

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 be 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.