0) 191 33 45044

Dr. Steffen Andersen, Department of Economics, Copenhagen Business School, Denmark, phone +45 3815 2591

Dr. Elisabet Rutstrom, Robinson College of Business, Georgia State University, phone (404) 413-7111

6. Caveat to accompany use of study results.

In all publications containing a discussion of the results of this study the following caveat shall appear: *This study is exploratory and experimental in nature. Results from this study will not be used as evidence in reports to Congress or in responses to Congressional testimony. Results from this study will not directly be used to alter programs or policies until followed up with further studies.*

**Supplementary section on matching research questions to survey instruments and choice tasks in the experiments.**

---

<sup>5</sup> Lascelles, A.E. (2008) "Alternative methods of eliciting individual willingness to pay for travel time savings a pilot study," *Master of Economics Thesis*, Department of Economics, College of Business Administration, University of Central Florida

QUESTION: Do risk attitudes and perceptions explain a large part of the variation in route choices of our participants during congestion conditions such as morning and afternoon commutes as road pricing varies?

We identify risk attitudes through two tasks: a pairwise prospect choice task and an isomorphic driving simulator route choice task. The former task is the most popular way to identify risk attitudes in a reliable and consistent way. We include the second task in order to test if the framing as a real time driving task affects these attitudes. If they do not then the much simpler and cheaper prospect instruments can be used in policy studies to understand the risk attitudes of the driving population.

We identify perceptions of travel times and congestion in two ways. For the field routes we use a scoring rule to reward respondents for guessing travel times under various conditions in such a way that they have incentives to tell the truth. For the simulator routes we can infer perceptions of the risk of congestion from their choices and test whether biases and variations in travel time perceptions are correlated with observable demographic characteristics.

Based on the observations of route choices we collect using the GPS units in the respondents' cars we can then test whether risk attitudes and perceptions affect these route choices. Apart from the observations on congestion conditions that we get from the GPS units we also collect information on incidents, construction, weather and other major traffic events that are available to the drivers and that may affect their choices.

In summary we collect field route choices using GPS recorders, we identify risk attitudes through risky prospect choices and simulator drives, we observe perceptions of field travel times using paid guessing task, and we identify variations in accuracy of perceptions of congestion using driving simulators. Together these observations will allow us to approach the question stated.

QUESTION: Does behavior observed among drivers in one region transfer to drivers in other regions?

If a large part of the variation in driving choices is explained by risk attitudes and perceptions, and if these are either similar across regions or directly observable, transferability is expected. If this is the case then it may be possible to observe drivers' reactions to congestion pricing in one region and use this information to make

informative predictions of drivers' reactions to congestion pricing in another region.

Even if risk attitudes and congestion perceptions do not in themselves explain the major part of route choice variations it may be that driver reactions to congestion and congestion pricing is sufficiently correlated with observable demographics that observing such correlations in one region may allow us to predict responses in another regions conditional on observing the same demographics in the latter. We study drivers in four different regions in order to approach this question and assess the extent to which route choices reflect primarily risk attitudes and perceptions. Two of these regions are from Orlando where we invite respondents who reside both on the east and the west side of downtown Orlando. We expect drivers on the east and west side of Orlando to share many driving habits since they live and drive in the same driving culture and on the same traffic networks. Since they live on different sides of downtown Orlando they do not, however, use the exact same commuter routes and we will observe them on different parts of the Orlando traffic network. Using drivers from regions that are very similar gives the possibility of transferability of findings across regions its best shot. If we do not find transferability in this case we are unlikely to find it when regions are less similar. We also, for the same reasons and also to provide a robustness test on the Orlando findings, include two regions in Atlanta: the northeast and the northwest. Finally, we test transferability across less similar regions by comparing behavior across Orlando and Atlanta. One difference between these two regions that may be important is the prior experience with road tolls and the significantly lower congestion levels that are present on the tolled roads that drivers in Orlando have. Atlanta drivers do not have this experience - all routes in and out of the downtown area are heavily congested during peak hours. Thus, by including respondents from four regions we can approach the question of transferability. If we find support for this phenomenon across the four regions then extending these tests to other regions will be warranted as a way of finding the boundaries for transferability.

QUESTION: Are results from using driving simulator experiments comparable to on-the-road choices?

Observations on respondents in driving simulators are combined with field driving observations on the same respondents. Transferability here will lend support to a decreased need for relying on large and costly field tests of new ways of solving congestion, since at least some understanding of the driving population can be reached through the much less costly simulator observations. In the simulators we will also be able to observe more detailed behavioral phenomena

than can be observed in the field, such as aggressiveness in acceleration and deceleration, and correlate this with both risk attitudes and route choices. The simulator also gives us the possibility of varying not just tolls but also congestion since the researcher has control of all other traffic in the simulator.

QUESTION: Does the distribution of risk attitudes, and perceptions of travel times of drivers explain important endogenous properties of the traffic system they are in?

In order to approach this question we use traffic simulators where many drivers can make independent route choice decisions at the same time. Thus we will move from studying the reactions of the individual driver to studying the reactions of a traffic system with multiple, independent drivers. The present use of traffic simulations to predict how traffic systems respond to various manipulations are all based on assumed behavioral assumptions among the drivers. We will contribute to this important methodology by providing empirical estimates of such behaviors.

QUESTION: Are there significant differences in reactions to congestion pricing across observable demographic segments of the population such as gender, age and household income?

We collect demographic information on all participants as well as data on their travel habits to correlate with their choices in the tasks.

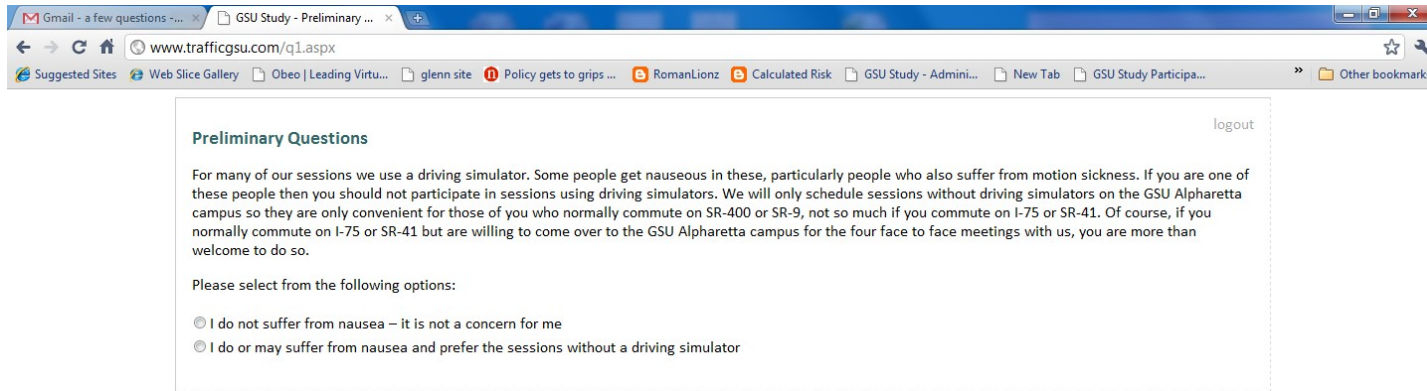
### Summary Table

Research Question	Observations collected to address question
Do risk attitudes and perceptions explain a large part of the variation in route choices of our participants during congestion conditions such as morning and afternoon commutes as road pricing varies?	Risky prospect choices, beliefs about travel times, perception biases from simulator driving, and field driving responses to varying road pricing
Does behavior observed among drivers in one region transfer to drivers in other regions?	Same tasks given to participants across four regions that differ by varying degrees.
Are results from using driving simulator experiments comparable to on-the-road choices?	Simulator driving choices and field driving choices with similar toll manipulations
Does the distribution of risk attitudes and perceptions of travel time among drivers explain important endogenous properties of the traffic system they are in?	Multi-driver traffic simulator experiments, risky prospect choices, and belief tasks
Are there significant differences in reactions to congestion pricing across observable demographic segments of the population such as gender, age and household income?	Demographic questionnaire





Supplement: Screen shot of online scheduling page screening for participants who may suffer from nausea.



Sample from log

where participants circle when they are not pressed for time:

In order for us to know if there were any days when you were particularly under time pressure we ask that you note down here a record of every drive when you WERE NOT under pressure:

First week: \_\_DATE FILLED IN HERE BY RESARCH ASSISTANTS

Drive 1 Please circle:        MON TUE WED THU FRI SAT  
   AM PM

Drive 2 Please circle:        MON TUE WED THU FRI SAT  
   AM PM

Drive 3 Please circle:        MON TUE WED THU FRI SAT  
   AM PM

ET

## **Supplementary Material to the Supporting Statement.**

The title of this information collection is “Experiments on Driving under Uncertain Congestion Conditions and the Effects on Traffic Networks from Congestion Pricing Initiatives”.

### Justification of time requirement per participant

Both the field driving and the simulator tasks need to be offered to the respondents. The fact that they are both included generates the data to support the main hypotheses. The core of the exploratory research project is to investigate the possibility of developing models of driver decision making under congestion conditions that can be used to predict driver responses to changes in congestion and congestion pricing across various populations and conditions. A big part of this is the comparison of the field driving and simulator driving. About half of the participant time is used directly in these two tasks. We show here and in the Supporting Document Parts A and B how the other tasks are necessary to construct and estimate the decision models for these driving data. Not controlling for these influences would severely confound the inferences drawn, greatly reducing the understanding of the determinants of driver decisions.

There are three major elements in such decision models: risk attitudes, uncertainty attitudes, and the properties of the perception of travel times and therefore the risk of delays. This is explained in Part B of the Supporting Statement. In addition, since the driving skills in the simulator will vary across participants and will affect the inferences drawn based on their choices, this needs to be controlled for, as will responses to the travel habit questionnaire. The need for a demographic questionnaire should be obvious. The study includes a standard demographic questionnaire to enhance FHWA’s understanding of who takes toll roads. One example is questions about household incomes intended to shed light on which income groups would be the heaviest users of tolled facilities.

None of the three major decision elements are directly observable, nor can they be ascertained in reliable ways from self reported assessments such as stated preference surveys. This is generally accepted in the literature, and has been known for some time. For example, in “Assessing the Construct Validity of Risk Attitude” (Joost M.E. Pennings and Ale Smidts, *Management Science*, Vol 46. No 10, Oct 2000) show that responses to tasks based on risky prospects that have actual monetary consequences are better predictors of actual market behavior among Dutch owner-managers of hog farms than are Likert scale responses with no actual consequences, monetary or otherwise. The bias that occurs in responses without actual consequences is also documented in “Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?” (Ronald G. Cummings, Glenn W. Harrison and E. Elisabet Rutstrom, *American Economic Review*, 1995).

The difference between risk and uncertainty is that in the former drivers would know the probabilities for various travel times but in the latter they would not. Drivers will generally have imprecise knowledge of travel times: they will not know exactly what the likelihood is of any one drive resulting in delays associated with various costly consequences. Therefore attitudes to uncertainty will be important and they need to be measured. In “The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation”, (Mohammed Abdellaoui, Aurelian Baillin, Laetitia

Placido, and Peter P. Wakker, American Economic Review, forthcoming) show that uncertainty attitudes do affect choices and they are different from risk attitudes. When observing research participants making driving choices, whether in a simulator environment or in a field environment, these elements have to be known and controlled for. The study thus has to measure each of these elements in addition to making observations on the driving for every participant in order to construct and estimate these models.

The alternative to building and estimating these structural decision models with their potential for enabling a new, more precise and less expensive methodology for predicting responses to future transportation policies of many kinds, is to either perform very large randomized trials or stated preference surveys. The former are very expensive with limited ability to generalize the findings, and the latter are very imprecise and biased. Neither approach improves the understanding of the motivations and concerns of the individual drivers. The purpose of this advanced exploratory research project is to advance the methodology available to FHWA to generate a better understanding of traveler choices and how road pricing can influence congestion.

The next sections contain an overview of how the time is allocated across various tasks: Core tasks, Control tasks, and non-task time.

Total estimated time per participant: 5 hours 25 minutes, see Table 12 in Supporting Statement. This total breaks down into three parts:

1. Measurements in Core tasks: 2 hours and 25 minutes

1.1 Measuring responses to variations in congestion and road pricing in the field: 1 hour

1.2 Measuring responses to variations in congestion, road pricing and value of time in the simulator including training: 1 hour 25 minutes

2. Measurements in Control tasks necessary to explain responses in core tasks: 1 hour 20 minutes

2.1 and 2.2 Measuring risk and uncertainty attitudes: 40 minutes

2.3 Measuring variations in simulator driving skills: 20 minutes

2.4 Measuring prior and posterior beliefs about field travel times: 25 minutes

2.5 Measuring demographics, travel habits and opinions: 35 minutes

3. Time for greetings, payments and breaks: 60 minutes

Below we detail a justification for each of these task groups.

1.1 Justifying the time requirement of the field driving study:

We collect data from three driving periods which is why four meetings are necessary, apart from the need to perform the complementing tasks. In meetings 1-3 instructions on the driving tasks is given. In meetings 2-4 the GPS data is downloaded. These three driving periods include a base line where no road pricing is manipulated against which the other observations are compared. Thus, we observe only two different tolls for each participant, necessitating a pooling of responses for estimation purposes. Reducing these observations even further would make it very difficult, if not impossible, to identify the confounding influence of unobservable variations in the driving circumstances of the individuals from those variations that depend on responses to pricing. In the field, as opposed to in the simulator lab, it is impossible to make

observations on all such influences, although we observe many of them by collecting traffic data as well as travel habit and travel experience data. Since risk attitudes have been shown to vary with income and stakes it is important to include more than one stake condition. (See for example, “Risk Aversion and Incentive Effects”, Charles A. Holt and Susan K. Laury, *American Economic Review*, 2002).

Reducing the number of stakes would also make it impossible to assess whether responses depend on whether the prices are increasing or decreasing. The possibility that responses may differ depending on whether prices are increasing or decreasing is related to the possibility that decision makers may be motivated by loss aversion. Evidence of loss aversion has been reported in a large number of studies, such as in “The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test” (Richard H. Thaler, Amos Tversky, Daniel Kahneman, and Alan Schwartz, *The Quarterly Journal of Economics*, 1997). Since much road pricing that is designed to combat congestion has variable pricing, drivers will be facing contexts where prices may sometimes be increasing and at other times decreasing. This context needs to be captured in the experiments.

### 1.1 Justifying the time requirement of the simulator driving study

It is important to understand how drivers learn about congestion and time delays and how their beliefs over travel time are updated as they gain experience through driving. The properties of such learning and updating can be measured simultaneously with observing their reactions to various congestion and pricing options in the simulator since the task allows the researcher to know the underlying likelihoods precisely as well as the value participants place on being on time, which is not possible in the field. We design a sequence of ten drives in the simulator in which we make these observations.

The most common number of task repetitions in experiments is ten. During the first 3-5 periods decisions are usually relatively noisy and not much convergence is observed. In order to have a reasonably high likelihood of convergence in driver learning 10 periods is considered a minimum in the experimental economics literature. There are many examples of experiments that run for at least 10 periods of repetitions. “Income Distributional Preferences: The Role of History,” (Laurie T. Johnson, E. Elisabet Rutström and J. Gregory George, *Journal of Economic Behavior and Organization*, 2006), “Testing Static Game Theory with Dynamic Experiments: A Case Study of Public Goods,” (Anabela Botelho, Glenn W. Harrison, Ligia Costa Pinto and E. Elisabet Rutstrom, *Games and Economic Behavior*, 2009), “Stated Beliefs Versus Inferred Beliefs: A Methodological Inquiry and Experimental Test,” (E. Elisabet Rutstrom and Nathaniel T. Wilcox, *Games and Economic Behavior*, 2009).

Evidence of belief-biases is reported in a large psychology literature, for example “The Domain Specificity and Generality of Belief Bias: Searching for a Generalizable Critical Thinking Skill”, (Walter C. Sa, Richard F. West, and Keith E Stanovich, *Journal of Educational Psychology*, 1999). Investigations into learning and Bayesian updating in experimental studies also indicate that there is a great degree of heterogeneity and that context matters importantly. “When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect”, (Gary Charness and Dan Levin, *American Economic Review*, 2005), “Experience-Weighted Attraction Learning in Normal Form Games”, (Colin Camerer and Teck Hua Ho, *Econometrica* 1999) are examples from this

literature. It is therefore important to measure how the beliefs about travel time change with time and experience.

Before this core data can be collected participants have to practice driving in the simulator. The fact that we give them control tasks in the simulator, in addition to the core ten driving tasks, will help here. It is crucial that participants have adequate time to get familiar with the simulator, how the accelerator and brake works, the feel of the wheel and interplay with the 'car' on the road. Therefore the task design includes 6 practice drives plus a video as baseline simulator training. The researchers are very respectful of peoples' time not only to minimize the burden but also as a key parameter in producing robust and accurate results – a bored participant is a poor participant.

### 2.1 Justifying the time needed for measuring risk attitudes:

The PIs have been involved in studies measuring risk attitudes on several field populations: Two panels in Denmark, one in 2003-2004 and one in 2009-2010. One panel in Florida, and several others in Ethiopia, India, Uganda and East Timor. "Eliciting Risk and Time Preferences," (Steffen Andersen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutstrom, *Econometrica*, 2008). "Virtual Experiments and Environmental Policy," (Stephen M. Fiore, Glenn W. Harrison, Charles E. Hughes and E. Elisabet Rutstrom, *Journal of Environmental and Economic Management*, 2009). "Choice Under Uncertainty: Evidence from Ethiopia, India and Uganda" (Glenn W. Harrison, Steven J. Humphrey and Arjan Verschoor, *The Economic Journal*, 2009). These demonstrate the need for multiple stakes and probabilities to identify risk attitudes, particularly the fact that risk attitudes vary with income and stakes. They also illustrate how important it is to control for risk attitudes when identifying other valuations and preferences, and how heterogeneous these attitudes and valuations are. It is therefore important not just to measure a risk attitude factor, but complete risk attitude functions, necessitating additional variations in stakes and probabilities.

To characterize an individual completely a series of 30 – 100 choice tasks would be needed, that vary in both likelihoods and stakes. ("Investigating Generalization of Expected Utility Theory using Experimental Data", John D. Hey and Chris Orme, *Econometrica*, 1994). By pooling observations across individuals it is possible to reduce this to 10-20 choice tasks per person and characterize not each individual but instead groupings of individuals identified by observable demographic characteristics such as gender, age and ethnicity. This has been the approach taken by most studies during the last 20 years. Further reductions in the number of observations per participant would limit the number of groupings that can be separately identified by risk attitudes. If only one choice task is presented to each participant it is impossible to understand the extent to which the participant has understood the task since no choice variation can be observed at all. In this study we present 4 choice tasks allowing us to vary both stakes and probabilities in the task itself but also to see how the estimated risk attitude varies as the total earnings the participant makes across all stakes vary. This allows us to capture at least a part of the heterogeneity of risk attitudes. The major part of the time that a participant spends on these tasks is during instructions on what the options are and the consequences that follow from various choices.



The literature on these risk attitude measurements uses very stylized and general instruments and it is unclear to what extent these provide measurements that explain choices in contextual tasks such as driving. No testing of this has been done in driving contexts. We therefore complement the four risk prospect tasks with contextual tasks relevant to the policy area of the study, driving under congestion conditions, also designed to measure risk attitudes. These contextual tasks are undertaken in the driving simulator. Evidence exists that the expressions of risk attitudes may vary with context. For example “A Domain-Specific Risk-Attitude Scale: Measuring Risk Perceptions and Risk Behaviors”, (Elke U. Weber, Ann-Renee Blais, and Nancy E. Betz, Journal of Behavioral Decision Making, 2009).

### 2.2. Measuring uncertainty aversion:

When probabilities over outcomes are not known, as is the case in most traffic circumstances, choices will depend on what attitudes drivers have over choices that vary in uncertainty, not just vary in risk. Uncertainty is the characterization of contexts where the agents do not know the probabilities over various outcomes, as opposed to risk where these probabilities are known. This is an additional element in the choice models, compounding the attitude to risk discussed above. Choice tasks to measure this have been designed to be as similar to those for measuring risk attitudes as possible so that there will be minimal instruction time. Incorporating any measure of uncertainty aversion is a major improvement in the data compared to current practices. Evidence that uncertainty aversion affects behavior has been cited above.

### 2.3 Measuring individual variations in simulator driving skills:

Due to the unfamiliarity with being in driving simulators we also include 4 driving tasks that measure variations across participants in how fast and reliably they perceive themselves as being able to drive under various congestion conditions in the simulator. These 4 drives vary in the degree of congestion and in the stakes involved, necessary variations to match these measures to the rest of the data. These 4 simulator drives are expected to take 20 minutes.

### 2.4 Measuring beliefs about the likelihood of various travel times

Participant's perception of the distribution of travel times is a major explanatory variable to driving choices. If they believe that local roads, with their traffic lights and speed limits, are always slower than expressways, no matter what the congestion conditions are, then this will obviously influence their choice of routes whether or not these beliefs are correct. Even more importantly, they may believe that local roads are less reliable than expressways and lack of reliability can lead to infrequent but extreme delays with unacceptable consequences. It is also important to understand how drivers learn about congestion and time delays and how their beliefs over travel time are updated as they gain experience through driving. We measure the respondents' beliefs about various travel times at different times of the days and on different routes both before they start the field driving in the study and at the very end. We include only routes and times of day during which respondents travel as part of the study. Without this information it would not be possible to explain route choices. We measure their learning in two ways: first by eliciting their beliefs about travel times in the field at the end as well as at the beginning of the study. Second, by observing how their route choices change over time in the simulator experiment as described under 1.1 above. The measures obtained as described in 1.1 can be used to understand the extent to which biases in the direct

measures of travel time, as described here, will or will not disappear over time for various individuals.

### 2.5 Measuring demographic characteristics, travel habits and experiences

In order to characterize the pattern of heterogeneity in responses to congestion and road pricing it is important to have demographic information on the respondents. We collect these through a demographic questionnaire which matches those we have used in other field projects reported above. In addition, much of the variation in responses regarding driving choices can be attributed to personal habits and experiences and these need to be controlled for. Documentation of the role of habits in driving choices can be found for example in “Choice of Travel Mode in the Theory of Planned Behavior: The Roles of Past Behavior, Habit, and Reasoned Action”, (Sebastian Bamberg, Icek Ajzen, and Peter Schmidt, Basic and Applied Social Psychology, 2003).

### Evidence that participants are not over-burdened

It is important for the purposes of this study to use field respondents and not just more convenient participant groups such as students. Student responses are useful they cannot capture the range of heterogeneity in the general population. While the study is not intended to provide observations that generalize to the full population, it is important in exploratory research of this kind to generate some understanding of the heterogeneity in responses that field populations exhibit.

*The use of two hour sessions.* In the field study reported in “Virtual Experiments and Environmental Policy,” (Stephen M. Fiore, Glenn W. Harrison, Charles E. Hughes and E. Elisabet Rutstrom, Journal of Environmental and Economic Management, 2009) participants from the general population in Florida participated in sessions involving simulators that lasted two hours. In addition to the simulator choice tasks, designed to assess how respondents value risk reducing activities, the respondents were also given risky prospect tasks in order to identify and control for their risk attitudes.

In the two Danish panel studies referred to above respondents participated in two sessions of 1 – 2 hours each. These participants were given a series of up to 80 tasks during a session, including risky prospect tasks to identify their risk attitudes.

In addition, the University of Florida conducted an experiment involving participants from the general community as well as students in the UCF driving simulator housed in the Center for Advanced Transportation Systems Simulation (CATSS). It is an I-Sim Mark-II system with a high driving fidelity and immersive virtual environments. 42 respondents in two age groups participated, 18 were younger than 26 and 24 were between 26 and 55 years of age. The experiment required participants to first attend an orientation session that lasted up to two hours. Upon arrival to the orientation, the subjects were given an informational briefing about the driving simulator and their driving task. Then, a practice course was programmed on the driving simulator. Participants then returned for 3 additional visits to the lab, totaling on average two additional hours per subject. The experiment included 8 experimental conditions. After completing all the driving tasks, participants also responded to a survey about their opinions of the proposed pavement marking and red-light running. “Impact of “Signal Ahead” Pavement Marking on Driver Behavior at Signalized Intersections”, (Xuedong Yan, Essam Radwan, Dahai Guo, and Stephen Richards, , Journal of the Transportation Research Part F: Traffic Psychology and Behavior, 2009).

*The use of multiple sessions.* In the two Danish panel studies, referred to above, respondents participated in two sessions of 1 – 2 hours each with very reasonable attrition rates.

As an additional example, the research group at Georgia Institute of Technology, Commute Atlanta, conducted a two-year panel observing drivers using GPS recording devices, requiring multiple sessions to retrieve data from GPS devices (typically 1 – 2 hours in length). “Variability in Traffic Flow Quality Experienced by Drivers: Evidence from Instrumented Vehicles”. (J. Ko, R. Guensler and M. Hunter, Transportation Research Record, forthcoming).

The UCF study, mentioned above, required participants to attend an orientation session for two hours and then return 3 more times to complete driving tasks in a high fidelity simulator.

The University of Iowa conducted a study “National Evaluation of a Mileage-Based Road User Charge” where respondents participated in recording their drives using GPS for a one-year period. This study required and received approval from OMB under the Paperwork Reduction Act. Their participants also participated in six surveys during this time. The Iowa study used a GPS technology that was linked to the respondents’ onboard vehicle computers and uploading travel data to a server. The study proposed here uses a less intrusive GPS technology to enhance the privacy comfort of the respondents but this requires the respondents meet with us and download the data on location. For participant compensation the current study uses the same basis for calculating the hourly value of time as the Iowa study, inflated to 2010 dollars. According to the supporting statement of the Iowa study’s OMB approval each respondent would spend 6 hours in the study. This is slightly more than in our proposed study.

The Supporting Statement of the Iowa study is appended to this burden justification.

#### Evidence of Payments to Participants

There is precedence on paying respondents more than token amounts of money as compensation for participating.

The proposed study estimates that average payments to participants will be \$350, consisting of \$100 compensation for attending the four sessions (\$25.00 per session), \$100 compensation for returning the GPS unit, and on average \$150 in earnings from the consequences of their choices.

In the field study reported in “Virtual Experiments and Environmental Policy,” (Stephen M. Fiore, Glenn W. Harrison, Charles E. Hughes and E. Elisabet Rutstrom, Journal of Environmental and Economic Management, 2009) participants from the general population in Florida were paid a fixed participation fee of \$50 for the session, plus earnings from the simulation and other tasks totaling a maximum of \$220 (including the fixed fee).

In the Danish panel studies participants were paid a fixed participation fee that varied from \$50 to \$100, plus additional earnings up to \$500 in the tasks performed.

In the University of Iowa study participants were paid a total of \$1,165. This was divided into a \$200 up front fee, \$65 per month for 11 months, and a \$250 completion fee. This is significantly more than our average payment of \$350, even more than the expected maximum of \$500.

### Due Diligence By Federal Highway Administration

In evaluating and approving the research plan of the University of Central Florida and Georgia State University FHWA utilized leading experts to ensure that public resources were properly employed. Dr. Karen White who is overseeing this research has a Ph.D. in economics and studied experimental economics as part of her graduate work at the University of Houston. Dr. Christopher Monk, FHWA lead research psychologist, is reviewing and contributing to the guidance of this research. Mr. Patrick DeCorla-Souza, leader of FHWA's Highway Pricing and System Analysis is also an active participant in ensuring the research contributes to, and extends FHWA's understanding of road pricing impacts.

The FHWA study team believes that the current scope is necessary to maximize the return on the government's research investment. Scaling back the research would ultimately cost the government additional resources since this research would not provide the full range of outcomes available under its present design. The current research as proposed represents the best value of the government's research resources.

