Complementery Online Lab Experiment for ride sharing preferences

March 19, 2025

1 Introduction

This document outlines the design of a complementary online lab experiment to our baseline field RCT (AEARCTR-0007019). The field experiment shows that fictitious women drivers and passengers are preferred to male ones on the long-distance carpooling platform Blablacar. Reduced form evidence suggests that the positive discrimination towards female users is driven by gender differentials in perceived risk. We will obtain a gender balanced sample from Qualtrics survey pool to select respondents who speak French and have used carpooling services before. First, participants respond to a choice experiment in which they choose between two types of fictitious passenger profiles in multiple configurations. The profiles combine randomized potential revenue for each passenger and their attributes: gender, age, rating, and comments from previous transactions. The goal is to measure gender differentials in willingness to pay to reduce the risk of harassment. Second, participants then respond to a series of questions related to their previous carpooling experiences and the determinants of their screening behavior. Below, we outline the survey design, key hypotheses, specifications and sampling restrictions for data quality.

2 Choice Experiment Mechanics

2.1 Attributes and Levels

The profiles in the choice experiment will have the following attributes and levels:

- **Name:** Chosen from the most popular names in our driver database. All names are unambiguously identified as typical male or female.

- **Age:** We choose 4 ages ranging from 20 to 40: 20, 26, 33,40

- **Rating:** From 1.8 to 5 stars in .4 increments

- **Comments:** "Ride went smoothly," "Made me uncomfortable" etc. We provide three different categories of comments: Comments saying the trip was fine, comments about lateness, and comments suggesting the passenger is risky or may tend to harass passengers.

- **Price:** Varying from 20 to 70 euros in 5 euro increments. This price range is wider than the actual range for similar types of trips, but we need this to recover willingnesses to pay for characteristics that are not regularly present on carpooling platform. For instance, willingnesses to pay for late passengers may be very high for some respondents, and so increasing the price range allows us to recover those.

2.2 Profile Design

Profiles are created by combining different levels of the attributes. For example:

- **Profile 1:** Laurence, 25, 4.2 stars, "Ride went smoothly.", 50 euros

- **Profile 2:** Alexandre, 30, 3 stars, "The trip was uncomfortable.", 70 euros

2.3 Choice Sets

Respondents will be presented with pairs of profiles and asked to choose one. Each choice set will contain two profiles.

3 Optimal Number of Attributes and Respondents

3.1 Number of Attributes and Levels

- **Attributes:** The literature shows that 4-6 attributes are recommended to balance detail and respondent cognitive load. We will have 5 attributes that vary across vignettes (see Section 2.1).

- **Levels per Attribute:** We plan to have some variation to maintain engagement, but to then collapse the attributes to a manageable level in the regressions. Some variables will be continuous and so easy to handle e.g. price or reputation. Others such as comments (risky, unreliable or positive) or names (based on gender) will be classified into clear categories.

3.2 Number of Choice Sets

We use the cjpowR calculator (available https://m-freitag.github.io/cjpowR_shiny/) to estimate ex ante what sample size is necessary to have 80 percent power when estimating an AIMCE between men and women. Gender differences in the RCT are of the order of 7 percentage points. We have a sample size of 1500 respondents, paid through qualtrics. Taking a conservative value of 0.05

for the interaction effect, and taking the maximum value for attributes (5), we find that we achieve 80% power when each respondent answers at least 10 tasks. Accordingly, each respondent will answer at least 10 tasks in the vignette.

We do not plan to explore heterogeneity in the attributes, except for the gender of the passenger. Individuals may interpret risky comments differently when they come from women relative to men.

Our measures of risk will be validated post-vignette by asking respondents to directly rate how risky they consider 8 profile on a scale of 1-5. We will also ask respondents directly whether they would accept the profile in their car for a sum of 50€. This amount is arbitrary but is on the high side relative to the vignette, which should help us find meaningful gender differences. To measure willingness to pay for risk, we need compensation amounts that are reasonably high since WTPS may be extremely high for some characteristics, e.g. very low reputation. This will help us identify gender differences in baseline rates of acceptance or refusal. For instance women may have higher rejection rates at all levels of risk relative to men, but react similarly to men when faced with risky passengers. In this case the conjoint study may not accurately capture gender differences, but this supplementary test may do so.

4   Regression Model - conjoint

We will rely on the 3 main methods people have used in similar conjoint studies.

4.1   Method 1 - Binary Logit Model

We will use a logistic regression model to estimate our outcome variable Y, whether the passenger A was chosen in the pair :

$$\text{logit}(P(Y=1)) = \beta_0 + \beta_1 \text{Diff Gender} + \beta_2 \text{Diff Age} + \beta_3 \text{Diff Rating} + \beta_4 \text{Diff Comment} + \beta_5 \text{Diff Price}$$

where the diff variables are the differences between passenger A and passenger B.

We will then follow Maestas et al (AER, 2024) and recover a willingness to pay based on the estimated coefficient for the price. We then plan to compare differences in willingness to pay for attributes by testing differences in WTP across gender of respondents. We will have limited info about the characteristics of respondents, but may reweigh answers if we see that these characteristics (rough measures of income and age) are important in explaining WTPs. Income might be of particular importance-men earn than women do on average, and individuals with higher income may be less sensitive to prices than those with less.

4.2   Method 2- Mixed Logit Model

Mixed logit models to account for individual-specific preference variation:

$$\text{logit}(P(Y=1)) = \beta_0 + \beta_1 \text{Gender} + \beta_2 \text{Age} + \beta_3 \text{Rating} + \beta_4 \text{Comment} + \nu_i$$

where (ν_i) captures random individual effects. We will use these estimates to compare our results to the regular logit.

4.3 Method 3- Marginal means

The dominant method in political science for similar experiments is the AMCE. This is essentially a regression estimating dummy variables for each level of the variable, (Hainmueller et al. (2015)). The method suffers from problems when comparing across subgroups since the probability of picking an attribute depends on the reference category(Leeper(2020)). For this reason, although we will show results using the AMCE, we plan to rely mainly on Method 4.

Marginal means just shows the mean probability of picking a passenger with a particular attribute. Leeper (2020) shows that this method helps in identifying the relevant baseline category across groups. We plan to follow Leeper(2020) in estimating differences across subgroups. He recommends estimating fully interacted regressions, where each attribute value is interacted with a subgroup dummy, in our case gender.

5 Regression Model - validation

We will also use the answers to the validation check for several complementary analyses.

1). Use the safety ratings to validate our experimental manipulation. We simply regress the rating on the characteristics of the profile. We expect profiles with worse reputation and risky comments to have worse safety ratings.

2). Measure gender differences in perceptions of safety by interacting the attributes with the gender of the respondent in the regression framework. We also expect women to have lower ratings on average (they feel more unsafe no matter the profile characteristics).

3). Measure how the attributes affect willingness to accept, and how this is affected by the safety ratings. We interact the attributes with the gender of the respondent to measure gender differences in the effect of safety.

4). We measure how the safety ratings given in the validation check affect the likelihood of the passenger being chosen in the conjoint study. We interact the safety rating with gender to measure gender differences in the effect of safety on likelihood of preferring a certain passenger over another.

6 Survey

6.1 Questions

After the conjoint study, each respondent will answer a brief questionnaire asking about the person's experience with ride-sharing apps. There are different ways to answer each question, we give an exhaustive list below along with each statement the respondent is tasked with answering.

The very first question after giving demographic information i.e. before the vignette, is a free-text question where we ask respondents directly how they would screen potential passengers on carpooling apps.

The respondents then have to answer the following questions after the vignettes.

Likert scale (1 - 5)

- On a 1-5 scale, what risks do you have in mind when carpooling as a driver: Organizational risk, risk to your car, risk of travelling with a dangerous passenger, risk of luggage theft.

- On a 1-5 scale, what risks do you have in mind when carpooling as a passenger: Organizational risk, risk of travelling in an unsafe car, risk of travelling with a driver with distressing behaviour, risk of travelling with a driver who is dangerous on the road. .

- On a 1-5 scale, what type of unsafe or threatening behaviour do you fear while carpooling, whether as a passenger or driver: a carpooler makes off-colour remarks, risk of assault, a carpooler has opinions too different from yours, discrepancy between carpooler appearance and profile picture, harassment risk

Likert scale (1 - 4: Never, rarely, sometimes, often), for carpoolers who report having been drivers or passengers,

- In your experience as a driver, how often have you been confronted to the following risks: unreliable passenger, damage to your car, travelling with a dangerous passenger, luggage theft.

- In your experience as a passenger, how often have you been confronted to the following risks: unreliable driver, travelling in an unsafe car, travelling with a driver with a distressing behaviour, dangerous driving

Multiple choice with multiple answers allowed: As a driver, if you had to accept a passenger for a carpool ride to which criteria would you give weight?

Choice of gender: Same gender as myself, of opposite gender, no preference

- As a driver, choose the option which corresponds the most: I prefer travelling with passengers who are..., I feel safer when travelling with passengers who are..., the most reliable passengers are, the trip will be more pleasant with passengers who are...

- As a passenger, choose the option which corresponds the most: I prefer travelling with drivers who are..., the safest drivers are ..., the most reliable drivers are, the fastest drivers are..., the trip will be more pleasant with a driver who is..., beyond driving I feel safer with a driver who is...

6.2 Regression

We explain briefly how we will estimate gender differences in responses. Essentially all our survey questions will be analyzed using the following regression framework:

$$Y_i = \beta_0 + \beta_1 \text{Gender} + \beta_2 X + \epsilon_i$$

where Y_i is the relevant dependent variable, gender is the gender of the respondent, X is a vector with the demographic characteristics we collected, and $epsilon_i$ is an error term. We will use a similar regression for the other categories of the free form question and for the multiple choice question.

Depending on the type of question, the dependent variable type will differ slightly.

Free-text response, multiple choice question: We transform the possible answers to the free-text question into multiple categories, depending on what criteria respondents say matters most: Safety, Profile, Hygiene, Gender, Reliability, Other, Nonsense. We then estimate the gender difference in the probability of the free-form answer being in any of these categories. We are particularly interested in the gender differences in those saying safety is the main criteria for selecting passengers. In the regression above, Y would be a dummy variable equal to 1 if the respondent's answer to the free-form question contained something related to safety and 0 if not.

For the Likert scale questions, the dependent variable will be the value of the Likert Scale, i.e. instead of the dependent variable being a dummy variable it will be a continuous variable with values ranging from 1 to 4 or from 1 to 5.

For the gender choice question we will be interested in two outcomes: the probability of preferring women and the probability of preferring one's own gender.

7   Sample selection

We ask Qualtrics to select respondents who speak French and have used carpooling services before. We have also asked them for a gender balanced sample. Qualtrics filters out respondents over 55 years old, which is not a concern for our study, as older respondents are unlikely to have used carpooling services before. Respondents who answer "Other" for their gender, or who do not give consent will be excluded.

Respondents who are inattentive or answer randomly can threaten the validity of our estimates. We introduced an attention check during the vignette, and those who fail this attention check are excluded from our sample and are not compensated.

Despite this check, there may still be respondents who randomly succeed in passing the attention check but still give nonsensical answers. We plan to show results without these low-quality answers by excluding respondents who answer our survey too quickly. We will show how our results differ when excluding

respondents who finish survey twice as fast as the median time. A second check will be to exclude those who write nonsensical answers to our free-text question.