Impact of online reputation on ethnic discrimination.*

Work in progress- current version here

Xavier Lambin† & Emil Palikot‡

October 29, 2018

*We are grateful to Charles Angelucci, Maria Elena Bontempi, Jacques Crémer, Daniel Garrett, Marc Ivaldi, Bruno Jullien, Yassine Lefouili, Nicolas Pistolesi, Patrick Rey, Mathias Reynaert, Timothy Simcoe and Steven Tadelis for their valuable comments at various stages of the paper. We also thank participants to Telecom ParisTech internal seminar sessions, IIOC, AFSE, the Toulouse School of Economics digital conference for useful comments and suggestions.

†Ph.D candidate at Toulouse School of Economics, xavier.lambin@tse-fr.eu

‡Ph.D candidate at Toulouse School of Economics, emil.palikot@gmail.com. Emil Palikot acknowledges support from the European Research Council under the Grant Agreement no. 340903

Abstract

The impact of reputation on the outcomes of ethnic minority users of a popular ride-sharing platform is studied in this paper. Using a large unique dataset, we find that minorities achieve lower economic outcomes: compared to non-minority users, their listings are less popular, they sell fewer seats and have lower revenue. We also show that reputation is instrumental in reducing the ethnic performance gap, which is concentrated during the first interactions of minority drivers, and reduces substantially after they establish a reputation. We develop a model of career concerns, which allows us to model markets' beliefs about the expected quality of service of drivers. We show that these beliefs are formed based on socio-demographic characteristics of drivers and updated with reviews. We show that a significant part of the initial performance gap is due erroneous, overly pessimistic beliefs about the expected quality of minority drivers. Finally, we conduct a counterfactual analysis of this market in the absence of a reputation system. Our research stresses the importance of a well-designed reputation system in alleviating discrimination.

1 Introduction

What is the source of online ethnic discrimination? The answer to this question has important implications for our understanding of ethnic inequality in access to online services. Facing this question is as important as ever due to increasing socio-economic differences across ethnic groups and their significant impact on economic growth (see e.g. Alesina et al. (2016)). The "collaborative economy" has embraced the notion of creating trust across social divides[1], and referrs to social capital created with their online reviewing systems as the "cornerstone" of success.[2] Online platforms typically match strangers to engage in activity with uncertain outcomes. As documented in the economic literature (see, e.g. Edelman and Luca (2014)), the lack of specific information about other users often leads to reliance on prejudice and ethnic bias. If priors based on ethnicity, gender or race, are not subject to change due to reputation, addressing online discrimination might require drastic changes in the design of many popular marketplaces (Edelman et al. (2017)). If, by constrast, online reputation allows minority users to correct market biases and narrow the gap – as we claim in this paper– then the right solution is to focus on improving the efficiency of reviewing systems.

This paper empirically tests whether the gap in economic performance across ethnic groups changes with reputation. Using a unique dataset collected with a web-crawler[3] on a popular ride-sharing platform, we find that drivers from ethnic minorities are achieving lower economic outcomes than non-minority users, especially upon entry to the platform. We show that only part of the difference is caused by statistical discrimination, which is a discriminatory treatment due to correct expectations formulated based on observed socio-demographic characteristics. Importantly, we claim that most of the difference in outcomes cannot be justified in terms of statistical discrimination and we attribute the difference to overly pessimistic (erroneous) beliefs about the quality of service provided by minority drivers. Importantly, establishing a reputation updates the and substantially decreases the performance gap.

Our analysis is performed on the popular French platform BlaBlaCar. Three features of the platform design and usage practice provide us with a unique opportunity to study

[1]BlaBlaCar claims that their reputation system recreates “a sense of trust almost comparable to the level of trust in friends” (Mazzella and Sundararajan (2016)), and that carpooling makes people “more open to others” (Blablacar (2018)).

[2]A TED talk by Joe Gebbia: https://www.youtube.com/watch?v=16cM-RFid9U , last accessed on October 8, 2018.

[3]Web-crawler is a bot that systematically browses and indexes a selected webpage.

the ethnic performance gap. First, BlaBlaCar is used by non-professional drivers (in contrast to the vast majority of Uber drivers) and does not require significant financial capital (e.g. to real estate, in the case of Airbnb), which results in a share of ethnic minorities roughly proportional to the general composition of French society. Second, the design of the website enables us to study the history of reviews for all drivers that we observe, and the high frequency of usage gives us the possibility to follow the "career" paths of drivers. Third, we observe all drivers available on a given trip, together with their economic outcomes and a detailed description of each listing, which allows us to analyze the performance of individual drivers while controlling for a rich set of observables including the specific characteristics of the drivers, listings and routes.

We first present reduced form evidence of a performance differential between minority and non-minority drivers upon entry to BlaBlaCar. We show that minority listings receive 3% fewer clicks, sell 7.5% fewer seats, and have on average 5.3% lower revenue, despite controlling for a rich set of variables. We also show that the gap significantly decreases as users collect reviews. We revisit the profiles of drivers a couple of months after the initial data collection and find no significant differences in exit and inactivity patterns. This information allows us to reject the change in the composition of the population of drivers as an explanation for our results. Reduced form evidence suggests that statistical discrimination, championed in the existing literature on this topic, is most likely not the only factor.

Statistical discrimination refers to the discriminatory treatment of an individual due to her socio-demographic characteristics. However, for such discrimination to be a persistent source of inequality it must be based on true differences in expected quality of service. If the gap in economic outcomes of different socio-demographic groups narrows over time, another mechanism must also play a role. In this paper, we propose updating of an erroneous prior on the distribution of quality of service of minority drivers as the explanation for the observed pattern of market outcomes.

To disentangle statistical discrimination from erroneous beliefs, we develop a structural model of career concerns, inspired by Holmstrom (1999), whereby users care about the reviews they receive and exert efforts to maximize life-long consumption. This process allows us to explicitly model belief formation and updating. Furthermore, from the market outcomes of minority entrants, we infer the market prior about the distribution of quality as well as the impact of reviews on updating of these beliefs. In this framework, we show that part of the gap is due to differences in expected quality, which we relate to statistical discrimination. The remaining part (2-4% of revenue), we attribute to erroneous, overly pessimistic beliefs about the distribution of quality.

The structural model enables us to perform counterfactual analyses. First, we calculate the value of a high grade received at the beginning of a career. We measure the discounted sums of predicted gains in revenue due to such a grade and compare them across drivers. Two key factors determine this value: first, a rating that exceeds expectations impacts positively the belief about the type of a driver – and in turn her expected profits. Second, an additional review reduces uncertainty about the true type. These elements combined result in a sizeable (more than four times) difference in the value of a good review across ethnic groups.

In the second counterfactual analysis, we compare market outcomes with and without the reputation system. We disentangle several forces that are at play. The reputation system forces drivers to exert effort, which leads to demand expansion. Then, we quantify the total cost of effort provision. Undertaking effort allows individual drivers to signal their quality. As long as initial beliefs about the distribution of quality were correct, this is a zero-sum game within a given socio-demographic group. Comparing across these groups, on average, minority drivers are providing more effort than non-minorities; therefore they bear the higher cost. Finally, we set initial beliefs about minority drivers as inferred from their market performance upon entry. We contrast a baseline scenario in which the beliefs are updated, but there is the cost associated with taking efforts, with the case in which minority drivers keep facing low demand due to overly pessimistic prior, but do not exert effort.

Relation to literature: The economic literature on ethnic and gender discrimination is vast, and we cannot do justice to all of it in this short literature review. We focus on the studies most closely related to our project. First, this paper is related to the abundant literature on statistical discrimination. While the concept was first described in the 1970s (see Phelps (1972), Arrow (1973) and later Altonji and Pierret (2001)) with numerous interesting applications in the labor market (see, e.g., Coate and Loury (1993), Farber and Gibbons (1996), Charles and Guryan (2011), Lang and Lehmann (2012)) and housing market (e.g., Ewens et al. (2014)), online markets and their relative anonymity lend themselves to such discrimination. As a result, online markets represent a new observation terrain for researchers. Edelman and Luca (2014), Edelman et al. (2017) and Laouenan and Rathelot (2017) present evidence of discrimination on the short-term house rental platform Airbnb. Castillo et al. (2013), Goddard et al. (2015), Ge et al. (2016) and Cook et al. (2018) study discrimination in transportation systems.

Closely related to our empirical analysis, Farajallah et al. (2016) shows minorities have a lower success rate on a carpooling platform. Our paper also identifies discrimination; however, it goes into greater detail by showing that this discrimination is only partly statistical and that differences in expected quality cannot rationalize a large part of the performance. [4] These observations are similar to those in Laouenan and Rathelot (2017), who in the context of Airbnb, show that the prices of ethnic minority listings are more responsive to reviews.

Second, sociological research has studied the potential of reputation systems to offset trust judgments (e.g., see Abrahao et al. (2017), Tjaden et al. (2018)), where the latter is particularly relevant to our research as it focuses on carpooling.

Third, recent economic and computer science literature has studied the effectiveness and design of reputation systems, some notable projects including Nosko and Tadelis (2015), Cabral et al. (2010), Bar-Isaac and Tadelis (2008), Liu and Skrzypacz (2014), Livingston (2005), Jolivet et al. (2016), Bolton et al. (2004), Mayzlin et al. (2014), Jullien and Park (2014) and Zervas et al. (2015). These studies focus on understanding how consumers react to the information provided. They aim at improving the accuracy of the reputation system by, for instance, reducing fraud or providing adequate information. Spagnolo (2012) and Butler et al.(2017) use laboratory experiments to show that a reputation system, if not designed wisely, may hinder the entry of new participants. Kovbasyuk and Spagnolo (2017) also find that a repeated game with limited records maximizes the number of trades. Randomly changing types provide the mechanism behind the identified effect.

Finally, we consider a model of moral hazard, which is related not only to seminal works in the field Baron and Myerson (1982), Holmstrom (1999),Chiappori et al. (1999), and Laffont and Tirole (1986) but also to the more recent works of Roger and Vasconcelos (2014) and Garrett and Pavan (2012). We extend the model of Holmstrom (1999) by considering multiple populations of drivers, which allows us to compare expected quality. We also introduce a bargaining game, where we show that similar dynamics of effort can be found in a model without a perfectly competitive labor market.

The rest of this paper is organized as follows: Section 2 describes the functioning of the carpooling platform we focus on, as well as our data gathering process. Section 3 provides reduced form results documenting an output gap between minority and

[4] Darity and Mason (1998) provide a more detailed review of the economic theories of discrimination.

non-minority drivers, which is followed by a study of the positive effect of reputation building using a cross-section, a panel and a matching method. Section 4 adopts a model of career concerns for our setting. Section 5 analyzes the counterfactual of a market with no reputation system. Section 6 concludes the study.

2 Empirical context and data collection

BlaBlaCar is an online marketplace for ride-sharing that was established in 2006 in France and today operates in 22 countries, mostly in Europe, but also Mexico, India, and Brazil. The platform has eight million active drivers and over 50 million passengers[5], which makes BlaBlaCar the largest ride-sharing platform in Europe. BlaBlaCar enables drivers to sell seats in their cars that would otherwise be empty. In the spirit of the sharing economy, BlaBlaCar attempts to limit usage of its platform by "professional" drivers. This restriction is reflected in the pricing rules: drivers receive price recommendations based solely on cost considerations. The suggested price is linear in distance: 0.062 EUR per km. Drivers can deviate from this price, but the price is capped at 0.082 EUR per km. Furthermore, the number of seats users can offer is limited. Having two passengers booked with BlaBlaCar should allow a driver to cover her fuel costs; however, the price cap should prevent drivers from running a profitable business.

A potential passenger looking for a ride between a pair of cities sees a list of all available drivers ranked, by default, by time of departure, together with basic information: a photo of the driver, her name, average rating, a few details about the ride and the price. To obtain more information and, in particular, the history of reviews, a prospective passenger has to click and visit the profile of the driver. Examples of profiles and listing pages are provided in Appendix B. The passenger chooses the listing that she finds the most attractive and sends a booking request. The selected driver then decides whether to accept. Roughly half of the drivers choose the automatic acceptance feature while posting a ride. Finally, payment is made upfront via the BlaBlaCar online system. BlaBlaCar fees are deducted from the price paid by the passenger. AppendixB discusses their magnitude.

Passengers and the driver are encouraged to leave a review that consists of a written comment and a number of stars from 1 to 5. The review system has a simultaneous re-

[5]https://techcrunch.com/2017/04/10/how-blablacar-faced-growing-pains-and-had-to-change-its-focus/

veal feature, which means that a user cannot access a received review unless she writes one herself. Only after both reviews are sent do they become available to other users.[6]

Data collection We have collected our dataset using a web-crawler designed by us specifically for https://www.blablacar.fr/, from 1.07.2017 to 29.09.2018. The program first randomly selects a route from a list of predefined trips between the largest cities in France. Trips start/end in Paris or its vicinity and have their other end in one of the 110 largest cities. Crawling through the site, the program gathered all the information available to prospective riders, including the price posted by the driver, time and date of the posting and planned departure, destination, origin, type of car, and whether pets or large luggage are allowed. The data also include information related to economic outcomes that the listing has already achieved, that is, the number of clicks it has received (views) and how many seats have already been sold. Clicking on the listing is necessary to book a trip, and clicking opens a detailed description of the ride. After visiting this page, the passenger can still change her mind at no cost. By calculating the product of sold seats and price, we determine revenue collected by drivers at the moment we visit their listings. Next, we open the individual profiles of all drivers available on a given route. For each driver, we observe her name, age, picture, a short biography, number of Facebook friends, etc. Most importantly, we collect the entire history of received ratings and written reviews. An important feature of our data collection process is that we simultaneously observe listings that have been available for different periods of time; in fact, some of the listings could have been just posted. This explains why many of our observations have zero sold seats and zero revenue. To account for this fact, we control for how long a given listing is available and how many hours are left until departure. Finally, listings that have not sold any seats have zero market shares in our demand estimation presented in the structural model section, which causes technical complications. To account for this factor, we have collected a supplementary, smaller, dataset through BlaBlaCar API, where we observe economic outcomes at the moment of departure – specifically, the number of seats finally sold. We observe that the vast majority of listings sell at least one seat. We therefore use this sample to perform robustness checks.

Additionally, we have matched this data with several other datasets. Gender and ethnicity have been established with two complementary methods. First, we use the ethnic

[6] Over the years, BlaBlaCar has introduced a few changes to their reputation system, which affected grading behavior. Appendix A discusses these changes.

origins of names database published by the French government and supplemented with some other publicly available sources.[7] Translations of names with foreign origins into French show considerable diversity. We phonetically encode our name lists and allow for small spelling mistakes to improve our classification. Second, to increase precision, we also use machine learning software[8] to confirm the matching procedure based on facial recognition. A detailed description of our gender and ethnic identification process is provided in Appendix C. Our definition of minority drivers is based on names with an Arabic or African origin or connotation: in this practice, we follow most of the existing literature. However, by considering both groups and using photo recognition together with name connotation, our approach constitutes an extension of the definition of minorities compared to prior investigations of discrimination on this platform, which restrict the definition of minorities to drivers with an Arabic-sounding name (see Farajallah et al (2016)). We proxy the quality of the car by approximating its value with the average price of the same type of car posted on eBay in Germany using data from a Kaggle data science challenge.[9] Fuel costs are potentially significant pricing factors. The fuel efficiency of cars is calculated by matching car names with a dataset of fuel consumption of cars at long distances (French environment and energy management agency – ADEME). We also collect data on city-level daily average fuel prices and highway tolls to construct instruments for prices. Distances and expected time by car or public transportation are calculated using google.maps API. We also include the suggested and maximum prices set by BlaBlaCar. Information specific to the city of destination/departure is included, such as population, median income, and index of crime (French government statistics INSEE). We also have data on strikes related to transportation services (in particular, railways) that occurred in Spring 2018 and created demand shocks. Descriptive statistics of select variables are shown in table 1.

The average price of a ride is 29 euros for a traveling distance of 370 km in a car worth 6000 euros. The average driver is 37 years old and has posted (successfully or not) almost 40 rides in the past. Most of the drivers are men (73%), and approximately 15% of drivers are from a minority. In our dataset, we have approximately 340.0000 observations. Each driver has a unique ID, and given that we observe some drivers multiple times, we can construct a panel dataset. Section 3.2.2 describes in detail how

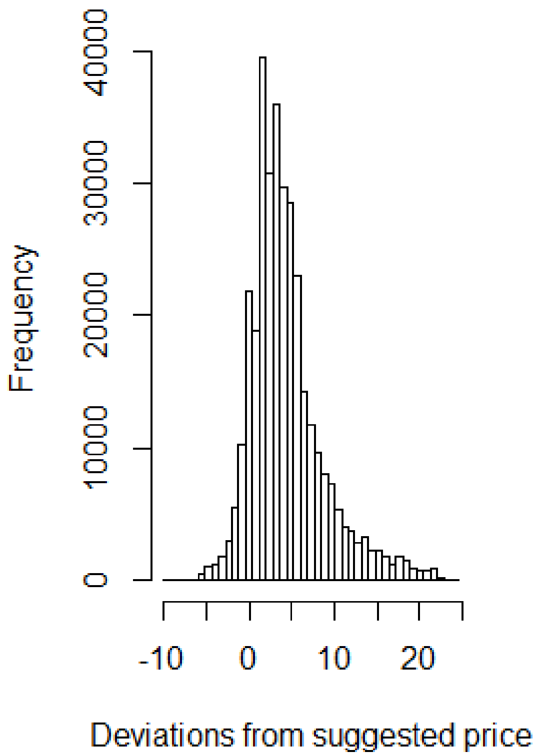
[7]https://www.data.gouv.fr/fr/datasets/liste-de-prenoms/ , http://www.signification-prenom.net/ , http://madame.lefigaro.fr/prenoms/origine/. The complete list of names and origins is available upon request.

[8]www.kairos.com

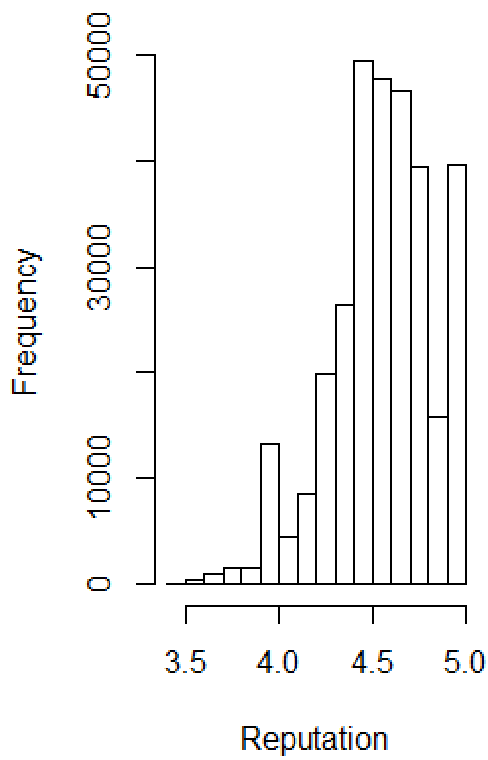
[9]https://www.kaggle.com/orgesleka/used-cars-database

Table 1: Descriptive statistics

Statistic	N	Mean	St. Dev.	Min	Max
price	333,572	28.79	14.50	6	72
age (driver)	338,338	37.20	12.60	18	68
reviews (#)	338,984	38.12	61.45	0	410
male	339,267	0.73	0.44	0	1
minority	342,429	0.15	0.36	0	1
picture	342,429	0.88	0.33	0	1
talkative	342,101	2.21	0.48	1	3
bio (#words)	338,815	14.66	16.64	0	79
ride description (# words)	338,988	25.69	38.16	0	186
reputation	315,419	4.60	0.27	3.60	5.00
published rides (total)	325,294	39.99	50.38	1	265
number of clicks	325,227	14.78	16.37	0	73
sold seats	342,429	0.27	0.58	0	4
revenue	338,878	6.12	14.42	0	76
posts per month	342,429	1.77	3.12	0.01	29.86
seniority (# months)	338,934	42.63	26.89	1	114
competition (# other drivers)	338,902	27.91	28.43	1	147
median revenue (city)	321,023	18,975.47	2,115.04	13,060.00	30,904.50
public transport (travel time)	331,271	3.75	2.27	0.14	15.24
train strike	342,429	0.06	0.25	0	1
value of car	281,696	6.03	5.04	0.60	24.40
fuel consumption	293,935	4.98	0.75	3.65	7.50
length (# km)	335,128	373.45	182.05	55.14	873.33
length (# hours)	342,429	13.29	4.60	0	23
hours until departure	321,124	129.42	167.73	−224.92	829.12
posted since	338,950	4.98	7.33	0	51
automatic acceptance	342,429	0.48	0.50	0	1
travel cost	277,206	59.75	29.15	1.00	135.93
weekday	342,430	0.65	0.48	0	1
weekend	342,430	0.35	0.48	0	1



(a) Histogram of deviation from the suggested price



(b) Histogram of reputation

this panel is constructed. Unfortunately, we have a number of missing observations for some variables; therefore, we typically have many fewer observations in the estimated regressions.

Key variables: Drivers' strategic decisions are, foremost, setting a price and the number of seats to offer. They receive a suggested price,from which they can deviate, based on distance. However, the maximum rate is bounded. Thus, the decision of drivers is really whether, and by how much, to deviate from the suggested price. Figure 1a shows the distribution of deviation from the suggested amount. Most drivers do not follow the suggested price, and typically deviations are in the range of 0-10 €. Some drivers set lower prices than recommended. This paper focuses on the impact of ethnic status and reputation on economic outcomes, but other, more standard, factors play important roles in these price differentials. Amongst others, fuel consumption and competition with other drivers and with public transportation have significant impacts. A characteristic feature of many online reputation systems is that most users leave very good reviews, and as a result, there is little variation in the data (Nosko and Tadelis (2015), Dellarocas and Wood (2008)). This is also the case with BlaBlaCar. Figure 1b shows the distribution of the average reputation of drivers. Low ratings are rare, and the vast majority or ratings falls within the 4-5 range (i.e., between "very good" and "perfect"). Finally, we collect several measures of economic outcomes. First, we have the number of clicks a listing has received, which serves as a proxy for the popularity of listings. The number of clicks is a softer measure of demand than sold seats and has more variation. Second, we observe sold seats, which allows us calculate revenue. These measures increase the number of zero values, which we account for by controlling for time since the listing has been posted and by providing additional robustness checks. Our dataset may miss some very successful rides that are no longer displayed when data are collected, which would lead to bias if the speed at which listings fill differs between minority and non-minority drivers. Since the most attractive listings are more likely to be those of majority drivers (see Appendix E for a discussion of a potential sampling bias), our estimates of the output gap should be seen as a lower bound.

3 What is the role of reputation?

We define the performance/output gap as the difference in economic outcomes achieved by minority and non-minority drivers, controlling for all available observables. To provide initial insight into the role played by reputation we present a set of simple reduced form models. A driver's career progresses as she collects reviews. Experienced users provide substantial individual information about the expected quality of service they provide, while expectations about new drivers are based more on socio-demographic information presented on their profiles. There are, necessarily, individual gains to having a good reputation, but as long as initial beliefs about the distribution of quality within a socio-demographic group are correct and the composition does not change over time, there should be no changes in differences in average group performance. If we observe that the difference between groups' average performance differs as drivers' careers progress, either the composition of drivers has changed, for example, due to the exit of less successful drivers, or the initial beliefs about expected quality were incorrect and have been corrected by the reviewing system. We argue for the latter explanation;

we will directly check and reject the former. We start by documenting the output gap, abstracting from the stage of the driver's career. Later, we study the impact of minority status for drivers who have just entered the platform and for experienced users.

3.1 The output gap

We start by providing a general assessment of the output gap without distinguishing between career stage. This process is similar to that in the studies documenting discrimination in digital marketplaces mentioned earlier. The raw data show that despite setting lower prices (29.1 EUR vs. 27.9 EUR), minority drivers achieve lower economic outcomes. Their listings receive fewer views (20.2 vs. 18.7), they sell fewer seats (0.27 vs. 0.26) and as a result, have lower revenue (7.2 EUR vs. 6.5 EUR). Part of the gap is due to differences in observables, as reported in table 2, which presents estimates of the regressions of various controls, including minority status, on the number of clicks received by listings, sold seats and revenue.

	Dependent variable		
	number of clicks	revenue	sold seats
minority	−0.417*** (0.095)	−0.488*** (0.092)	−0.014*** (0.004)
age	−0.044*** (0.003)	−0.014*** (0.003)	−0.001*** (0.0001)
reviews	0.026*** (0.002)	0.037*** (0.002)	0.002*** (0.0001)
(reviews)2	−0.00005*** (0.00000)	−0.0001*** (0.00000)	−0.00000*** (0.00000)
talkative	0.250*** (0.068)	0.065 (0.065)	0.003 (0.003)
male	−0.988*** (0.075)	−0.149** (0.072)	−0.002 (0.003)
seniority	−0.014*** (0.001)	−0.009*** (0.001)	−0.001*** (0.0001)
hours untill ride	−0.026*** (0.0002)	−0.015*** (0.0002)	−0.001*** (0.00001)
posted since	1.082*** (0.005)	0.279*** (0.005)	0.012*** (0.0002)
posts per month	−0.453*** (0.016)	−0.143*** (0.015)	−0.009*** (0.001)
length bio	0.002 (0.002)	0.004** (0.002)	0.0001 (0.0001)
car price	−0.018*** (0.007)	−0.014** (0.006)	−0.001*** (0.0003)
competition	0.027*** (0.001)	0.018*** (0.001)	0.001*** (0.0001)
median revenue	0.001*** (0.0001)	0.0004*** (0.0001)	0.00001** (0.00001)
public transport ratio	136.648 (157.910)	−907.637*** (150.883)	−7.425 (5.991)
km	0.006** (0.003)	−0.006** (0.002)	−0.0003*** (0.0001)
day	0.324** (0.126)	0.481*** (0.122)	0.016*** (0.005)
night	−0.438** (0.198)	−1.207*** (0.189)	−0.055*** (0.008)
train strike	4.266*** (0.175)	2.574*** (0.167)	0.119*** (0.007)
ride description	0.020*** (0.001)	0.009*** (0.001)	0.0003*** (0.00003)
picture	0.639*** (0.107)	0.476*** (0.104)	0.013*** (0.004)
automatic acceptance	−1.347*** (0.067)	2.070*** (0.065)	0.095*** (0.003)
weekday	−0.258** (0.132)	−0.564*** (0.127)	−0.035*** (0.005)
fuel consumption	0.390*** (0.045)	0.282*** (0.044)	0.016*** (0.002)
day*weekday	0.585*** (0.156)	0.139 (0.150)	0.012** (0.006)
night*weekday	0.444* (0.245)	0.173 (0.235)	0.004 (0.009)
Constant	0.901 (2.346)	5.423** (2.244)	0.094 (0.089)
Observations	201,619	209,974	212,274
Trip FE	YES	YES	YES
Time FE	YES	YES	YES
R^2	0.237	0.091	0.089
Adjusted R^2	0.236	0.090	0.088
Residual Std. Error	14.348 (df = 201355)	14.102 (df = 209710)	0.564 (df = 212010)
F Statistic	237.732*** (df = 263; 201355)	79.500*** (df = 263; 209710)	78.758*** (df = 263; 212010)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2: Output measures regressed over driver and ride characteristics. IV estimates controlled for endogeneity of price and quantity are in Appendix D

First and foremost, minority status has a negative coefficient and is highly statistically significant for all measures of economic outcomes. Second, reputation, measured by the number of reviews, has a positive impact and is highly statistically significant in all regressions. Finally, several other patterns are consistent across all regressions: younger, female drivers receive better economic outcomes. After we control for number of reviews, seniority on the platform has a negative coefficient. Drivers with profiles that include extended descriptions and a picture receive higher outcomes.

3.2 The reputation effect

We have shown that, on average, minority drivers receive lower economic outcomes than non-minority driver. In this section, we investigate how this gap differs with reputation. Reputation can be considered from at least two perspectives: quantity of reviews or quality of reviews.Passengers likely consider both dimensions. However, it is a priori not clear how to compare drivers across these two variables (e.g., is a driver with a single 5-star rating better than someone with two 5s and a 4?). In the reduced form analysis, we focus on the quantity dimension. We integrate both dimensions in the structural model in Section 4.

3.2.1 Cross-section

We divide our sample into three subsamples: listings of drivers with no reputation, i.e., drivers with five or fewer reviews; listings with some reputation (between 6 and 49 reviews); and listings with an established reputation (more than 50 reviews).[10] We estimate standard OLS regressions with the same set of controls as in table 2 for drivers with different levels of experience. The coefficients associated with minority status are presented in table 3; full results are in Appendix F. Minority status has a significant

Table 3: Reputation effect, coefficients of minority status

	Minority coefficients:		
reviews(#)	1:5	6:49	50:450
number of clicks	-1.890^{***} (0.374)	-0.700^{***} (0.204)	0.042 (0.259)
seats sold	-0.031^{***} (0.009)	-0.013^{**} (0.006)	-0.001 (0.006)
revenue	-0.806^{***} (0.251)	-0.486^{***} (0.150)	-0.149 (0.200)

*p<0.1; **p<0.05; ***p<0.01

impact at the beginning of a career. Minority drivers with few reviews (fewer than 5) receive significantly fewer clicks, sell fewer seats and make less revenue. Crucially,

[10]The selection of these thresholds is ad hoc. However, the reputation effect remains for small changes.

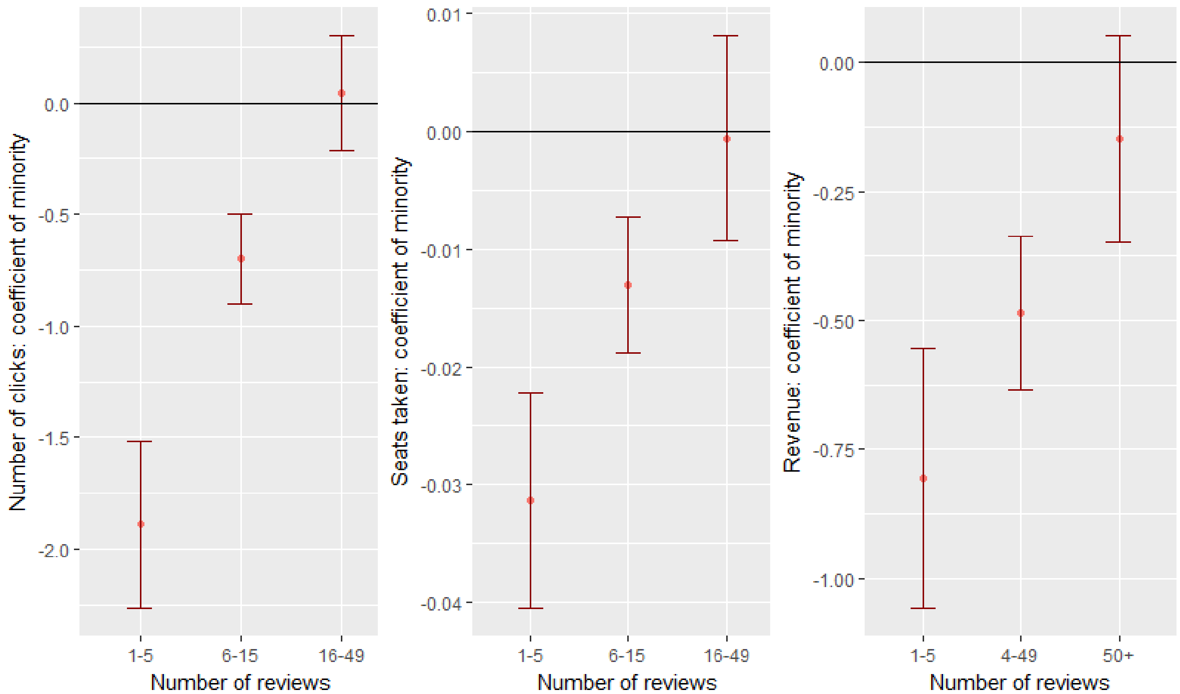


Figure 2: Coefficients of minority status at different stages of experience

this effect decreases relatively fast and, in some cases, once the reputation is built, is overcome entirely. Figure 2 illustrates this result. Furthermore, most of the other driver- or listing-specific characteristics remain unchanged. Finally, discrimination could also be gender specific, and indeed, listings posted by male drivers receive fewer visits. In this case, reputation does not appear to play a crucial role since the effect is only slightly diminished for experienced drivers. However, we do not draw any clear-cut conclusions, as this effect would be sensitive to the choice of output measure and the definition of the cut-offs.[11]

Size of the gap depends on where we draw the line between entrants and experienced drivers. Controlling for other observables, in terms of the number of clicks, the gap is 4% for drivers with one review, 3.99% with five, and 3% with ten. Regarding sold seats, the initial gap is 12.8%, it decreases to 11.58% with five reviews and to 7.53% at ten. The gap appears to be most persistent for revenue: the initial gap is 10.81%, drops to 10.7% with five reviews and to 8.98% with ten reviews.[12] Throughout this paper, we

[11] The quantity of seats has a high number of zeros. In the Appendix D, we also present estimates of a hurdle model.

[12] Calculated based on the results from cross-section controls, as in table 2.

define an entrant driver as a driver with five or fewer reviews.

3.2.2 Panel data analysis

Thousands of drivers are active on BlaBlaCar at any moment; thus, every time we collect data, we observe only a fraction of all available listings. As a consequence, we see most drivers only once. However, in some cases (22.800 drivers), we see the driver at least twice, which gives us a panel with almost 56.800 observations. However, this sample is unbalanced, with drivers being observed between 2 and 30 times. We use several standard models that allow us to compare the gap associated with being a minority entrant or incumbent entrant. Reduction in the sample size results in lower significance of our estimates. However, the signs and point estimates appear to confirm our hypothesis. Generally, we estimate the following model:

$$y_{it} = \alpha + X_{it}\beta + Z_i\gamma + c_i + \epsilon_{it}$$

where i denotes an individual driver; t represents her seniority, defined as the number of reviews; y is the variable of interest (i.e., the number of clicks the listing received, sold seats or the revenue); x_{it} is a vector of time-varying explanatory variables (number of reviews, trip-specific attributes); z_i is the time-invariant explanatory variables (gender, minority status); α is the intercept; c_i are individual fixed effects; and ϵ_{it} is an idiosyncratic error term. Here, we report only the variables of main interest; full results are presented in Appendix H.

	Pooled	**Between**	**Random**
Dependent variable:		*number of clicks*	
minority	0.288 (0.202)	0.409 (0.275)	0.317 (0.236)
minority*entrant	-0.678^{*} (0.353)	-0.692 (0.449)	-0.717^{*} (0.387)
Dependent variable:		*sold seats*	
minority	0.002 (0.009)	0.016 (0.011)	0.002 (0.009)
minority*entrant	-0.035^{**} (0.016)	-0.041^{**} (0.018)	-0.035^{**} (0.016)
Dependent variable:		*revenue*	
minority	-0.334 (0.213)	0.022 (0.275)	-0.272 (0.228)
minority*entrant	-0.680^{*} (0.372)	-0.866^{*} (0.448)	-0.741^{*} (0.387)
Observations	56,760	22,794	56,760

Table 4: Panel data results, entrant is a driver with less than 15 reviews, full results presented in the Appendix H

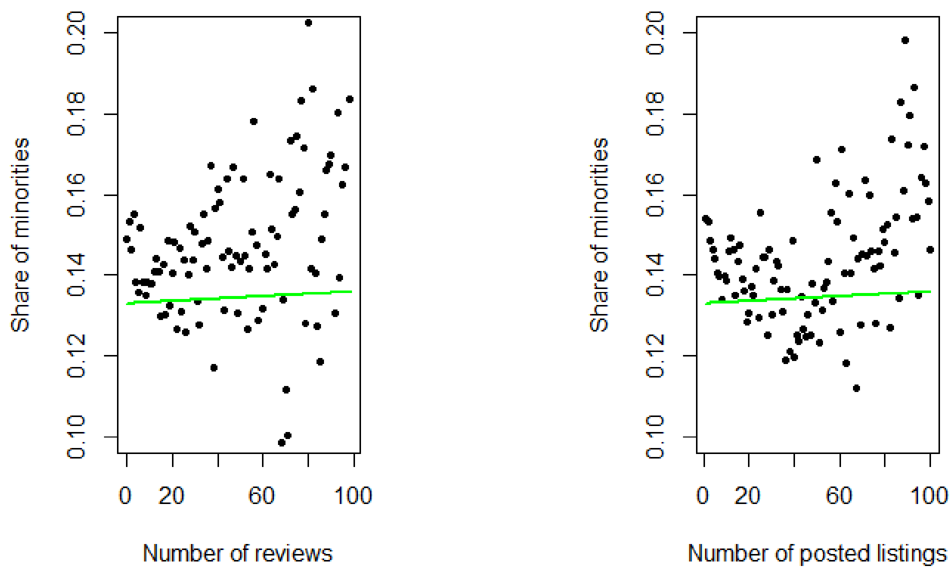
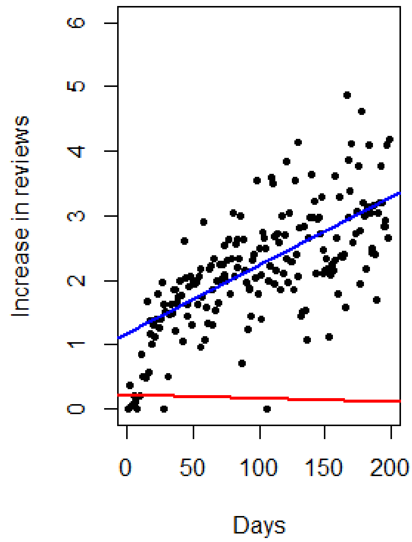


Figure 3: Share of minority drivers at different career levels

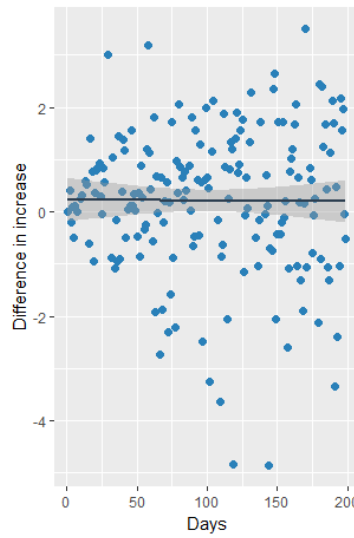
We present minority dummies and the products of minority and entrant dummies. Similarly to the cross-sectional analysis in Section 3.2.1, we conclude that upon entering the market, minority drivers receive lower outcomes; however, this effect disappears as drivers receive reviews. The reputation effect is significant for all measures of economic performance.

3.3 Alternative explanation: selection

An alternative mechanism explaining the observed pattern of economic outcomes is the exit of underperforming minority drivers. If after entering, minority drivers observe that it is harder for them to sell rides, then their churn rate should also be higher. Alternatively, if they are aware upon entering that they need to collect a few good reviews to achieve higher outcomes, they would not exit, but exert effort to gather good reviews. We provide two complementary sources of evidence showing that a change in composition of drivers is not the cause of the narrowing output gap. If minority drivers are facing harder selection, their share in the population of drivers should be a decreasing function of driver seniority. Figure 3 presents the share of minority drivers at different levels of experience, measured either by the number of reviews, panel (a), or by



(a) Average increase



(b) Difference in average increase across ethnicities

Figure 4: Increase in number of reviews over time

the number of posted listings, panel (b). The share is relatively stable, which suggests that selection cannot explain the reputation effect. This reasoning assumes that entry patterns are constant over time. A second way to look at the problem without relying on this assumption is to revisit the profiles of drivers that we have seen in the past and determine whether majority drivers have collected more reviews since the first time we saw them. In April 2018, we randomly selected a subset of almost 13 thousand drivers that we had seen in the past; we loaded their profiles to determine the increase in the number of reviews or posted listings. Figure 4 shows the average results. The observed increase clearly depends on the time elapsed since the first time we visited their profiles. Some of the drivers had been seen only a few weeks, so the increase in the number of reviews is small. In other cases, more than half a year might have elapsed. The left panel shows the difference in gains over time. The horizontal axis is the number of days between the two visits to the driver's profile. The vertical axis is the increase in the number of reviews between the two visits. The blue line shows the trend over time. The red line shows the difference between non-minority and minority drivers. The right panel focuses on the difference between non-minority and minority drivers, and shows that the difference is not statistically significant. We would expect the sign to be positive because of the smaller number of seats sold, and thus the fewer reviews received, by minority entrants. However, we observe a small, insignificant difference that is consistent with the reputation effect playing a role relatively fast. These two arguments lead us to the conclusion that the results shown earlier are not due to a change in the composition of drivers but to the causal impact of reputation building. This is an important observation; it suggests that drivers are aware of the *reputation effect* and know that after a couple of periods of underachievement, their outcomes will improve. Finally, from these results, we can draw several conclusions regarding the cost structure of using BlaBlaCar. While posting rides does not bear a significant cost, there is a potential fixed cost, paid upon entering (installing the app, learning how to use it, setting up a profile). If this is true, there could be some selection before entering. If potential entrants are aware of these market dynamics, on average, better minority drivers will join the platform. Unfortunately, we are not able to test this claim with our current dataset, and we leave it for future research.

3.4 Reduced form evidence of exerting effort to build reputation

Reputation is valuable for users; it allows them to signal their individual quality. It has an additional benefit for minority drivers because it allows them to mitigate discrimination. The more valuable is the reputation, the more effort drivers will exert to establish it. In our dataset, we do not observe efforts directly (we will estimate them with a structural model in the next section), but we have a few variables that can suggest an increased effort. Making detours, taking big luggage and accepting pets is costly for the drivers, but appreciated by the passengers. Table 5 shows that drivers with few reviews are more likely to exert additional efforts and that it is more the case for minority drivers, for whom reputation matters more.

	Dependent variable:		
	detour	luggage	pet
minority	0.669*** (0.039)	−0.008 (0.046)	−0.556*** (0.028)
reviews (#)	−0.003*** (0.0002)	−0.001*** (0.0002)	−0.001*** (0.0002)
minority*reviews	−0.001*** (0.0004)	−0.0004 (0.0005)	−0.004*** (0.0005)
Observations	74,155	74,446	131,679
Log Likelihood	−39,691.410	−25,060.570	−63,486.190
Akaike Inf. Crit.	79,446.820	50,185.140	127,048.400
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: Proxies for effort, selected variables, logistic regressions, full results in Appendix I

All three regression have a binary dependent variable that indicates whether a driver is willing to make a detour for the convenience of a passenger, take particularly big luggage and accept a pet in the car. We see that users with no reviews are more likely to take these efforts, which is consistent with the interpretation of reputation building investments. Furthermore, this effect is more significant for minority drivers. We interpret it by noting that minority drivers value reputation more, and are therefore more likely to make this additional effort.

4 Driver's Incentives Problem: a model of career concerns

The reduced form results show that the outputs achieved by drivers from ethnic minorities are lower than those of non-minority users. A sizable initial gap in performance exists that narrows as reputation is built. This result suggests that the reputation system is crucial for drivers to signal their quality and for passengers to update their beliefs. Nevertheless, the reduced form results do not allow us to separate differences in the intrinsic quality of drivers from the efforts actively taken by them; thus, we are not able to observe potential differences in the quality of service provided by minority drivers due to some form of selection on entry or difference in preferences. In this section, we adapt the canonical model of career concerns of Holmstrom (1999) using ratings as performance measures. Our goal is to decompose the amount of quality of service due to efforts taken by drivers from their intrinsic quality.

We first, discuss the supply side of the market. We present a problem faced by a driver and derive her optimal effort schedule. Using reviews obtained by all drivers and information available on their profiles, we derive a measure of expected quality. Second, we turn to the demand side of the market. Our goal here is to study the usefulness of our measure of expected quality and whether its variation can explain the differences in performance between minority and non-minority drivers. The part that is left unexplained by quality differentials (or by other observables) we attribute to erroneous beliefs. We will exploit regressions presented in the reduced form and develop a simple logit demand system to further reinforce our results.

4.1 Supply side

Theoretical model: Consider a driver operating in a competitive market[13]. In our context, this means that there are more passengers than available seats. The driver, aside from having an empty seat in her car, is endowed with a unit of labor that she wants to sell in exchange for consumption. Selling labor means exerting effort, which is costly, but benefits passengers. A passenger derives utility from the output created by the driver, that is, the sum of the driver's intrinsic quality and exerted effort. No contingent contracts are available; the transaction is made via the centralized system and cannot be conditioned on the quality of service offered by the driver.

Neither in a static game nor in a dynamic game of complete information do drivers exert effort. However, in a game of incomplete information, the driver might exploit uncertainty about her intrinsic quality to increase future payoffs.

Let η_i be a measure of the driver's talent. Suppose that talent is fixed and incompletely known both to the market and to the driver herself. However, some characteristic of the driver (gender, age, ethnicity, etc.) are publicly observed. A combination of these characteristics is indexed by i and denotes the subpopulation to which a driver belongs (for example, young, female, non-minority driver with a long biography, who frequently uses the service). The market (here, the passengers) have an initial belief about the intrinsic quality of driver j from population i because the distribution of types in the population is known. We assume that the quality in population i is distributed normally with mean m_i and precision (inverse of variance) h_i (i.e., $\eta_j \sim N(m_i, h_i)$). The output

[13] In the appendix, we provide an extension that introduces a bargaining game that allows for different levels of bargaining power; importantly, as long as bargaining power differs only across drivers and not in time, all insights remain unchanged.

of driver j from population i in period t is given by:

$$y_{t,i,j} = \eta_j + a_{t,i} + \epsilon_t,\ \ t = 1, 2, ...$$

where $a_{t,i} \in [0, \infty]$ is a measure of exerted effort (input of labor), and ϵ_t is a stochastic noise term, which is assumed to be independent from types and efforts and distributed normally with mean zero and precision h_ϵ. Drivers maximize the discounted sum of consumption:

$$U(c, a) = \sum_{t=1}^{\infty} \beta^{t-1} \left[c_t - g(a_t)\right] \tag{1}$$

where $g(\cdot)$ measures the cost of effort and is assumed to be increasing and convex. The utility function is publicly known. Consumption is a function of payments obtained from passengers and is assumed to be proportional to expected quality.

$$c_{j,t,i} \propto w_{j,t,i}(y^{t-1}) = \mathbf{E}\left[y_{j,t}|y_j^{t-1}\right] = \mathbf{E}\left[\eta_j|y_j^{t-1}\right] + a_{t,i}(y^{t-1})$$

Expected quality is a function of past outputs (y_j^{t-1}) and population characteristics. Given the information structure, passengers can deduce the optimal effort that a driver should take $(a^*_{t,i}(y_j^{t-1}))$; hence, observing $y_{t,i,j}$ is equivalent to observing:

$$z_{t,i,j} \equiv \eta_j + \epsilon_t = y_{t,i,j} - a^*_{t,i}(y^{t-1})$$

This observation allows us to characterize the posterior beliefs about η_j

$$m_{t+1,i,j} = \frac{h_{t,i} m_{t,i,j} + h_\epsilon z_{t,i,j}}{h_{t,i} + h_\epsilon} = \frac{h_i m_i + h_\epsilon \sum_{s=1}^{t} z_{s,i,j}}{h_i - t h_\epsilon}$$

$$h_{t+1,i} = h_{t,i} + h_\epsilon = h_i + t h_\epsilon$$

where $s \in \{1, t\}$ denotes the previously observed outputs. As t goes to infinity (i.e., drivers accumulate reviews), η is fully revealed. Notably, $h_{i,t}$ is the same for every driver from population i at time t because it is a function of the variance of types in the population and the error term. However, $m_{t+1,i,j}$ is specific to an individual since individual performance reports impact the expectations about the drivers' type.

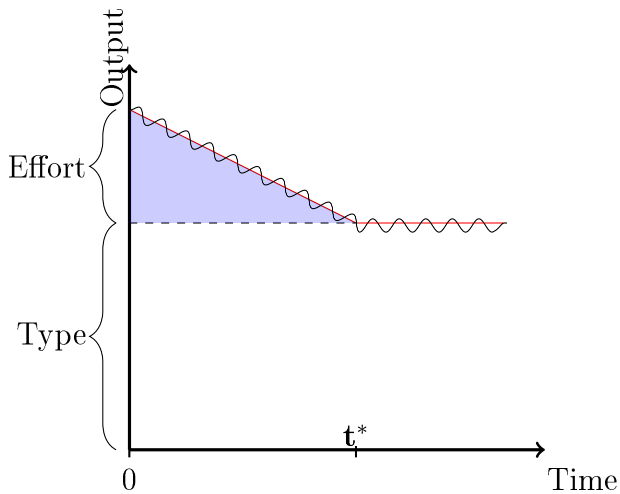


Figure 5: Effort as a function of time

Inserting updated beliefs into the expected output function, we obtain:

$$\mathbf{E}\left[w_{t,j,i}(y^{t-1})\right] = \frac{h_i m_i}{h_{t,i}} + \frac{h_\epsilon}{h_{t,i}} \sum_{s=1}^{t-1} \left(m_i + a_{s,i} - \mathbf{E}\left[a^*_{s,j,i}(y^{s-1})\right]\right) + \mathbf{E}\left[a^*_{t,i}(y^{t-1})\right] \tag{2}$$

Combining equations (1) and (2) and taking the first-order condition, we obtain the following formula for optimal effort:

$$g'(a^*_{t,i}) = \sum_{s=t}^{\infty} \beta^{s-t} \frac{h_\epsilon}{h_{t,i}} \tag{3}$$

The effort schedule is a sequence declining in time that is highest at the beginning of a career and gradually approaches zero. The amount of effort taken is characterized by the degree of uncertainty about the driver's type: the higher the variance in the population (lower precision) is, the more effort a driver has to exert. The precision of the reports plays a crucial role: when reports are imprecise, the updating process is slow and (discounted) returns are lower, which results in less effort being exerted. Perfectly accurate reporting $h_\epsilon = \infty$ results in effort being made in all periods, whereas no precision $h_\epsilon = 0$ results in no effort. Figure 5 depicts the evolution of effort over time.

Description of the data and recovering parameters of the model: Data from the reputation system is characterized by high frequency of performance reports, which

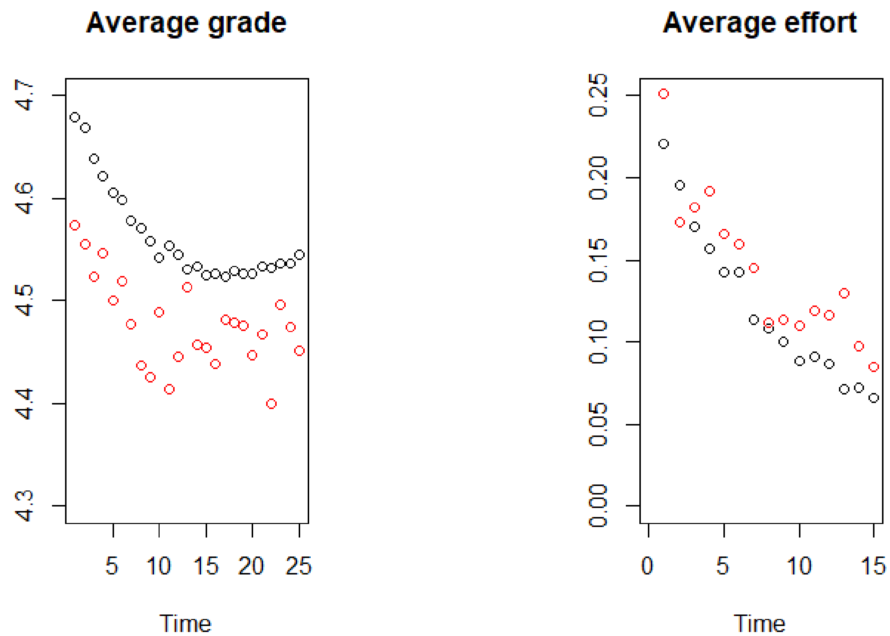


Figure 6: Non-minorities: black dots; minorities: red dots

allows for a precise study of effort dynamics. For all 350.000 drivers in our dataset, we have a full history of ratings obtained from their first ride until the moment we collected the observation. Hence, we can study changes in ratings received depending on the level of experience. After restricting the sample to only drivers who have at least 30 reviews (so that we see all drivers in each stage of their career), we are left with 1.5 mln observations. Figure 6 (left panel) shows the average ratings at different stages of the drivers' careers. The first point on the left side is the average first rating. We see that the ratings are high at the beginning and stabilize at a certain level as more ratings are collected. We also see that ratings obtained by minority drivers are lower at every stage of their careers.[14]

Following the logic of the theoretical model presented previously, we recover the parameters in the following way:

- The intrinsic quality (type) of an individual driver is the average of their ratings after they stabilize (in practice, after the 15th rating).
- The distribution of the error term is normal with mean zero, and the variance is given by the total variance of the ratings after they have stabilized.

[14]They are also noisier, which may be due to fewer observations.

- The effort of an individual driver during their $t-th$ ride is the difference between her type and the $t-th$ rating received.

Figure 6 (right panel) shows the efforts taken by drivers at different points of their career. Effort is indeed highest initially and gradually goes to zero. Additionally, the average effort taken by minorities is higher than that of non-minorities.
In our approach, 15 reviews are sufficient for the market to deduce the true type of any driver. Before 15 reviews are collected, passengers must form expectations of quality based on the publicly available information. When looking at a driver with no reviews, a passenger will judge her quality as the mean quality of the population that this driver comes from. In subsequent interactions, passenger will also take into account past reviews. We choose the following characteristics to define subpopulations: ethnicity status, gender, age, length of bio, and number of posts per month. We categorize the three last values as above or below the mean.[15] We have 32 populations; we assign all the drivers to their populations and calculate the mean and variance of each. Table 6 presents the impact of the characteristics on the mean and variance of the types of a population.

In our dataset, ratings are densely clustered around 4.5, hence so are our types.

Dependent variable	mean	variance
minority	−0.074	0.011
driver's age	−0.034	0.005
male	−0.008	0.011
bio (# words)	0.024	−0.004
posts per month	0.018	0.010
Constant	4.544	0.059
Observations	780,340	780,340

Table 6: Mean and variance of the types of populations, impact of driver's characteristics

Nevertheless, the impact of the characteristics that we consider on the distribution of types is significant. Minority, young, male drivers with a short bio who do not post a large number of rides have a low expected intrinsic quality, whereas minority, old, male drivers with a short bio who post many rides have a high variance of types.

[15]The choice of the covariates and cutoffs is ad hoc and was driven by the size of all groups and ease of interpretation.

Estimation of the cost of effort function: Having identified the parameters of the model, we are ready to estimate the effort function. Recall that in the theoretical model, this relationship is characterized as:

$$a^*_{t,i} = g^{-1,'}\left(\sum_{s=t}^{\infty} \beta^{s-t} \frac{h_\epsilon}{h_{t,i}}\right)$$

The optimal effort has been directly recovered from the observation of ratings. h_ϵ is the inverse of the variance of the error term, also calculated from the data, and $h_t = h_i + th_\epsilon$ is a combination of the precision in population i and the error term. We estimate this function with two models: first, assuming that $g(\cdot)$ is a quadratic function, and second, approximating this function with a polynomial.[16] This theoretical model uses an infinite horizon decision. For computational reasons, we estimate several models with different bounded horizons and show that the cost of effort and the discount factor converge to a specific level. We estimate the following equation:

$$a_{i,t,j} = \gamma\left(\sum_{s=t}^{n} \beta^{s-t} \frac{h_\epsilon}{h_{i,t,j}}\right) + \varepsilon_{t,j,i}$$

where $a_{i,t,j}$ is the effort of driver j from population i at time t. The polynomial regression takes the following form:

$$a_{i,t,j} = \sum_{z=1}^{\omega}\left\{\gamma_k^z\left(\sum_{s=t}^{n} \beta^{s-t} \frac{h_\epsilon}{h_{i,t,j}}\right)\right\} + \varepsilon_{t,j,i}$$

where z measures the degree of the polynomial and goes from 1 to ω. After the 5th-degree, the improvement in fit is small, and the estimation becomes computationally demanding. This is a nonlinear regression; we use the package NLS in R with the algorithm port. Figure 25(a) presents the estimates. The blue line shows the cost of effort γ, and the discount factor is in red. The circles are the estimates of a quadratic function, and the diamonds are the estimates of the polynomial function. The estimates stabilize when we consider a horizon of more than 15 periods. In practice, we use the estimates of the 30-period horizon.

Figure 25(b) compares the predictions of the two models. The horizontal axis shows

[16] We are currently working on a nonparametric estimation using local linear method improvements of the nonparametric model. Our current procedure is computationally complex, and we are unable to use the whole dataset.

the career stage of the driver, and effort is on the vertical axis. Black, red, and blue dots correspond to the quadratic model, polynomial model, and the data. We conclude that the two models give similar predictions, but the polynomial model offers a slightly better fit.

The expected quality: Recall that the output is described by the following equation:

$$y_{t,i,j} = \eta_j + a_{t,i} + \epsilon_t,\ \ t = 1, 2, ...$$

Using the estimates presented above, we are now in position to calculate the expected output of a driver. Equation (4) describes the measure of expected quality used in the rest of this section.

$$\mathbf{E}\left[w_{t,j,i}(y^{t-1})\right] = \frac{h_i m_i}{h_{t,i}} + \frac{h_\epsilon}{h_{t,i}} \sum_{s=1}^{t-1} \left(m_i + a_{s,i} - \mathbf{E}\left[a^*_{s,i}(y^{s-1})\right]\right) + \mathbf{E}\left[a^*_{t,i}(y^{t-1})\right] \tag{4}$$

By observing the output, passengers learn drivers' types: by seeing a report of y_t and knowing the optimal effort that the driver should have taken at time t, prospective passengers can deduce $\eta + \epsilon_t$. With the history of all ratings and the estimates of the optimal effort for all drivers at every stage of their careers, we can directly obtain equation (5), which is equivalent to (4).

$$w_{t,j,i}(y^{t-1}) = \frac{h_i m_i}{h_i + th_\epsilon} + \frac{h_\epsilon}{h_i + th_\epsilon} \sum_{s=1}^{t} \left(y_{s,j,i} - a^*_{s,j,i}(y^{s-1})\right) + a^*_{t,j,i}(y^{t-1}) \tag{5}$$

We compare the predicted efforts from the polynomial and quadratic functions and see that the polynomial regression offers a slightly better fit to the data. Hence, for the rest of this paper we focus on the polynomial model. Figure 8 presents a histogram of expected quality.

4.2 The demand side: Do differences in expected quality explain the output gap?

The quality measure derived above represents a passenger's best guess about the next performance of any given driver. We calculate the quality for all drivers in our sample based on the population characteristics and past reviews. Therefore, the quality measure provides additional information to that embedded in the covariates used in

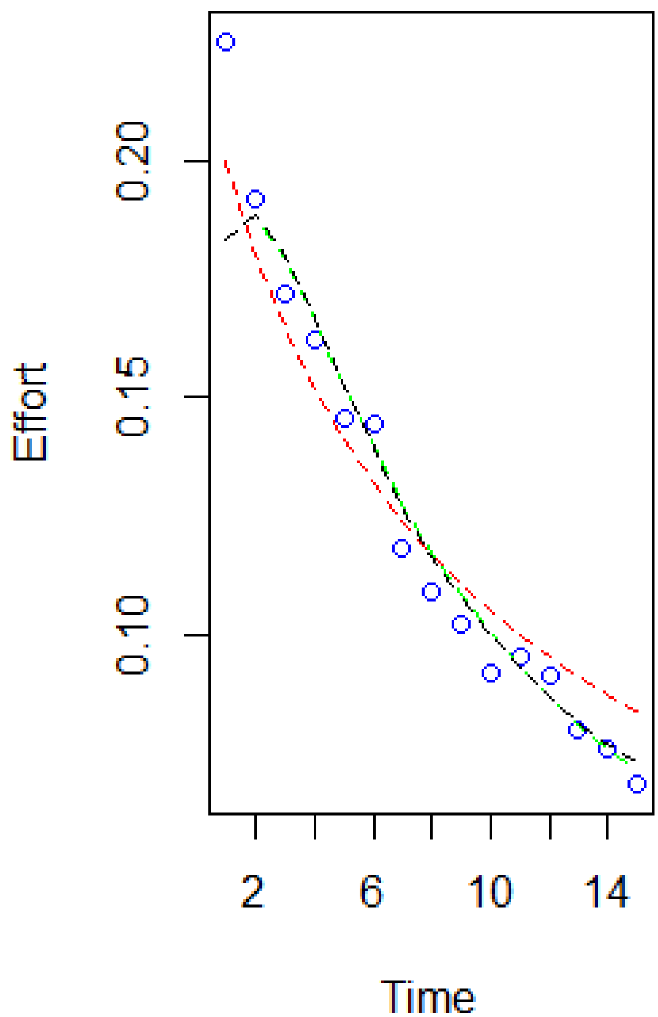


Figure 7: Predicted efforts: red, quadratic; green, polynomials up to the 4th degree; black, polynomials up to the 5th degree; blue, data

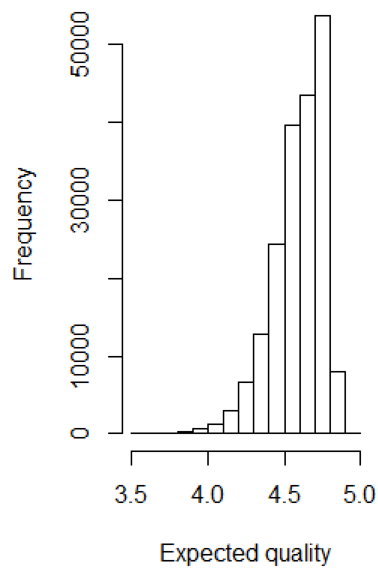


Figure 8: Histogram expected quality

the reduced form regressions. We are interested in analyzing whether this new measure is a good predictor of demand and if variations in this parameter can help to explain the minority gap. The first two moments of the distribution play a role in forming expectations. The part of the gap that is more persistent can be explained by the fact that, on average, minority drivers receive lower ratings, which suggests they have lower expected quality. The part of the gap that we cannot explain by differences in quality or by any other covariate, and that narrows over time, we attribute to erroneous beliefs of passengers about the quality of service of minority drivers. These beliefs are updated as reviews are collected until the true type is established. In this section, we develop a model of demand to study the impact of expected quality in greater detail. Our measure of quality is also a strong predictor in reduced form models, and we can draw analogous conclusions from models with revenue and sold seats as dependent variables. These models are presented in the Appendix.

Simple demand system: *work in progress*

Our dataset is constructed by conducting a search for a given route and checking all the available listings; thus, we have the entire choice set available to a passenger, together with the output measures for all available drivers. Therefore, we can calculate market share, which we use to estimate a logit demand system with imperfectly informed consumers. Equipped with this information, we repeat the exercise from the previous paragraphs. We define the market on a route and day basis, which gives us many markets (more than 40.000) with few drivers per market. Consumer i receives utility from choosing driver j in market m at time t, given by:

$$u_{i,j,t} = \alpha w_{j,t} + \mathbf{X_{j,t}}\beta + \gamma p_{j,t} + \xi_j + \epsilon_{i,j,t}$$

where $w_{j,t}$ is our measure of expected quality, $\mathbf{X}$ is a vector of covariates, ξ_j is a driver characteristic unobserved by the econometrician, and $p_{j,t}$ is the price. The error term is assumed to be iid extreme value type 1. In this simple model, we assume that passengers have the same valuations for characteristics and prices. The mean utility is denoted by:

$$\delta_{j,t} = \alpha w_{j,t} + \mathbf{X_{j,t}}\beta + \gamma p_{j,t} + \xi_j$$

Following Berry (1994), we can equate $\delta_{j,t}$ with $log(\frac{s_{j,t}}{s_{0,m(j),t}})$, where $s_{j,t}$ is the market share of driver j in time period t. We define the market share as $\frac{q_{j,t}}{M_{m(j),t}}$, where $q_{j,t}$ is the number of seats driver j sold in period t on a given route, and $M_{m(j),t}$ is the number of potential buyers in market j at time t. As a proxy for market size, we use the maximum number of seats ever sold on the route. The term $s_{0,m(j),t} \equiv \left(1 - \frac{\sum_{j \in M} s_{j,t}}{M_{m(j),t}}\right)$ is the share of consumers choosing the "outside option" of not traveling on a given day with any of the drivers.[17]

To address the standard problem of the endogeneity of price, we employ two instrumenting strategies. First, we have a number of cost-shifters. Over time, the price of gas changes, and we can observe the average price at gas stations in any given city on any given day (https://www.prix-carburants.gouv.fr). These prices will change over time (because of general oil price fluctuations) and geographically (e.g., due to varying intensities of competition). Additionally, the level of highway tolls varies across routes. Moreover, since we observe the characteristics of all drivers available in a given market,

[17] There are many small markets in our dataset; we have approximately 40000 markets, with fewer than 5 drivers per market. Moreover, we have many drivers who have not sold any of their offered seats at the time of observation; hence, we observe a large number of zero market shares. As noted by Gandhi et al. (2013), a typical "fix" in such a case is to add a small ϵ to all market shares or drop observations with zero market share, which effectively lumps them with the outside option. Unfortunately, both methods lead to biased estimates. In the present report, we add ϵ to the market shares of all drivers and acknowledge the problem. We are working on solving this problem with a subsample of rides observed after completion.

we can derive measures of isolation in the spirit of Berry (1994) or the BLP model, which were defined earlier.

Empirical results: Table 7 presents estimates using expected quality as a covariate.[18] In columns 1 and 2, we first estimate reduced form models using our measure of quality, which is positive and highly significant. Columns 3-5 show the logit demand regressions with quality measured using the quadratic cost function (column 3) and with the polynomial approximation (columns 4 and 5). Price in regressions 1 to 5 is instrumented using measures of isolation in characteristic space and cost shifters; time fixed effects are not presented. All the results indicate that our quality measure is rel-

	Dependent variable:				
	taken seats	revenue	mean utility		
quality	0.038*** (0.011)	1.535*** (0.272)	0.169** (0.086)		
quality I(4)				0.174** (0.086)	
quality I(5)					0.171** (0.086)
price	−0.012*** (0.004)		−0.119*** (0.026)	−0.119*** (0.026)	−0.118*** (0.026)
reviews	0.006*** (0.001)	0.177*** (0.020)	0.036*** (0.006)	0.036*** (0.006)	0.036*** (0.006)
reviews 2	−0.0001*** (0.00003)	−0.003*** (0.001)	−0.001*** (0.0002)	−0.001*** (0.0002)	−0.001*** (0.0002)
hours until ride	−0.001*** (0.00001)	−0.013*** (0.0003)	−0.004*** (0.0001)	−0.004*** (0.0001)	−0.004*** (0.0001)
posted since	0.010*** (0.0003)	0.246*** (0.007)	0.087*** (0.002)	0.087*** (0.002)	0.087*** (0.002)
car price	−0.0002 (0.0003)	−0.018** (0.009)			
competition	0.001*** (0.0001)	0.019*** (0.001)	0.028*** (0.0004)	0.028*** (0.0004)	0.028*** (0.0004)
median revenue	0.00001*** (0.00000)	0.0002*** (0.00002)	0.00004*** (0.00001)	0.00004*** (0.00001)	0.00004*** (0.00001)
public transport ratio	−0.287 (0.214)	−29.228*** (5.582)	11.423*** (1.627)	11.392*** (1.627)	11.395*** (1.627)
km	0.001*** (0.0003)	0.007*** (0.0003)	0.010*** (0.002)	0.010*** (0.002)	0.010*** (0.002)
day	0.014** (0.006)	0.440*** (0.163)	0.031 (0.048)	0.031 (0.048)	0.031 (0.048)
night	−0.043*** (0.010)	−1.266*** (0.262)	−0.393*** (0.076)	−0.392*** (0.076)	−0.393*** (0.076)
train strike	0.101*** (0.008)	2.193*** (0.218)	1.078*** (0.063)	1.077*** (0.063)	1.078*** (0.063)
ride description (# words)	0.0002*** (0.00005)	0.007*** (0.001)	0.002*** (0.0004)	0.002*** (0.0004)	0.002*** (0.0004)
picture	0.009 (0.005)	0.261* (0.141)			
automatic acceptance	0.076*** (0.007)	2.142*** (0.086)	0.467*** (0.050)	0.469*** (0.050)	0.469*** (0.050)
weekday	−0.026*** (0.007)	−0.496*** (0.173)	−0.223*** (0.051)	−0.224*** (0.051)	−0.223*** (0.051)
consumption (fuel)	0.014*** (0.002)	0.271*** (0.059)			
talkative	−0.002 (0.003)	−0.018 (0.089)	−0.025 (0.026)	−0.026 (0.026)	−0.027 (0.026)
day*weekday	0.001 (0.008)	0.013 (0.203)	0.094 (0.059)	0.095 (0.059)	0.095 (0.059)
night*weekday	−0.006 (0.013)	−0.018 (0.328)	0.094 (0.095)	0.095 (0.095)	0.095 (0.095)
Constant	−0.123** (0.059)	−10.666*** (1.428)	−11.799*** (0.462)	−11.828*** (0.462)	−11.813*** (0.463)
Observations	104,505	104,899	118,785	118,778	118,777
Time FE	YES	YES	YES	YES	YES
R^2	0.065	0.062	0.117	0.117	0.117
Adjusted R^2	0.065	0.061	0.116	0.116	0.116
Residual Std. Error	0.513 (df = 104471)	13.443 (df = 104866)	4.156 (df = 118755)	4.156 (df = 118748)	4.156 (df = 118747)
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 7: Mean utility, regressed over driver and ride characteristics

evant and is a strong predictor of the different measures of demand. For columns 3-5,

[18] The logit demand model without the quality measure but with ethnic status instead is in the Appendix, where we show that minority drivers experience a performance gap in terms of expected market share.

the elasticity of demand with respect to price varies from 3.38% to 3.41% depending on the model.[19] The elasticity of our quality measure is 0.77%.

The patterns of performance and reduced form regressions suggest that our measure of expected quality should explain the persistent part of the minority gap, which is due to factual differences in quality, as measured by reviews. However, our expected quality measure should fail to account for the part of the gap that changes. Therefore, upon entry to the market, we expect minority drivers to underperform compared to their expected quality. Table 8 presents the predicted market shares and actual performance for minority entrants. Upon entry, minority drivers underperform compared to their

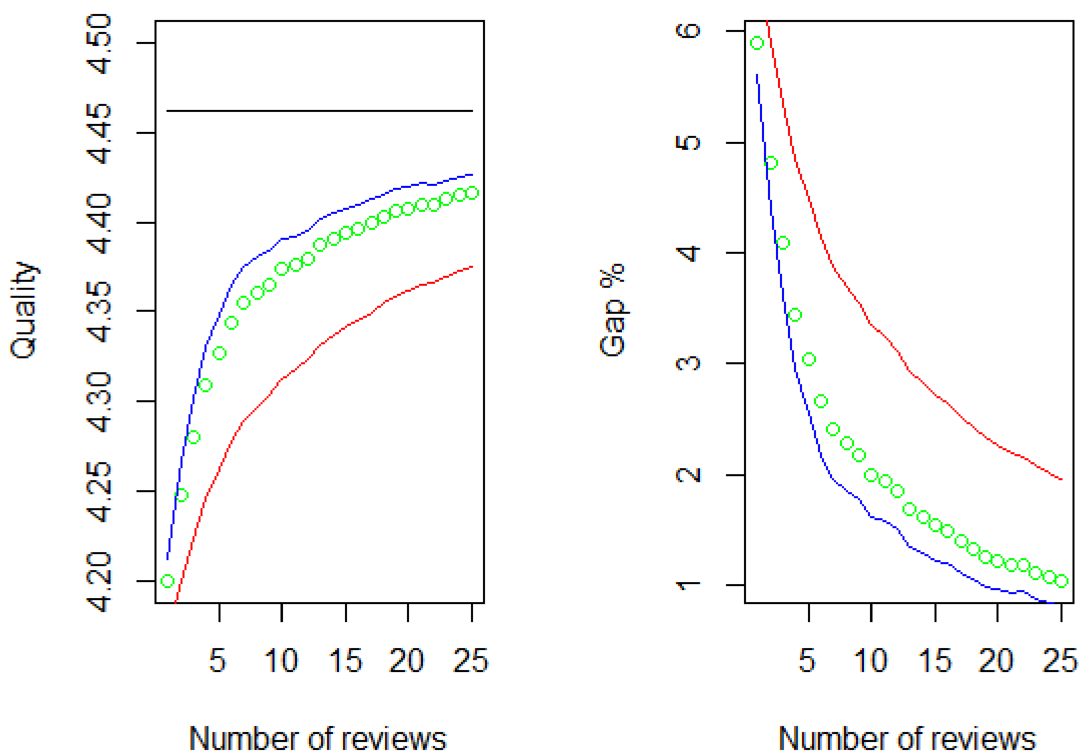
	Predicted and actual market shares	
	Mean	Mean error
quality	0.0155	-5.58%
quality I(3)	0.0153	-4.68%
quality I(5)	0.0153	-4.64%
market shares	0.0146	

Table 8: Minority entrants underperform upon entry

factual quality by approximately 4.6%-5.6%. If we attribute the entire gap to beliefs about expected quality, we can recover the market's belief about expected quality and the initial distribution of types in the population of minorities. The intrinsic quality in the population of minority drivers has a mean of 4.462 and is distributed between 4.408 and 4.753, whereas the quality distribution that would justify the gap in market shares is distributed between 4.076 and 4.42, with a mean of 4.130. Given the high concentration of types in our sample, these values represent a large difference. In fact, with the estimated measure of quality and informativeness of the reputation system (ϵ), we can simulate how long, on average, it takes for minority drivers to signal their true quality.

We now set the initial beliefs about expected type as implied by the actual market shares of minority entrants, and we update these beliefs based on reviews. We assume that a driver receives an average review for a minority driver at the given stage of their career. The green dots show the updating for a population with the mean variance, blue represents a population with the lowest variance, and red represents a population with the highest variance. The horizontal line shows the true quality. Panel

[19] $\eta_j = -\alpha p(1 - s_j)$



(a) Updating beliefs

(b) Part of the gap that remains

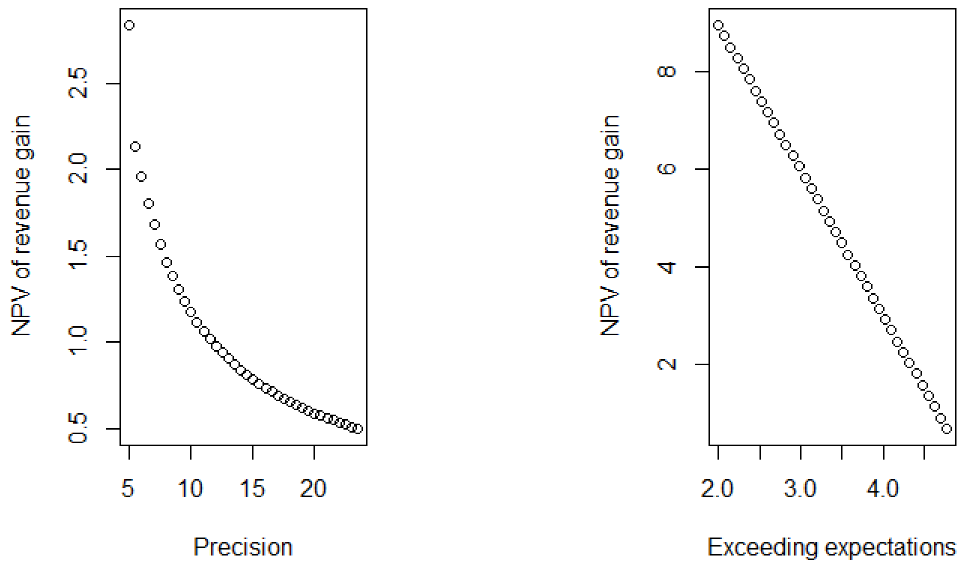
Figure 9: Time required to update beliefs. The green dots, blue line and red line show the updating of populations with the mean variance, lowest variance and highest variance, respectively

(a) compares the belief about the quality of a driver with the belief estimated by our model. Panel (b) shows how the updating process closes the gap. We observe that the first reviews are the most important in correcting beliefs. However, the large differences in the variances of types across our populations imply that for some drivers, considerably more time is required to catch up. Finally, panel (a) confirms the pattern identified in the reduced form regressions. Even if the expected quality of drivers across ethnic status is the same, initial erroneous beliefs about the quality of minority drivers plays an important role, despite decreasing as the number of collected reviews increases.

We are now equipped to calculate the value that different drivers derive from a good review. Consider equation (6), which describes predicted quality.

$$q_{t,j,i}(y^{t-1}) = \frac{h_i m_i}{h_i + t h_\epsilon} + \frac{h_\epsilon}{h_i + t h_\epsilon} \sum_{s=1}^{t} \left(y_{s,j,i} - a^*_{s,j,i}(y^{s-1}) \right) + a^*_{t,j,i}(y^{t-1}) \tag{6}$$

A good review will have a different impact on drivers depending on the level of uncertainty about their type and the expected rating. Suppose that a driver receives a 5-star review instead of the expected value as a first rating. Figure 10 provides some comparative statistics. Both panels present the discounted sum of gains in revenue due to a 5-star review. Panel (a) fixes the expected rating (at 4.74, which is the average expected rating) and varies the level of precision (h_i) between values characterizing different populations in our dataset. We see that the differences in gains are large. Panel (b) fixes the precision at the average level and varies the expected rating. We observe that a driver whose expected rating is 2 exceeds expectations the most and has the highest gain, whereas a driver with an expected rating of 5 meets her expectations and her revenues are not affected. Non-minority drivers have both higher precision and higher expectation; therefore, their gain due to a good review is relatively low. On the basis of the discount factor from the estimation of the cost of effort (0.919), the discounted sum of gains from 30 future rides adds up to 53 cents, whereas for a typical minority driver, the sum of gains is 2.5 eur, which is almost 5 times higher. Furthermore, if we set the initial beliefs (and hence expectations) to the level implied by the performance estimated by the demand model, this gain is 3.93 eur.



(a) Differences in h_i

(b) Differences in expected rating

Figure 10: Value of a good review

5 No reputation system: a counterfactual analysis *work in progress*

We have shown that the reputation system allows individual drivers to signal their quality and increase revenue. However, the reputation system forces all drivers to exert costly effort, even when the resulting signal is low. In this section, we study the impact of the reputation system on entire populations of drivers. We need to estimate a counterfactual revenue when drivers do not exert effort. Two effects must be weighed against each other. First, due to the efforts taken by drivers, there is a higher aggregate demand targeted at entrants. Second, effort is costly, so whether drivers, on average, actually benefit from the reputation system is unclear. We compare revenues with and without a reputation system for different populations. Next, non-minority drivers enter the platform with no prejudice, which this means their perceived quality corresponds to the actual quality. However, minority drivers start with overly pessimistic beliefs. To account for this difference, we add a third scenario, in which drivers are not able to signal high quality and are stuck with the same level of revenue as they achieve in their first interaction. We then discuss the biggest winners and losers. Finally, we use the logit demand system, described in section 4.2, to calculate the change in consumer surplus due to the effort exerted by drivers.

Aggregate net effect of reputation system: With the reputation system, some drivers will be shown to be of higher (lower) quality than others, which will lead to their individual gain. However, as we assume passengers to be risk neutral, no aggregate gains are obtained from the fact that passengers now know the quality of each driver. Therefore, any expansion in demand must be due to effort, which is are non-negligible. Small differences in average quality exist between populations: the best has a mean of 4.75 while the worst has a mean of 4.41. Hence, average efforts of 0.11 (0.20 on average in the first period) can lead to a significant increase in demand. On the basis of estimates of the elasticity of demand with respect to quality for the logit demand presented in the previous section, we expect demand for entrants (no reviews) to increase by 12.5-18.94% due to exerted effort. The net present value of revenues when drivers expend effort ranges from 46EUR to 62EUR based on the discount factor estimated with a quadratic specification in Section 4.1.[20] As a counterfactual, we predict the revenue achieved by drivers when they do not expend effort, that is, the perceived quality is the mean quality in a given population. In such a scenario, the population with the lowest mean should expect a discounted sum of revenues of 32EUR, whereas that for the population with the highest mean should be 62EUR. Figure 11 shows histograms of these revenues.

[20] The discount factor estimated with the polynomial regression leads to higher net present value of revenues; nevertheless, as we use the same discount factor for all revenues/profits, the relations between them are maintained.

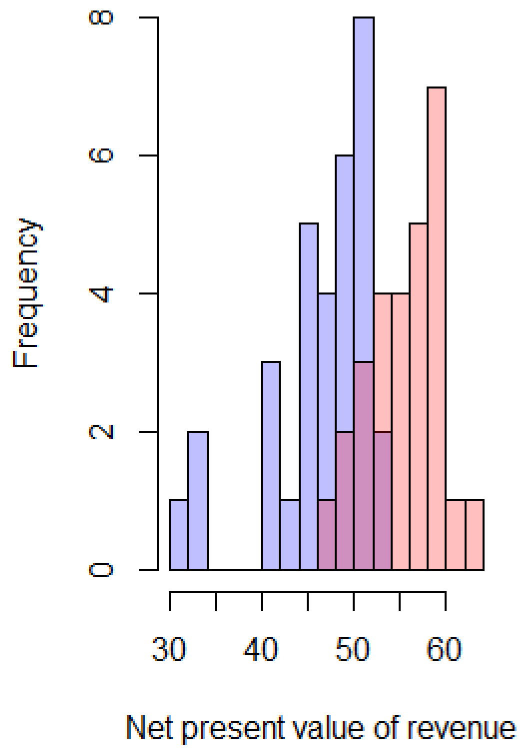


Figure 11: Histogram in blue: discounted sum of revenues with no effort; pink: with effort

Effort leads to significantly higher revenues. The magnitude is related to the fact that most of the effort occurs at the beginning of a career, which means quick returns are expected. However, effort is costly. The per period profit of a driver is given by:

$$\pi_{i,t} = R_{i,t}(m_t + a_t) - g(a_t)$$

where $R_{i,t}(\cdot)$ is the expected revenue of a driver from population i at time t, which is a function of her expected type m_t and the effort taken. The arguments presented in Section 3.3 (selection) indicate that once a driver has entered the market, differences in the level of her revenue are unlikely to influence the decision to exit. This suggests that there are costs related to entering the platform (creating a profile, learning how to use the service, etc.), but there are no significant costs related to continued use of the platform, other than exerting effort. Therefore, we can assume that the following inequality determines the decision to enter the platform:

$$\Pi_i = \sum_{t=1}^{\infty} \pi_{i,t}\delta^{t-1} - F \geq 0$$

where F is the cost of entry, which is assumed to be identical for all populations. To estimate F and calculate the profits of drivers, we impose a zero-profit entry condition on the lowest type of drivers. To determine which population achieves the lowest profit, we need to compare the total expected revenue and the amount of required effort. This ratio is the lowest for the population of older, male, minority drivers who have a long bio and do not post frequently. We set the profit of this population to zero, which gives us the cost of a unit of effort of 36EUR. We repeat the same procedure in the situation where there is no reputation system and compare the profits. We set the profits of the lowest population under both regimes to zero, which results in an important limitation. If the difference between the total revenue increase due to effort and the cost of effort is significantly positive, then some populations might not be present when no reputation system exists; otherwise, new populations might appear if the difference is significantly negative.

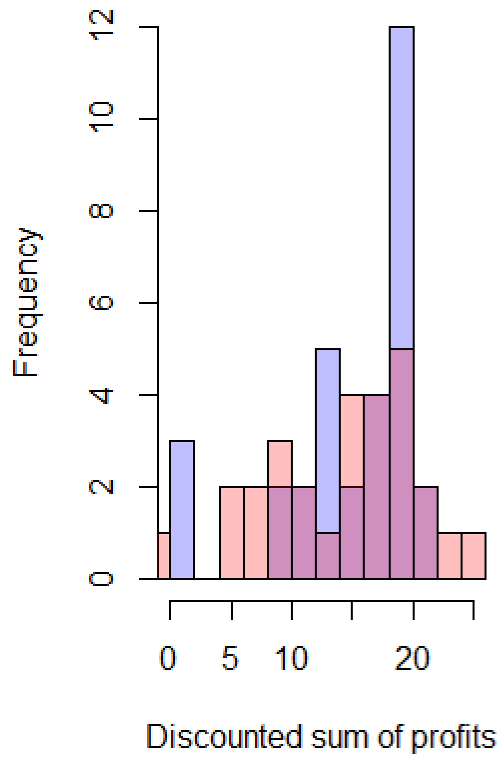


Figure 12: Histogram in blue: discounted sum of revenues with no reputation system; in pink: with reputation system

Figure 12 shows the histograms of profits. Despite the significant increase in revenue, drivers are worse off when there is a reputation system; however, the difference is small, i.e., less than 1EUR on average (mean profit with reputation is 14.28). This difference is due to the costly effort required. Finally, and more importantly, a sizable increase in the dispersion of profits is observed. In the no-reputation-system regime, the profit of a driver at the 25th quantile of the population is 12.56 EUR, whereas in the reputation system, the profit is 8.86EUR. Moreover, with the reputation system, the best population receives, on average, 24.48EUR, whereas in the no reputation system, the best population receives 21.32EUR. These differences are due to the significant variation in the amount of effort expected from different populations. In table 9, the first regression shows which drivers are winners/losers in the reputation system. Drivers belonging to the most homogeneous populations gain the most from the reputation

	Dependent variable: change in profits	
	rational beliefs	real beliefs
	(1)	(2)
minority	-3.643^{***} (0.018)	2.505^{***} (0.043)
male gender	-2.702^{***} (0.014)	-3.806^{***} (0.033)
driver's age	-0.045^{***} (0.0005)	-0.243^{***} (0.001)
posts per month	-0.295^{***} (0.003)	-1.303^{***} (0.007)
bio (#words)	0.018^{***} (0.0004)	0.066^{***} (0.001)
Constant	3.344^{***} (0.021)	22.127^{***} (0.050)
Observations	195,252	195,252
R^2	0.380	0.358
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 9: Winners and loser in the reputation system

system because they do not need to exert effort to signal their high quality. Non-minority, young, female drivers who have a lengthy biography but do not post a large number of listings are the largest beneficiaries of the reputation system.

An important point raised in the previous section is that the part of the output gap that cannot be explained by differences in expected quality is associated with erroneous beliefs. Therefore, the expected revenue based on actual quality is not a valid point of comparison for minority drivers because, absent the reputation system, they are stuck with the prejudiced belief about their quality. To calculate the counterfactual profits when no reputation system exists and drivers are stuck with the initial, erroneous beliefs, we calculate the mean revenue per population from the first ride posted on BlaBlaCar and assume that drivers continue to receive the same revenue. We use the same fixed cost of entry as in the previous analysis, i.e., the minimum average (per population) profit when no effort is exerted[21]. Figure 13 compares histograms of the discounted sum of profits per population under rational and "erroneous" beliefs. In blue, we see the same profits as in the previous figure, i.e., the predicted sum of revenues based on rational beliefs with no reputation system. Revenues based on "erroneous" beliefs are shown in pink.

[21] This is likely a lower bound on the difference because without a reputation system, drivers would not exert effort.

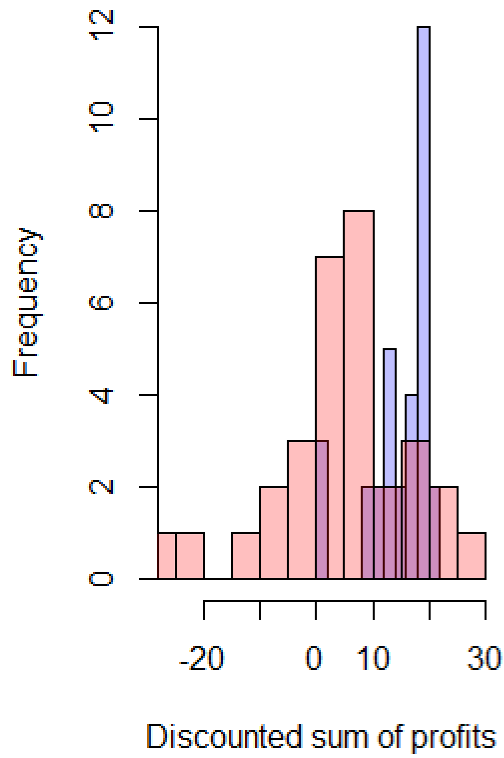


Figure 13: Histogram in blue: discounted sum of profits with rational beliefs; in pink: initial beliefs

When drivers are stuck with the initial beliefs, some populations do not break even. Hence, the benefits of the reputation system do not only consist of increased revenue due to effort but also the correction of beliefs. The second regression in table 9 shows who gains from the reputation system when both factors are considered. In this case, minorities are indeed the largest beneficiaries of the reputation system because they start with overly pessimistic beliefs about their quality and are given an opportunity to change the erroneous beliefs.

Consumer welfare: The reputation system also has an impact on consumer surplus: effort exerted by drivers when they enter immediately benefits passengers. To investigate this source of increase in consumer surplus, we estimate logit demand systems from the previous section. We compare consumers' surplus in a regime with a reputation system with that in a counterfactual system where drivers do not exert effort. The consumer surplus of passenger j in market m is defined by the following equation:

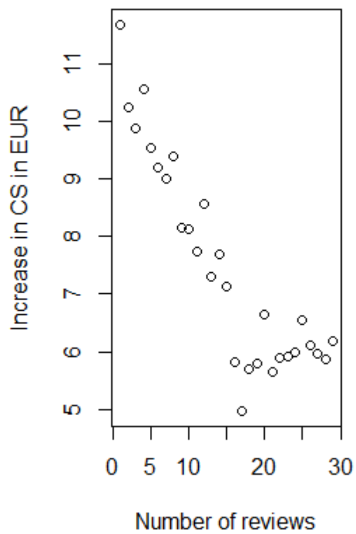
$$\mathbf{E}[CS_{j,m}] = -\frac{1}{\alpha}\log\left(\sum_{j\in m}\exp(\delta_m)\right) + C$$

here α is the coefficient of price in the demand system, and δ_m is the mean utility in market m. C is a constant that is not identified but is differentiated out when making the comparisons. The change in consumer surplus is defined as:

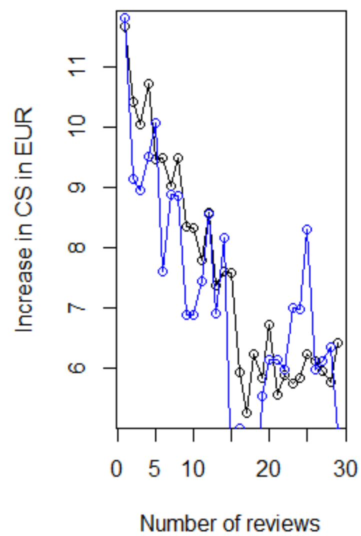
$$\Delta\mathbf{E}[CS] = \mathbf{E}[CS^R] - \mathbf{E}[CS^{NR}] = -\frac{1}{\alpha}\left(\sum_{j\in m}\exp(\delta_m^R) - \sum_{j\in m}\exp(\delta_m^{NR})\right)$$

where R stands for surplus under the reputation system, and NR is the surplus without a reputation system. We assume that the prices in both cases are the same; hence, the only change is the absence of effort when no reputation system is implemented. Thus, we calculate the value, in euros, of the additional quality provided to passengers. The mean additional value is 7.5EUR, which given the average price per ride of 28.8EUR, is a significant increase for consumers. This value also changes over time, as entrants provide the highest effort immediately upon entry. Figure 25(a) shows the average increase in consumer surplus over time, and panel 25(b) introduces a distinction between minority and non-minority drivers. A sizable gain for passengers is observed due to efforts taken by drivers; however, no significant differences exist between ethnic groups.

We conclude that the reputation system has two main impacts on aggregate welfare. On the one hand, the reputation system serves as a mechanism to elicit additional effort from drivers who want to signal high quality. On the other hand, it can counter prevailing erroneous beliefs by revealing the true qualities of drivers. We show that passengers benefit from the presence of the reputation system and that minority drivers, even though they must exert the most effort, are likely to be the greatest beneficiaries of the reputation system because they are able induce prospective passengers to update their initial pessimistic beliefs.



(a) Mean change



(b) Black lines: non-minority drivers; blue lines: minority drivers

Figure 14: Change in consumer surplus due to drivers' efforts

6 Conclusion

While minority users have been documented to face discrimination in a number of online marketplaces, the role of reputation systems in overcoming discrimination has been less studied. Our empirical analysis uses unique data on listings on a popular online car-pooling platform to show that minority users achieve lower economic outcomes. Their listings are less popular, they sell fewer seats, and they receive lower revenue. However, this effect is concentrated during the first interactions on the platform. Building a reputation helps minority drivers to narrow the gap. A minority driver with several reviews receives similar economic outcomes as a non-minority driver. We obtain this result via various econometric techniques and show that these market effects are not due to underperforming minority drivers exiting the market. We also provide a study of career dynamics, enabling us to separate statistical discrimination from erroneous beliefs, and find that statistical updating of rationally held beliefs can explain only part of the gap.

Passengers are willing to change their minds about minority drivers when they see reviews. This observation highlights the importance of a well-designed reputation system for revealing information and indicates that the early stages of reputation building are particularly important. A platform aiming to alleviate discrimination should, therefore, concentrate its efforts on the quality of the reputation system during the first interactions.

In future research, we will aim to further exploit our dataset. Preliminary results show a significant degree of social homophily in the data. Indeed, passengers tend to give higher ratings to drivers from the same population. While this factor is unlikely to alter the insights of the present paper, measuring homophily could, in itself, be an interesting research project. Additionally, we are aware that more recent reviews are more prominent than older reviews. We believe this may encourage drivers to continue exerting effort, even if they have already accumulated many reviews. The impact of the prominence of recent reviews may, therefore, require further investigation.

As noted previously, this is an ongoing research project, and the results presented in this paper may be modified in the future.

References

Abadie, A. and Imbens, G. W. (2016). Matching on the Estimated Propensity Score. *Econometrica*.

Abrahao, B., Parigi, P., Gupta, A., and Cook, K. S. (2017). Reputation offsets trust judgments based on social biases among airbnb users. *Proceedings of the National Academy of Sciences*, 114(37):9848–9853.

Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). Ethnic inequality. *Journal of Political Economy*, 124(2):428–488.

Altonji, J. G. and Pierret, C. R. (2001). Employer Learning and Statistical Discrimination EMPLOYER LEARNING AND STATISTICAL DISCRIMINATION*. *Source: The Quarterly Journal of Economics*, 116(1):313–350.

Arrow, K. (1973). The Theory of Discrimination.

Bar-Isaac, H. and Tadelis, S. (2008). Seller Reputation. *Foundations and Trends® in Microeconomics*, 4(4):273–351.

Baron, D. P. and Myerson, R. B. (1982). Regulating a monopolist with unknown costs. *Econometrica: Journal of the Econometric Society*, 50(4):911–930.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Blablacar (2018). Bringing People Bringing People. Technical report.

Bolton, G. E., Katok, E., and Ockenfels, A. (2004). How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation. *Management Science*, 50(11):1587–1602.

Cabral, L. L., Hortacsu, A., and Hortaçsu, A. (2010). The dynamics of seller reputation:Evidence from eBay. *The Journal of Industrial Economics*, LVIII(1):54–78.

Castillo, M., Petrie, R., Torero, M., and Vesterlund, L. (2013). Gender differences in bargaining outcomes: A field experiment on discrimination. *Journal of Public Economics*, 99:35–48.

Charles, K. K. and Guryan, J. (2011). Studying Discrimination: Fundamental Challenges and Recent Progress. *Annual Review of Economics*, 3(1):479–511.

Chiappori, P.-A., Salanie, B., and Valentin, J. (1999). Early starters versus late beginners. *Journal of Political Economy*, 107(4):731–760.

Coate, S. and Loury, G. C. (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes ? *The American Economic Review*, 83(5):1220–1240.

Cook, C., Diamond, R., Hall, J., List, J. A., Oyer, P., et al. (2018). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *Upubliceret paper. Tilgængelig på: https://web. stanford. edu/˜ diamondr/UberPayGap. pdf. Besøgt*, pages 26–04.

Darity, W. A. and Mason, P. L. (1998). Evidence on Discrimination in Employment: Codes of Color, Codes of Gender. *Journal of Economic Perspectives*, 12(2):63–90.

Dellarocas, C. and Wood, C. A. (2008). The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. *Management Science*, 54(3):460–476.

Edelman, B., Luca, M., and Svirsky, D. (2017). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics*, 9(2):1–22.

Edelman, B. G. and Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *SSRN Electronic Journal*.

Ewens, M., Tomlin, B., and Wang, L. C. (2014). Statistical discrimination or prejudice? a large sample field experiment. *Review of Economics and Statistics*, 96(1):119–134.

Farajallah, M., Hammond, R. G., and PPnard, T. (2016). What Drives Pricing Behavior in Peer-to-Peer Markets? Evidence from the Carsharing Platform BlaBlaCar. *SSRN Electronic Journal*.

Farber, H. S. and Gibbons, R. (1996). LEARNING AND WAGE DYNAMICS. *Quarterly Journal of Economics*, 111(4):1007–1047.

Gandhi, A., Lu, Z., and Shi, X. (2013). Estimating demand for differentiated products with error in market shares. Technical report, cemmap working paper, Centre for Microdata Methods and Practice.

Garrett, D. F. and Pavan, A. (2012). Managerial Turnover in a Changing World. *Journal of Political Economy*, 120(5):879–925.

Ge, Y., Knittel, C., MacKenzie, D., and Zoepf, S. (2016). Racial and Gender Discrimination in Transportation Network Companies. *NBER Working Paper Series*, (22776):1–38.

Goddard, T., Kahn, K. B., and Adkins, A. (2015). Racial bias in driver yielding behavior at crosswalks. *Transportation Research Part F: Traffic Psychology and Behaviour*, 33:1–6.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*.

Holmstrom, B. (1999). Managerial Incentive Problems: A Dynamic Perspective. *Review of Economic Studies*.

Iacus, S. M., King, G., and Porro, G. (2009). cem : Software for Coarsened Exact Matching. *Journal of Statistical Software*.

Jolivet, G., Jullien, B., and Postel-Vinay, F. (2016). Reputation and prices on the e-market: Evidence from a major French platform. *International Journal of Industrial Organization*, 45:59–75.

Jullien, B. and Park, I. U. (2014). New, like new, or very good? Reputation and credibility. *Review of Economic Studies*, 81(4):1543–1574.

Laffont, J.-J. and Tirole, J. (1986). Using Cost Observations to Regulate Firms. *Journal of Political Economy*, 94(3):614–641.

Lang, K. and Lehmann, J.-Y. K. (2012). Racial Discrimination in the Labor Market: Theory and Empirics. *Journal of Economic Literature*, 50(4):959–1006.

Laouenan, M. and Rathelot, R. (2017). Ethnic Discrimination on an Online Marketplace of Vacation Rentals. *working paper*.

Liu, Q. and Skrzypacz, A. (2014). Limited records and reputation bubbles. *Journal of Economic Theory*, 151(1):2–29.

Livingston, J. A. (2005). How Valuable Is a Good Reputation? A Sample Selection Model of Internet Auctions. *Review of Economics and Statistics*, 87(September):453–465.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation.

Mazzella, F. and Sundararajan, A. (2016). Entering the Trust Age. Technical report.

Nosko, C. and Tadelis, S. (2015). The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment. *NBER Working Paper Series*, page 20830.

Phelps, E. S. (1972). The Statistical theory of Racism and Sexism. *American Economic Review*, 62(4):659–661.

Roger, G. and Vasconcelos, L. (2014). Platform Pricing Structure and Moral Hazard. *Journal of Economics and Management Strategy*, 23(3):527–547.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*.

Sarsons, H. (2017). Interpreting Signals in the Labor Market: Evidence from Medical Referrals. *Job Market Paper*.

Spagnolo, G. (2012). Reputation, competition, and entry in procurement. *International Journal of Industrial Organization*, 30(3):291–296.

Tjaden, J. D., Schwemmer, C., and Khadjavi, M. (2018). Ride with me -ethnic discrimination, social markets, and the sharing economy. *European Sociological Review*, 34(4):418–432.

Zervas, G., Proserpio, D., and Byers, J. (2015). A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average. *Where Every Stay is Above ...*, pages 1–22.

A Changes in the BlaBlaCar reputation system

In our study of the evolution of ratings, we have abstracted from the potential changes in the design of the reputation system of BlaBlaCar. Some drivers in our sample have been BlaBlaCar customers since December 2008, and others have joined since April 2018. These drivers were operating under different market characteristics. See Figure 15 for the evolution of the average rating over time. Until the end of 2013, ratings were either 0 or 1, a binary system. Ratings were later translated to the current system (1s became 5s), but the vast majority of ratings were 1s. Later, in February 2016, the wording of the ratings was changed: *excellent* became *tres bien* and *extraordinaire* became *parfait*. The impact of this change on the average rating is clear. People are more likely to call a ride *parfait* than they were to call it *extraordinaire*. Finally, these changes influenced the informativeness of the reputation system; see Figure 16. The dotted black line shows HHI (which is a measure of dispersion and, hence, the informativeness of the classifiers): the smaller the HHI is, the more informative the classifier. The ratings in the period 2014-2016 were the most informative. Dark green, green, orange, pink, and red represent the shares of 5s, 4s, 3s, 2s and 1s, respectively. Initially, there is a considerable noise because we have very few observations: fewer than 100 per month before October 2009 and more than 30.000 per month starting in 2017.
These changes are important because affected the ratings that we study, but they also show how important the design of the review system is. We were concerned that some

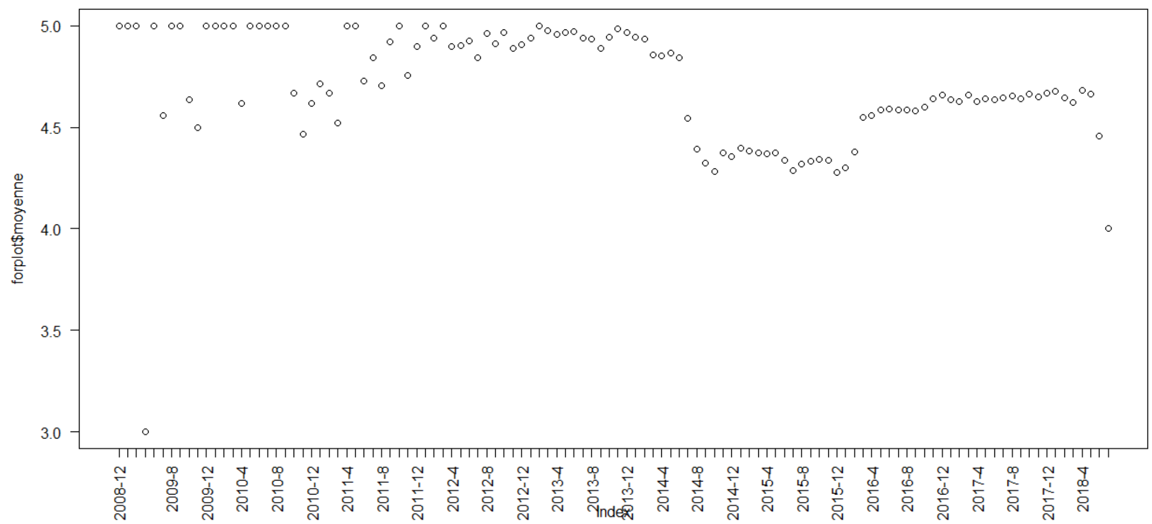


Figure 15: Average rating for drivers with more than 30 reviews

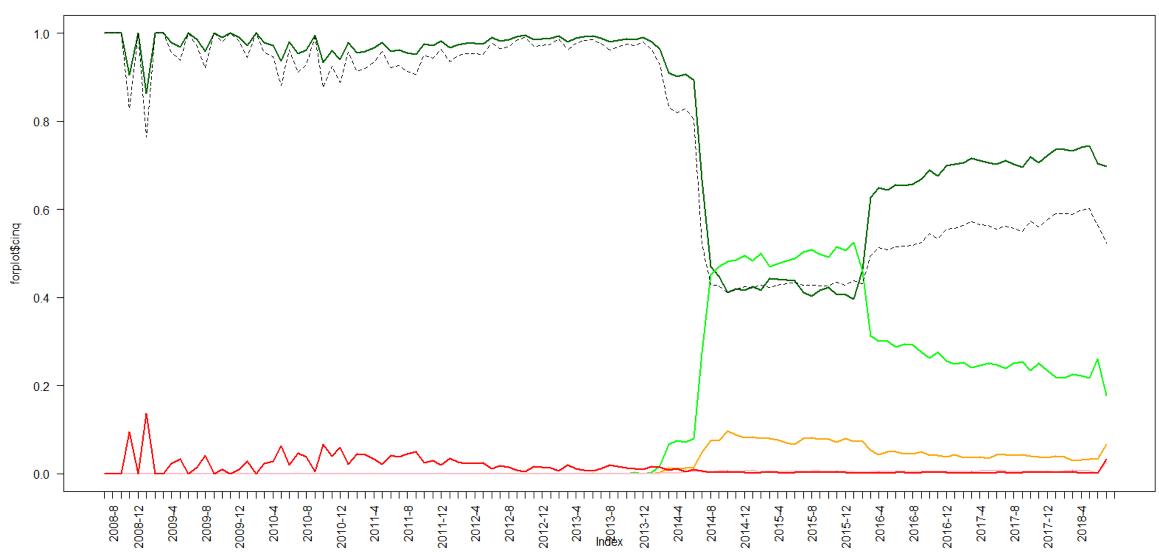


Figure 16: Informativeness of the reputation system

of the decline in effort that we characterized could be due to changes in the reputation system. In a sample restricted to drivers who joined after all the changes in the reputation system were made, we can reproduce the same patterns of behavior; however, we lose a considerable number of observations. Thus, we argue that the evolution of ratings throughout the career of a driver on BlaBlaCar is due to the economic logic of career concerns rather than exogenous changes in the reputation system.

B Navigation on Blablacar.fr

First, users type in the origine, destination and date of the ride they are seeking. They then see a list of rides meeting their request (figure 17). They may then click on specific postings to have more details about the ride (figure 18). Finally they may either see the profile of the driver (figure 19) or proceed directly to payment. BlaBlaCar service fees are a function of the price posted by the driver. These fees are shown on figure 20.

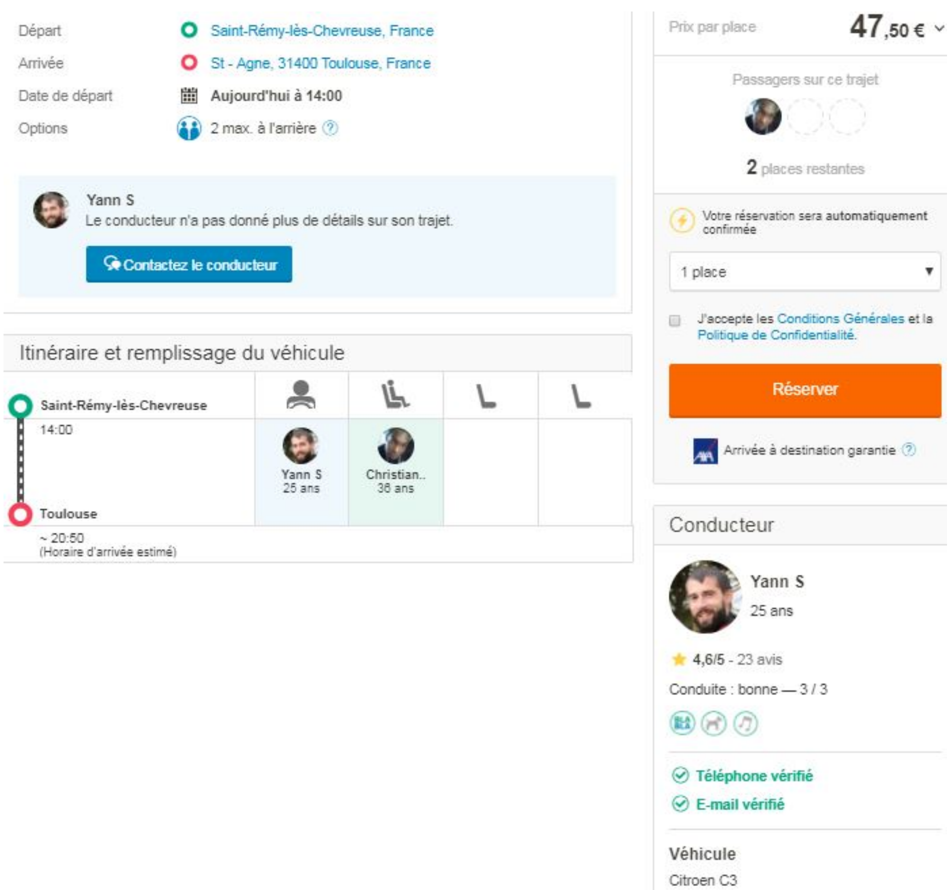


Figure 17: Listing offered on a given route

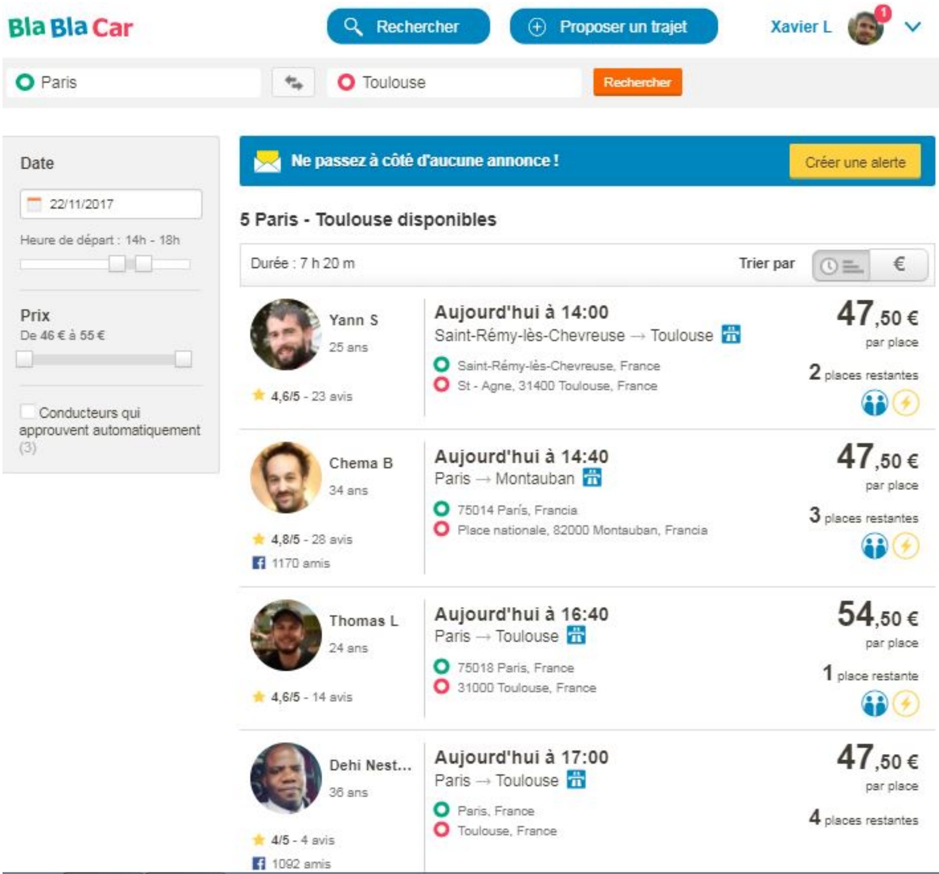


Figure 18: Details of a posting

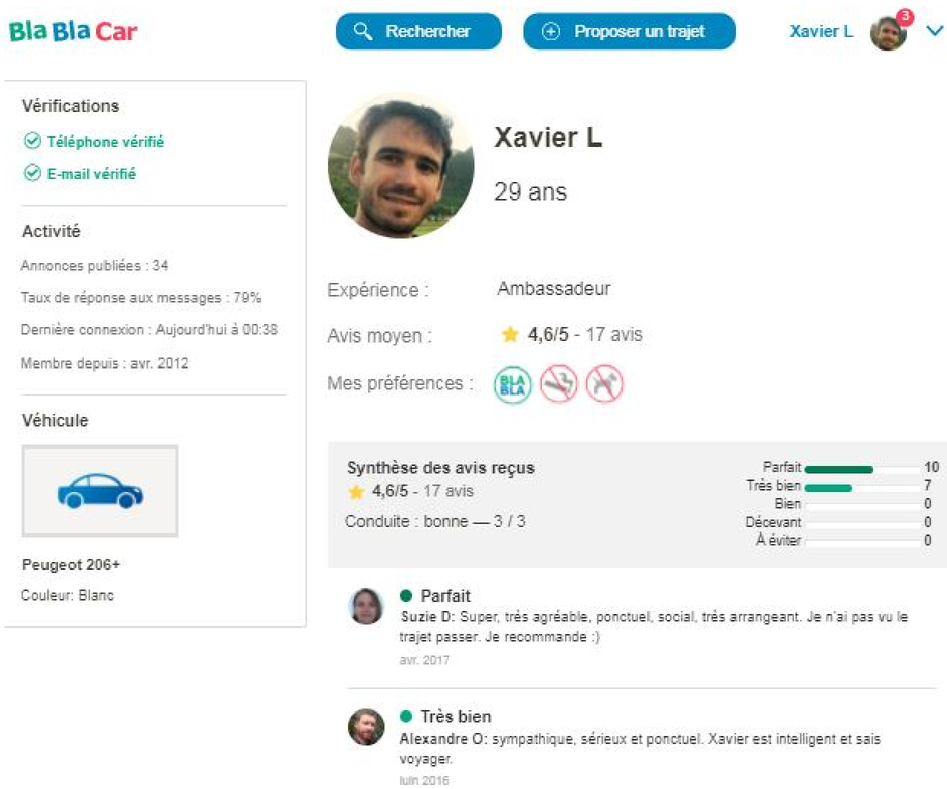


Figure 19: A driver's profile

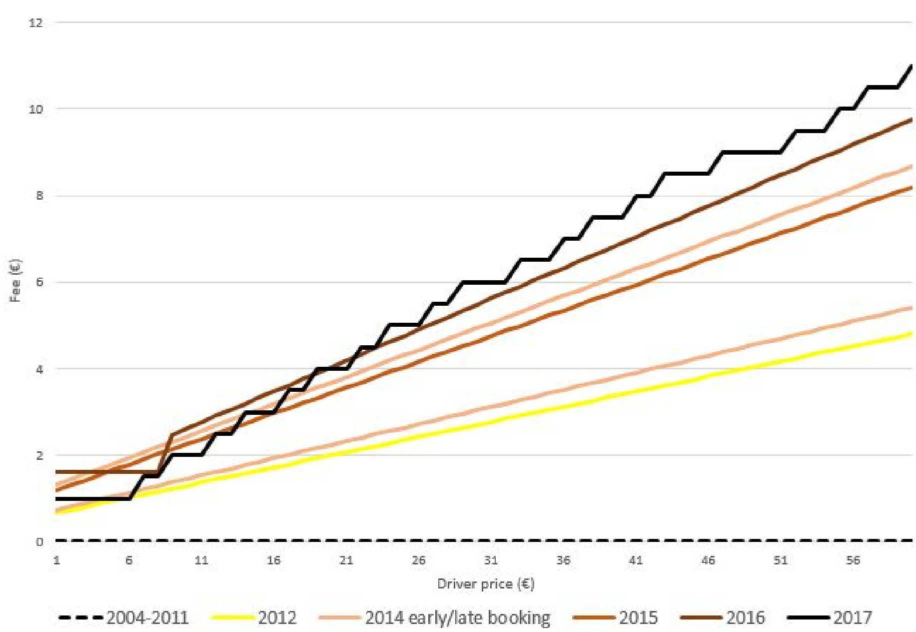


Figure 20: Evolution of service fees on BlaBlaCar over time

C Classification method for gender and ethnicity

Driver-specific characteristics are key determinants in our model. Hence, the drivers' type must be identified as accurately as possible. Specifically, gender and ethnicity are critical to our analysis. To identify these characteristics, both prospective riders and the econometrician consider two relevant sources of information: the first name and the profile picture. We use both information to infer gender and ethnicity.

C.1 Classification of gender

As a first source of information, we use the name of the driver. We match our dataset of driver names with those of various sources relating first names with ethnicity. The French Government repository of names (https://www.data.gouv.fr/fr/datasets/liste-de-prenoms/) constitutes our main source of information. We complement it with data from other sources (http://www.signification-prenom.net/, http://madame.lefigaro.fr/prenoms/origine/). These data enable us to identify the gender of almost 80% of drivers, along with 3% unisex names.

We then use facial recognition to identify gender whenever a picture is available. This process also enable us to identify 80 % of the dataset. By combining these two processes, we can directly identify gender for 95% of the dataset.

Then, we use facial recognition to enrich and correct our name database. Rare or misspelled names (either because the driver registered under a nickname or because of translation variations if the name is not originally French) can be re-classified. This process can identify the gender of some drivers whose names are not listed in our inventories and who do not have a picture (or for pictures where gender is not easily identified) because other drivers with the same name may have posted identifiable pictures. This method brings our gender identification to 99%. Figure 21 summarizes our identification process.

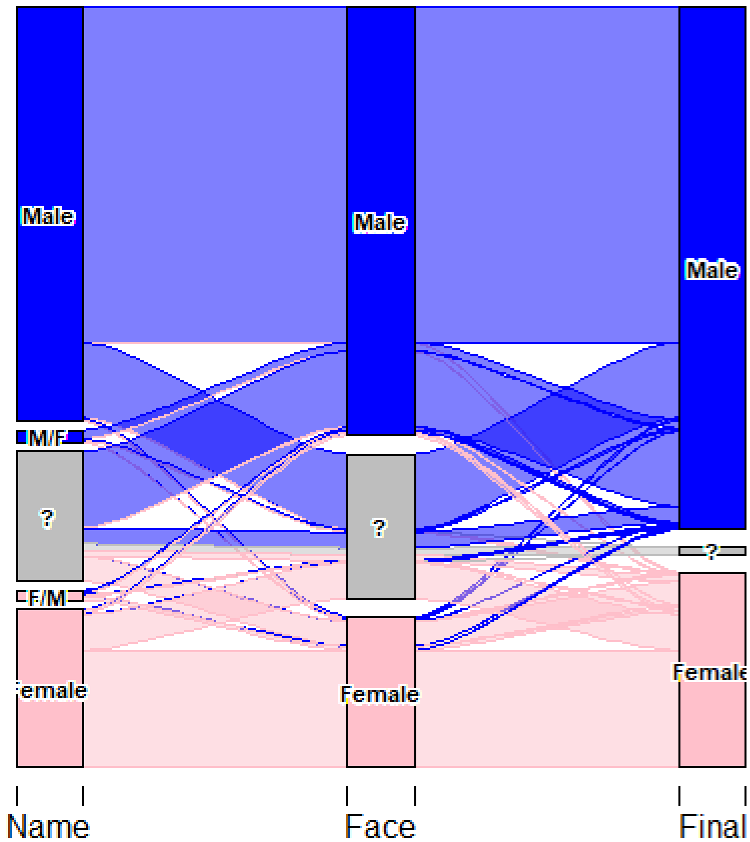


Figure 21: Classification process for gender: by name (left), by facial recognition (center) and final classification (right)

C.2 Classification of ethnicity

Our methodology for the identification of ethnicity follows the same steps and uses the same sources as those for gender classification. First, we collect the origins of names from the data sources mentioned above. This provides the ethnicity of approximately 81 % of our sample. However, names might not be a perfect indicator of ethnicity. Indeed, many visible minorities have a French name for various historical reasons or because they have foreign origins but were born in France. In that case, a simple name analysis would classify them as non-minorities while they might belong to a minority on the basis of their skin color.

Hence, we use facial recognition to identify ethnicity whenever a picture is available. The algorithm proposes an ethnicity for 80 % of the dataset. However, only "white", “black", “Asian" , and “Latino" ethnicities are proposed. People of Arabic origin are classified as “white". Hence, facial recognition is useful only to identify black people with a French or unidentified name.

We also use facial recognition to enrich and correct our name repository and to better identify ethnicity. Overall, facial recognition reclassifies 2.5% of drivers with a French name and 5% of drivers with Arabic names (predominantly Muslim names) into Sub-Saharan ethnicity. Including facial recognition increases the sample size for minorities from 11% to 15% of our sample. Figure 22 summarizes our identification process.

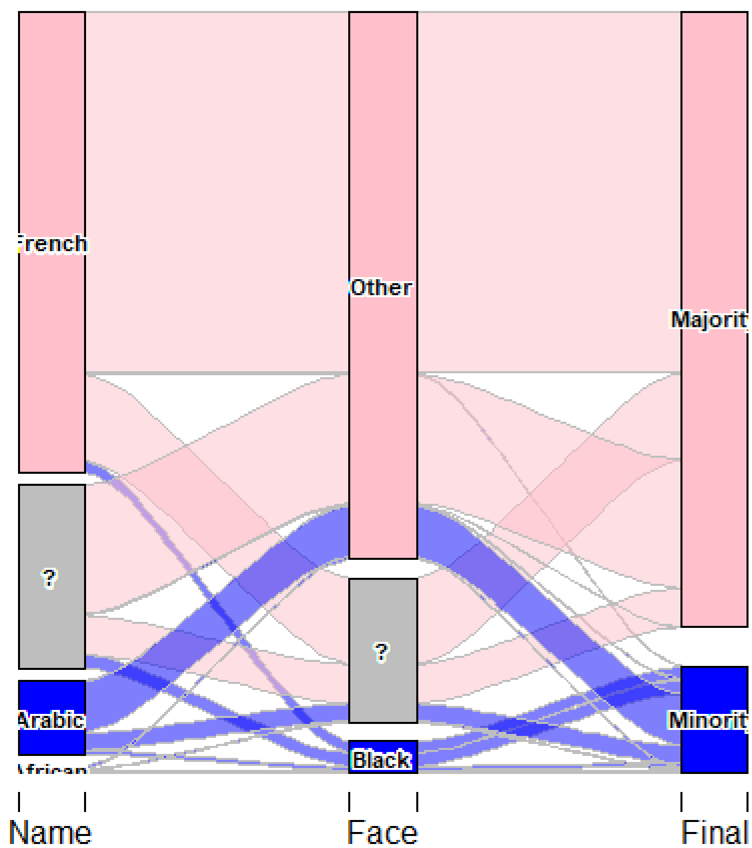


Figure 22: Classification process for ethnicity: by name analysis (left), by facial recognition (center) and final classification (right)

D Cross section: additional regressions

Controlling for number of views and price in quantity regressions, instrumental variables

Table 10: Additional controlls: full sample

	Dependent variable:			
	Revenue (OLS)	Revenue (IV)	Sold seats (OLS)	Sold seats (IV)
minority	−0.296*** (0.098)	−0.391*** (0.116)	−0.002 (0.004)	−0.007* (0.004)
driver's age	0.0005 (0.003)	0.016*** (0.006)	0.0001 (0.0001)	0.0002 (0.0002)
reviews	0.029*** (0.002)	0.027*** (0.003)	0.001*** (0.0001)	0.001*** (0.0001)
(reviews)^2	−0.0001*** (0.00000)	−0.0001*** (0.00001)	−0.00000*** (0.00000)	−0.00000*** (0.00000)
talkative	0.071 (0.069)	−0.032 (0.084)	0.003 (0.003)	−0.001 (0.003)
male	0.234*** (0.075)	0.419*** (0.135)	0.013*** (0.003)	0.017*** (0.005)
seniority (# months)	−0.006*** (0.001)	0.0003 (0.003)	−0.0004*** (0.0001)	−0.0003** (0.0001)
hours until ride	0.001 (0.002)	0.013*** (0.002)	0.0002*** (0.0001)	0.0005*** (0.0001)
posted since	0.159*** (0.051)	−0.047 (0.189)	0.010*** (0.002)	0.007 (0.007)
posts per month	−0.015 (0.017)	0.059 (0.066)	−0.003*** (0.001)	−0.002 (0.002)
bio (#words)	0.002 (0.002)	−0.0001 (0.002)	−0.0001 (0.0001)	−0.0001* (0.0001)
car's price	−0.006 (0.007)	−0.010 (0.007)	0.0001 (0.0003)	−0.0003 (0.0003)
competition	0.008*** (0.001)	0.005** (0.002)	0.0003*** (0.00005)	0.0003*** (0.0001)
median revenue	0.0001 (0.0001)	0.0002*** (0.00002)	0.00000 (0.00001)	0.00001*** (0.00000)
duration public transport	−2.368*** (0.498)	−0.069*** (0.019)	−0.005 (0.019)	−0.001* (0.001)
km	0.015*** (0.003)	0.007*** (0.0005)	0.001*** (0.0001)	−0.0002*** (0.00002)
hour	0.016** (0.008)	0.019 (0.011)	0.001*** (0.0003)	0.001*** (0.0004)
day	0.430*** (0.076)	0.352*** (0.107)	0.015*** (0.003)	0.014*** (0.004)
night	−1.294*** (0.121)	−1.737*** (0.133)	−0.062*** (0.005)	−0.070*** (0.005)
sncf strike	0.923*** (0.198)	−0.404 (0.618)	0.047*** (0.008)	0.018 (0.024)
ride descrpition (# words)	0.002*** (0.001)	−0.005 (0.003)	−0.00001 (0.00003)	−0.0001 (0.0001)
notice	−0.210*** (0.051)	−0.297*** (0.073)	−0.013*** (0.002)	−0.015*** (0.003)
automatic acceptance	2.464*** (0.068)	3.084*** (0.229)	0.100*** (0.003)	0.125*** (0.008)
number of views	0.225*** (0.002)	0.426*** (0.089)	0.010*** (0.0001)	0.013*** (0.003)
ride price			−0.009*** (0.0004)	
Constant	0.180 (2.076)	−9.414*** (1.876)	0.077 (0.081)	−0.057 (0.069)
Observations	175,142	175,142	174,732	176,937
R^2	0.173	0.087	0.195	0.173
Adjusted R^2	0.172	0.087	0.194	0.173
Residual Std. Error	13.550 (df = 174881)	14.235 (df = 175104)	0.530 (df = 174470)	0.535 (df = 176899)
F Statistic	141.190*** (df = 260; 174881)		161.847*** (df = 261; 174470)	
Note:				*p<0.1; **p<0.05; ***p<0.01

Hurdel model, seats taken:

E Oversampling of minorities for short-notice rides

Due to our scraping method, it cannot be excluded that our sample provides a slightly biased representation of listings. Indeed, the program takes snapshots of listings displayed on the website at a given point time. However, rides that are already full are no longer displayed on the platform. This means our data collection may undersample the particularly interesting rides that would sell out very fast, or those corresponding to times when demand is much higher than supply. This wouldn't be an issue if both

	Model 1
(Intercept)	-1.00^{***}
	(0.03)
driver's age	-0.00
	(0.00)
reviews	0.00^{***}
	(0.00)
talkative	0.04^{***}
	(0.01)
minority	-0.12^{***}
	(0.01)
male	-0.05^{***}
	(0.01)
bio (# words)	0.00^{**}
	(0.00)
post's per month	-0.04^{***}
	(0.00)
posted since	0.16^{***}
	(0.00)
notice	-0.11^{***}
	(0.00)
competition	0.00^{***}
	(0.00)
sncf strike	0.59^{***}
	(0.02)
ride (# words)	0.00^{***}
	(0.00)
km	0.00^{***}
	(0.00)
ride price	-0.07^{***}
	(0.00)
AIC	338835.93
Log Likelihood	-169387.96
Num. obs.	266699

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 11: Hurdle model: count model not included

minorities and non-minorities were affected the same way by this sampling bias. However, as we show in this paper the minority status does impact the attractiveness of a given listing. Therefore, minorities who may be perceived as posting less interesting rides remain longer on display and may therefore be over-represented in our sample. Therefore, our minority gap estimates should be understood as lower bounds. Indeed, minorities are compared to a pool constituted of non-minorities that are not so good as to have sold out their seats extremely fast. Table 23 shows that minority drivers represent a specially high share of rides that are posted on a short notice, a possible sign that non-minority drivers have sold their seats faster. For trips posted with more notice, we believe our sample is indeed representative of the actual participants on blablacar. Indeed, most of the rides –either from minorities or not – still have more than one empty seat, which means that most listings and indeed collected.

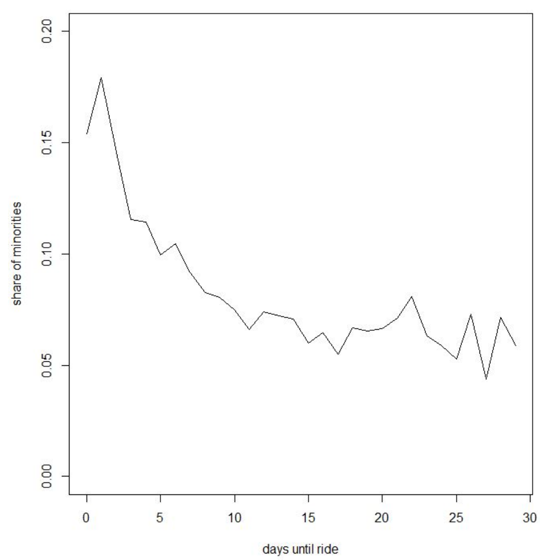


Figure 23: Share of minorities in sample as a function of number of days between posting and departure

This is true despite the fact minorities tend to allow for automatic confirmation more frequently than non-minorities (12% of drivers with automatic confirmation are minorities, while they represent only 8% of the drivers with manual confirmation).

Table 12: Number of clicks regressed over driver and ride characteristics

	Dependent variable:		
	number of clicks		
reviews(#)	1:5	6:49	50:450
driver's age	−0.074*** (0.010)	−0.061*** (0.005)	−0.014* (0.008)
reviews	0.161 (0.429)	0.050** (0.025)	0.025*** (0.005)
$(\text{reviews})^2$	0.005 (0.072)	−0.0003 (0.0005)	−0.00003** (0.00001)
talkative	0.232 (0.261)	0.353** (0.143)	0.672*** (0.181)
minority	−1.890*** (0.374)	−0.700*** (0.204)	0.042 (0.259)
male	−1.199*** (0.265)	−1.316*** (0.147)	−0.563** (0.233)
seniority months	−0.016*** (0.006)	−0.020*** (0.003)	−0.048*** (0.004)
hours util ride	−0.011 (0.008)	−0.015*** (0.004)	0.002 (0.006)
posted since	2.034*** (0.193)	2.165*** (0.107)	2.303*** (0.134)
post per month	−0.276*** (0.087)	−0.573*** (0.039)	−0.698*** (0.036)
picture	1.670*** (0.346)	0.942*** (0.225)	0.317 (0.287)
bio (#words)	−0.002 (0.008)	0.001 (0.004)	−0.008 (0.005)
car price	0.010 (0.024)	0.025* (0.014)	−0.036** (0.017)
competition	0.012*** (0.004)	0.015*** (0.002)	0.024*** (0.003)
median revenue	−0.0001 (0.0005)	0.001** (0.0003)	0.001** (0.0004)
duration public transport	−1.390 (1.682)	2.109** (1.017)	1.827 (1.670)
km	0.003 (0.009)	−0.0001 (0.005)	0.004 (0.007)
hour	0.116*** (0.028)	0.065*** (0.016)	0.143*** (0.020)
day	0.089 (0.284)	0.683*** (0.155)	1.801*** (0.202)
night	1.254*** (0.453)	0.550** (0.260)	−0.534* (0.314)
sncf strike	6.027*** (0.659)	6.970*** (0.409)	6.799*** (0.592)
notice	−0.576*** (0.192)	−0.569*** (0.106)	−1.003*** (0.133)
automatic acceptance	−1.176*** (0.251)	−2.273*** (0.137)	−3.388*** (0.192)
ride (# words)	0.039*** (0.004)	0.032*** (0.002)	0.026*** (0.002)
Constant	22.563*** (7.172)	2.164 (4.197)	−1.988 (7.112)
Observations	29,788	90,450	45,647
R^2	0.255	0.264	0.271
Adjusted R^2	0.249	0.262	0.267
Residual Std. Error	20.527 (df = 29529)	19.860 (df = 90190)	18.804 (df = 45388)
F Statistic	39.244*** (df = 258; 29529)	124.828*** (df = 259; 90190)	65.382*** (df = 258; 45388)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 13: Sold seats regressed over driver and ride characteristics

	Dependent variable:		
	sold seats		
reviews(#)	1:5	6:49	50:450
driver's age	−0.001*** (0.0002)	−0.001*** (0.0002)	−0.0003 (0.0003)
reviews	−0.004 (0.010)	0.004*** (0.001)	0.001*** (0.0002)
$(\text{reviews})^2$	0.002 (0.002)	−0.00004*** (0.00001)	−0.00000*** (0.00000)
talkative	−0.003 (0.006)	0.002 (0.004)	0.019*** (0.006)
minority	−0.031*** (0.009)	−0.013** (0.006)	−0.001 (0.009)
male	−0.007 (0.006)	0.010** (0.004)	0.003 (0.008)
seniority months	−0.0003** (0.0001)	−0.001*** (0.0001)	−0.001*** (0.0001)
hours until ride	0.0001 (0.0002)	0.0001 (0.0001)	0.0005** (0.0002)
posted since	0.024*** (0.005)	0.029*** (0.003)	0.047*** (0.005)
posts per month	0.004* (0.002)	−0.008*** (0.001)	−0.016*** (0.001)
picture	0.021** (0.008)	0.009 (0.006)	0.010 (0.010)
bio (#words)	−0.0002 (0.0002)	−0.0002 (0.0001)	−0.0003* (0.0002)
car price	−0.001 (0.001)	−0.0001 (0.0004)	0.0003 (0.001)
competition	0.0002* (0.0001)	0.0005*** (0.0001)	0.001*** (0.0001)
median revenue	0.00000 (0.00001)	0.00001 (0.00001)	0.00001 (0.00001)
duration public transport	0.002 (0.041)	0.014 (0.029)	0.039 (0.056)
km	−0.0004* (0.0002)	−0.0002 (0.0001)	−0.0001 (0.0002)
hour	0.002*** (0.001)	0.002*** (0.0005)	0.003*** (0.001)
day	0.019*** (0.007)	0.017*** (0.004)	0.042*** (0.007)
night	−0.045*** (0.011)	−0.058*** (0.007)	−0.070*** (0.011)
sncf strike	0.094*** (0.016)	0.118*** (0.012)	0.151*** (0.020)
notice	−0.014*** (0.005)	−0.017*** (0.003)	−0.033*** (0.004)
automatic acceptance	0.094*** (0.006)	0.089*** (0.004)	0.094*** (0.006)
ride (# words)	0.001*** (0.0001)	0.0003*** (0.0001)	0.0003*** (0.0001)
Constant	0.418** (0.176)	0.137 (0.118)	−0.004 (0.237)
Observations	30,162	91,262	45,986
R^2	0.078	0.078	0.102
Adjusted R^2	0.070	0.075	0.096
Residual Std. Error	0.505 (df = 29903)	0.563 (df = 91002)	0.635 (df = 45727)
F Statistic	9.764*** (df = 258; 29903)	29.603*** (df = 259; 91002)	20.032*** (df = 258; 45727)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 14: Revenue regressed over driver and ride characteristics

	Dependent variable:		
	revenue		
reviews(#)	1:5	6:49	50:450
driver's age	−0.021*** (0.007)	−0.009** (0.004)	−0.007 (0.006)
reviews	0.019 (0.288)	0.077*** (0.018)	0.017*** (0.004)
$(\text{reviews})^2$	0.037 (0.048)	−0.001* (0.0004)	−0.00002** (0.00001)
talkative	−0.107 (0.174)	0.116 (0.104)	0.344** (0.140)
minority	−0.806*** (0.251)	−0.486*** (0.150)	−0.149 (0.200)
male	−0.333* (0.178)	0.099 (0.108)	−0.048 (0.180)
seniority months	−0.012*** (0.004)	−0.021*** (0.002)	−0.020*** (0.003)
hours until ride	−0.003 (0.005)	−0.005 (0.003)	0.009** (0.004)
posted since	0.487*** (0.129)	0.586*** (0.078)	0.992*** (0.104)
posts per month	0.100* (0.058)	−0.140*** (0.029)	−0.246*** (0.028)
picture	0.582** (0.233)	0.305* (0.165)	0.507** (0.222)
bio (#words)	0.002 (0.006)	−0.001 (0.003)	−0.007 (0.004)
car price	−0.025 (0.016)	−0.004 (0.010)	0.010 (0.013)
competition	0.008*** (0.003)	0.012*** (0.002)	0.014*** (0.002)
median revnue	−0.0001 (0.0003)	0.0003 (0.0002)	0.00003 (0.0003)
duration public transport	−2.063* (1.122)	−1.635** (0.745)	−4.790*** (1.334)
km	0.012* (0.006)	0.013*** (0.004)	0.023*** (0.005)
hour	0.039** (0.019)	0.030** (0.012)	0.039** (0.016)
day	0.256 (0.191)	0.575*** (0.114)	0.882*** (0.156)
night	−1.119*** (0.303)	−1.214*** (0.190)	−1.128*** (0.242)
sncf strike	2.313*** (0.441)	2.573*** (0.300)	2.971*** (0.460)
notice	−0.240* (0.129)	−0.280*** (0.077)	−0.673*** (0.103)
automatic acceptance	2.239*** (0.168)	1.927*** (0.101)	1.809*** (0.148)
ride (# words)	0.012*** (0.003)	0.008*** (0.001)	0.008*** (0.001)
Constant	10.735** (4.806)	0.012 (3.080)	9.728* (5.696)
Observations	29,864	90,317	45,394
R^2	0.077	0.086	0.115
Adjusted R^2	0.069	0.084	0.110
Residual Std. Error	13.782 (df = 29605)	14.530 (df = 90057)	14.519 (df = 45135)
F Statistic	9.578*** (df = 258; 29605)	32.805*** (df = 259; 90057)	22.757*** (df = 258; 45135)
Note:			*p<0.1; **p<0.05; ***p<0.01

F Reputation effect

G Coarsened Exact Matching

This project, likewise most in the literature, uses non-experimental data for evaluating the impact of minority status. Hence, estimates of the impact of being a minority may suffer from a bias of the selection on the non-observables. There is a growing, mostly theoretical, literature on the use of matching techniques to address this issue. Rosenbaum and Rubin (1983) and Heckman et al. (1997) demonstrate that this bias can be greatly reduced by use of various matching techniques. Some of their properties are discussed by Abadie and Imbens (2016), (2016). A similar methodology has been applied in Sarsons (2017). [22]

The objective of matching exercise is to test the robustness of results from the standard OLS of Section 3.2. We will firstly estimate propensity scores for each of the observations and discard these with extreme values. Secondly, we will perform matching of the minority and non-minority subsamples on driver-specific variables. We will execute both exact matching and coarsened matching. Finally, we will regress model using the matched sample, controlling for listing-specific characteristics.

The propensity score is a logistic regression with minority status being dependent variables and following controls: the price of a car, driver's age, number of posts per month, picture dummy, length of biography, gender, fuel consumption of the car and whether the driver is talkative. Minority drivers are more likely to be a young male and to enjoy conversations. They have on average more expensive cars that consume more fuel; their profiles are also shorter. We delete 5% smallest and 5% largest propensity scores, in this way we delete observations for which we are unlikely to find a counterpart.

Exact matching is performed on all driver's characteristics for which we have estimated logistic regression. In our sample, it means that we have 8809 minority drivers matched with 22617 non-minority drivers. As entrants, we will label minority drivers with less than five reviews and as incumbent's (experienced users) these with more than 50 reviews. In the case of exact matching, the definition of an incumbent is extended to drivers with more than 30 reviews so as to increase the size of the group. From table 15 we can see that even after the matching procedure, minority entrant drivers are fac-

[22]We use matching software developed by Iacus et al. (2009).

Table 15: Economic outcomes of entrants, exact matching

	Dependent variable		
	number of clicks	Revenue	Taken seats
minority	-1.2546*** (0.341)	-0.69684** (0.221)	-0.0258*** (0.007)
hours until ride	-0.0107 (0.009)	-0.0109 (0.006)	-0.0005* (0.0002)
posted since	2.0561 *** (0.225)	0.2721 (0.145)	0.0059 (0.005)
competition	0.0249 ***(0.005)	0.003*** (0.569)	0.0005*** (0.0001)
day	0.3792 (0.349)	0.5492* (0.226)	0.0121 (0.012)
night	0.8578 (0.517)	-1.2120*** (0.335)	-0.0492 *** (0.005)
notice	−0.4786* (0.122)	-0.0027 (0.145)	0.0031 (0.023)
Matched Observations	19,112		
Note: Trip fixed effects not reported			*p<0.05; **p<0.01; ***p<0.001

ing discrimination. We repeat the same process for drivers with reputation. Table 16 reports the results. We observe the reputation effect; minority status is insignificant for the number of clicks and seats taken. In case of revenue, there is a positive relationship, which could suggest a selection effect.

Table 16: Economic outcomes of incumbent drivers, exact matching

	Dependent variable		
	number of clicks	Revenue	Taken seats
minority	-0.2876 (0.383)	0.1409 (0.303)	0.0106 (0.012)
hours untill ride	-0.0340** (0.0117)	0.0112 (0.009)	0.0009* (0.0004)
posted since	1.8896*** (0.282)	1.341*** (0.223)	0.0626 *** (0.009)
competition	0.0168 **(0.005)	0.0076 (0.004)	0.0003 (0.0002)
day	0.1916 (0.394)	0.0322 (0.311)	0.0177 (0.012)
night	0.5573 (0.634)	-1.9742*** (0.500)	-0.0829*** (0.019)
notice	-0.1398 (0.280)	-0.8167 *** (0.221)	-0.0436 *** (0.008)
Matched Observations	12314		
Note: Trip fixed effects not reported			*p<0.05; **p<0.01; ***p<0.001

Coarsened Matching is a method used to increase the number of matched observations. We introduce bins in which we will match non-binary covariates: age of the driver, the price of a car, number of posts per month, length of bio and fuel consumption of the car. Choice of cutoffs influences the precision of matching procedure as well as the number of matched observations; we match within a quantile for each of the variables. In this way, we match 14146 minority drivers with 45959 nonminority ones, which is almost a twofold increase. We present only coefficient on minority (table 17). In the coarsely matched sample, we also see a clear reputation effect. Minority entrants

Table 17: Economic outcomes entrants and incumbents, coarsened matching

	Dependent variable		
	number of clicks	Revenue	Taken seats
minority (entrant)	-0.9276 *** (0.271)	-0.6032*** (0.180)	-0.0191** (0.006)
minority (incumbent)	-0.3883 (0.303)	0.0330 (0.229)	-0.0084 (0.010)
Matched Observations (both models)	57,853		

*p<0.1; **p<0.05; ***p<0.01

have lower economic outcomes, however after they build reputation the effect goes entirely away. These results depend on cut-offs for labeling as entrants/ incumbents, as well as on the selection of bins for coarsened matching; they are however robust to local changes.

We have provided evidence of reputation effect on all available economic outcomes measures: number of clicks, taken seats and revenue. Minority drivers face discrimination at the beginning of their career, this effect, however, disappears as soon as they develop a reputation. We have implemented a matching technique that made minority drivers more comparable with non-minorities. Still, the reputation effect was visible.

H Detailed panel data results

	Dependent variable: number of clicks		
	Pooled	Between	Random
minority	0.288 (0.202)	0.409 (0.275)	0.317 (0.236)
entrant	−0.995*** (0.143)	−0.811*** (0.179)	−0.764*** (0.155)
minority*entrant	−0.678* (0.353)	−0.692 (0.449)	−0.717* (0.387)
driver's age	−0.036*** (0.005)	−0.038*** (0.006)	−0.036*** (0.006)
talkative	0.220* (0.123)	0.363** (0.156)	0.282** (0.141)
male	−1.074*** (0.142)	−1.105*** (0.171)	−1.128*** (0.159)
hours until ride	−0.028*** (0.0005)	−0.023*** (0.001)	−0.029*** (0.0005)
posted since	1.136*** (0.010)	1.068*** (0.016)	1.172*** (0.010)
bio (# words)	−0.002 (0.004)	−0.003 (0.005)	−0.002 (0.004)
car price	−0.018 (0.012)	−0.031** (0.015)	−0.021 (0.014)
competition	0.036*** (0.002)	0.035*** (0.003)	0.034*** (0.002)
median revenue	−0.00002 (0.00003)	−0.0001* (0.00004)	−0.00000 (0.00003)
public transport ratio	−0.909 (7.222)	−1.761 (10.861)	−2.131 (7.765)
km	0.007*** (0.0004)	0.006*** (0.001)	0.006*** (0.0004)
day	0.538** (0.231)	0.574 (0.364)	0.462** (0.231)
night	−0.605* (0.357)	−1.134* (0.581)	−0.763** (0.358)
train strike	3.269*** (0.325)	3.049*** (0.538)	3.545*** (0.319)
ride (# words)	0.018*** (0.002)	0.021*** (0.002)	0.020*** (0.002)
picture	0.246 (0.201)	0.494* (0.260)	0.496** (0.230)
automatic acceptance	−1.334*** (0.122)	−1.299*** (0.164)	−1.307*** (0.132)
weekday	−0.018 (0.236)	−0.457 (0.387)	0.112 (0.237)
consumption	0.278*** (0.084)	0.377*** (0.106)	0.303*** (0.095)
day*weekday	0.465 (0.284)	0.925** (0.460)	0.389 (0.284)
night*weekday	−0.018 (0.444)	1.577** (0.746)	0.094 (0.443)
Constant	11.748*** (0.880)	10.465*** (1.234)	11.477*** (0.948)
Observations	56,760	22,794	56,760
R^2	0.244	0.220	0.262
Adjusted R^2	0.244	0.219	0.261
F Statistic	495.292*** (df = 37; 56722)	173.642*** (df = 37; 22756)	543.306*** (df = 37; 56722)
Note:			*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: sold seats		
	Pooled	Between	Random
minority	0.002 (0.009)	0.016 (0.011)	0.002 (0.009)
entrant	−0.060*** (0.011)	−0.058*** (0.012)	−0.059*** (0.011)
minority*entrant	−0.035** (0.016)	−0.041** (0.018)	−0.035** (0.016)
male	0.005 (0.008)	0.004 (0.009)	0.004 (0.008)
driver's age	−0.0004* (0.0002)	−0.0003 (0.0002)	−0.0004* (0.0002)
talkative	0.001 (0.005)	0.003 (0.006)	0.001 (0.006)
hours until ride	−0.001*** (0.00002)	−0.001*** (0.00003)	−0.001*** (0.00002)
posted since	0.016*** (0.0004)	0.012*** (0.001)	0.016*** (0.0004)
bio (# words)	−0.0001 (0.0002)	−0.0003 (0.0002)	−0.0001 (0.0002)
car price	−0.0003 (0.001)	−0.001 (0.001)	−0.0004 (0.001)
competition	0.001*** (0.0001)	0.001*** (0.0001)	0.001*** (0.0001)
median revenue	0.00000*** (0.00000)	0.00001*** (0.00000)	0.00000*** (0.00000)
public transport ratio	−0.146 (0.318)	−0.566 (0.440)	−0.147 (0.322)
km	−0.00002 (0.00002)	−0.0001** (0.00002)	−0.00002 (0.00002)
day	0.015 (0.010)	0.004 (0.015)	0.015 (0.010)
night	−0.048*** (0.016)	−0.062*** (0.023)	−0.048*** (0.016)
train strike	0.126*** (0.014)	0.110*** (0.022)	0.128*** (0.014)
ride (# words)	0.0004*** (0.0001)	0.0004*** (0.0001)	0.0004*** (0.0001)
picture	0.002 (0.009)	0.015 (0.010)	0.003 (0.009)
automatic acceptance	0.109*** (0.005)	0.108*** (0.007)	0.109*** (0.005)
weekday	−0.045*** (0.010)	−0.059*** (0.016)	−0.045*** (0.010)
consumption	0.020*** (0.004)	0.021*** (0.004)	0.020*** (0.004)
minority*entrant	−0.035** (0.016)	−0.041** (0.018)	−0.035** (0.016)
entrant*male	−0.016 (0.013)	−0.022 (0.014)	−0.015 (0.013)
day*weekday	0.019 (0.012)	0.034* (0.019)	0.019 (0.013)
night*weekday	−0.020 (0.019)	0.012 (0.030)	−0.020 (0.020)
Constant	0.180*** (0.039)	0.158*** (0.050)	0.175*** (0.039)
Observations	59,359	23,076	59,359
R^2	0.089	0.085	0.088
Adjusted R^2	0.088	0.083	0.087
F Statistic	152.507*** (df = 38; 59320)	56.087*** (df = 38; 23037)	150.026*** (df = 38; 59320)
Note:			*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: revenue		
	Pooled	Between	Random
minority	−0.334 (0.213)	0.022 (0.275)	−0.272 (0.228)
entrant	−1.387*** (0.150)	−1.452*** (0.179)	−1.308*** (0.155)
minority*entrant	−0.680* (0.372)	−0.866* (0.448)	−0.741* (0.387)
driver's age	−0.006 (0.005)	−0.002 (0.006)	−0.005 (0.005)
talkative	0.020 (0.129)	0.065 (0.155)	0.026 (0.137)
male	−0.201 (0.148)	−0.307* (0.170)	−0.240 (0.156)
hours untill ride	−0.018*** (0.0005)	−0.016*** (0.001)	−0.019*** (0.0005)
posted since	0.371*** (0.010)	0.290*** (0.014)	0.375*** (0.010)
bio (# words)	−0.001 (0.004)	−0.005 (0.005)	−0.001 (0.004)
car price	−0.007 (0.013)	−0.022 (0.015)	−0.010 (0.013)
competition	0.024*** (0.002)	0.024*** (0.003)	0.024*** (0.002)
median revenue	0.0002*** (0.00003)	0.0003*** (0.00004)	0.0002*** (0.00003)
public transport ratio	−33.375*** (7.569)	−40.181*** (10.934)	−33.318*** (7.835)
km	0.013*** (0.0004)	0.011*** (0.001)	0.013*** (0.0004)
day	0.410* (0.243)	0.445 (0.367)	0.400 (0.244)
night	−1.341*** (0.373)	−2.091*** (0.579)	−1.300*** (0.376)
train strike	2.367*** (0.339)	1.779*** (0.543)	2.429*** (0.338)
ride (# words)	0.007*** (0.002)	0.008*** (0.002)	0.007*** (0.002)
picture	0.087 (0.212)	0.366 (0.260)	0.170 (0.225)
automatic acceptance	2.064*** (0.128)	2.012*** (0.164)	2.104*** (0.133)
weekday	−0.847*** (0.249)	−1.142*** (0.390)	−0.828*** (0.251)
consumption	0.315*** (0.088)	0.340*** (0.106)	0.325*** (0.093)
minority*entrant	−0.680* (0.372)	−0.866* (0.448)	−0.741* (0.387)
day*weekday	0.317 (0.299)	0.446 (0.465)	0.290 (0.300)
night*weekday	−0.215 (0.465)	0.938 (0.745)	−0.232 (0.467)
Constant	−1.089 (0.926)	−2.287* (1.239)	−1.200 (0.957)
Observations	58,621	23,018	58,621
R^2	0.095	0.093	0.094
Adjusted R^2	0.094	0.091	0.093
F Statistic	165.658*** (df = 37; 58583)	63.359*** (df = 37; 22980)	163.313*** (df = 37; 58583)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 18: Propensity score table

	Dependent variable:
	f minority
car price	0.024^{***} (0.001)
driver age	-0.033^{***} (0.001)
post per month	0.058^{***} (0.003)
picture	0.078^{***} (0.023)
length bio	-0.013^{***} (0.0005)
f gender	0.952^{***} (0.019)
consumption	0.140^{***} (0.010)
driver blabla	0.329^{***} (0.014)
Constant	-2.965^{***} (0.063)
Observations	195,333
Log Likelihood	−75,005.150
Akaike Inf. Crit.	150,028.300
Note:	*p<0.1; **p<0.05; ***p<0.01

I Efforts reduced form

	Dependent variable:		
	detour	luggage	pet
minority	0.669*** (0.039)	−0.008 (0.046)	−0.556*** (0.028)
reviews (#)	−0.003*** (0.0002)	−0.001*** (0.0002)	−0.001*** (0.0002)
minority*reviews	−0.001*** (0.0004)	−0.0004 (0.0005)	−0.004*** (0.0005)
driver's age	−0.013*** (0.001)	−0.002** (0.001)	−0.025*** (0.001)
talkative	0.179*** (0.019)	0.164*** (0.025)	0.451*** (0.014)
male	0.377*** (0.020)	0.606*** (0.026)	−0.688*** (0.015)
seniority (# months)	−0.008*** (0.0004)	−0.004*** (0.001)	−0.002*** (0.0003)
hours untill ride	−0.0001** (0.0001)	−0.0004*** (0.0001)	0.00005 (0.00005)
posted since	−0.002** (0.001)	−0.011*** (0.002)	−0.002* (0.001)
posts per month	−0.015*** (0.004)	−0.029*** (0.005)	0.021*** (0.003)
bio (# words)	−0.0002 (0.001)	0.0001 (0.001)	0.007*** (0.0004)
car price	0.0004 (0.002)	−0.011*** (0.002)	−0.023*** (0.002)
competition	−0.001*** (0.0004)	−0.001** (0.0005)	−0.002*** (0.0003)
median revenue	−0.00001 (0.00000)	0.00001 (0.00001)	−0.00001** (0.00000)
public transport ratio	−3.696*** (1.234)	−1.298 (1.725)	3.356*** (0.879)
km	0.001*** (0.0001)	0.00001 (0.0001)	0.0001*** (0.00004)
day	0.056 (0.035)	0.012 (0.047)	0.017 (0.027)
night	−0.041 (0.054)	−0.011 (0.073)	0.171*** (0.041)
ride (# words)	−0.006*** (0.0002)	−0.005*** (0.0003)	0.0004** (0.0002)
weekday	−0.070** (0.036)	0.001 (0.048)	−0.029 (0.029)
consumption	0.006 (0.012)	0.394*** (0.018)	0.130*** (0.010)
picture	0.162*** (0.027)	0.052 (0.037)	0.106*** (0.027)
day*weekday	0.044 (0.043)	0.011 (0.057)	0.008 (0.034)
night*weekday	0.186*** (0.067)	0.009 (0.090)	0.056 (0.051)
Constant	1.642*** (0.254)	−0.261 (0.292)	−1.482*** (0.105)
Observations	74,155	74,446	131,679
Log Likelihood	−39,691.410	−25,060.570	−63,486.190
Akaike Inf. Crit.	79,446.820	50,185.140	127,048.400
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 19: Proxies for effort, logistic regressions, trip fixed effects not reported

J Introducing bargaining power

An important, yet somewhat implicit, assumption in Holmstrom (1999) is that drivers have all the bargaining power, and capture the entire surplus of interaction at all times. This seems inconsistent with most online platform markets and especially Blablacar, where prospective riders choose amongst potentially numerous drivers, who are therefore competing against one another. In order to account for different levels of bargaining power we will consider two types of markets. Driver's market which is exactly the same as in Holmstrom (1999) and passenger's market, which is when passenger makes a take-it-or-leave-it offer, by which she leaves no surplus for the driver. In passenger market, driver exerts effort only to benefit from it in future when the odds change. Before each turn starts, *nature* will draw whether market is driver's or passenger's. The former happens with probability q and the latter case happens with probability $1-q$. Suppose we start from a driver's market; effort choice problem of the driver can be now written as:

$$\max_{a_t} y_t - g(a_t) + q\left(\sum_{t=1}^{\infty}\beta^{t-1}\left[\mathbf{E}[w_t(y^{t-1})] - \mathbf{E}[g(a_t(y^{t-1}))]\right]\right)$$
$$+(1-q)\left(\sum_{t=1}^{\infty}\beta^{t-1}\left[0 - \mathbf{E}[g(a_t(y^{t-1}))]\right]\right)$$
$$\Longleftrightarrow$$
$$\max_{a_t}\left\{y_t - g(a_t) + \left(\sum_{t=1}^{\infty}\beta^{t-1}\left[q\mathbf{E}[w_t(y^{t-1})] - \mathbf{E}[g(a_t(y^{t-1}))]\right]\right)\right\} \tag{7}$$

Driver gets to receive a positive wage with probability q, however the underlying process goes on all the time; note that as in baseline Holmström model it is argued that whatever the equilibrium level of effort is a_t^*, market (passanger) deduces:

$$z_t = \eta + \epsilon_t = y_t - a_t^*$$

and updates driver's type

$$m_{t+1} = \frac{h_1 m_1 + h_\epsilon \sum_{s=1}^{t} z_s}{h_1 + t h_\epsilon}$$

The first order condition of the above maximization problem writes:

$$a_t^* = g^{-1'}\left(q\sum_{s=t}^{\infty}\beta^{s-t}\frac{h_\epsilon}{h_s}\right)$$

In expectation, drivers receive a positive wage in a share q of periods. Nevertheless, she has to take positive effort in all periods in order to induce the market into thinking she is a high type, thereby increasing her future expected wages.

K Evidence using reduced form estimates

Results discussed earlier, highlighted importance of the reputation system as well as of the demographic characteristics of drivers, in this section we study, first, whether the measure of expected quality from the model presented above is strong predictor of demand and, secondly, whether it can explain the minority gap. Table **??** presents results using sold seats (1 and 2) and revenue (3 and 4) as dependent variables. Taken seats regressions use a set of instruments in the spirit of Berry (1994), which are calculated as $z_{jtm} = \sum_{j' \neq j} x_{j',t,m}$ in order to account for the endogeneity of price and quantity. Models of oligopoly suggest that the more isolated a firm is in a product space, the higher should be it's margin. Indeed such an instrument is very strong in our case. We see that the number of seats sold is strongly correlated with changes in the expected quality, the price has a negative impact. Columns 3 and 4 present regressions with revenue as the dependent variable. Regressions 1 and 3 use quality estimated with a quadratic cost function, while 2 and 4 polynomial the 5th degree.

Our measure of quality is strongly associated with the number of sold seats and with revenue. In order to investigate what is the implide output gap between minorities and non-minorities based on quality and other covariates, we compare the outcomes predicted by the model with the ones achieved in reality. Model predicts that outcomes for minorities should be higher than they really are, however, a large part of the gap is explained. Minorities underperform , conditioned on their quality, by 5.27%(4.4% in polynomial model) in the case of sold seats and by 6.93% (5.58%) for revenue. Hence, we conclude that part (roughly half) of the gap is due to differences in expected quality.

L Demand side:

M Robustness check: No zero-market shares

Our data is collected by choosing a route at random and collecting data on all available listings in a given moment. Therefore, in our sample there are listings that have been posted just a couple of seconds before we have seen it, and some that have been available

	Dependent variable:		
	mean_utility	revenue	taken_seats
	(1)	(2)	(3)
f_minority	−0.104*** (0.039)		
f_gender	−0.074*** (0.027)		
driver_age	−0.003*** (0.001)		
post_per_month	−0.025*** (0.008)		
ride_price	−0.167*** (0.026)		−0.012*** (0.004)
w_nls5		1.534*** (0.272)	0.038*** (0.011)
number_avis	0.035*** (0.006)	0.179*** (0.020)	0.006*** (0.001)
number_avis2	−0.001*** (0.0002)	−0.003*** (0.001)	−0.0001*** (0.00003)
seniority_months	−0.004*** (0.001)	−0.008*** (0.002)	−0.0003*** (0.0001)
posted_since	0.088*** (0.002)	0.246*** (0.007)	0.010*** (0.0003)
car_price		−0.018** (0.009)	−0.0002 (0.0003)
competition	0.028*** (0.0004)	0.019*** (0.001)	0.001*** (0.0001)
revenu_median	0.00004*** (0.00001)	0.0002*** (0.00002)	0.00001*** (0.00000)
public_transport_ratio	11.440*** (1.625)	−29.233*** (5.582)	−0.287 (0.214)
km	0.014*** (0.002)	0.007*** (0.0003)	0.001*** (0.0003)
night_dayday	0.039 (0.047)	0.440*** (0.163)	0.014** (0.006)
night_daynight	−0.375*** (0.076)	−1.266*** (0.262)	−0.043*** (0.010)
sncf_strike	1.103*** (0.063)	2.193*** (0.218)	0.101*** (0.008)
length_ride	0.002*** (0.0004)	0.007*** (0.001)	0.0002*** (0.00005)
picture		0.260* (0.141)	0.009 (0.005)
hours_till_ride	−0.005*** (0.0001)	−0.013*** (0.0003)	−0.001*** (0.00001)
acceptation_automatique	0.401*** (0.051)	2.142*** (0.086)	0.076*** (0.007)
weekday	−0.202*** (0.051)	−0.496*** (0.173)	−0.026*** (0.007)
consumption		0.271*** (0.059)	0.014*** (0.002)
driver_blabla	−0.026 (0.026)	−0.018 (0.089)	−0.002 (0.003)
December2017	0.989*** (0.088)	2.216*** (0.217)	0.064*** (0.012)
February2018	−0.778*** (0.061)	−1.112*** (0.173)	−0.068*** (0.008)
January2018	−0.634*** (0.074)	−1.238*** (0.185)	−0.080*** (0.010)
July2017	−0.320** (0.132)	0.594* (0.304)	−0.009 (0.018)
July2018	0.760 (0.820)	3.582 (2.644)	0.173* (0.101)
June2018	−0.341*** (0.113)	−2.217*** (0.380)	−0.049*** (0.015)
March2018	−0.917*** (0.052)	−1.647*** (0.166)	−0.081*** (0.007)
November2017	0.075 (0.085)	0.889*** (0.195)	0.008 (0.011)
October2017	0.107 (0.085)	0.558*** (0.171)	−0.005 (0.011)
September2017	0.631*** (0.135)	1.474*** (0.402)	0.020 (0.018)
night_dayday:weekday	0.075 (0.059)	0.014 (0.203)	0.001 (0.008)
night_daynight:weekday	0.069 (0.095)	−0.018 (0.328)	−0.006 (0.013)
Constant	−10.552*** (0.173)	−10.670*** (1.432)	−0.123** (0.059)
Observations	119,904	104,899	104,505
R^2	0.111	0.062	0.065
Adjusted R^2	0.111	0.061	0.065
Residual Std. Error	4.169 (df = 119870)	13.443 (df = 104866)	0.513 (df = 104471)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 20: Impact of expected quality. Standard errors not bootstrapped, to be added

Table 21: Impact of expected quality

	Dependent variable:			
	taken_seats		revenue	
	OLS	*instrumental variable*	*OLS*	*instrumental variable*
w_nls3	0.067*** (0.010)	0.045*** (0.017)	2.003*** (0.272)	1.562*** (0.579)
ride_price	−0.007*** (0.0004)	−0.006 (0.022)		
number_avis	0.003*** (0.0002)	0.003*** (0.001)	0.079*** (0.005)	0.078*** (0.006)
driver_blabla	−0.001 (0.003)	−0.001 (0.003)	−0.005 (0.085)	−0.010 (0.086)
seniority_months	−0.0003*** (0.0001)	−0.0003*** (0.0001)	−0.009*** (0.002)	−0.008*** (0.002)
hours_till_ride	0.0002*** (0.0001)	0.0002*** (0.0001)	0.004** (0.002)	0.004** (0.002)
posted_since	0.003 (0.002)	0.014 (0.011)	0.114*** (0.043)	0.328 (0.250)
car_price	0.0003 (0.0003)	0.0001 (0.001)	−0.003 (0.008)	−0.006 (0.009)
competition	0.0003*** (0.00005)	0.0004*** (0.0001)	0.007*** (0.001)	0.010*** (0.003)
revenu_median	0.00001*** (0.00000)	0.00001*** (0.00000)	0.0002*** (0.00002)	0.0002*** (0.00002)
duration_public_transport	−0.0001 (0.001)	−0.0004 (0.001)	−0.052** (0.021)	−0.054** (0.022)
km	0.0004*** (0.00003)	0.0003 (0.002)	0.006*** (0.0003)	0.007*** (0.001)
hour	0.002*** (0.0003)	0.002*** (0.001)	0.043*** (0.009)	0.053*** (0.015)
night_dayday	0.011*** (0.003)	0.013*** (0.005)	0.360*** (0.092)	0.413*** (0.111)
night_daynight	−0.050*** (0.006)	−0.049*** (0.006)	−1.416*** (0.154)	−1.378*** (0.161)
sncf_strike	0.052*** (0.007)	0.080*** (0.026)	1.124*** (0.192)	1.637*** (0.623)
length_ride	−0.0002*** (0.00004)	0.00001 (0.0002)	−0.003** (0.001)	0.0003 (0.004)
notice	−0.008*** (0.002)	−0.012*** (0.004)	−0.219*** (0.042)	−0.286*** (0.088)
acceptation_automatique	0.092*** (0.003)	0.086*** (0.029)	2.435*** (0.082)	2.260*** (0.218)
number_of_views	0.012*** (0.0001)	0.006 (0.006)	0.296*** (0.003)	0.178 (0.136)
Constant	−0.370*** (0.051)	−0.211* (0.111)	−15.165*** (1.369)	−11.941*** (3.972)
Observations	102,143	100,683	103,012	103,012
R^2	0.182	0.153	0.154	0.139
Adjusted R^2	0.182	0.153	0.154	0.139
Residual Std. Error	0.466 (df = 102110)	0.474 (df = 100650)	12.611 (df = 102980)	12.720 (df = 102980)
F Statistic	710.955*** (df = 32; 102110)		603.772*** (df = 31; 102980)	
Note:				*p<0.1; **p<0.05; ***p<0.01

for days. Furthermore, some of the listings which have been available for a given route, are not observed, because all seats have been sold. As a consequence, most of the listings that we have observed haven't sold a single seat so far. There are two importnat consequences of this, firstly we have a selected sample, where we expect to see listings that are on average less attractive than those that sold out (we have discussed this in Appendix E). Secondly, our demand model approximated choice probability with market share. Arguably, rides that have been just posted have a higher than zero choice probability. We have dealt with it two-fold, either by adding a small number to the market share, or by lumping all zero market shares to the outside option. Here, we provide another robustness check by using a supplementary dataset. Our objective is to argue that we observe zero market shares because of our data collection process rather than because some drivers have too low quality to be able to sell even at very low price.

We have used Blablacar API[23] to collect data on all rides that have been available on a given route in a given day, and their final performance measures, ie. how many seats have been sold overall by a given driver. API does not allow us to open drivers profiles, however, some of them we have observed earlier and have then in our core dataset. We select these drivers. Summary statistics are in table We can observe that average number of seats sold is 2.61, which is drastically higher than in our core dataset. In fact, less than 1% of drivers haven't sold any seat. Last three rows are measure of expected quality from the supply side, first, one with quadratic costs, the two latter ones with an approximation by a polynomial. We proceed by estimating a logit of demand, in an exactly the same way as before. Table 23 present estimates, with three measures of quality, we use measures of distance in characteristics space as instruments.

Point estimates of quality and price have expected signs; elasticities are higher than in the case of our core model,2.6% for the price and 7% for the quality, this is expected, as the market shares are much higer. Figure 24 shows predicted market shares for different leves of quality.

N Consumer welfare

The reputation system also has an impact on consumer surplus; efforts that are taken by drivers when they enter immediately benefit passengers. To take a look into this source

[23]Application Programming interface: https://dev.blablacar.com/

Table 22: Summary stats API data

Statistic	N	Mean	St. Dev.	Min	Max
ride price	4,042	29.45	13.94	6	76
driver's age	4,129	36.39	12.89	18	68
reviews (#)	4,129	9.10	6.87	1	30
talkative	4,129	2.18	0.46	1	3
listings (#)	4,129	17.94	24.77	2	471
minority	4,129	0.09	0.29	0	1
reputation	4,129	4.67	0.29	3.60	5.00
seniority (# months)	4,078	35.10	24.42	1	113
post per month	4,129	1.00	1.75	0.02	24.00
picture	4,129	0.95	0.21	0	1
bio (#words)	4,129	13.38	16.84	0	79
male	4,129	0.67	0.47	0	1
car price	3,278	5,632.79	4,702.20	902.30	24,403.34
fuel	3,384	4.91	0.71	3.65	7.40
competition	4,129	22.37	27.59	1	218
median revenue	3,941	19,066.05	2,004.57	13,263.48	26,214.00
duration public transport	3,970	217.78	125.62	26.00	914.19
km	4,087	375.24	179.14	41.21	881.65
notice	3,932	8.62	6.93	0	34
train strike	4,129	0.41	0.49	0	1
ride (#words)	3,957	15.55	24.11	0	126
automatic acceptance	4,129	0.46	0.50	0	1
travel cost	3,327	64.22	29.61	3.33	204.82
sold seats	4,129	2.61	0.79	0	4
total clicks	4,091	72.88	53.38	2	310
market share	4,034	0.22	0.13	0.0000	0.67
revenue	4,004	73.96	39.28	0	212
size	4,129	55.07	71.36	4	518
mean pop	4,129	4.53	0.03	4.41	4.75
variance pop	4,129	0.07	0.01	0.04	0.10
quality	4,086	4.64	0.14	4.19	4.87
quality I(3)	4,087	4.65	0.14	4.18	4.86
quality I(5)	4,087	4.65	0.14	4.18	4.86

Table 23: Logit demand, API dataset

	Dependent variable: mean utility		
quality	0.478** (0.234)		
quality I(3)		0.531** (0.233)	
quality I(5)			0.531** (0.233)
ride price	−0.189 (0.116)	−0.195* (0.117)	−0.195* (0.117)
competitors	−0.091*** (0.005)	−0.091*** (0.005)	−0.091*** (0.005)
seniority (# months)	−0.004*** (0.001)	−0.004*** (0.001)	−0.004*** (0.001)
median revenue	0.00003 (0.00002)	0.00003 (0.00002)	0.00003 (0.00002)
duration public transport	0.0002 (0.0004)	0.0002 (0.0004)	0.0002 (0.0004)
km	0.013 (0.009)	0.014 (0.009)	0.014 (0.009)
picture	0.183 (0.134)	0.183 (0.135)	0.183 (0.135)
notice	0.013** (0.006)	0.013** (0.006)	0.013** (0.006)
talkative	−0.041 (0.070)	−0.036 (0.070)	−0.036 (0.070)
automatic acceptance	−0.170* (0.097)	−0.174* (0.097)	−0.174* (0.097)
Constant	−3.603*** (1.167)	−3.869*** (1.160)	−3.870*** (1.160)
Observations	1,597	1,598	1,598
R^2	0.141	0.131	0.131
Adjusted R^2	0.135	0.125	0.125
Residual Std. Error	1.220 (df = 1585)	1.227 (df = 1586)	1.227 (df = 1586)
Note:			*p<0.1; **p<0.05; ***p<0.01

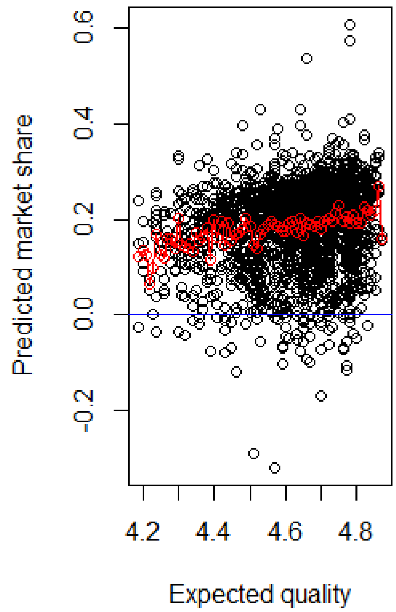


Figure 24: Predicted market share for different levels of quality

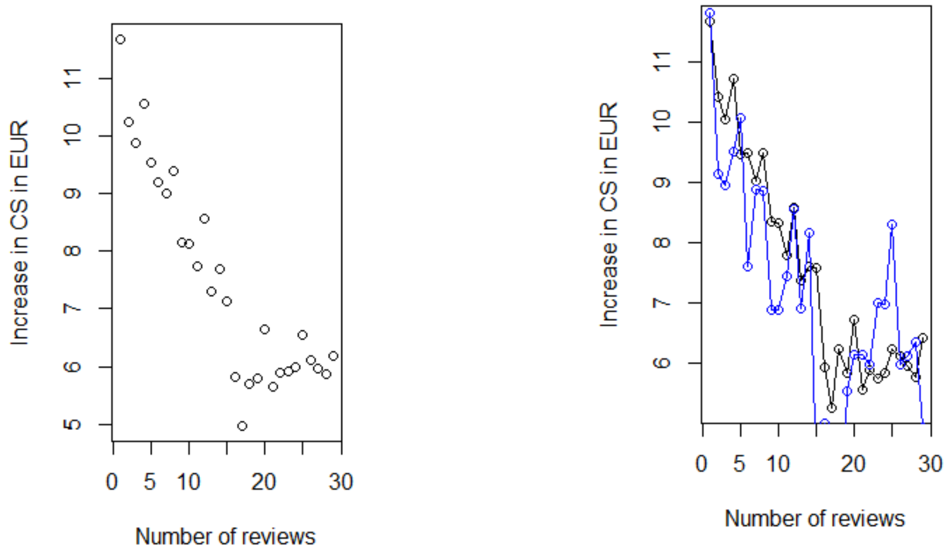
of increase in consumer surplus, we estimate logit demand systems from the previous section. We will compare consumers' surplus in a regime with a reputation system against a counterfactual one when drivers are not taking effort. Consumer surplus of a passenger j in market m is described with the following equation:

$$\mathbf{E}[CS_{j,m}] = -\frac{1}{\alpha}\log\left(\sum_{j\in m}\exp(\delta_m)\right) + C$$

,where α is a coefficient of price in the demand system, δ_m mean utility in market m. C is a constant that is not identified, but we will differentiate it out while making comparisons. Change in consumer surplus is described by:

$$\Delta\mathbf{E}[CS] = \mathbf{E}[CS^R] - \mathbf{E}[CS^{NR}] = -\frac{1}{\alpha}\left(\sum_{j\in m}\exp(\delta_m^R) - \sum_{j\in m}\exp(\delta_m^{NR})\right)$$

,where R stands for surplus under the reputation system and NR without one. We assume that prices in both cases are the same; hence the only change is the absence



(a) Mean change

(b) Black lines non-minority drivers; blue lines minority

Figure 25: Change in consumer surplus due to drivers' efforts

of effort when there is no reputation system. Thus, we calculate the value in euros of the additional quality provided to the passengers; its mean is 7.5EUR, which given the average price per ride of 28.8EUR is a significant increase for consumers. This also changes in time, as entrants provide the highest effort just after they have entered. Figure 25(a) shows the average increase of consumer surplus in time, while the panel 25(b) introduces a distinction between minority and non-minority drivers. There is a sizeable gain for passengers due to efforts taken by drivers; however, there are no significant differences between ethnic groups.