Discrimination on Online Markets: Evidence from a Field Experiment
Pre-Analysis Plan

January 2021

1 Introduction

We study the extent and type of ethnic and gender discrimination that exists on two of the most widely used "P2P" online marketplaces in France: an online classifieds platform and a ride sharing platform. The first phase involves data scraping in which we collect large amounts of data on real ads and rides that are publicly available through the platforms' APIs. This will allow us to better understand the characteristics of the two markets and provide correlational evidence on the existence of discrimination and its magnitude in different submarkets. As is typical in the literature using French data, we will categorize the first names (in the case of the online classifieds platform) to proxy for the gender and/or minority status (supposed non-French origin) of the buyers and sellers, as it is illegal to categorize ethnic status in France. Using insights from this observational data, we will then implement a randomized controlled trial (RCT) in which we create sets of fictitious profiles (both buyers and sellers of fictitious goods for the online classifieds platform and drivers for the ride sharing platform car) that are precisely matched on all observable characteristics but differ on the supposed minority/gender status. This will allow us to provide causal evidence on whether discrimination exists in these online marketplaces. Furthermore, by experimentally manipulating the fictitious buyer and seller profiles on other dimensions and exploiting existing market heterogeneity, we aim to better understand the ways in which online markets can be designed in order to minimize the negative manifestations of discrimination.

This mimeo outlines our conceptual framework that will guide the empirical analysis. In the appendix, we provide results on statistical power and outline the technical aspects of profile and advert creation.

On the supply side (sellers) our baseline specification is

$$y_i = a + \beta maj_i + \gamma min_i + X_i + C_i + T_t + e_i \tag{1}$$

where y is an outcome for advertisement i posted. maj and min are indicator variables of majority and minority status of the seller. X is vector of advert heterogeneity orthogonal to status and T a vector of time effects (time of day, day of week, month of posting). C is a vector of the random variations in adverts introduced so that our fictitious adverts do not appear suspicious. The average treatment effect of majority and minority status in comparison to the anonymous pseudonym are β and γ, respectively; and $\gamma - \beta$ measures the effect of minority in comparison to majority status.[1]

In this project we use K dimensions of experimental and baseline heterogeneity in X to explore the nature and extent of discrimination on these two online platforms as well as estimate potential policy parameters.

$$y_i = a + \theta min_i + \sum_k^K (\lambda_k min_i * x_k + \delta_k x_k) + C_i + T_t + e_i \tag{2}$$

where we focus on minority status in comparison to majority status for ease of exposition. See the Appendix and AEA registry for further details on the dimensions of heterogeneity we exploit.

2 A test for statistical discrimination

Following Aigner and Cain (1977) and using notation from Autor (2009), suppose buyers in these markets believe that the mean quality of goods and services is distributed as,

$$\eta_m \sim N\left(\bar{\eta}_m, \sigma_\eta^2\right) \tag{3}$$

where $\bar{\eta}_1$ and $\bar{\eta}_0$ are mean beliefs on good/service quality for minorities $(m = 1)$ and majorities $(m = 0)$, respectively. σ_η^2 is assumed identical for both groups. Average quality may be the same or differ, but the dispersion of quality is the same for both groups. We can write

$$\eta_i = \bar{\eta}_m + \varepsilon_i$$

but buyers only see a noisy signal of quality that we can allow to be more or less informative by type:

$$\tilde{\eta}_i = \eta_i + \iota_i \text{ where } \iota \sim N\left(0, \sigma_{\iota m}^2\right), \text{ with } \sigma_{\iota m}^2 > 0. \tag{4}$$

Hence:

$$\tilde{\eta}_i = \bar{\eta}_m + \varepsilon_i + \iota_{im}$$

and $E\left(\tilde{\eta}_i \mid \eta_i\right) = \eta_i$, meaning that the signal is unbiased. The expectation of quality (η) given the signal $(\tilde{\eta})$ is simply the regression equation of quality on the signal and a constant for both groups:

[1] Our specification will be very similar for measuring bias against buyers, where X and C are adapted to reflect fictitious buyers. We will also fit count data and binary response models depending on distribution of outcome y.

$$E(\eta_1 \mid \tilde{\eta}_1) = \bar{\eta}_1(1-\gamma_1) + \tilde{\eta}_1\gamma_1$$
$$E(\eta_0 \mid \tilde{\eta}_0) = \bar{\eta}_0(1-\gamma_0) + \tilde{\eta}_0\gamma_0$$

where

$$\gamma_x = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_{\iota m}^2} = \frac{Var(\eta)}{Var(\eta) + Var(\tilde{\eta}_m)} = \frac{Cov(\eta, \tilde{\eta})}{Var(\tilde{\eta}_m)} = \frac{Cov(\eta, \tilde{\eta}_m)^2}{Var(\eta)Var(\tilde{\eta}_m)} = r_x^2$$

which is bounded between 0 and 1 and is the squared correlation coefficient. We see that when the signal becomes more and more precise more and more weight is put on the signal and less on the group average.

2.1 Difference in variances

Suppose now that average quality is the same, $\bar{\eta}_1 = \bar{\eta}_0 = \bar{\eta}$. The regression equation can be expressed as

$$\begin{aligned} E(\eta_i \mid \tilde{\eta}_i, m) &= \bar{\eta} * (1-\gamma_1) * m + \bar{\eta} * (1-\gamma_0) * (1-m) \\ &\quad + \tilde{\eta}_i\gamma_1 * m + \tilde{\eta}_i\gamma_0 * (1-m) \\ &= \bar{\eta} * (1-\gamma_0) + (\gamma_0 - \gamma_1) * (\bar{\eta} * m) + \gamma_0 * \tilde{\eta}_i + (\gamma_1 - \gamma_0) * (\tilde{\eta}_i * m) \end{aligned}$$

where

$$\begin{aligned} &\bar{\eta} * (1-\gamma_0) = \text{ weight put on mean majority average} \\ &(\gamma_0 - \gamma_1) * (\bar{\eta} * m) = \text{ minority difference of weight put on mean} \\ &\gamma_0 * \tilde{\eta}_i = \text{ weight put on majority signal} \\ &(\gamma_1 - \gamma_0) * (\tilde{\eta}_i * m) = \text{ minority difference on weight put on signal} \end{aligned}$$

- There is no discrimination when true quality equals the group average: $\eta_i = \bar{\eta}_m$ because the signal is unbiased = no discrimination on average
- Discrimination simply hurts or helps the seller depending on whether true quality is above or below the group average

The difference in expected quality between majority and minority sellers given the same signal is

$$E(\eta_i \mid \tilde{\eta}_i = \tilde{\eta}, m = 1) - E(\eta_i \mid \tilde{\eta}_i = \tilde{\eta}, m = 0) = (\gamma_0 - \gamma_1)(\bar{\eta} - \tilde{\eta})$$

- If minorities have a more noisy signal, then high/low quality minorities are differentially hurt/helped by discrimination compared to majorities
- Whether it hurts or helps simply depends on if i's signal quality is above or below the market average

2.2 Difference in means

Now assume $\bar{\eta}_1 < \bar{\eta}_0$ i.e. minorities sell inferior quality products/services on average but the variances of the quality and the signal of quality are the same $\gamma_0 = \gamma_1 = \gamma$. We find that,

$$E(\eta_i \mid \tilde{\eta}_i = \tilde{\eta}, m = 1) - E(\eta_i \mid \tilde{\eta}_i = \tilde{\eta}, m = 0) = (\bar{\eta}_1 - \bar{\eta}_0)(1 - \gamma) < 0$$

- If the signal is uninformative ($\gamma = 0$) then the expectation is just the difference in means for all i's = discrimination for all minorities with above average quality, yet on average there is no discrimination
- With $\gamma = 1$ the signal is perfect and expected quality conditional on the signal is equal to true quality
- Or unless the special case of $\tilde{\eta}_i = \bar{\eta}_m$ holds

2.3 Experimental variation in treatments

We will experimentally treat adverts with more or less precise information about the good/service/seller: $\sigma^{T=1}_{\iota i m} < \sigma^{T=0}_{\iota i m}$.

Equating with our empirical equation we can set $k = \tilde{\eta}^T$ which is a dummy indicating that the signal is more precise. If average quality is the same, $\bar{\eta}_1 = \bar{\eta}_0 = \bar{\eta}$ but the baseline variance of the signal is higher for minorities than majorities ($\gamma_1^{T=0} < \gamma_0^{T=0}$), the coefficient on the interaction term then captures the "difference in differences" quantity,

$$E(\eta_i \mid \tilde{\eta}_i^{T=1}, m = 1) - E(\eta_i \mid \tilde{\eta}_i^{T=0}, m = 1) - \big(E(\eta_i \mid \tilde{\eta}_i^{T=1}, m = 0) - E(\eta_i \mid \tilde{\eta}_i^{T=0}, m = 0)\big)$$

which is equal to

$$(\gamma_0^{T=1} - \gamma_1^{T=1})(\bar{\eta} - \tilde{\eta}_i^{T=1}) - (\gamma_0^{T=0} - \gamma_1^{T=0})(\bar{\eta} - \tilde{\eta}_i^{T=0})$$

Some points and caveats:

- *ex post* difficult to interpret sign. Relates to the Heckman critique: a noisier minority signal and a sufficiently low threshold for buyers' purchase price means that we could see negative effects of the information treatment on minorities

- One solution is to directly estimate the relative variance of unobservables, as suggested in Neumark(2012). As he points out, "a higher variance of unobservables for one group implies a smaller effect of observable characteristics on the probability that an applicant meets the standard for hiring." Taking our offer of goods as a similar treatment to a correspondence study, we can identify the ratio of variances if we can vary characteristics that significantly affect the buying probability of consumers (e.g. price or quality of goods). In that case, we can infer from the reactions of buyers the ratio of variances using a heteroskedastic probit. If the ratio of variances is equal to 1, then we can assume that observed differences between minority and majority sellers come from differences in means - in which case increasing information and giving a more precise signal of the good's quality should decrease discrimination.
- Otherwise, with different variances, we need to *ex ante* accurately define signals (profiles, adverts, etc.) that would be above or below average quality in market (harder)
- If we can plausibly assume/do all this, then for above average quality signals:
  - a positive estimated diff in diff ($\hat{\lambda}_{\tilde{\eta}^T} > 0$) would be evidence that improving the signal reduces discrimination and improves outcomes for high quality sellers.
- For below average quality signals:
  - a positive estimated diff in diff ($\hat{\lambda}_{\tilde{\eta}^T} > 0$) would be evidence that improving the signal reduces discrimination, but hurts these "below average" minorities.
- Policy would have no effect when i's signal equals group average: $\tilde{\eta}_i = \bar{\eta}_m$

Now, what if it is that the average quality of good/service is indeed lower, $\bar{\eta}_1 < \bar{\eta}_0$ but the variances only differ by treatment status and not by group, $\gamma_1^{T=0} = \gamma_0^{T=0} = \gamma^{T=0} < \gamma^{T=1}$. The differences in differences would capture the quantity,

$$(\bar{\eta}_1 - \bar{\eta}_0)(\gamma^{T=0} - \gamma^{T=1})$$

- This quantity will always be positive if treatment makes the signal less noisy for both groups.
- Much more easy to interpret: information reduces discrimination because the counter party puts more weight on the signal thus expected quality is closer to i's quality, on average.

2.4 Taste-based discrimination

Becker's model predicts manifestations of distaste or animus against minorities cannot hold in perfectly competitive markets because it presents a profit opportunity for non discriminating individuals who arbitrage away the profits of taste discriminators. Thus a standard test would be to use baseline heterogeneity in the competitiveness of the submarkets to test differences in impacts of minority status:

$$y_i = a + \theta min_i + \lambda_{comp}(min_i * competitive) + \delta_k competitive + e_i \qquad (5)$$

The competitiveness index, or market "thickness" can be constructed using the data scraped from the respective websites. Using the mean time for goods to disappear from the platform, and the number of sellers, we can create thickness indices both for specific locations and specific types of goods. Since we can also recover the number of clicks on our fictitious ads, we will be able to construct more precise indices, and possibly a tightness measure, based on this measure once the experiment has started.

Becker's model also predicts that distaste or animus against minority groups will be manifested at the point of the marginally biased counter party (if for some reason discrimination is not arbitraged away). Charles and Guryan (2008) show that the Black White wage gap is highly correlated with this "marginal discriminator" which is defined as the bias of Whites at a given quantile of the share of Black workers in the work force. Applied to our setting we can run,

$$y_i = a + \theta min_i + \lambda_{marg}(min_i * marginal) + \delta_k marginal + e_i \qquad (6)$$

Where *marginal* is the local vote percentage for the extreme right political party at the p percentile of the proportion of minorities (classified by names) in the geographic submarket.

2.5 Insurance mechanisms

We will explore two insurance mechanisms as additional policy instruments that we believe may affect the manifestation of discrimination on the two markets.

The first is a formal option, called "Secure payment," and offered by the online classifieds platform for certain categories of goods. It allows the transaction to pass exclusively by mail and provides recourse (dedicated service team who help buyers unhappy with the good). We will experimentally vary this option for our fictitious sellers. An interesting aspect of this insurance mechanism is that it is costly for the buyer: 4% of the price of good + shipping costs (which vary with the distance between the two parties). This may allow us to put an actual price tag on discrimination.

The second is informal and relates to the existing composition of the car for rides on the ride sharing platform. At the moment when our fictitious rider requests a seat in a car, we are able to collect information on the composition of the car: the names (giving us minority and gender status) of the other passengers and, of course, the driver. We hypothesize that discrimination may be contextual in that it may depend on the existing gender/minority proportion of the car: the gender/minority effect will change whether they are the first rider, the only minority/female in the car, the only majority/female in the car, etc. We will also create fictitious rides, where we will subtly signal the racial composition of the car in the car ride ad. Our hypothesis is that there might be more demand for cars composed of white females, and less demand for cars where the passengers are minority men. Our goal is to test the demand for these "homogeneous" car compositions relative to mixed rides.

An initial formal channel by which these insurance schemes may work could be that this simply reduces noise in the quality signal (higher γ), hence buyers/passengers put more weight on the signal and less on the average. But it may also be correlated to tastes in that it changes the way people interact during the transaction/ride.

2.6 Group specific demand curves

By experimentally varying the price of goods around the market averages we can test whether minorities face more or less elastic demand. The existing correspondence study literature, including studies on discrimination on online marketplaces have implicitly modeled discrimination as minority sellers facing a demand curve that is shifted down/in due to a discrimination term D. In our context, this means that majority sellers face demand as a function of price $Q(p)$ and minorities face demand $Q(p+D)$, where D could be a Beckerian cost incurred by buyers interacting with minorities or a term capturing increased uncertainty or differences in beliefs about the mean quality of the goods that consumers price in when deciding whether to buy. If our outcome $y_i = Q_i$ then the parameter β_1 in equation 1 clearly measures D. But this does not tell us whether minorities face more or less elastic demand than majorities which can have implications for producer surplus when aggregate labor market conditions change.

Examining pricing as our dimension of heterogeneity of interest $x_k = p$ we model a linear demand curve as:

$$Q_i = a + \theta min_i + \lambda_p min_i * p + \delta_p p + e_i \tag{7}$$

Clearly the slope of the demand curve for majorities is $\frac{dQ}{dp} = \delta_p$ while for minorities it is $\frac{dQ}{dp} = \lambda_p + \delta_p$. A log ($ln$) linear demand curve would give the different point elasticities faced by the different sellers.

$$-\frac{dQ}{dp}\frac{p}{Q} = -(\lambda_{lnp} min + \delta_{lnp}) \tag{8}$$

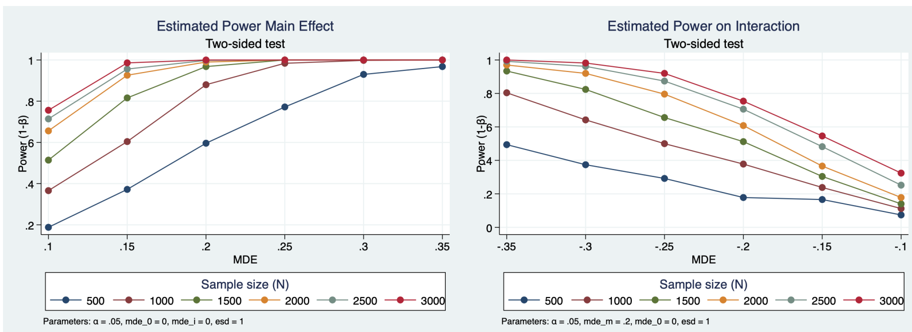
Why is this interesting? Imagine an aggregate shock where market supply of the good or service becomes more scarce or the government imposes a tax on the good or service. If $\lambda_{lnp} < 0$ it means minorities face more elastic demand (flatter demand curve). Thus, in addition to any level differences in demand captured by θ_0, minorities would have less pricing power or would face sharper drops in volume *relative* to majorities after the shock/tax.

Appendix

A Power

The following figures simulate the study's statistical power over effect sizes of the θ and λ_k parameters:

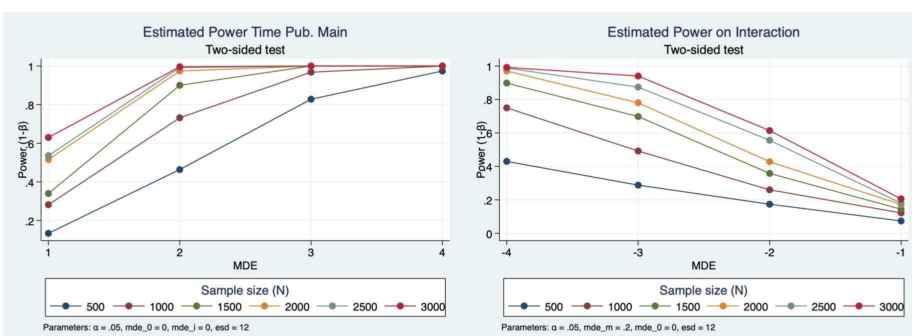
Assignment probabilities to majority status as to secondary treatment are 0.5.



(a) Main majority effect (b) Interaction effect

Figure 1: Statistical power for standardized main and interaction effect: $e \sim (0, 1)$

The next figures simulate power for the effect on days since publication using the estimated standard deviation from scraped data on cars sold on the online classifieds platform. Thus, $e \sim (0, \hat{\sigma})$ where $\hat{\sigma} \approx 12$ and $\hat{\beta}_0 \approx 8$. The estimated standard deviation was obtained from our fully specified model.



(a) Main majority effect (b) Interaction effect

Figure 2: Statistical power for main and interaction effect in days: $e \sim (0, \hat{\sigma})$

B Experiments - Le Bon Coin

B.1 The relevant market for online ads

- We will focus our study on large Urban Units defined by INSEE. Focusing on these dense urban areas ensures that ads will attract enough attention as they will be posted in relatively dense areas and where goods can be collected via public transports or driving.
- The demand can come from broader areas and we will analyze the responses received to identify whether our assumption seems reasonable

B.1.1 Selection of relevant markets for the experimentation

- Our experiment will take place in a limited number of urban units.
- When identifying discrimination on the demand side. It is important to avoid being detected and to limit our influence on the whole market. We will thus focus on large urban units with different market tightnesses.
- Most of our experiment will take place in the largest urban unit: the urban unit of Paris. This methodological choice is justified by several facts:
  - 13% of the online ads posted on LBC are located in Ile De France, the Paris urban area.
  - Moreover, a large portion of minorities, migrants and descendants of migrants live in this urban unit. Therefore, 28% of the online ads identified as coming from a minority are issued in Ile de France.
  - Finally, this urban unit offers a broad diversity of municipalities.
- As Ile de France might be characterized by a relatively high market tightness, Experiment 4 and additional robustness checks will be performed in smaller urban units.

B.1.2 Heterogeneity of the territory within the relevant market

We randomize within municipalities making sure that one ad is not posted simultaneously in different places (every 15 days).

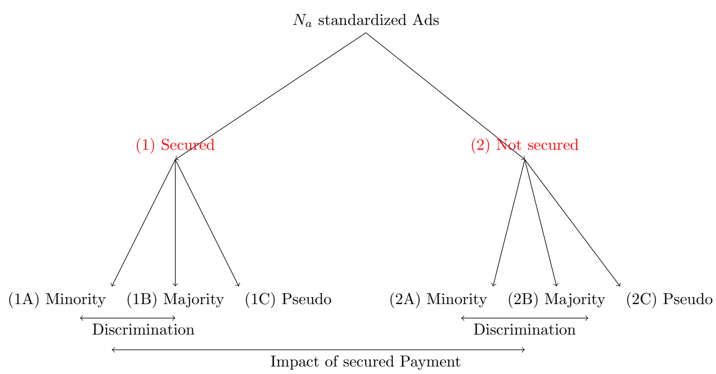
B.2 Creation of fictitious sellers

B.2.1 Experiment 1: identification of discrimination following the availability of secured payment

The first experiment aims at identifying a) the existence or not of discrimination b) the impact of secured payment. To do so we proceed in two steps:

1. We will focus on a small (yet to be defined) number of categories i to reduce variation. First, we will create $N_{a,i}$ standardized ads for each category i with 6 random variations that will be controlled for in the analysis. The random variation might affect, the structure of the text (inclusion of hello or not, order of the information provided) or some features of the good (relatively close colors as dark grey/ black, white/ cream , etc...) to ensure the comparability of the goods without being suspicious. The price of the 6 variations of the ads will be the same. The total number of observations in the RCT will be given by $Obs = \sum_{i=1}^{I} \sum_{a=1}^{A} N_{a,i} \times 6$

2. Second, we also create N_p triplets profiles with similar characteristics (geographical location, age, ratings) but different ethnicity (minority, majority, pseudo).



The characteristics of the Ads and profiles are summarized in Table 1

	Profile					Ads			
Name Random	Ethnicity Random	Municipality Random (15 days)	Type of good Fixed	Colour Random (black/grey)	State Fixed	Price Fixed ? Or +/-5%	Title Fixed	text Random (Hello,Reasons of sales etc..)	secured Random
Jean	Caucasian	Neuilly	Samsung G5	Black	Good	150	Samsung G5	Vends samsung	No
Mohammed	Arabic	Puteaux	Samsung G5	Grey	Good	150	Samsung G5	bonjour, je Vends mon samsung	No
Jojo92	pseudo	Courbevoie	Samsung G5	Dark grey	Good	150	Samsung G5	Vends samsung Merci	No
Yves	Caucasian	Levallois Perret	Samsung G5	Black	Good	150	Samsung G5	Bonjour, Vends samsung	yes
Ahmed	Arabic	Neuilly	Samsung G5	Grey	Good	150	Samsung G5	je Vends mon samsung en double	Yes
Patou92	pseudo	Puteaux	Samsung G5	Dark grey	Good	150	Samsung G5	Vends samsung	Yes

Table 1: Example of 6 ads for a Samsung G5

This experiment on secured payment will be tested on a limited set of homogeneous municipalities.

B.2.2 Experiment 2: identification of the role of information

In this second experiment, the randomisation will be on the number of characteristics informed. For the first group, the information provided will be the minimum of its category while for the second group most of the information will be filled. The idea is to have goods with similar signals (ideally with the same picture) but one ad will have much more detail on the characteristics of the good while the other one won't. The randomization will be on whether the required information is filled (T=0) or the maximum information is filled for the category (T=1). There are many characteristics that can be filled for cars for instance. There won't be partial treatment. Besides close colors, few randomization on the characteristics should be allowed.

B.2.3 Experiment 3: identification of the role of the territory

In this third experiment, the randomisation will be on the set of homogeneous municipalities as defined previously. We plan to randomise which municipalities are treated within urban units, to check whether minorities, who often live in different municipalities within urban units, could be affected by segmented demand due to their different localisation.

B.2.4 Experiment 4: identification of the role of the market tightness

In this third experiment, the randomisation will be on the urban unit. Within these urban units, the profiles should live in the largest municipality of the urban unit.

B.3 Creation of fictitious buyers

As previously (and to be consistent), we will send some requests regarding ads a located in a limited number of urban units and within the selected categories i.

Concerning the Experiment 1 (identification of discrimination following the availability of secured payment), we should find two ads that post the exact same good but one features a secured payment and the other one does not. Three fictitious buyers (majority, minority, pseudo) will contact the seller of each ad. We proceed the same way for each ad, therefore the total number of ads will be given by $\sum_{i=1}^{I} \sum_{a=1}^{A} N_{a,i} \times 2$ and the number of requests will be : $\sum_{i=1}^{I} \sum_{a=1}^{A} N_{a,i} \times 2 \times 3$.

Concerning the Experiment 2 (identification of the role of information), we will identify one ad for which we will send 6 requests : 3 potential buyers (majority, minority, pseudo) providing some info on why he/she wants to buy the good and 3 other potential buyers (majority, minority, pseudo) providing no info. We proceed the same way for each ad, therefore the number of requests will be : $\sum_{i=1}^{I} \sum_{a=1}^{A} N_{a,i} \times 6$. To reduce the level of detection, we can run this experiment between goods rather than within goods (2 goods X 3 buyers).

Concerning the Experiment 3 (identification of the role of the territory), we will identify one ad which is surrounded by municipalities with different standards of living. We will send 6 requests : 3 potential buyers (majority, minority, pseudo) from a deprived municipality and 3 other potential buyers (majority, minority, pseudo) from a wealthy municipality. We proceed the same way for each ad, therefore the number of requests will be : $\sum_{i=1}^{I}\sum_{a=1}^{A} N_{a,i} \times 6$.

Concerning the Experiment 4 (identification of the role of the market tightness), we can easily have a measure of market tightness (ad X location) with our scraped dataset of ads. We will send 3 requests from potential buyers (majority, minority, pseudo) to ads belonging to a tight market and 3 other requests to ads belonging to a non-tight market. The tightness of the market is defined by the couple (ad X location) : we will compare the callback rates for both similar goods but different location and different goods but different goods.

B.4 Experiments - real buyers and sellers

As long as the sale and purchase are fictitious, it remains difficult to exactly quantify the level of discrimination. Indeed, discrimination in the request for information does not necessarily translate into discrimination in purchases. In order to solve this issue, we plan to have a part of the experimentation dedicated to the purchase and sale of real goods. This experimentation will first involve the actual purchase of some of the goods (within the selected categories), by the fictitious buyers mentioned above. In order to limit the overall cost, we will focus on low-cost goods (e.g. books) and a well-defined area (Ile de France). This first step will allow us to compare our results and check their robustness on the discrimination against buyers when the transaction is effective. In a second step, the acquired goods will be resold on Le Bon Coin, which will allow us to estimate the discrimination against sellers. These goods will be resold at a price equal to or lower than the acquisition price.

C Experiments on the ride sharing platform (BlaBlaCar)

The businesses are slightly different and so are their market design. Here we describe the steps of the experiment on the ride sharing platform, and we highlight in what respects it may differ from that of classified ads platform.

C.1 Dimensions to be tested

In accordance with the technical brief, we explore three of the four hypotheses developed for the classified ads framework:

- Hypothesis H1: The information provided by the driver (passenger) influences the discriminatory behavior of the passenger (driver).
- Hypothesis H2: Competition reduces discrimination.
- Hypothesis H3: Location affects discrimination.

Because a transaction on the ride sharing platform induces a lot of interpersonal interactions, we extend the analysis to gender discrimination, on top of ethnic discrimination. There is however no possibility of posting pseudos.

This means there are in total 5 binary characteristics we want to test (minority, gender, information, competition, location).

C.2 Creation of fictitious drivers

C.2.1 Documenting discrimination in "base markets"

To facilitate the smooth integration of our ads, we will post them primarily on the most frequented routes. We will again create N_a standardized ads. To study all possible interactions, we should create $2^4 = 16$ variations of a similar ad. However, to increase our statistical power, we focus on the main ones. All hypothesis except H2 can be tested within markets.

We will post $N_a * 2^2$ ads on the most frequented routes, during regular weekdays, representing all possible combinations of gender-ethnicity.

Base markets: These route-weekdays are our "base markets". The base markets will cover at least the 10 largest routes on the ride sharing platform. The number of weekdays experimented will be such that our ads never represent more than 10% of the rides posted in a given 2 hour window.

C.2.2 Experiment 1: the role of information

We test H1 only on ethnic discrimination. This means we post N_a minority and N_a non-minority additional ads with detailed information in our base markets.

C.2.3 Experiment 2: the role of Competition

To test H2, we will post $N_a * 2$ additional ads on the same routes as the base markets, but specifically when the traffic is particularly dense (Friday evenings). Again, we will test the effect of competition on the minority effect, meaning there will be N_a minority and N_a non-minority ads.

In total we will post a minimum of $8N_a$ ads (black leaves of the tree), down from $16N_a$ if we include all interactions.

C.2.4 Experiment 3: the role of location/territory

All these postings will occur in a set of "base markets". If the size of the sample allows, we may derive route-specific estimates and relate it to city-specific characteristics.

C.3 Creation of fictitious passengers

As previously (and to be consistent), we will send some requests regarding ads located on the "base markets" defined above.

The experiment consists in creating fictitious passenger profiles. These profiles will vary according to four dimensions : apparent ethnicity (name and photo), apparent gender (name and photo), level of information (more or less detailed description of the passenger) and location. These fictitious passengers

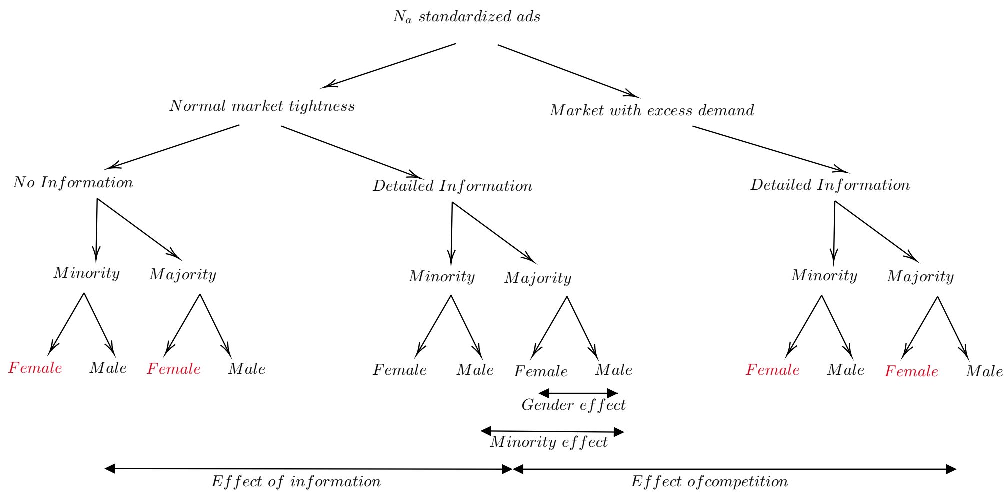


Figure 3: Variations to be tested on the ride sharing platform

will send precision requests to the drivers, allowing us to analyze the response rate of the latter according to these four dimensions. We can also analyze the response rate according to market conditions (e.g. high or low demand, trips with a high representation of minority drivers), according to the number of seats still available and according to the number of seats that the fictitious passenger wants to book.