The Price of Trust, Female Participation and Ethnic Sorting in P2P Markets. Evidence from BlaBlaCar

Ignacio Berasategui Mugica *

March, 2026

Abstract

This paper explores the implications of profile information on female participation and ethnic sorting. In the context of the world's leading peer-to-peer car-sharing platform, BlaBlaCar, I leverage a novel data set that contains detailed information on users and transactions in the routes connecting eight of France's largest cities. Using a structural framework that accommodates pricing and sorting decisions for both market sides, this paper shows that women prefer to travel with other women and that there exists a substantial degree of ethnic-based homophily. This paper also provides evidence that alternative designs limiting the sorting abilities of users do not necessarily benefit ethnic minorities and that women from the ethnic majority tend to be the population segment whose participation and welfare reduces the most when anonymous marketplaces are imposed.

Keywords: Two-sided markets, Marketplace design, Ethnic Sorting, Female access.

JEL codes: C51, C54, D80, D82, J15, J16.

---

*Berasategui Mugica: ignacio.berasategui@psemail.eu. I am deeply grateful to Susanna Esteban and Gerard Llobet, for their guidance, as well as to Pedro Mira, Nezih Guner and Guillermo Caruana, whose help and advice have been essential. This paper also benefited from helpful comments from Mónica Martinez-Bravo, Dmitry Arkhangelsky, Manuel Arellano, and Samuel Bentolila. The construction of the data set at the core of the paper has been only possible with the initial support of Junhui Yang and due to the invaluable help of the families Berasategui Mugica, Shalaf Arrizabalaga, Párraga Estraviz and Leizea Nieves. Finally, I thank the time and information provided by the management team of BlaBlaCar Iberia.

1 Introduction

Policy-makers and platform designers determine the information market participants can observe about potential counterparts. From picture-less CVs to reputation systems, various tools are used to limit the characteristics on which agents can base their decisions. In labor markets and educational settings, for example, a common intervention is to suppress identifying information to reduce sorting based on observable characteristics. The logic is immediate: if agents cannot observe traits such as gender or ethnic origin, they cannot condition their choices on them, improving outcomes for disadvantaged groups.

However, when such measures are implemented in decentralized markets, the equilibrium effects extend beyond sorting. Specifically, prices adjust when agents cannot directly select their preferred counterparts. This adjustment seems to be particularly important in two-sided markets, as both buyers and sellers care about the characteristics of their trading partners. In such environments, restricting information affects not only who matches with whom, but also the prices required to clear the market. Whether more anonymous market designs ultimately improve outcomes for the groups they are intended to protect should be, thus, regarded as an open empirical question.

This question is especially relevant in peer-to-peer markets, where individuals frequently interact offline with unknown counterparts. Platforms for home sharing, ride sharing, and car sharing typically display profile information—including names and photographs—that allows users to infer characteristics such as gender or ethnic origin. While these features may facilitate discriminatory behavior, they may also help users reduce uncertainty and coordinate with preferred counterparts. As a result, policies that make profiles more anonymous may affect not only sorting patterns, but also market participation and welfare across groups. These concerns are not merely theoretical: following an enforcement action by the California Department of Fair Employment and Housing, Airbnb modified its platform design to limit the prominence of users' names and profile pictures.

This paper studies how observable characteristics shape sorting in peer-to-peer markets and how platform designs with limited sorting tools affect participation and welfare. I focus on two characteristics that are both salient and observable in these environments: gender and ethnic origin. Importantly, I study how preferences depend on the characteristics of *both* sides of the market. This allows me to evaluate policies that alter the information environment in a genuinely two-sided setting. Also importantly, I evaluate the effects of these policies highlighting the interaction between prices and information.

The analysis uses novel web-scraped transaction-level data from BlaBlaCar France, the largest long-distance car-sharing platform in the country. BlaBlaCar allows non-professional drivers to offer seats in their vehicles to prospective passengers and to freely set the price for the trip. The data contain detailed information on published trips and include the gender and ethnic origin of both drivers and accepted passengers. This feature of the data makes it possible to study sorting jointly on both sides of the market, an aspect still understudied in the empirical literature.

An important institutional feature of the platform is that drivers can choose between two booking modes when publishing a trip. In the *automatic* mode, passengers can immediately book a seat without driver approval. In the *manual* mode, passengers must first request to join the trip and the driver decides whether to accept or reject the request.

I develop and estimate a structural model that accommodates the main strategic decisions of drivers and passengers. On the driver side, the model endogenizes the booking mode, pricing, and sorting choices. On the passenger side, individuals decide whether to request to join a trip, anticipating both the probability of rejection and the associated penalty. The model highlights the two-sided nature of the market: when drivers choose the manual mode, a match occurs only if both the passenger requests the trip and the driver accepts the request.

Estimating preferences in this environment presents a central identification challenge: transaction data only reveal passengers who are ultimately accepted into a trip. I address this challenge by exploiting variation in drivers' booking modes. Differences in the composition of realized trips across the automatic and manual modes provide information on drivers' selection behavior and allow me to recover preference parameters for both passengers and drivers.

The estimates reveal substantial heterogeneity in preferences across users. On the passenger side, women prefer female drivers, and passengers across all groups display a substantial degree of ethnic homophily. Specifically, women are willing to pay 7% more on average to travel with a female driver, and both majorities and minorities are willing to pay, on average, 20% more to travel with a driver of the same ethnic group. On the driver side, there is an overall preference for female passengers and for passengers belonging to the ethnic majority, although the latter preference is weaker among minority drivers.

I then use the estimated preference parameters to conduct counterfactual exercises that evaluate how alternative platform designs affect market outcomes. The results show that reducing information about gender and ethnic origin significantly changes participation patterns. In particular, women on both sides of the market experience the largest reductions in participation and surplus when blind profiles are imposed and/or when drivers are forced into the automatic booking mode. Moreover, policies that remove identifying information do not systematically improve outcomes for ethnic minority users. Specially in those routes in which the share of minority passengers is highest, the price increase by majority drivers reduces, and in some extreme cases nets-out, the welfare-gains from reduced sorting.

I then use the estimated preference parameters to conduct counterfactual exercises that evaluate how alternative platform designs affect market outcomes. Removing profile information and forcing the automatic mode reduce the scope for direct sorting but also induce equilibrium price adjustments. In particular, when the automatic mode is imposed, majority drivers increase prices in response to the higher expected probability of traveling with minority passengers. These price responses are strongest on routes where the share of minority passengers is highest. As a result, more opaque marketplace designs generate substantial welfare losses for women on both sides of the market. Since they are the population segment with starker preferences and a lower probability of rejection, they experience the largest reductions in participation and surplus: their acceptance rates vary little, they face higher prices, and they cannot sort less-desirable counterparts. For ethnic minorities, the effects are more nuanced: while reduced sorting may generate direct welfare gains, the accompanying price increases can substantially reduce these gains and, in some markets with a high share of minority passengers, fully offset them.

These results highlight a central trade-off in the design of information environments. Limiting observable characteristics may reduce some forms of sorting, but it may also discourage participation among users who rely on such characteristics to mitigate safety concerns or other forms of uncertainty. The effects of anonymity are therefore not uniform across neither groups nor markets. More broadly, the results show that evaluating policies that restrict information requires accounting for the preferences and strategic responses of both sides of the market.

Related literature. This paper relates to three strands of literature. First, it contributes to the literature studying how preferences over counterpart characteristics affect both sorting patterns and equilibrium prices. Starting with the seminal work of Becker (1957), economists have emphasized that discrimination and sorting in markets may affect not only who trades with whom, but also the prices at which transactions occur. A subsequent structural literature has studied how heterogeneous preferences and search frictions jointly determine matching patterns and wage dispersion in labor markets (Burdett & Mortensen, 1998; Shimer & Smith, 2000; Postel-Vinay & Robin, 2002; Eeckhout & Kircher, 2011). In these environments, equilibrium prices adjust as agents compete for preferred matches. This paper contributes to this literature by showing that when information about observable characteristics is restricted, equilibrium adjustments may occur not only through changes in sorting but also through price responses that reshape participation and welfare across groups.

Second, the paper contributes to the growing empirical literature documenting discrimination and sorting in peer-to-peer markets (Pope & Sydnor, 2011; B. G. Edelman & Luca, 2014; Tadelis, 2016; Laouenan & Rathelot, 2017; Kakar et al., 2016; B. Edelman et al., 2017; Tjaden et al., 2018; Lambin & Palikot, 2018; Farajallah et al., 2019; Ivaldi & Palikot, 2020). These studies show that observable characteristics such as gender and ethnic origin affect outcomes on platforms such as housing- and ride-sharing services. Existing work, however, typically focuses on the characteristics of one side of the market—most often suppliers—while abstracting from the characteristics of the counterpart. By observing both drivers and passengers, this paper shows how sorting depends on the interaction between the characteristics of the two agents and how changes in information disclosure affect not only matching patterns but also prices, participation, and welfare.

Third, the paper relates to the literature on sorting and matching in two-sided markets (Burdett & Coles, 1999; Choo & Siow, 2006; Adachi, 2003; Hitsch et al., 2010; Browning et al., 2014). The most closely related paper is Hitsch et al. (2010), which estimates preferences in a dating market where both requests and acceptance decisions are directly observed. In contrast, I study a platform environment in which agents also choose prices and booking rules, and where explicit rejection decisions are only partially observed. I therefore develop an identification strategy that recovers preferences from variation in booking modes and the composition of realized matches. By doing so, the paper brings the analysis of two-sided sorting to a setting in which platform design directly shapes both the information available to participants and the equilibrium prices that emerge in the market.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting and the information available to users on the platform. Section 3 presents the data and variable construction. Section 4 provides descriptive evidence on sorting patterns. Section 5 introduces the model. Section 6 describes the estimation strategy. Section 7 presents the main results. Section 8 evaluates counterfactual platform designs that limit profile information. Section 9 concludes.

2 Overview of BlaBlaCar

BlaBlaCar is the leading car-sharing platform in Europe and France. Through BlaBlaCar, non-professional drivers and passengers meet and share a trip between a common origin and destination. Although BlaBlaCar presents a dominant market position in multiple European countries (Spain, Portugal, Germany, Italy,...), it is France, its birthplace, where the platform has found the most success. In this country, BlaBlaCar has 20 million members and is used by 60% of the population between the ages of 18 and 35[1]. Moreover, the platform has been promoted by the French ministry itself and encouraged during pandemic times as a safe mean of transport.

BlaBlaCar emerged in early 2006, as an alternative to traditional means of transport for medium-length trips (c. 400 kilometers on average). The platform promotes the values of efficiency and environmentally friendly practices as core principles of its business model.

The platform allows two types of agents to operate through its system, drivers and passengers. To participate in the platform, both drivers and passengers need first to generate a personal profile. These profiles contain three main pieces of information. The first block refers to the personal information of the user, including the name and the age. Moreover, the platform encourages heavily users to provide an updated profile picture of themselves. In case the user decides not to upload a personal picture, the platform generates a default picture containing an avatar that identifies his or her gender. The second block shares the series of checks that the platform performs about the veracity of the profile. These checks refer to the verification of the mobile phone and mail address, and are publicly available in the profile of the user. The third block summarizes the review system, which feeds on the opinions and grading of other users based on previous experiences regarding trips arranged through the platform. Aside from the detail of all the reviews, the main profile page of the user displays a summary containing the number of received reviews and the average grading, which is established on a scale from one to five.

On the supply side, drivers publish their trips, including the place of origin, destination and time of departure. Drivers choose the number of seats they publish through the platform and they decide whether to leave one space between seats at the back (which the platform signals establishing that the driver has chosen the comfort mode). Each driver chooses a unique price for all the published seats. Finally, drivers have to choose between two booking modes: the manual booking mode and the automatic booking mode.

[1] https://blog.blablacar.com/newsroom/news-list/blablacar-reaches-100-million-members-for-its-15th-anniversary

In the latter, if any passenger decides to request to join the driver, the user automatically becomes a part of the trip. When the booking mode is manual, the driver can reject at will any of the received requests. Although the time that the platform provides to the driver so as to reject a request varies depending on the proximity to the date of departure, it is considerably large and, in most cases, extends until this date.

Passengers, once they have generated a profile, can contact any of the publishing drivers and make a request for any of the available trips. To do so, they have access to all the previous information about the trip and the publishing drivers. The platform provides a one-to-one private chat, that allows prospective passengers to clarify and detail, with each driver, any information regarding the trip that is not available: the exact collection point, the amount of luggage allowed, the number of stops, the route and the usage of toll roads or the possibility of eating in the car, to mention a few. Once a request has been sent and accepted, the passenger incurs a penalty if he decides to cancel it. If the cancellation takes place twenty-four hours before departure, the passenger loses the full purchase price. If the cancellation takes place prior to this time frame, the penalty varies depending on the time between the cancellation and the departure.

3 Data set

The data set contains information on all the trips taking place through BlaBlaCar between 8 of the biggest French cities (Paris, Lyon, Strasbourg, Marseille, Nice, Nantes, Toulouse and Bordeaux). The transactions refer to trips with a date of departure between the 27th of October 2020 and the 24th of March 2021[2]. It contains information on all the trips taken place on these routes and during these times, even if all the seats remained empty at the time of departure. To develop the data set, BlaBlaCar´s API and a series of different web scraping techniques were employed.

Regarding the existing literature in the field of ethnic discrimination in peer-to-peer markets, this data set presents a remarkable feature. This feature is crucial to analyze the influence, in the decisions of passengers and drivers, of the interaction between the gender and ethnic origin of the user herself/himself and the gender and ethnic origin of the counterpart. In addition to detailed information about all the published trips (price,

[2] The data set refers to transactions taking place in a time and space affected by the Covid pandemic. Although most limitations on free movement had already been removed and despite the French minister of transport ensuring the safety of BlaBlaCar, it is a very particular period that may have conditioned the actions of the different population segments in various ways. In Appendix F, new data on transactions taking place in the same time frame in the year 2022 is analyzed. I find that, although the number of trips increases more than twice, neither the average fraction of seats sold nor the composition of the sampled drivers changes significantly in the comparable time frame.

booking mode, number of seats originally published, comfort mode, car brand, exact place of origin, exact destination and exact date and time of departure), this is the first data set to contain the profile information (name, age, profile picture, mail and phone verification, number of reviews and average rating) of the participating users from both market sides. That is, the resulting database contains the profile information, on the one hand, of all the drivers that have published a trip in the referred routes and time frame (even if no passenger joins them) and, on the other hand, of all the passengers that have effectively joined any of these drivers in their trips and the conditions under which they joined them. In this sense, I have access to the same information set that users on each side of the market face when they take a decision.

This feature entails a significant development over existing data sets, not only in the particular setting of BlaBlaCar, in which researchers have been able to gather real market information exclusively on the identity of the driver and the trip (Lambin & Palikot, 2018; Farajallah et al., 2019; Ivaldi & Palikot, 2020), but over the majority of the peer-to-peer market literature (B. G. Edelman & Luca, 2014; Kakar et al., 2016; Laouénan & Rathelot, 2022), as these papers are also based on transaction information (the posted prices) of exclusively one side of the market. With regards to B. Edelman et al. (2017), in the context of Airbnb the authors make use of experimental variation in the names of artificially generated profiles to test for racial differences in the rejection probability. Hence, they focus, in a *quasi* experimental setting, only on the choices of one side of the market. Even within the supplying agents, they consider exclusively part of the sorting decisions of a fraction of this side of the market: the acceptance/rejection of the manual hosts.

Gender and ethnic origin classification is performed following a similar procedure to that employed in Lambin & Palikot (2018). In terms of gender, users are classified as men or women. In terms of ethnic origin, users are classified as of the ethnic majority if they belong to northern, southern and central Europe, or of the ethnic minority, if they belong to any other country (although in my data set a compelling majority of those classified as of the ethnic minority are of African and Middle Eastern origin[3]).

To determine the gender and ethnic origin of the user, firstly, the declared name is used. The French ministry provides a list of common names in France, with the associated gender and ethnic origin. If the name matches any of the existing ones in the list, the user is classified with the associated gender and ethnic origin. When it does not match, the gender and ethnic origin is individually verified using the profile picture of the user.

[3]Shortly, as the data set expands over a larger time frame and I have access to more accurate artificial intelligence classification technology, this research will hopefully be able to distinguish amongst different minority groups.

If no profile picture is available, the ethnic origin is determined using similar names of other users with profile pictures, alternative name lists and Google searches. When no information is available (less than one percent of the total number of users), users are classified as of the ethnic minority. The residual classification rule of users to the ethnic minority is informed by the construction and detail of the name list[4].

As in Lambin & Palikot (2018), facial recognition software is used over a sample of users with profile pictures. Due to the generally low quality of the images, the probability of incorrect classification is very high (more than twenty-five percent of the pictures are erroneously classified), forcing the classification of the pictures to be performed individually by the author of the paper. To test the validity of the name classification, a random sample of 100 individuals for each group of interest is extracted (men and women of the ethnic majority and minority) and I manually classify them according to the gender and ethnic origin observed in their profile pictures. None of the four random samples presents an error larger than three percent. The only exception is the gender classification of women of the ethnic minority (error of eight percent). The closeness between certain masculine and feminine name variants and the larger prevalence of minority names amongst women of the ethnic majority are the reasons behind this error.

Table 1 contains the summary of key statistics. On the supplying side, men of the ethnic majority are the most active, accounting for more than 50% of the published trips. Women of the ethnic minority, on the opposite, have a very small presence, representing less than 3% of the total population of published trips. Booking mode-wise, women and ethnic majorities make the most frequent use of the manual mode. The average price is similar across all driver types, although slightly higher amongst the ethnic minority.

In terms of the average passenger presence per published seat, Table 1 shows that female passengers of the ethnic majority are more prevalent, on average, in cars driven by women and by ethnic majorities. Male passenger distribution seems to follow a similar pattern, although these differences, and especially the gender gap, are attenuated. Women of the ethnic minority, also on the demanding side of the market, have an overall limited appearance, and they are present evenly across the different segments of drivers. The presence of men of the ethnic minority follows the opposite pattern to that of women of the ethnic majority: they are more prevalent in cars driven by ethnic minorities and by men. Finally, it is worthy of notice that the overall proportion of empty seats is higher for minority drivers than for their majority counterparts, as a result of the smaller prevalence

[4] French and other central European names are covered in detail, with their multiple variations, detailing the specific region, within the country, to which they are original. Ethnic minority names often do not have a uniformly accepted translation and consequently, can be seen written in various ways. Moreover, they are covered with inferior precision in terms of the variants and regions of origin.

Table 1: Summary of key statistics for all routes and dates

	Fema - Maj Driv.	*Male - Maj Driv.*	*Fema - Min Driv.*	*Male - Min Driv.*
Number of trips	27,741	50,969	2,163	14,072
Manual booking mode	0.74	0.67	0.65	0.58
Price (euros)	23.26 (9.19)	24.56 (9.55)	27.07 (10.70)	28.28 (10.19)
Trip price/km (euros)	0.66 (0.12)	0.66 (0.12)	0.70 (0.14)	0.68 (0.13)
Fema-Maj Pas. per seat	0.24 (0.31)	0.21 (0.29)	0.15 (0.26)	0.12 (0.23)
Male-Maj Pas. per seat	0.19 (0.28)	0.18 (0.27)	0.12 (0.23)	0.11 (0.22)
Fema-Min Pas. per seat	0.03 (0.12)	0.03 (0.12)	0.04 (0.13)	0.04 (0.12)
Male-Min Pas. per seat	0.05 (0.15)	0.06 (0.16)	0.07 (0.17)	0.08 (0.18)
Empty per seat	0.49 (0.45)	0.51 (0.46)	0.62 (0.39)	0.65 (0.38)

Notes: The table summarizes various statistics for the four different population segments of drivers (females of the ethnic majority, males of the ethnic majority, females of the ethnic minority and males of the ethnic minority), reporting one column for each segment. The upper part of the table reports the number of published trips, the proportion of the manual mode users and the price, in absolute (measured in euros) and relative terms (measured in euros per kilometer). The lower part of the table reports the average proportions of passengers belonging to each population segment and empty seats. Standard deviations are reported in parenthesis.

of majority passengers in the trips of the former.

4 Motivating evidence

Table 1 presents evidence supporting the idea that the gender and ethnic origin of the passenger and the driver plays a role in the choices of both agents. Leaving endogeneity concerns aside, this evidence suffers from three types of limitations. Firstly, it does not control for other systematic sources of variation that could be correlated to the gender and/or ethnic origin of the users (e.g., female passengers of the ethnic majority are less prevalent in those cars driven by minority drivers because they have low reputation and experience in the platform). Secondly, it does not account for likely route composition effects. In those routes connecting cities with a larger minority population, drivers and passengers belonging to the ethnic minority will have a larger representation than in the rest of the routes. Mechanically, this will raise the proportion of passengers of the ethnic minority that travel with ethnic minority drivers, even if they have the same preferences towards majority riders. Thirdly, the previous exercise does not consider that the proportions of passengers that join each population segment of drivers are the result of the request of the passenger and the acceptance decision of the driver, if the latter has chosen the manual mode (e.g., male passengers of the ethnic minority may be less prevalent in cars driven by female drivers of the ethnic majority because they make fewer requests or because these drivers accept them in a smaller proportion).

This section aims to correct the first two limitations and motivate two main points. On the one hand, both users on the demanding and supplying sides choose to request and reject based, amongst others, on the gender and ethnic origin of their counterparts. On the other hand, their decisions are modulated by their own gender and ethnic origin and the interaction between these features and those of the users on the opposing market side.

Subsection 5.1. focuses on the features of matching passengers and drivers irrespective of the booking mode employed by the latter, assessing the prevalence of each passenger population segment, which is denoted by k, in terms of the features of the offered trips and their drivers (including the gender and ethnic origin of the latter). Subsection 5.2. extends the analysis so that it also considers the booking mode choice of the driver and its interaction with the features of the trip and the driver.

Resulting from the interaction of gender and ethnic origin, four segments of passengers are distinguished: women of the ethnic majority, women of the ethnic minority, men of the ethnic majority and men of the ethnic majority.

4.1 Prevalence analysis without booking mode

To assess the relationship between the prevalence of each passenger segment and the different features of drivers and trips, I propose the following specification,

$$Y_{j,k,t} = X_{j,t}^{P\prime}\beta_k + \rho_{k,t} + \xi_{j,k,t}. \tag{1}$$

The unit of observation is the published trip j undertaken in a particular route and time t. For each passenger segment k, a different regression is performed. Each regression analyses the variation in the proportion of accepted passengers belonging to a specific population segment, given the features of each trip and publishing driver. $Y_{j,k,t}$ is the ratio, per published seat, of passengers of segment k that join trip j in route and time t.[5] $X_{j,t}^{P\prime}$ is a vector of the features considered by the passenger to make a request, both features of the trip (including price, seats, comfort mode, car value and distance from the city center) and of the driver (gender, ethnic origin, age, average rating and number of received reviews). $\rho_{k,t}$ are route and day-fixed effects specific to each passenger segment.

One of the main concerns of this type of analysis is its ability to overcome the natural sorting of driver and passenger population segments across the different routes. For example, if drivers and passengers of the ethnic minority tend to be present on longer routes (perhaps because they lack the financial possibilities to use more expensive and better-suited means of transport for longer trips) or if ethnic minority drivers are more prevalent on routes that connect cities with a higher proportion of minority citizens, trips published by minority drivers will be filled, on average in a higher proportion, by minority passengers. That is, even if these passengers do not have any preference towards the ethnic origin of the driver, they will naturally have a higher prevalence amongst trips published by drivers of similar features. To overcome this concern, route-fixed effects are allowed to change with the passenger population segment. Therefore, when passengers belonging to a certain segment are present in a higher proportion at a certain route, the route fixed effect parameter will capture the difference with respect to the mean of all the routes, allowing the coefficient of the ethnic minority condition of the driver to capture exclusively the prevalence of this segment in cars driven by users of the ethnic minority, all else equal.

Table 2 shows a well-defined pattern in terms of the prevalence of each passenger segment, given his/her own gender and ethnic origin and those of the driver. From a gender perspective, women passengers tend to be significantly more prevalent in cars driven by women. Men of the ethnic majority tend to be slightly more prevalent in cars driven by women than by men, while men of the ethnic minority present the opposite pattern.

Regarding the ethnic origin of the driver, even after controlling for route composition differences, minority passengers tend to be more prevalent in cars driven by minority drivers. Passengers of the ethnic majority present the exact opposite pattern, as they have a higher representation in cars driven by men and women of the ethnic majority. Described empirical evidence suggests that ethnic-based homophily informs majority and minority passenger choices. Additionally, these results point to gender playing a crucial role in passenger choices, acting both as a modulator of ethnic homophily and as a feature that has opposite perceptions for the different population segments.

[5] Other outcomes such as the total number of passengers are analyzed in alternative specifications, remaining the main results unaltered.

Table 2: Matching Evidence. Without Booking Mode

	Fema-Maj Pas. (1)	*Male-Maj Pas.* (2)	*Fema-Min Pas.* (3)	*Male-Min Pas.* (4)
Price	-0.0024*** (0.0001)	-0.0012*** (0.0000)	-0.0006*** (0.0001)	-0.0009*** (0.0001)
Female Driv.	0.0191*** (0.0020)	0.0065*** (0.0019)	0.0034*** (0.0001)	-0.0027*** (0.0012)
Minority Driv.	-0.0523*** (0.0025)	-0.0398*** (0.0023)	0.0030*** (0.0011)	0.0145*** (0.0014)
Reputation Driv.	0.0284*** (0.0009)	0.0226*** (0.0009)	0.0049*** (0.0004)	0.0078*** (0.0005)
Experience Driv.	0.0001*** (0.0000)	0.0001*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.0494	0.0720	0.0107	0.0257
N	94,945	94,945	94,945	94,945

Notes: The unit of observation is the trip. The four dependent variables, in Columns (1)-(4), are the proportions of accepted passengers belonging to each population segment (females of the ethnic majority, males of the ethnic majority, females of the ethnic minority and males of the ethnic minority), with respect to the initial number of published seats. The variable price is in euros, including the platform's commission. The variable female driver is a dummy that takes value one if the driver is identified as a female and zero otherwise. The variable minority driver is a dummy that takes value one if the driver belongs to the ethnic minority. The variable reputation is the driver's rating, measured on a scale from 1 (very bad) to 5 (excellent). The variable experience is the number of reviews the driver has received through the platform from previous trips. The set of controls includes other information on the trip and the driver: driver's age, availability of personal picture, car value, comfort mode, number of initially published seats, the time between publication and departure, distance from the city center in the city of departure and arrival, and others. The four regressions include route and time-fixed effects. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.2 Prevalence analysis with booking mode

The previous exercise analyses the observed proportions of traveling passengers in terms of the features that passengers face when choosing to join a trip, that is, characteristics of the trip and the driver, and the passenger's own gender and ethnic origin. Nevertheless, the observed proportions of traveling passengers are the result of two conditions: the request of the passenger and, when the manual booking mode has been established, the acceptance decision of the driver. Thus, for those observations in which the manual booking mode has been chosen, covariates referring to the features that the manual driver may consider in his/her decision need to be controlled.

Along these lines, I propose the following specification,

$$Y_{j,k,t} = Manual_{j,t} \cdot X_{j,t}^{D\prime}\lambda_k + \eta_{k,t} + X_{j,t}^{P\prime}\beta_k + \rho_{k,t} + \xi_{j,k,t}. \tag{2}$$

In contrast to the previous exercise, two new types of parameters emerge. Given that only manual drivers can choose whether to accept or reject a request, the newly introduced covariates interact with a dummy variable $Manual_{j,t}$, which takes value one if the manual booking mode has been chosen in trip j at time and route t, and zero otherwise. These covariates mirror those features that influence the passenger choice but for the driver. In particular, $X_{j,t}^{D\prime}$ contains the features of the trip, same as in the passenger case (including price, seats, comfort mode and car value), interacted by the features of the population segment to which the driver belongs (defined by the gender and ethnic origin). $\eta_{k,t}$ are route and time-fixed effects interacted by the manual dummy variable.

In terms of passenger prevalence in relation to trip and driver features aside from the booking mode, Table 3 presents almost the same conclusions as Table 2. The only remarkable difference is that men passengers seem to be equally prevalent in cars driven by men and by women.

In what refers to the terms interacted by the booking mode of the driver, coefficients show that all segments of users but women of the ethnic majority are less prevalent in those trips in which the driver has chosen the manual booking mode. These coefficients also show that manual drivers from the ethnic majority tend to present a smaller proportion of male passengers of the ethnic minority than comparable manual minority drivers, suggesting that drivers' preferences do also depend on their own gender and ethnic origin.

To summarize, the analysis of the prevalence of passengers, given their features and those of the drivers they join, provides suggestive evidence in line with the two initial conjectures: users on both sides of the market have a preference for their counterparts'

Table 3: Matching Results. With Booking Mode

	Fema-Maj Pas. (1)	*Male-Maj Pas.* (2)	*Fema-Min Pas.* (3)	*Male-Min Pas.* (4)
Price	-0.0018*** (0.0001)	-0.0018*** (0.0000)	-0.0011*** (0.0001)	-0.0020*** (0.0001)
Female Driv.	0.0221*** (0.0038)	-0.0000 (0.0034)	0.0027* (0.0016)	-0.0063*** (0.0021)
Minority Driv.	-0.0518*** (0.0041)	-0.0402*** (0.0038)	0.0035** (0.0018)	0.0147*** (0.0023)
Manual	0.1275*** (0.0294)	-0.0482** (0.0270)	-0.0217** (0.0126)	-0.0606*** (0.0167)
Manual x Female Driv.	-0.0171 (0.0154)	0.0136 (0.0141)	0.0078 (0.0066)	0.0000 (0.0087)
Manual x Minority Driv.	0.0141 (0.0200)	-0.0171 (0.0154)	-0.0065** (0.0032)	0.0430*** (0.0114)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.0506	0.0730	0.0118	0.0278
N	94,945	94,945	94,945	94,945

Notes: The unit of observation is the trip. The four dependent variables, in Columns (1)-(4), are the proportions of accepted passengers, belonging to each population segment (females of the ethnic majority, males of the ethnic majority, females of the ethnic minority and males of the ethnic minority), with respect to the initial number of published seats. The variable price is in euros, including the platform's commission. The variable female driver is a dummy that takes value one if the driver is identified as a female and zero otherwise. The variable minority driver is a dummy that takes value one if the driver belongs to the ethnic minority. The variable manual is a dummy that takes value one if the driver has chosen the manual booking mode and zero otherwise. The set of controls includes other information on the trip and the driver: driver's age, availability of personal picture, car value, comfort mode, the number of initially published seats, time until departure, distance from the city center, and others. The four regressions include route and time-fixed effects. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

gender and ethnic origin, and these preferences are modulated by their own gender and ethnic origin.

Despite encouraging, this evidence *per se* is not enough to produce any claim concerning the specific preferences of the agents or the consequences of a more opaque marketplace. Moreover, as it has become apparent in this section, from the information contained in the data set, disentangling the decisions and preferences of both market sides is a complicated task. To do so, developing a properly micro-founded structural model becomes critical. Understanding the relationship between the multiple endogenous choices (booking mode, pricing, acceptance/rejection) taking place through BlaBlaCar can only be done with a structure that accommodates all these strategic interactions in a unified framework. Such a framework will enlighten the design of the counterfactuals and will clarify the role that preferences play in the multiple strategic decisions adopted by each market side.

5 The model

5.1 Setting

This is a partial equilibrium model with two types of users: drivers D and passengers P. Each user belongs to a population segment k.

The timing of the model is as follows. In each market t, one seat is published by driver j.[6] In the first stage of the game, the driver chooses the booking mode $bm_{j,t}$, which can be either manual MAN or automatic $AUTO$ (stage 1). The manual booking mode entails a driver-specific fixed cost $fc_{j,t}$, reflecting the additional knowledge and effort that the driver requires to make the manual mode fully functional and, consequently, to benefit from the most advanced features enabled by the platform.

After choosing the booking mode, in the second stage, the driver sets the price $p_{j,t}$ (stage 2). For exposition purposes, the booking mode and pricing decisions are assumed to be undertaken sequentially by the driver.

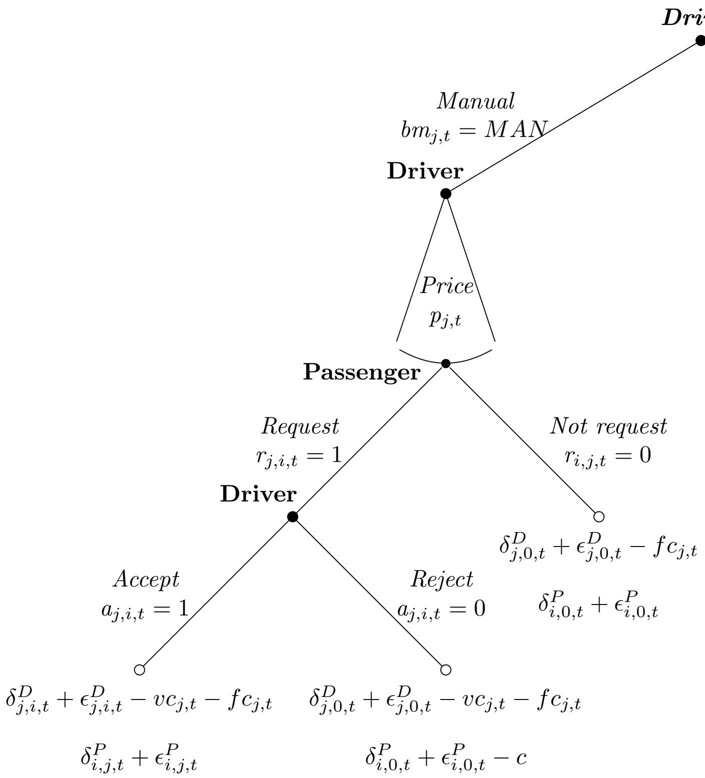
Once the price and booking mode have been established, one passenger i enters the market.[7] With probability $q_{k,t}$ the passenger belongs to population segment k. This exogenous probability $q_{k,t}$ resembles the fact that different routes present differing proportions of each passenger segment in the total population. Knowing the features of the driver and the trip, including the price and booking mode, in the third stage of the game, the passenger decides whether to make a request $r_{j,i,t} = 1$ to the driver or choose the outside option $r_{j,i,t} = 0$. The latter consists of a composite good that bundles the possibility of staying at home or using an alternative transportation mean (stage 3).

If the passenger makes the request and the driver has chosen the manual booking mode,

[6] The model assumes there is no entry decision for the driver. I expand on the consequences of endogenizing this decision in the counterfactual section.

[7] The "one passenger per published seat" assumption avoids the estimation problems that stem from the rationing rules resulting from binding capacity constraints. Other potential modeling routes have been evaluated, but they are computationally more demanding. In particular, a model of sequential passenger entry has been estimated with almost identical results in terms of the sign and relative size of the preferences. The inclusion of route and time-fixed effects in the preferences of users captures the differences in market sizes, alleviating concerns of estimation biases with regard to this assumption.

Figure 1: Setting



in the fourth stage, the driver decides whether to accept $a_{i,j,t} = 1$ or reject $a_{i,j,t} = 0$ the petition (stage 4). Rejection results in the driver making the trip with the seat empty and the passenger being forced into the outside option.

Prior to the rejection/acceptance decision of the manual driver, each driver faces variable cost $vc_{j,t}$. This cost resembles the time and wealth expenditures that the driver needs to incur in order to have the seat and passenger fully prepared to make the trip. Hence, it expands from the time cost of answering the clarifying questions of prospective passengers to the expenditures required to have the seat clean and the trunk cleared.

If making the trip together, the driver and the passenger obtain utility $u^D_{j,i,t}$ and $u^P_{i,j,t}$, respectively. That is, if the passenger makes a request to an automatic driver or if the passenger makes a request to a manual driver and the driver accepts the request, they will receive $u^D_{j,i,t}$ and $u^P_{i,j,t}$, respectively. The utility is composed of two additive terms, the mean utility, denoted as δ, that depends on the features of the trip (price, route, seats, car value, comfort mode, and others that will be detailed in the following subsections), and the features of the deciding user and the counterpart (in particular, their gender and ethnic origin)[8]; and the taste shock, denoted by ϵ, which is specific to the user, the counterpart and the market,

$$u^D_{j,i,t} = \delta^D_{j,i,t} + \epsilon^D_{j,i,t}, \tag{3}$$

$$u^P_{i,j,t} = \delta^P_{i,j,t} + \epsilon^P_{i,j,t}. \tag{4}$$

When the trip does not take place, it can be due to the passenger deciding not to make the request or the driver deciding not to accept the request. In case of the passenger deciding not to make the request, the agents obtain the utilities associated with their outside options $u^D_{j,0,t}$ and $u^P_{i,0,t}$,

$$u^D_{j,0,t} = \delta^D_{j,0,t} + \epsilon^D_{j,0,t}, \tag{5}$$

$$u^P_{i,0,t} = \delta^P_{i,0,t} + \epsilon^P_{i,0,t}. \tag{6}$$

When the passenger is forced to the outside option due to the rejection of the manual driver, in addition to the mean utility of the outside option and the taste shock, the passenger incurs a cost c. This cost reflects the associated penalties that the passenger may suffer from the rejection, such as the hold-up cost until knowing the decision of the driver (e.g., the loss of a reservation for a hotel room or the negative implications of arriving late to an important meeting). As the driver is making the trip anyways, the utility of the driver remains as if no passenger had made the request,

$$u^D_{j,0,t} = \delta^D_{j,0,t} + \epsilon^D_{j,0,t}, \tag{7}$$

$$u^P_{i,0,t} - c = \delta^P_{i,0,t} + \epsilon^P_{i,0,t} - c. \tag{8}$$

The choice of the seat as the basic unit of analysis of the model responds to the need to produce a tractable and computationally feasible model while capturing the essence of

[8]While the gender and ethnic origin of the counterpart is a feature that may be valued by the user as any other feature of the trip, the driver's gender and ethnic origin acts as a modulator of this preference. Hence, although majority and minority passengers may value the gender of the driver, this valuation is allowed to change depending on their ethnicity.

the empirical findings documented in Appendix A.[9]

The remaining parts of this section contain a detailed description of each of the four stages described above. Given that the model is solved by backward induction, stages are reported in reverse order.

5.2 Stage 4. Manual driver acceptance/rejection decision.

Only in the case of the driver having chosen the manual booking mode, stage 4 takes place. If the driver had chosen the automatic booking mode, once the passenger decides to make the request, the user automatically becomes part of the trip, and no decision is left to the driver.

At market t, driver j, if travelling with passenger i, receives the expected utility

$$E[u^D_{j,i,t}] = u^D_{j,i,t} = \delta^D_{j,i,t} + \epsilon^D_{j,i,t}, \tag{9}$$

composed by the shock $\epsilon^D_{j,i,t}$, specific to the driver, passenger and market, and the mean utility

$$\delta^D_{j,i,t} = p_{j,t}\alpha^D_{k_j} + X^D_{j,i,t}{}'\beta^D_{k_j} + \rho^D_t \tag{10}$$

where $X^D_{j,i,t}{}'$ is a vector containing a constant term, the features of the trip (including the number of published seats, the car's value, and the comfort mode) and the passenger i (gender and ethnic origin); and ρ^D_t are route and time fixed effects. The preference parameters that determine the marginal utility of the price $\alpha^D_{k_j}$ and the marginal utility of the rest of the covariates $\beta^D_{k_j}$ are allowed to depend on the population segment defined by the gender and ethnic origin of the driver, k_j.

If the driver rejects the passenger, the former receives expected utility $E[u^D_{j,0,t}]$. As no price is perceived, but no cost associated with the features of the trip and the passenger is incurred, the driver will obtain the mean utility associated with making the trip with the empty seat and the taste shock of the outside option,

[9] If the whole car was to be considered as the unit of observation, the formulation of the driver problem would become much more complex, containing a dynamic decision of bundle formation. To partially control this issue, the utility of the driver is allowed to depend on the total number of published seats. Furthermore, Appendix A documents the correlation between the number of published seats and the proportion of passengers of each kind. These correlations show that the proportions of men and minorities are not correlated to the number of seats initially published. Consequently, it seems fair to assume that drivers decide the acceptance and rejection of each passenger individually, with synergies across seats and passengers playing little or no role in their choices.

$$E[u^D_{j,0,t}] = u^D_{j,0,t} = \delta^D_{j,0,t} + \epsilon^D_{j,0,t}. \tag{11}$$

In both cases, whether the driver accepts or rejects the passenger, the expected utility is equivalent to the specific utility that the driver obtains from the trip. Once the passenger has made the request, the realization of the taste shocks both toward the passenger and the outside option takes place for the driver. This reflects the fact that the driver has been able to go through the profile of the passenger and, thus, that the former has acquired a better knowledge of whether the specific individual that has produced the request is a good match. Moreover, it also represents the fact that the date of departure is closer and, consequently, that the driver has a better understanding of the features and conditions in which the trip will take place if traveling with an empty seat.

The driver will choose to accept the request $a_{j,i,t} = 1$ if the expected utility from traveling with the passenger, equation (9), is higher than the expected utility of making the trip with the empty seat, equation (11), $E[u^D_{j,i,t}] > E[u^D_{j,0,t}]$ or $u^D_{j,i,t} > u^D_{j,0,t}$. Thus, the driver will accept the request with probability $pa_{j,i,t}$,

$$pa_{j,i,t} = prob(\epsilon^D_{j,i,t} - \epsilon^D_{j,0,t} > \delta^D_{j,0,t} - \delta^D_{j,i,t}) \equiv pa(\delta^D_{j,i,t} - \delta^D_{j,0,t}). \tag{12}$$

When the expected utility of making the trip with the passenger, equation (9), is smaller than the utility of the outside option, equation (11), $E[u^D_{j,i,t}] < E[u^D_{j,0,t}] = u^D_{j,i,t} < u^D_{j,0,t}$, the driver will choose to reject the request $a_{j,i,t} = 0$.

5.3 Stage 3. Passenger request decision.

In stage 3, the passenger, who has already entered the market and observed the features of the driver and the existing trip, decides whether to make the request or directly choose the outside option.

At market t, passenger i, if requesting to travel with driver j, receives expected utility $E[u^P_{i,j,t}]$, which can be decomposed as follows

$$E[u^P_{i,j,t}] = pa_{i,j,t}u^P_{i,j,t} + (1 - pa_{j,i,t})(u^P_{i,0,t} - c) = pa_{j,i,t}(\delta^P_{i,j,t} + \epsilon^P_{i,j,t}) + (1 - pa_{j,i,t})(\delta^P_{i,0,t} + \epsilon^P_{i,0,t} - c). \tag{13}$$

The expected utility from making a request is the sum of the utility of making the trip $u^P_{i,j,t}$, pondered by the probability of acceptance $pa_{j,i,t}$, and the utility of being rejected

$u^P_{i,0,t} - c$, pondered by the probability of rejection $1 - pa_{j,i,t}$.

When traveling together, the passenger receives the shock $\epsilon^P_{i,j,t}$, specific to each driver, passenger and market combination, and the mean utility of the trip,

$$\delta^P_{i,j,t} = p_{j,t}\alpha^P_{k_i} + X^P_{i,j,t}{}'\beta^P_{k_i} + \rho^P_t + \xi^P_{k_i,k_j,t} \tag{14}$$

where $X^P_{i,j,t}{}'$ is a vector containing a constant term, the features of the trip (including distance, car value, number of originally published seats and comfort mode) and the features of the driver, including gender, ethnic origin and some additional characteristics (age, rating, experience, phone and mail verification, personal picture). ρ^P_t is again used to denote route and time-fixed effects. As in the case of the driver, price and other trip-specific parameters, α^P and β^P, are allowed to depend on the population segment to which the passenger belongs k_i. This feature of the model is key for capturing the modulating influence on the preferences of the passenger, of the interaction between his/her gender and ethnic origin, and those of the driver.

It is important to note that the unobserved quality component $\xi^P_{k_i,k_j,t}$ is not specific to the driver but to the population segments to which the driver and passengers belong, and to the route and time in which the trip takes place. This unobserved component aims to capture the changes in perception, both across time and regions, that the different population segments have in favor or against other population segments. For illustrative purposes, consider the following example. Imagine a trip whose city of origin has been recently suffering from a cascade of sexual abuse cases towards women. The perception of women towards unknown men is likely different than if the region is well-known for having none of these cases or if there has been a long period of time in which these types of crimes have not been committed.

If the driver publishing the available trip has chosen the automatic booking mode, the passenger knows with certainty that any request will be accepted. Given that $pa_{j,i,t} = 1$ in this case, the expected utility of the passenger is equivalent to the utility of making the trip, and the previous expression collapses to

$$E[u^P_{i,j,t}] = u^P_{j,i,t} = \delta^P_{i,j,t} + \epsilon^P_{i,j,t}. \tag{15}$$

If the passenger decides to choose the outside option, the user obtains, with certainty, $u^P_{i,0,t}$

$$E[u^P_{i,0,t}] = u^P_{i,0,t} = \delta^P_{i,0,t} + \epsilon^P_{i,0,t}. \tag{16}$$

Regarding the structure and timing of the shocks, three features need to be highlighted. Firstly, in stage 3, the driver has not received any requests. Hence, the driver has not been able to analyze the profile of the entering passenger. Consequently, the taste shocks of the driver have not been realized, and the passenger needs to formulate expectations in terms of the likelihood of the acceptance and rejection decisions. Secondly, in stage 3, the taste shocks of the passenger are realized; this implies that when the passenger decides to make a request or choose the outside option, the passenger already knows the precise value of making the trip with a specific driver. This condition reflects the fact that the passenger has been able to scrutinize the profile of the driver, and thus, the former has been able to develop an accurate opinion on the specific quality of the match with the latter. Thirdly, note that the taste shocks of the passenger are specific to the two different results of the game, traveling with the driver or not traveling with the driver. In other words, there is no specific taste shock attached to the decision of requesting or not requesting. This feature of the model ensures the consistency of the decision of the passenger, making a passenger with a close to zero probability of acceptance never willing to make a request if c is positive.

The passenger will make the request $r_{i,j,t} = 1$ to the driver if the expected utility of making the request, equation (13), is larger than the expected utility of not doing so, equation (16), $E[u^P_{i,j,t}] > E[u^P_{i,0,t}]$. Thus, the passenger will make the request with probability $pr_{i,j,t}$

$$pr_{i,j,t} = prob(\epsilon^P_{i,j,t} - \epsilon^P_{0,j,t} > \delta^P_{i,0,t} - \delta^P_{i,j,t} + c(pa^{-1}_{j,i,t} - 1)) \equiv pr(\delta^P_{i,j,t} - \delta^P_{i,0,t} - c(pa^{-1}_{j,i,t} - 1)). \quad (17)$$

Now, as pointed out by the last term in the equivalence, the probability of request is a function of the standardized mean utility that the passenger perceives if making the trip and an additional component. This component is not present in the case of the automatic booking mode and reflects the anticipation of the passenger of the cost that will be incurred if rejected. When the probability of acceptance equals one, the term vanishes, and the request probability for the manual driver is identical to the one faced by the automatic driver. As the probability of acceptance decreases, the likelihood of being rejected and incurring the rejection cost c increases, making the passenger more eager to initially choose the outside option. In the limit, when the probability of acceptance tends to zero, the passenger will always choose the outside option if c is positive.

The previous expression shows that, under this specification of the pay-off functions, the passenger does not care about the booking mode choice of the driver but through the probability of rejection and its associated cost.[10] Consequently, when the cost c is zero, the passenger does not lose anything for making the request. Hence, the user will be willing to make the request with the same probability as if the driver had chosen the automatic booking mode. When the cost is positive, the passenger ponders the expected negative impact of the rejection and will choose to make a request with a lower probability than if, all else equal, the driver had decided to establish the automatic booking mode.

The passenger will choose the outside option, not producing any request to the driver $r_{i,j,t} = 0$, whenever the expected utility of the outside option, equation (16), is larger than the expected utility of requesting to join the trip, equation (13), $E[u^P_{i,j,t}] < E[u^P_{i,0,t}]$.

5.4 Stage 2. Pricing decision.

At the moment in which the pricing decision is to be taken, no passenger has entered the market, and thus, no taste shock has been realized. Thus, irrespective of the booking mode, the expected utility of making the trip with an empty seat is

$$E[u^D_{j,0,t}] = \int (\delta^D_{j,0,t} + \epsilon^D_{j,0,m}) f_\epsilon(\epsilon^D_{j,0,m}) d\epsilon^D_{j,0,m} = \delta^D_{j,0,t} + \gamma, \tag{18}$$

where γ denotes the expectation of the preference shock of the driver. That is, when drivers have to choose the price at which the seat is offered, the expected utility of the outside option is equivalent to the mean utility of the outside option plus the expected value of the associated taste shock.

If the driver has chosen the automatic booking mode and a passenger decides to make a request, his expected utility for the trip will equal the mean utility of making the trip with the passenger plus the expectation of the taste shock,

$$E[u^D_{j,i,t}] = \int (\delta^D_{j,k,t}(p_{j,t}) + \epsilon^D_{j,i,t}) f_\epsilon(\epsilon^D_{j,i,t}) d\epsilon^D_{j,i,t} = \delta^D_{j,k,t}(p_{j,t}) + \gamma. \tag{19}$$

If selecting the manual mode, the driver anticipates that if a passenger makes a request, the latter will only be accepted if the utility of traveling with an empty seat is smaller.

[10] It could be argued that the booking mode of the driver may signal specific unobserved quality components to the passengers. One could think that drivers choosing the manual mode are more caring in general about their counterparts, and thus, they are more eager to provide a good service. It could also be argued that drivers that choose the manual mode may be less polite to those passengers that present certain observable traits, and this circumstance is anticipated by those passengers. To shed some light on this possibility, Appendix B presents the relation between received ratings by drivers and the booking mode choice of the driver. In fact, it is shown that there is no significant correlation between the received ratings and the booking mode choice, even if being interacted with the gender and ethnic origin of the driver and the passengers.

Hence, the *a priori* expected utility when receiving a request can be written as follows,

$$E[\max(\delta^D_{j,k,t}(p_{j,t}) + \epsilon^D_{j,i,t}, \delta^D_{j,0,t} + \epsilon^D_{j,0,t})] = log(1 + exp(\delta^D_{j,k,t}(p_{j,t}) - \delta^D_{j,0,t})) + \delta^D_{j,0,t} + \gamma. \quad (20)$$

The last equality only holds if the taste shocks are drawn from independent and identical Gumbel distributions with location parameter zero and scale parameter one.

The driver knows that with a certain probability $q_{k,t}$, the arriving passenger belongs to population segment k. Moreover, the driver knows the probability with which the passenger will decide to make a request $pr_{i,j,t}$ or will choose the outside option $1 - pr_{i,j,t}$.

Prior to the acceptance/rejection decision, the driver also incurs the variable cost $vc_{j,t}$. This cost reflects the time and wealth expenditure that drivers face in order to have the seat and passenger ready to perform the trip. As such, the cost is only incurred if a request is produced by the passenger.[11] For example, the driver will have to answer questions clarifying the availability of the seat and the exact point of departure. Moreover, the driver will have to prepare the car and the trunk so that the potential passenger can make proper use of it. Note that these variable costs may vary across drivers; that is, some drivers may have more time available to answer questions or to prepare the car for the trip than others. As it happens with profit-maximizing firms in traditional Industrial Organization settings, variable costs affect the pricing choice of the driver: all else equal, those drivers with a higher cost will charge a higher price to compensate for the additional disutility.

The cost $vc_{j,t}$ is assumed to have the following functional form,

$$vc_{j,t} = X^{D,v\prime}_{j,t}\beta^{D,v} + \rho^{D,v}_t + \xi^{D,v}_{j,t} \quad (21)$$

where $X^{D,v\prime}_{j,t}$ is a vector of covariates containing a constant term, features of the trip and features of the driver (including the gender, ethnicity, reputation and experience); $\rho^{D,v}_t$ are route and time fixed effects and $\xi^{D,v}_{j,t}$ is the unobserved cost component only known to the driver, the source of price heterogeneity and the econometric error in the estimation process.

As a final note about the variable cost $vc_{j,t}$, at the time of the acceptance/rejection decision by the driver, it has already been incurred, and thus, it becomes sunk, playing

[11] An alternative interpretation may consider that the cost is only incurred if the passenger produces a request which is accepted by the driver. Modeling this strategy is equivalent to establishing a reduction in the variable cost that the manual driver faces.

no role in the acceptance/rejection decision of the manual driver.

Taking into consideration all factors, the automatic driver will choose the price that maximizes the total expected utility,

$$\sum_{k\in K} q_{k,t}(pr_{k,j,t}(p_{j,t})(E[\delta^D_{j,k,t}(p_{j,t}) + \epsilon^D_{j,i,t}] - vc_{j,t}) + (1 - pr_{k,j,t}(p_{j,t}))E[\delta^D_{j,0,t} + \epsilon^D_{j,0,t}]). \quad (22)$$

Rearranging terms in equation (22) and exploiting the equivalences stated in equations (18) and (19), the expected utility of the automatic driver $U^{AUT}_{j,t}(p_{j,t})$ can be written as

$$U^{AUT}_{j,t}(p_{j,t}) = \sum_{k\in K} q_{k,t}pr_{k,j,t}(p_{j,t})(\delta^D_{j,k,t}(p_{j,t}) - \delta^D_{j,0,t} - vc_{j,t}) + \delta^D_{j,0,t} + \gamma. \quad (23)$$

This expression shows that the pricing problem of the automatic driver is no different from that of a classical monopolistic firm. Drivers may receive requests from passengers belonging to different segments, and the probability with which they receive a request is affected by the price they choose. Similarly, monopolists face markets in which consumers have a certain probability of purchasing their product, which is affected by their features, including the price. In the car-sharing setting, requests result in sharing the trip, and thus, obtaining the utility of the price, incurring the cost of sharing the vehicle and time with another person, and losing the utility of making the trip with an empty seat. Moreover, drivers face the variable cost associated with having the seat and passenger prepared for the trip. In parallel fashion, firms derive a benefit from the purchase, stemming uniquely from the price, and incur a variable cost attached to the production and delivery of the sold unit.

Similarly, manual drivers will choose the price that maximizes their total expected utility,

$$\sum_{k\in K} q_{k,t}(pr_{k,j,t}(p_{j,t})(E[Max(\delta^D_{j,k,t}(p_{j,t})+\epsilon^D_{j,i,t}, \delta^D_{j,0,t}+\epsilon^D_{j,0,t})]-vc_{j,t})+(1-pr_{k,j,t}(p_{j,t}))E[\delta^D_{j,0,t}+\epsilon^D_{j,0,t}]). \quad (24)$$

Applying the equivalences of equations (18) and (20) to the expression in (24), the total expected utility of the manual driver $U^{MAN}_{j,t}(p_{j,t})$ can be simplified to the following expression,

$$U_{j,t}^{MAN}(p_{j,t}) = \sum_{k\in K} q_{k,t}pr_{k,j,t}(p_{j,t})(log(1+exp(\delta_{j,k,m}^{D}(p_{j,t}) - \delta_{j,0,t}^{D})) - vc_{j,t}) + \delta_{j,0,t}^{D} + \gamma. \quad (25)$$

Equation (25) highlights the difference between the manual and the automatic pricing problem. In the manual case, the driver knows that the requesting passenger will be accepted only if the utility of making the trip with an empty seat is smaller than the utility derived from making the trip with the specific passenger. Otherwise, the driver will resort to the outside option. However, in the automatic case, any requesting passenger is accepted, even if it turns out for the driver that the utility of traveling with this passenger is smaller than the utility of making the trip with the empty seat.

To characterize the optimal price $\overline{p}_{j,t}^{\cdot}$, I take derivatives of equations (23) and (25) with respect to the price and equate them to zero. For the automatic driver, the optimal pricing rule can be written as

$$\sum_{k\in K} q_{k,t}\frac{\partial pr_{k,j,t}(p_{j,t})}{\partial p_{j,t}}\Big|_{\overline{p}_{j,t}^{AUT}}(\delta_{j,k,t}^{D}(\overline{p}_{j,t}^{AUT}) - \delta_{j,0,t}^{D} - vc_{j,t}) + q_{k,t}pr_{k,j,t}(\overline{p}_{j,t}^{AUT})\frac{\partial \delta_{j,k,t}^{D}(p_{j,t})}{\partial p_{j,t}}\Big|_{\overline{p}_{j,t}^{AUT}} = 0. \quad (26)$$

In the manual case, the optimal price satisfies the following condition,

$$\sum_{k\in K} q_{k,t}\frac{\partial pr_{k,j,t}(p_{j,t})}{\partial p_{j,t}}\Big|_{\overline{p}_{j,t}^{MAN}}(log(1+exp(\delta_{j,k,t}^{D}(\overline{p}_{j,t}^{MAN}) - \delta_{j,0,t}^{D})) - vc_{j,t})$$
$$+ q_{k,t}pr_{k,j,t}(\overline{p}_{j,t}^{MAN})pa_{j,k,t}(\overline{p}_{j,t}^{MAN})\frac{\partial \delta_{j,k,t}^{D}(p_{j,t})}{\partial p_{j,t}}\Big|_{\overline{p}_{j,t}^{MAN}} = 0. \quad (27)$$

Optimal pricing rules balance two opposing forces. On the one hand, a decrease in the expected utility of the driver results from the reduction in the probability of request due to a marginally higher price. On the other hand, the immediate increase in the mean utility of the driver stems from a marginal increase in the price. When the manual booking mode has been chosen, the latter effect is modulated by the probability of acceptance, as the increased mean utility will only be enjoyed when the driver receives a request that is accepted.

When comparing the equilibrium prices of the manual and the automatic modes, two effects need to be addressed. The first effect refers to the reduced amount of requests that comparable manual drivers receive whenever the cost of rejection takes a positive value $c > 0$. As the cost of rejection increases, the passenger anticipates the increasingly negative consequences of rejection, and the probability of receiving a request decreases. To increase the probability of acceptance of the passenger and thus the probability of request, the manual driver will have incentives to set a higher price than in the automatic case. The second effect refers to the increased expected utility when receiving a request that stems from the direct ability of the driver to sort undesired passengers. This effect is embodied in two conditions. On the one hand, the driver will travel, in expected terms, more often with passengers belonging to population segments that provide a higher utility. On the other hand, within a particular population segment, the driver will choose to accept only those passengers that provide a utility higher than that of the outside option. In both cases, the increased utility of the trip associated with the expected request will encourage the driver to set lower prices and increase the probability of the request.

5.5 Stage 1. Booking mode decision.

The ability to reject and accept passengers allows the manual driver to have an additional chance that the automatic driver does not have. But this advantage comes at a cost. For its proper usage, drivers that choose the manual mode need to have a deep understanding of the platform: they need to know where the request is received and stored, how to check the identity and profile of the requesting passenger, and how to finally decide to accept or reject the request. Moreover, in order to be notified of the incoming requests by the means they are accustomed to (mail, SMS, or directly through the platform), they need to appropriately adapt their personal settings in the application. This knowledge and time requirements are, in fact, entry barriers that limit the usage of this mode. To reflect this circumstance, the manual mode is assumed to entail a fixed cost $fc_{j,t}$ that makes specific drivers prefer the automatic booking mode.

The fixed cost $fc_{j,t}$ is assumed to have the form

$$fc_{j,t} = \delta_{j,t}^{D,f} + \mu_{j,t}^{D,f} = X_{j,t}^{D,f}{}'\beta^{D,f} + \rho_t^{D,f} + \xi_{k_j,t}^{D,f} + \mu_{j,t}^{D,f} \tag{28}$$

where $X_{j,t}^{D,f}{}'$ is a vector of covariates containing a constant term, the gender and ethnic origin of the driver, and other features of the driver (including the age, average reputation and experience) and the trip (number of published seats, comfort mode and car value); $\rho_t^{D,f}$ are route and time fixed effects; and $\mu_{j,t}^{D,f}$ is the associated cost shock of the manual mode for driver j in market t and the econometric error for the estimation of the fixed cost.

Note that the unobserved fixed cost component $\xi^{D,f}_{k_j,t}$ is specific to the driver population segment, the route and the time period. This feature aims to reflect unobserved elements that may make specific population segments in some routes or points in time better suited to use the manual mode. For illustrative purposes, consider the following example. If certain municipalities offer free courses in the usage of new technologies or specific online platforms, population segments for which these initiatives are intended will likely face fewer problems when employing the manual mode.

Driver j in market t will choose the manual mode if the expected utility associated with this booking mode modality is greater than the expected utility of the automatic booking mode,

$$E[U^{MAN}_{j,t}(\overline{p}^{MAN}_{j,t}) - fc_{j,t}] \geq E[U^{AUT}_{j,t}(\overline{p}^{AUT}_{j,t})]. \tag{29}$$

The probability with which a driver chooses the manual mode pm is,

$$pm_{j,t} = prob(\mu^{D,f}_{j,t} < U^{MAN}_{j,t}(\overline{p}^{MAN}_{j,t}) - U^{AUT}_{j,t}(\overline{p}^{AUT}_{j,t}) - \delta^{D,f}_{j,t}). \tag{30}$$

The probability of using the manual mode increases as the difference between the manual and the automatic expected utilities grows. Consequently, drivers with starker preferences towards the different passenger segments will use the manual booking mode in a higher proportion.

6 Estimation procedure

The estimation process is built upon the information available at the seat level: the profile information of the posting rider and, if any, of the joining passenger, the price, the booking mode and the rest of the features of the trip. This process focuses on three population segments k: men M of the ethnic minority B, women F of the ethnic majority W, and men M of the ethnic majority W.[12]

Estimation takes place in three phases. First, the parameters governing trip mean utilities, both for drivers and passengers, are estimated. These parameters inform the de-

[12]Women of the ethnic minority are not included in the analysis because they represent a small fraction of the total market, both as passengers and as drivers. Given that identification requires the aggregation of similar cars due to the limited amount of seats per trip, precise estimation of specific preferences of female drivers and passengers of the ethnic minority is unfeasible with existing data.

cisions undertaken in stages 3 and 4, and as such, they are recovered from moment conditions built by equating the probabilities of request and acceptance to the observed proportions of passengers in each car. Secondly, using the estimates of stages 3 and 4, the parameters governing drivers' variable costs are estimated. To this purpose, the optimal pricing rules are inverted and variable costs are recovered. Finally, given the trip preference parameters of drivers and passengers and the variable costs of the former, driver counterfactual expected utilities, associated with the alternative booking mode choice, are computed and compared to the expected utilities associated with the chosen booking mode. Matching the observed proportions of manual drivers in each population segment, route and time frame to the predicted probabilities of choosing the manual booking mode by the model, fixed cost parameters are recovered.

The following three subsections contain explain the main elements in the estimation process of the various structural parameters.

6.1 Passenger and driver trip preferences

The acceptance and requesting decisions of drivers and passengers exclusively rely on the utilities of making the trip together and the respective outside options. When these actions are undertaken, the rest of the interactions taking place through the platform, and the associated costs, have already occurred. For this reason, the observed proportions of passengers joining each trip should help identify the preference parameters informing the utilities that drivers and passengers derive from traveling.

In stages 4 and 3, the agents directly decide, given the conditions of price, booking mode, other trip features, and their own personal traits and those of their counterparts, to make a request and to accept or reject the request. The data set does not contain requests $r_{i,j,t}$ and acceptance/rejections $a_{j,i,t}$, it contains the effective passenger, if any, that joins each seat and makes the trip $r_{i,j,t}a_{j,i,t}$. That is, I observe those passengers that have made the request only if they have been accepted.

In the case of the automatic driver, as requests are immediately accepted $a_{j,i,t} = 1$, the proportion and identity of the passengers that join the driver exclusively result from the decisions of the former $r_{i,j,t}a_{j,i,t} = r_{i,j,t}$. Thus, in the automatic case, requests are observed.

In the case of manual trips, the proportion and identity of the passengers that join them result both from the decisions of passengers and drivers. To disentangle their actions and preferences, I propose an *ad hoc* identification strategy that, in broad terms, takes advantage of the differences between the observed proportions of passengers in comparable manual and automatic groups, to understand the probability of rejection and to estimate the preferences of the drivers towards each passenger kind.

The rest of this subsection is going to address five main components of the estimation procedure: the constructions of the moment conditions, the identification of the different parameter types, the weighting of each observation, the discussion of potential endogeneity concerns and the construction of instruments, and the computational process.

Moment conditions. For each seat, I observe if any passenger has filled it, in which case, I also observe the gender and ethnic origin of the user. As previously stated, for seats published by manual drivers, $r_{i,j,t}a_{j,i,t}$ is observed and, for seats published by automatic drivers, $r_{i,j,t}$ is observed.

To overcome the limited information contained in each trip, aggregation needs to take place. In particular, aggregation units are built as combinations of specific routes, time intervals (two-week periods), booking mode modalities, and gender and ethnic origin of the drivers (e.g., manual female drivers of the ethnic majority in the route Paris-Lyon between 01.01.2021 and 15.01.2021). For each aggregation unit g, moment conditions are defined using the observed proportion of traveling passengers belonging to each population segment.

For each aggregation unit, a number of moments equivalent to the total number of passenger population segments are built. Moment conditions are represented by equation (31). The left-hand side is the sum of observed passengers $a_{j,k,t}r_{k,j,t}$ of a certain population segment k joining any of the published seats s by any of the drivers j belonging to the aggregation unit g, divided by the total number of published seats in the aggregation unit. To put it simply, the left-hand side contains the average proportion of passengers of a specific population segment that join the drivers in g. The right-hand side is the structural counterpart: the average proportion of passengers that the model predicts will join those drivers belonging to the aggregation unit. For each k passenger population segment, the following condition holds,

$$\frac{1}{\sum_{j\in g} S_{j,t}} \sum_{j\in g} \sum_{s\in S_{j,t}} a_{j,k,t}r_{k,j,t} = \frac{1}{\sum_{j\in g} S_{j,t}} \sum_{j\in g} \sum_{s\in S_{j,t}} pa_{j,k,t}pr_{k,j,t}q_{k,t}, \tag{31}$$

where $S_{j,t}$ is the set of published seats in route and time t by driver j. When drivers in the aggregation unit use the automatic booking mode, the previous condition collapses to

$$\frac{1}{\sum_{j\in g} S_{j,t}} \sum_{j\in g} \sum_{s\in S_{j,t}} r_{k,j,t} = \frac{1}{\sum_{j\in g} S_{j,t}} \sum_{j\in g} \sum_{s\in S_{j,t}} pr_{k,j,t} q_{k,t}. \tag{32}$$

Note that the aforementioned conditions add the probabilities of request and acceptance of all the published seats, even if any of these end up not being filled by any passenger. In a different vein, it is also important to note that the decisions to join the seats published by two different drivers belonging to the same aggregation unit are affected by features that may differ across them (the price, the amount of experience of the driver, the average rating of the rider or the number of kilometers of the selected route, for example). In other words, the requesting and accepting decisions referring to the seats published by different drivers within the same aggregation unit are random variables following binomial distributions with different expectations, those resulting from equations (12) and (17), and variances. Thus, for the previous equalities to hold, Chebyshev's Law of Large numbers for independent and not identically distributed variables is used. The proof of this law can be found in Appendix C. This law of large numbers ensures the convergence provided that random variables are independent, which is a reasonable assumption to adopt in this setting, as the decisions are taken by different drivers and passengers without any coordination amongst themselves. As a result, embedded in the estimation process is the assumption that driver and passenger trip taste shocks are independent across drivers, passengers and markets. It is also assumed that driver and passenger taste shocks are independent between them.[13]

Precisely because some features may differ across drivers within the same aggregation unit, I distinguish in this section those features that are identical to all the drivers in the aggregation unit (the constant term, the gender and the ethnic origin of the driver), which are denoted by $X^{P,G}_{i,j,t}$, from those covariates that may differ across the grouped drivers (including price, car value, number of initially published seats, comfort mode, rating, experience, phone and mail verification, and personal picture), which are denoted by $X^{P,NG}_{i,j,t}$. Each of these covariates has its own set of preference parameters, such that,

$$X^{P}_{i,j,t}{}'\beta^{P}_{k_i} = X^{P,G}_{i,j,t}{}'\beta^{P,G}_{k_i} + X^{P,NG}_{i,j,t}{}'\beta^{P,NG}_{k_i}. \tag{33}$$

[13]This feature is different from the independence of passenger taste shocks across passengers and drivers and to the independence of driver taste shocks across drivers and passengers. It refers to the independence between driver and passenger taste shocks, implying that if a passenger obtains a large taste shock towards a certain driver, the distribution of driver taste shocks remains unaltered. There is no "mutual love at first sight" effect. Consequently, the probabilities of request and acceptance are independent objects, and the probability of a passenger making the trip with a certain driver is merely the product of both.

The choice of the aggregation unit is not trivial. On the one hand, aggregation units need to cover a sufficiently large time frame to comprehend a sufficient amount of seats so that the previous law of large numbers operates.[14] In other words, the sample size needs to be sufficiently big within each aggregation unit. On the other hand, as it will become apparent from the discussion in the following paragraphs, the identification of the parameters related to covariates that differ across drivers within the same aggregation unit fosters the formation of aggregation units as granular as possible. This tension motivates the previous choice, which covers the smallest possible time frame, two-week periods, to ensure a sufficiently large number of seats per aggregation unit, 30 on average.

To facilitate the computation of the model, in addition to previous assumptions, taste shocks of drivers and passengers $\epsilon^D_{j,i,t}, \epsilon^D_{j,0,t}, \epsilon^P_{i,j,t}, \epsilon^P_{0,j,t}$ are assumed to follow Gumbel distributions. Additionally, the mean utility of the outside option, traveling with an empty seat for the drivers, and the composite good of staying at home and using alternative means of transport for the passengers is normalized to zero.[15]

Parameter identification. For the automatic case, two types of parameters are identified from the previous moment conditions. Due to the way in which they enter the moment generation condition, I distinguish between linear and non-linear parameters.[16]

The first set of parameters $\{\beta^{P,G}_k, \rho^P_t\}$ are those defining the preferences towards the features that are identical to all the drivers in the aggregation unit. These are the preferences of making the trip (the constant term), the preferences towards the features defining the population segment of the driver (the gender and ethnic origin), and the features associated with making the trip in a specific route and time frame (the route and time fixed effects). The identification of these parameters follows the standard logic

[14] The number of seats a car possesses is limited. At most, van drivers are allowed to publish five seats. Nevertheless, the median number of published seats is 3. Hence, the aggregation of the decisions taking place in each of the seats within a car does not provide enough information to understand the true probabilities of request and acceptance. In other words, given that only one passenger per seat, or a limited amount of them, is deciding to request, $\sum_{s \in S_{j,t}} r_{i,j,t} a_{j,i,t} = \sum_{s \in S_{j,t}} pr_{i,j,t} pa_{j,i,t}$ cannot be established.

[15] As it can be seen from the decisions taking place in the different stages of the model, these normalizations do not have any implication in prior stages: they do not affect the formation of prices and, given they are the same irrespective their booking mode, they do not affect this decision either. They only increase or decrease by the same amount the average marginal utility of making the trip for the passengers and the average utility of sharing the car with a passenger for the driver, but the utility difference between making the trip and not making it or traveling with the passenger and traveling with the empty seat is the same.

[16] For further details on this distinction, Appendix D.6 decomposes the moment conditions, showing how the different parameters interact with the unobserved quality component.

of a traditional demand estimation problem built upon a logistic discrete-choice model.[17]

The remaining passenger preference parameters are called non-linear $\{\alpha_k^P, \beta_k^{P,NG}\}$, precisely because they relate non-linearly to the unobserved quality component. These parameters govern the preference of the passenger towards features of the drivers in which they differ from other drivers of the same aggregation unit. Identification of these parameters is not immediate as it requires sufficient variation across similar aggregation units of the distribution of the feature within the aggregation units. To build intuition, think that, in addition to the linear parameters, passengers only care about the price of the trip. If female drivers tend to establish on average lower prices than men, and no heterogeneity across different groups of female drivers exists in terms of price (for example, if female drivers across all routes establish an average price that is lower in the same proportion to that established by men), identification will be poor or unfeasible. In this example, it will not be possible to tell if female passengers prefer to travel with female drivers due to them being women or because they have a more acute sensitivity towards prices. Thus, the identification of these parameters requires enough variation of the associated average covariates across similar aggregation units. This condition fosters a definition of aggregation units as granular as possible.

In a sense, the identification logic of the model presents a certain degree of resemblance to that of a random coefficient model in the context of demand estimation. The main difference is that, given the passenger's gender and ethnicity, preference parameters do not change across individuals, but the value of the covariates considered by the passengers in each aggregation unit may change because they refer to drivers and trips with differences in certain features. Consequently, identification does not rely on the varying proportions of each population segment across markets but on the heterogeneity of these features across similar aggregation units, encouraging the choice of the smallest feasible

[17]Identification is immediate and stems directly from the choice of the aggregation units. To build intuition, consider the case in which passengers only cared, in their requesting decision, about these features. In this particular case, the decision of any passenger of the same kind to make a request to a driver of a certain kind in a certain route and time frame would be a random variable with identical mean and variance. Thus, given the taste shock independence assumption and the unobserved component assumption, the aggregation taking place in the moment condition would be equivalent to having the same driver repeat the trip for all the different passengers or using a bus with lots of empty seats. From the observed proportion of requests, the mean utility would be directly recovered. Consequently, if female passengers tend to be more prevalent in cars driven by female drivers, this would translate, all else equal, into a positive parameter of female passengers towards the women condition of the driver.

aggregation unit.[18]

When the booking mode is manual, the moment conditions contain four different types of parameters. As in the automatic case, via the probability of request, estimation-wise, the moment conditions become functions of the linear passenger preference parameters (those referring to the marginal utility of the covariates that are equal to all the drivers within the same aggregation unit) and the non-linear passenger preference parameters (those aiming to capture the marginal utility associated to covariates that vary across drivers within the same aggregation unit). Besides these two types of parameters, also within the probability of request, a new component emerges controlling the cost of rejection c. This parameter determines the anticipation of the rejection possibility by the passenger. Thus, the higher the cost c, the less likely it is that the passenger will make a request to the manual driver.

The fourth type of parameter $\{\alpha^D_{k'}, \beta^D_{k'}, \rho^D\}$ determines the utility of the driver towards the different features of the trip and the passenger. These parameters are distinguished from all the rest because of the way in which they interact with the unobserved quality component. In fact, these parameters are also non-linearly related to the unobserved quality component but in a different fashion from the non-linear passenger preference parameters.

Identification of the first two types of parameters follows the same logic as in the automatic case. Identification of the remaining two is more complex. For this reason, let us follow a simplifying approach. Consider first that the cost c takes value zero and, thus, that the passenger, all else equal, is going to be indifferent between two identical drivers that differ exclusively in their booking mode choice. In other words, as there is no cost associated with the rejection, the passenger has nothing to lose if not accepted. Hence, the user will make the request with the same probability if the driver has chosen the manual or the automatic booking mode. Under this condition, the probability of acceptance of the driver, the mean utility, and the parameters of which it is composed will be identified through the comparison of the composition of manual groups and similar automatic aggregations.

To further illustrate this case, consider what would happen if drivers of the same kind

[18] An alternative that avoids the non-linear specification would be to assume that all drivers of a certain kind, in a certain route and time frame, present the same features. This alternative would consider that all drivers in the same aggregation unit present the same price, reputation and the rest of the features (a weighted average of all the drivers that compose the aggregation unit). Although it is common to find papers in which average prices are used to estimate demand preferences, and although it would simplify the estimation procedure as no non-linear parameters would arise, the underlying assumption is remarkably strong for this setting, especially if considering its interaction with the manual condition of certain drivers.

were identical in the rest of their covariates but differed exclusively in their booking mode. If automatic drivers of a particular gender and ethnicity present, per published seat, a larger number of passengers of a particular population segment than the equivalent manual drivers, the model would identify these differences as a decision of the manual driver that stems from a disutility when sharing a trip with the latter population segment.

As it has already been hinted, the lack of knowledge about the natural outcomes of the model, rejections and requests, comes at a price.[19] Precisely, from the observed proportions of accepted passengers, disentangling the preferences of the driver and the anticipated negative consequences suffered by passengers when rejected is unfeasible without an additional restriction. For illustrative purposes, let us continue building on the example provided in the previous paragraph. The reason why passengers belonging to a particular population segment are in a lesser proportion present in the cars of specific drivers can be due to various combinations. The drivers may have a great disutility of traveling with the passenger, and the passengers may not have a high cost of rejection, as in the original example. It could also be that drivers have a small disutility if traveling with these specific passengers and that passengers have a remarkable cost in case of rejection, or any other intermediate scenario.

Therefore, an additional restriction is required to disentangle the preferences of the driver and the cost of rejection that diminishes the probability of request of the passenger. In this case, after talking directly with BlaBlaCar, the management team provided information regarding the total average probability of acceptance in BlaBlaCar France during the year 2017. This information resulted in an average probability of acceptance of around 91-92 percent. With this additional piece of information and the moment conditions of manual and automatic drivers, I simultaneously estimate the preferences of all passengers and the manual drivers.

The probability $q_{k,t}$ with which a passenger of a particular population segment k arrives at the market in route and time t is computed using the equilibrium proportions of each passenger kind that are present in the route. These proportions are used because the available data from the INSEE (Institut National de la Statistique et des Études Économiques), only contains information on the percentage of migrants and their origin, as it cannot reveal information about the ethnic origin of French nationals. Comparing the equilibrium composition of the cities of origin with the data from the INSEE, it can be noted that the percentage of minorities in each city is positively correlated to the

[19]As in (Hitsch et al., 2010), observing rejected passengers would allow me to directly estimate the parameters that govern the preferences of the drivers and to compute the probabilities of acceptance that are input to the passenger choices. Then, passenger preferences would be estimated using the observed request proportions.

proportion of migrants, but that INSEE figures are smaller by almost one-half.

Weights. As previously argued, in the estimation process, aggregation needs to take place. This aggregation is based on a unit of observation built upon an unobserved quality component, the econometric error term, which is specific to the population segments to which the driver and the passenger belong, and to the route and time frame in which the trip takes place. Drivers and passengers of each kind are present in different proportions on the different routes, e.g., naturally, passengers of the ethnic minority will likely be present in a smaller number than female passengers of the ethnic majority. Therefore, different aggregation units contain more accurate information than others.

Nevertheless, in an unweighted estimation process, all aggregation units are considered to contribute equally to the identification of the parameters of interest. To address this circumstance, and given the route and population segment of the passenger, each observation is weighed according to w_g. This weight is the product of the number of seats that the drivers in the aggregation unit have published and the average equilibrium passenger per seat in the route,

$$w_g = \sum_{j \in g} S_j \frac{\sum_{j \in t} 1\{passenger\ in\ group\ g\}}{\sum_{j \in t} S_j}. \tag{34}$$

Using the weights of each aggregation unit, the diagonal matrix W is built.

Instrumental variables. The only endogenous variable in the preferences of drivers and passengers is the price.[20] As it can be read from equations (26) and (27), in stage 2, prices are strategically set considering, amongst others, the unobserved quality components affecting the utility of the passengers making the trip. If the driver faces a positive passenger unobserved component, all else equal, higher prices will be established, yet the supplier may face a higher probability of request than if enduring a lower unobserved quality component. Intuitively, the direct correlation between the price and the unobserved quality component of the passenger is going to downplay the price sensitivity parameter of the demanding users. For similar reasons, it may also impact estimates of the price sensitivity of the driver.

Instruments are built using a similar approach to that in Reynaert & Verboven (2014). Following Chamberlain (1987), the author develops optimal instruments in the context of a classical demand random coefficient model *a la Berry*. Optimal instruments are optimal combinations of exogenous regressors built to minimize the sum of squared errors. These

[20]See Appendix B for evidence supporting the assumption that the booking mode of the driver does not signal to passengers the quality of the trip or their prospective match.

instruments stem from the gradient of the econometric objective function, and thus, their main component is the derivative of the econometric error with respect to the different types of covariates. In my setting, leaving the cost of rejection c aside, there exist three types of covariates and, thus, three different types of instruments.

In light of their work, trip-level instruments are weighted, within each aggregation unit, equally. Furthermore, to construct instruments for the features considered by the drivers in their decisions to accept and reject a request, drivers that choose the manual mode are distinguished from those drivers that choose the automatic, with the latter being assigned zero-valued instruments.[21]

At the trip level, price instruments are built using the average features of the drivers that have published a trip in the same route and a sufficiently close time interval (four-hour intervals, in this exercise). The preferred interpretation is that drivers in these routes and time frames share unobserved variable cost components. For example, if the trip is set to take place on a Sunday afternoon or another rest day, the driver will have more time to answer questions from passengers than if it is programmed for Thursday morning. Given the lower variable cost, more drivers will be willing to enter the market, and they will set lower prices. A second possible interpretation is that, despite the substantial capacity limitations, there exists a sufficient degree of competition amongst drivers in these routes and time frames that push them to reduce prices. Under this interpretation, the market structure of suppliers would affect the pricing choice of the drivers. Following this second interpretation, it should be the case that in those routes and times in which a large number of drivers arrive at the market, the overall price level is lower due to fiercer competition.

Computational methods. Given c, the computational procedure takes place in two phases. The first phase provides values to the non-linear parameters of the model, both for drivers and passengers. Given these values, the probabilities of acceptance for each driver and passenger are computed, and the contribution of the non-linear components to the marginal utilities of the passengers are quantified. Then, making use of a similar contraction mapping to that employed in Berry et al. (1995), the linear component of passenger utilities is calculated, and the associated parameters are estimated in the second phase.

In the second phase, linear preference parameters are recovered via weighted IV regression. The first phase follows an iterative process that provides guesses using Matlab's built-in algorithm for the minimization of the following objective function,

[21] For further detail on the construction of optimal instruments, see Appendix D.3.

$$\xi^{P\prime}WZ(Z'WZ)^{-1}Z'W\xi^{P}, \tag{35}$$

where Z is the matrix of instruments, W is the diagonal matrix containing the weight of each aggregation unit, and ξ^P is the vector of unobserved quality components of the drivers perceived by the passengers. c is calibrated so that the predicted average probability of acceptance by the model matches the average probability of acceptance of 91-92%.

As it is common in the demand estimation literature, those observations that present unfeasible proportions of passengers, those containing proportions equal to zero or equal or greater than one, are removed. These observations present a small proportion of the total published seats (c. 3.01%). Estimation-wise, they contain limited amounts of information, and they present more noise, which, given the logarithmic transformation embedded in the estimation process, encourages their removal.

6.2 Driver platform interaction costs

Given the trip utility parameters, driver variable costs can be recovered from the optimal pricing rules. Inverting the first-order conditions of the pricing problems of the automatic drivers, equation (19), variable costs are computed

$$vc_{j,t} = \frac{\sum_{k \in K} q_{k,t} \alpha_k^P pr_{k,j,t}(\overline{p}_{j,t}^{AUT})(1 - pr_{k,j,t}(\overline{p}_{j,t}^{AUT}))\delta_{j,k,t}^D(\overline{p}_{j,t}^{AUT}) + q_{k,t} pr_{k,j,t}(\overline{p}_{j,t}^{AUT})\alpha_j^D}{\sum_{k \in K} q_{k,t}(\alpha_k^P) pr_{k,j,t}(\overline{p}_{j,t}^{AUT})(1 - pr_{k,j,t}(\overline{p}_{j,t}^{AUT}))}. \tag{36}$$

The same line of reasoning applies to manual drivers. In their case, variable costs are recovered from equation (20) and have the following form,

$$vc_{j,t} = \frac{\sum_{k \in K} q_{k,t}(\alpha_k^P pr_{k,j,t}(\overline{p}_{j,t}^{MAN})(1 - pr_{k,j,t}(\overline{p}_{j,t}^{MAN}))log(.) + pr_{k,j,t}(\overline{p}_{j,t}^{MAN}) pa_{j,k,t}(\overline{p}_{j,t}^{MAN})\alpha_j^D)}{\sum_{k \in K} q_{k,t}(\alpha_k^P) pr_{k,j,t}(\overline{p}_{j,t}^{MAN})(1 - pr_{k,j,t}(\overline{p}_{j,t}^{MAN}))}. \tag{37}$$

Equilibrium prices are functions of the unobserved cost component of the drivers. Therefore, using equilibrium prices in the aforementioned conditions would likely lead to biased estimates of the variable cost parameters. To overcome this concern, prices need to be instrumented. Note that in this case, instruments cannot be of the kind proposed for the estimation of stages 3 and 4 precisely because, in the preferred interpretation, these instruments can be correlated with the driver-specific unobserved component of the variable cost $vc_{j,t}$. Passenger unobserved quality components $\xi^P_{k,k',t}$ are used instead. Given that drivers set prices strategically, they should establish higher prices whenever and wherever unobserved passenger quality components are larger. In other words, instruments for prices are the estimated quality unobservables.

Using instrumented prices, marginal costs are recovered from (36) and (37) and regressed over the covariates that inform the marginal cost $c_{j,t}$.

Identification, in this case, is immediate. The features of the drivers that, all else equal, present higher prices will result in parameters that increase the cost. On the contrary, those features systematically prevalent amongst comparable drivers that choose lower prices will result in cost parameters of a negative sign.

In econometric terms, the information provided by the pricing rule of each driver presents the same weight. Therefore, no weighting matrix is necessary, and the resulting parameters are obtained from standard regression techniques.

6.3 Driver manual booking mode cost

After estimating drivers' and passengers' trip preferences and drivers' variable costs, the expected utility associated with the observed booking mode choice and the counterfactual price and expected utility of the alternative booking mode can be computed for each driver. More explicitly, if the driver has chosen the manual mode, the expected utility associated with this mode is computed, and the price and expected utility associated with the automatic booking mode are also calculated by simulating the consequences of the alternative choice.

Given the shock associated with the manual mode cost and the limited amount of information that each individual booking mode choice contains, the decision undertaken by each driver in a specific market will not coincide with the true probability of choosing the manual mode for that driver.

As it happened with stages 3 and 4, aggregation needs to take place. Booking mode decisions are aggregated across drivers of the same gender and ethnic origin for each route and two-week period,

$$\frac{1}{J_g}\sum_{j\in g} 1\{bm_{j,t} = MAN\} = \frac{1}{J_g}\sum_{j\in g} pm_{j,t}. \tag{38}$$

Equation (38) equates the proportion of manual drivers in each group, on the left-hand side, with the average predicted probabilities of using the manual mode, on the right-hand side.

The moment condition in (38) is built upon the law of large numbers for independent but not identically distributed random variables. As it happened in the passenger request decision, the booking mode choice is a random variable whose expected value follows the form established in equation (30). Accordingly, for the law to hold, fixed cost taste shocks need to be independent across drivers and markets.

Furthermore, to enhance the computational capacity of the model, fixed cost shocks are assumed to follow a logistic distribution with zero mean. Under these assumptions, the previous moment condition (38) can be written as

$$\frac{1}{J_g}\sum_{j\in g} 1\{bm_{j,t} = MAN\} = \frac{1}{J_g}\sum_{j\in g} \frac{exp(U_{j,t}^{MAN}(\overline{p}_{j,t}^{MAN}) - U_{j,t}^{AUT}(\overline{p}_{j,t}^{AUT}) - \delta_{j,t}^{D,f})}{1 + exp(U_{j,t}^{MAN}(\overline{p}_{j,t}^{MAN}) - U_{j,t}^{AUT}(\overline{p}_{j,t}^{AUT}) - \delta_{j,t}^{D,f})}. \tag{39}$$

Similarly to the first estimation phase, although in the same aggregation unit, drivers may differ in certain features that can potentially impact the manual fixed cost. I denote these features as $X_{j,t}^{D,f,NG}$ and I distinguish them from the covariates that are equal to all the drivers in the aggregation unit (the constant term, the gender and the ethnic origin of the driver), which are denoted as $X_{j,t}^{D,f,G}$. Accordingly, each of these covariate groups is multiplied by a different set of parameters such that,

$$X_{j,t}^{D,f}{\prime}\beta^{D,f} = X_{j,t}^{D,f,G}{\prime}\beta^{D,f,G} + X_{j,t}^{D,f,NG}{\prime}\beta^{D,f,NG}. \tag{40}$$

The identification of the econometric model relies on different assumptions for each type of parameter. If the parameter refers to the features that are identical to all drivers within the same aggregation unit $\{\beta^{D,G,f}, \rho_t^{D,f}\}$, identification is immediate: all else equal, those features more prevalent amongst aggregation units with higher proportions of manual users will be attributed a smaller fixed cost. In the extreme case in which the manual mode fixed cost is only determined by these parameters, all the members within an aggregation unit will present the same fixed cost and, taking logarithms, parameters would be directly derived from a standard regression. If the parameter refers to features that differ across drivers within the same aggregation unit, variation across the mean features of similar aggregation units is required for the parameters to be identified. This condition encourages the usage of aggregation units as granular as possible, while the scarce information contained in the decision of each driver forces the aggregation of the booking mode decisions of a sufficiently large number of drivers.

As it happened in the estimation of stages 3 and 4, each aggregation unit is composed of a different number of drivers. Consequently, in an econometric sense, each aggregation unit contains a different amount of information. To factor this circumstance in the estimation process, each observation is weighted according to the number of drivers of which it is composed.

In computational terms, the estimation process is similar to that of the preferences associated with traveling. In the outer loop, values to the parameters governing the cost associated with the covariates that differ across trips within the same aggregation unit are guessed. Given the values of these parameters, the part of the fixed cost referring to the components that are identical across trips within the same aggregation unit is recovered, and with it, linear parameters are estimated *via* weighted least squares. Guesses in the outer loop are generated to minimize the objective function,

$$\xi^{D,f\prime} W^f \xi^{D,f}. \tag{41}$$

where W^f is the diagonal matrix containing the weight of each aggregation unit and $\xi^{D,f}$ is the vector of unobserved fixed cost components of the drivers. As happened with the estimation of trip preferences and for the same reasons, aggregations containing unfeasible proportions of manual drivers are removed.

7 Estimation Results

This section reports the main estimation results. Under this rubric, I present the main coefficients of interest and comment on the effects and potential mechanisms underlying the agents' preferences. To facilitate comprehension, a similar structure to that of the previous section follows. First, I analyze the preference parameters which define the utility of the trip for drivers and passengers. Second, I share and explore the variable cost parameters. Third, I discuss the preference parameters informing the manual mode fixed cost.

7.1 Passenger and driver trip preferences

Recall that the specific functional form of driver and passenger mean utilities, equations (10) and (14), is

$$\delta^P_{i,j,t} = p_{j,t}\alpha^P_{k_i} + X^P_{i,j,t}{}'\beta^P_{k_i} + \rho^P_t + \xi^P_{k_i,k_j,t}$$

$$\delta^D_{j,i,t} = p_{j,t}\alpha^D_{k_j} + X^D_{j,i,t}{}'\beta^D_{k_j} + \rho^D_t.$$

User preference parameters α, β depend on the user's type: the driver or passenger condition. Furthermore, they also depend on the population segment to which the user belongs. That is, they are allowed to vary with ethnic origin and gender. To facilitate the analysis of significant differences across the preferences of the different population segments, I use the following specification,

$$\theta^{P/D}_k = \theta^{P/D} + \theta^{P/D}_F \cdot 1\{\text{user is a woman}\} + \theta^{P/D}_B \cdot 1\{\text{user is a minority}\}$$

where θ is a generic notation for any of the preference parameters of drivers and passengers $\{\alpha^D, \alpha^P, \beta^D, \beta^{P,G}\, \beta^{P,nG}\}$. To make things even more clear, consider the following example. In the case of generic passenger i, his price sensitivity preference parameter is

$$\alpha^P_{k_i} = \alpha^P + \alpha^P_F \cdot 1\{i \text{ is woman}\} + \alpha^P_B \cdot 1\{i \text{ is minority}\},$$

where α^P is the price sensitivity of male passengers belonging to the ethnic majority, $\alpha^P + \alpha^P_F$ is the price sensitivity of female passengers of the ethnic majority, and so on. α^P_F and α^P_B represent the differences in price sensitivity between the baseline group, men of the ethnic majority, and the rest of the population segments, women of the ethnic majority and men of the ethnic minority, respectively.

Given this specification, for $c = 0$, the preference parameters are recovered. Table 4 and Table 5 report the preferences for the main features of interest, gender and ethnic origin, for passengers and drivers, respectively.

From Table 4, two conclusions arise. First, it becomes apparent that female passengers of the ethnic majority prefer to travel with female drivers. Second, there exists a substantial degree of ethnic-based homophily. That is, passengers of the ethnic majority prefer to travel with drivers of the ethnic majority, and passengers of the ethnic minority prefer to travel with drivers of the ethnic minority. In the case of female passengers of the ethnic majority, their preference towards ethnic in-group members is even more prominent.

At a secondary level, it is interesting to note that all passengers present similar and negative preferences toward the price set by the driver. Thus, trips with higher prices are less preferred and will obtain fewer requests irrespective of the gender and ethnicity of

Table 4: Passenger GMM Preference Estimation for $c = 0$

	Passenger Preferences (1)		
	Male-Majority Pas.	Δ *Female* Pas.	Δ *Minority* Pas.
Price	-0.1235*** (0.0062)	-0.0070*** (0.0021)	-0.0074** (0.0031)
Female Driv.	0.0227 (0.0722)	0.1541* (0.0896)	-0.0795 (0.1158)
Minority Driv.	-0.3110*** (0.0829)	-0.4777*** (0.0972)	0.8687*** (0.1491)
Controls		Yes	
Fixed Effects		Yes	
R^2		0.89	
Observations		8,004	

Note: The unit of analysis is the aggregation unit. The table reports the preferences of passengers towards the price of the trip and the gender and ethnic origin of the driver. Male-Majority Pas. reports the preferences of male passengers of the ethnic majority towards these features. ΔFemale Pas. and ΔMinority Pas. report the differences in the preferences between male passengers of the ethnic majority, female passengers of the ethnic majority and male passengers of the ethnic minority, respectively. The explanatory variable price is the final price paid by the passenger. The explanatory variable female driv. is a dummy that takes value one if the driver is a woman, and zero otherwise. The explanatory variable minority driv. is a dummy that takes value one if the driver belongs to the ethnic minority, and zero otherwise. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the passenger. Price elasticities of demand are in line with literature estimates for non-business travelers, around 1.8 on average. For further detail on the whole distribution of price elasticities, please see Appendix D.1.

To further illustrate the preferences of the different population segments towards the gender and ethnic origin of the drivers, I assess the price and reputation equivalents. For female passengers of the ethnic majority, traveling with a minority driver is equivalent to facing a price of 23.81 percent higher than the average price or sharing a car with a driver with 2.6 standard deviations less reputation. Men of the ethnic minority show the opposite pattern, as they are willing to pay 16.79 percent more, or share the car with a driver that has a reputation 1.89 standard deviations higher, in order to travel with a driver belonging to the ethnic minority.

Driver preferences are more homogenous. All drivers prefer to travel with female passengers, and all favor passengers of the ethnic majority. Nevertheless, these effects are attenuated amongst drivers of the ethnic minority.

Table 5: Driver GMM Preference Estimation for $c = 0$

	Driver Preferences (1)		
	Male-Majority Driv.	Δ*Female Driv.*	Δ*Minority Driv.*
Price	0.5826*** (0.0076)	-0.0224*** (0.0091)	0.0260 (0.0182)
Female Pas.	6.0738*** (0.5780)	0.2974*** (0.1119)	-1.0210*** (0.1382)
Minority Pas.	-2.1705*** (0.3117)	-0.1146 (0.1428)	0.6101*** (0.2296)
Controls		Yes	
Fixed Effects		Yes	
R^2		0.89	
Observations		8,004	

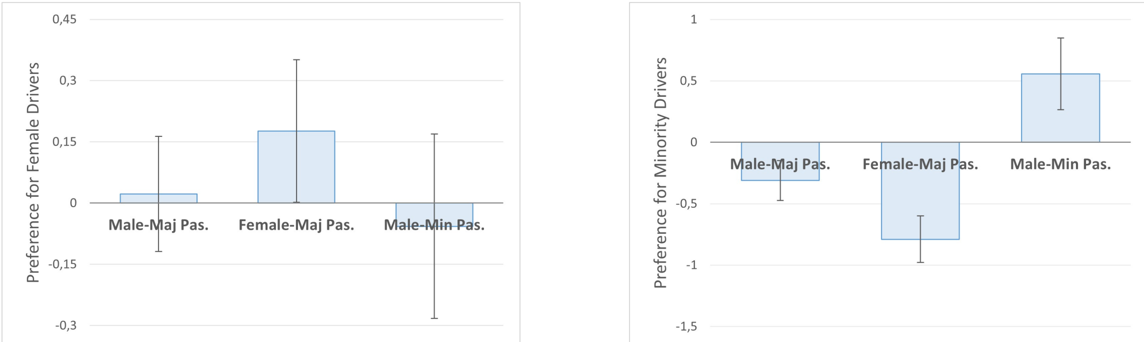
Notes: The unit of analysis is the aggregation unit. The table reports the preferences of drivers towards the price of the trip and the gender and ethnic origin of passengers. Male-Majority Driv. reports the preferences of male drivers of the ethnic majority towards these features. ΔFemale Driv. and ΔMinority Driv. report the differences in the preferences between male drivers of the ethnic majority, female drivers of the ethnic majority and male drivers of the ethnic minority, respectively. The explanatory variable price is the final price paid by the passenger. The explanatory variable female pas. is a dummy that takes value one if the passenger is a woman, and zero otherwise. The explanatory variable minority pas. is a dummy that takes value one if the passenger belongs to the ethnic minority, and zero otherwise. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I also analyze the price equivalent of drivers' preferences towards the gender and ethnic origin of their counterparts. I find that drivers of the ethnic majority are willing to be paid a price 41 percent lower than the average price to share their car with a woman passenger, and they require a price 14.48 percent higher to share their car with ethnic minorities. With this regard, female and male preferences are very similar. Minorities, nonetheless, present a downplayed version of these equivalents as they only require a price 10 percent higher than the average price to share their car with another minority.

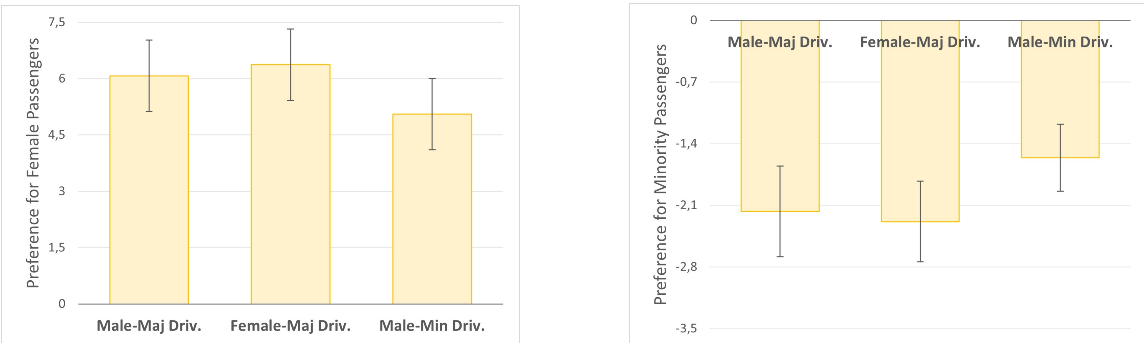
Regarding price sensitivity, coefficients are significantly positive. When a driver is confronted with the problem of accepting and rejecting a request, all else equal, the larger the price, the more willing he/she will be to accept the passenger. Interestingly, drivers belonging to an ethnic minority react more to the price than their majority equivalents.

To facilitate the comparison between driver and passenger preferences towards gender and ethnic origin, Figure 2 is built. This figure confirms that driver preferences towards women and minorities are more homogeneous: they prefer women and majorities, even if they belong to the ethnic minority. Passengers show preference patterns heavily influ-

Figure 2: Total Passenger and Driver Preferences Toward Women and Minorities



Passenger Preferences Toward Female (left) and Minority (right) Drivers



Driver Preferences Toward Female (left) and Minority (right) Passengers

Notes: Total preferences for passengers, on top (in blue), and drivers, at the bottom (in red); toward women (left) and minorities (right); of each population segment: male of the ethnic majority (left bar), female of the ethnic majority (center bar) and male of the ethnic minority (right bar). Bars report preferences, and brackets report 95% confidence intervals.

enced by their own gender and ethnicity. This leads to utilities, towards traveling with a woman or a minority driver, of differing signs for each population segment.

Finally, c is calibrated so that the predicted average probability of acceptance by the model matches the average probability of acceptance of 91-92%. In this case, it coincides that, for $c = 0$, the resulting average predicted probability of acceptance by the model is 91.47%. The cost of rejection c represents the loss of value that the outside option experiences after rejection. This study focuses on routes between well-connected big cities, with alternative means of transport that enjoy high frequencies and multiple possibilities (train, bus, plane, in addition to other trips posted through the platform). Therefore, from this perspective, the loss of value that the outside option experiences if rejection takes place is arguably irrelevant. Furthermore, psychological harm stemming from the rejection should be minor as this circumstance is only known by the requesting passenger and the rejecting driver. This result is in line with previous work in the field of dating apps by Hitsch et al. (2010). Exploiting observed rejection probabilities, they find rejection costs are not significantly different from zero.

Appendix D.5 reports alternative estimation results and the average acceptance probability for different rejection costs c. For a wide range of calibrated values of c, the main results and preference patterns hold identical. As previously argued, the main difference is that, as c increases, the observed differences between comparable manual and automatic drivers stem from the decision of the passenger to not request. Consequently, as c increases, the average probability of acceptance grows.

The mechanism behind ethnic homophily seems in line with the existence of a homogeneous minority group (essentially, of Arab and northern African origin). Cultural and language similarities may facilitate interactions across users of the same ethnic group. In this sense, positive preferences towards in-group members seem natural, especially considering the efforts of communication and coordination required for making a trip (gathering at a specific location at a particular time, communicating fluently, and agreeing on multiple other nuances that affect the overall comfort of the trip).

The mechanism behind female preferences is not evident. Female passenger preferences towards female drivers may stem from various reasons, from shared interests or opinions to specific safety issues. Given that driver and passenger preferences do not mirror each other and that the driver condition provides substantial control over multiple trip features, I hypothesize that safety concerns play an important role. Appendix E assesses the prevalence of female passengers on routes and times closer to events that raise awareness about sexual violence. In particular, I exploit the events taking place in the ten days prior to the audience in the so-called *Affaire Julie*, finding that, in those cities that were more exposed to the case, the proportion of female passengers significantly decreased in all trips but those posted by female drivers.

7.2 Driver platform interaction cost

Recall that driver variable costs of interaction through the platform, specified in equation (21), are a linear function of various trip and driver features,

$$vc_{j,t} = X_{j,t}^{D,v\prime}\beta^{D,v} + \rho_t^{D,v} + \xi_{j,t}^{D,v},$$

where $X_{j,t}^{D,v}$ is a vector of covariates that contains a constant term, features of the trip (number of seats, comfort mode, car value, amongst others) and features of the driver (including gender, ethnic origin, reputation and experience); $\rho_t^{D,v}$ are route and time fixed effects and $\xi_{j,t}^{D,v}$ is the unobserved quality component and the econometric error whose minimization drives the estimation of the parameters of the model.

Table 6: Driver Variable Cost Estimation for $c = 0$

	Variable Cost (1)
Female Driv.	-0.6673*** (0.0647)
Minority Driv.	2.4097*** (0.0833)
Reputation Driv.	-0.8213*** (0.0306)
Age Driv.	0.0493*** (0.0033)
Controls	Yes
Fixed Effects	Yes
R^2	0.51
Observations	94,945

Notes: The unit of analysis is the trip. The dependent variable is the variable cost $vc_{j,t}$. The explanatory variable Female Driv. is a dummy that takes value one if the driver is a woman and zero otherwise. The explanatory variable Minority Driv. is a dummy that takes value one if the driver belongs to the ethnic minority, and zero otherwise. The explanatory variable Minority Driv. is a dummy that takes value one if the driver belongs to the ethnic minority, and zero otherwise. The explanatory variable Reputation Driv. is the average rating received by the driver in previous trips, in a scale from 1 to 5. Age Driv. is the age of the driver measured in years. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

According to equations (36) and (37), variable costs are recovered for each trip and driver. The average variable cost is close to 0.7 utils, which is remarkably low if compared to the estimated average driver cost of sharing the trip with a passenger, around 9 utils (the latter cost is reflected by the coefficient interacting the constant term of drivers' trip utilities). This sharp difference suggests that marginal costs prior to the trip represent a minimal portion of the costs drivers face in the whole car-sharing process.

Results are contained in Table 6. Minorities face significantly higher costs than the rest of the drivers, perhaps because they have less time, are less accustomed to interactions through the platform, or need more effort to prepare the car thoroughly. As drivers build a better reputation, their marginal costs reduce, which seems in line with the fact that they become more accustomed to the whole trip process (having the car prepared for the trip or developing a more profound knowledge of the interface and tools to answer the questions of prospective passengers). Women also have lower costs, suggesting that they are better suited to interact through the platform (they may have a better understanding of the application itself or they are keener to interact with prospective passengers in a completely safe environment) or that the state of their car is better prepared to accommodate a new passenger. Older drivers present substantially higher variable costs, which is consistent with older people having more difficulties to interact through the platform and prepare their cars for the trip.

Table 7: Manual Mode Fixed Cost Estimation for $c = 0$

	Fixed Cost (1)
Female Driv.	-0.0007 (0.0025)
Minority Driv.	0.0011 (0.0028)
Reputation Driv.	-0.1845*** (0.0076)
Age Driv.	0.0038*** (0.0001)
Controls	Yes
Fixed Effects	Yes
R^2	0.62
Observations	1,610

Notes: The unit of analysis is the trip. The dependent variable is the fixed cost $fc_{j,t}$. The explanatory variable Female Driv. is a dummy that takes value one if the driver is a woman and zero otherwise. The explanatory variable Minority Driv. is a dummy that takes value one if the driver belongs to the ethnic minority, and zero otherwise. The explanatory variable Reputation Driv. is the average rating received by the driver in previous trips, in a scale from 1 to 5. Age Driv. is the age of the driver measured in years. Standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7.3 Driver manual mode cost

The specification for manual mode costs is stated in equation (28) as a linear function of different driver and trip parameters,

$$fc_{j,t} = \delta_{j,t}^{D,f} + \mu_{j,t}^{D,f} = X_{j,t}^{D,G,f\prime}\beta^{D,G,f} + X_{j,t}^{D,NG,f\prime}\beta^{D,NG,f} + \rho_t^{D,f} + \xi_{k_j,t}^{D,f} + \mu_{j,t}^{D,f}.$$

Following the aforementioned estimation process, fixed cost parameters are recovered.

Table 7 reports the estimates of the fixed cost parameters. First, the table shows that neither gender nor ethnic origin plays a significant role. Second, the average grading received by the driver is negatively related to the booking mode cost. Rating in this case may signal a deeper knowledge on the platform and its various tools and, consequently, a more profound understanding of the advanced booking mode. Third, in line with variable cost estimates, older drivers face larger manual mode costs. These drivers are not equally suited as their younger equivalents for the usage and development of the advanced manual mode and thus, they are forced into the automatic mode.

8 Counterfactual Exercises

The proposed counterfactual exercises aim to assess the implications of profile information and sorting tools in the participation and surplus of the different population segments on both market sides.

In this section, I distinguish driver surplus and driver-related outcomes (pricing and booking mode choice) from passenger surplus and their decisions (participation). Amongst driver outcomes, I analyze the percentage change in the manual mode usage and the variation of average equilibrium prices. I also report the variation in the proportion of minority passengers with whom they share the trip, both in relation to the seats they publish and to the number of passengers they accept. Amongst passenger outcomes, I measure the average change in the number of requests generated by each population segment and the average change in the number of effectively undertaken trips. The former and the latter may differ because, under specific counterfactual scenarios, drivers maintain the ability to screen and reject passengers.

The current state of the model takes as exogenous the entry decision of the drivers. This simplification is not inconsequential to the counterfactual assessment. As it will later become apparent, when information is limited or not available, agents have to generate expectations on the gender and ethnic origin of their counterparts. If entry is given, the proportions of drivers belonging to each population segment remain unaltered. Nevertheless, their utilities may shift in different directions. Those drivers suffering from utility losses will be prone to leave the market. This reaction presents two implications. First, available capacity will be lost. Second, the proportions of drivers operating through the platform and belonging to a specific population segment will change. This second condition will affect the decisions of passengers *via* their expectations. To build intuition on the latter effect, two panels for each counterfactual exercise are reported. Panel A shows the outcomes and surplus variation in all the routes. Panel B displays the outcomes and surplus variation only for those routes with the highest proportions of male drivers belonging to the ethnic minority.

Table 8: Passenger and Driver Surplus. Forced Automatic Mode

Panel A. All routes.

	Women Majority	Men Majority	Men Minority	Total
Manual mode	-100.00%	-100.00%	-100.00%	-100.00%
Price	+1.69%	+1.47%	+1.68%	+1.57%
Driver Surplus	-5.95%	-5.48%	-8.65%	-5.93%
Minority Passengers (over total pass.)	+16.06%	+12.48%	+6.09%	+12.30%
Participation (before driver choice)	-1.81%	-2.10%	-3.58%	-2.30%
Participation (after driver choice)	-1.68%	+2.43%	+15.50%	+3.29%
Passenger Surplus	-2.72%	-0.06%	+6.37%	-0.15%

Panel B. Top 10% routes with the highest proportion of ethnic minority drivers

	Women Majority	Men Majority	Men Minority	Total
Manual mode	-100.00%	-100.00%	-100.00%	-100.00%
Price	+2.37%	+1.87%	+2.04%	+2.01%
Driver Surplus	-9.28%	-8.75%	-14.37%	-9.93%
Minority Passengers (over total pass.)	+19.02%	+12.91%	+6.48%	+12.11%
Participation (before driver choice)	-4.20%	-3.92%	-6.85%	-4.63%
Participation (after driver choice)	-3.61%	+2.44%	+15.51%	+2.03%
Passenger Surplus	-5.61%	-1.57%	+1.37%	-2.66%

Notes: Panel A reports the differences between counterfactual and observed markets in all routes. Panel B the differences between counterfactual and observed markets only for those routes in the upper decile in terms of minority driver presence. The upper part of each panel reports driver-related decisions and surplus: manual mode usage, price choices, driver surplus, and the proportion of minority passengers with whom drivers share the trip (over the total number of accepted passengers). The lower part of each panel reports passenger-related decisions and surplus: request decisions (to participate in the platform before the driver choice), accepted requests (to participate in the platform after the driver choice), and passenger surplus.

8.1 Drivers forced into the automatic mode

Under this scenario, drivers cannot reject requests from passengers. In other words, the platform imposes a unique booking mode for all drivers, the automatic booking mode. Passengers still see driver profiles, and, given the gender and ethnic origin of the driver, they freely choose to make a request. In a sense, BlaBlaCar drivers become long-distance offline taxi operators who can freely set prices.

Table 8 contains the results of the first counterfactual exercise. As drivers can no longer accept or reject requests from their preferred passengers, they use prices to compensate for the increasing cost. Drivers with the starkest preferences, and those more likely to receive the least attractive passengers, will suffer the most. Although they have a higher probability of being requested by a woman of the ethnic majority, precisely due to their preferences, female drivers of the ethnic majority face harsher consequences than their male equivalents. Alongside women, minority drivers suffer the most from this measure. As they receive the highest share of minority passengers, this group endures the highest average cost. Hence, when rejection is no longer possible, their welfare reduces.

The overall price increase negatively impacts passenger requests. Women are notably affected because their preferred drivers, women of the ethnic majority, increase prices the most. Moreover, given their generally high probability of acceptance, their effective participation is also reduced. The only population segment whose surplus increases are men of the ethnic minority. Even though they see their utility decrease with the price rise, the universal acceptance rule has a net positive effect, boosting their effective participation and total welfare.

The described effects, both for drivers and passengers, are escalated in those routes with a higher proportion of drivers belonging to the ethnic minority. Given that these routes also present the highest proportion of passengers of the ethnic minority, drivers increase prices even further to compensate for the average utility loss, while passengers have to face the negative consequences of this price rise.

8.2 Driver blind profiles

This subsection aims to quantify the consequences of a policy oriented toward anonymizing the information drivers share through their profiles. More precisely, the aim is to understand the implications attached to removing any profile feature (personal picture and name) that could be used to identify the gender and ethnic origin of the user. This scenario is the closest, in spirit, to the Airbnb display change adopted in 2017, as from that moment on, neither the name nor the personal picture of the host was available on the main search page.

If established, the marketplace managed by BlaBlaCar would become closer to that of Uber, as the features of the driver, aside from experience and average rating, would no longer be available. As in the case of Uber, drivers could reject requests from passengers. The only difference would be that an algorithm would not set the price. Drivers would freely set them instead.

Model-wise, in this scenario, passengers have to form expectations about the exact identity of the counterpart given the prevalence of each kind of driver on each route.

Table 9 reports the results of the second counterfactual exercise. On the rider side, only drivers from the ethnic minority benefit from the measure. Intuitively, passengers can no longer screen these drivers, and, despite losing their advantage over minorities, drivers belonging to an ethnic minority gain more from having access to counterparts belonging to the ethnic majority. Due to the size difference between minorities and majorities, on average, ethnic minority drivers set higher prices and experience a parallel surplus increase.

Table 9: Passenger and Driver Surplus. Driver Blind Profiles

Panel A. All routes.

	Women Majority	Men Majority	Men Minority	Total
Manual mode	+0.75%	+0.22%	-1.27%	+0.16%
Price	-3.07%	-0.95%	+2.51%	-0.80%
Driver Surplus	-2.61%	-0.88%	+10.30%	-0.40%
Minority Passengers (over total pass.)	+11.50%	+6.26%	-28.54%	+1.97%
Participation (before driver choice)	+0.39%	+0.75%	+3.16%	+0.89%
Participation (after driver choice)	+0.22%	-0.87%	+0.64%	-0.08%
Passenger Surplus	-0.16%	-0.73%	+0.61%	-0.29%

Panel B. Top 10% routes with the highest proportion of ethnic minority drivers

	Women Majority	Men Majority	Men Minority	Total
Manual mode	+1.40%	+0.53%	-1.05%	+0.25%
Price	-2.65%	-0.98%	+0.7%	-0.81%
Driver Surplus	-3.87%	-2.30%	+7.62%	-0.83%
Minority Passengers (over total pass.)	+20.44%	+14.56%	-29.43%	+3.55%
Participation (before driver choice)	+0.16%	+0.87%	+4.92%	+1.15%
Participation (after driver choice)	-0.24%	-1.14%	+1.22%	-0.27%
Passenger Surplus	-1.14%	-1.51%	+3.03%	-0.49%

Notes: Panel A reports the differences between counterfactual and observed markets in all routes. Panel B the differences between counterfactual and observed markets only for those routes in the upper decile in terms of minority driver presence. The upper part of each panel reports driver-related decisions and surplus: manual mode usage, price choices, driver surplus, and the proportion of minority passengers with whom they share the trip (over the total number of accepted passengers). The lower part of each panel reports passenger-related decisions and surplus: request decisions (to participate in the platform before the driver choice), accepted requests (to participate in the platform after the driver choice), and passenger surplus.

The opposite occurs for drivers belonging to the ethnic majority: given that they become less attractive to ethnic majority passengers as they can no longer identify them, they have to reduce prices. To minimize the higher likelihood of traveling with minority passengers, majority drivers are more prone to using the manual mode. Despite adjusting prices and the booking mode, ethnic majority drivers are worse off. The negative implications are more prevalent amongst women of the ethnic majority, as they favor traveling with female counterparts, and they are heavily preferred by female passengers.

Overall, passengers of the ethnic majority are worse and participate less when driver profiles are blind. This result is a direct consequence of the preferences of each passenger kind towards the gender and ethnic origin of the driver, even after internalizing the price reduction. The starker the preferences towards these features, the stronger the negative implications for the population segment. Moreover, the higher the likelihood of the trait (the more prevalent it is amongst the population of drivers) that induces a decrease in utility, the more it will impact the decision of the passenger to make a request. Consequently, women of the ethnic majority are again the population segment suffering the most regarding participation and overall surplus.

These results, although in line with the estimated preferences, force the adoption of new conclusions in terms of the extended positive perception linked to the removal of profile information. In particular, these results show that diminishing profile information presents negative asymmetric consequences for the different population segments, substantially harming the surplus of female drivers.

Finally, two conclusions emerge from the comparison between panels A and B. First, total driver surplus decreases even though prices do not change in such an abrupt fashion across population segments. Routes with a high proportion of male drivers of the ethnic minority also present a higher proportion of passengers of the ethnic minority. Due to the higher average cost, drivers belonging to the ethnic majority are more willing to choose the manual mode and, if choosing the automatic mode, less willing to reduce prices, which, at the same time, decreases even further the probability of receiving a request. Second, on the other side, passengers from the ethnic majority face higher prices than on the rest of the routes. They update their expected utility downwards, as they are more prone to send a request to a driver belonging to an ethnic minority. These two effects lead to a sharper decline in passenger surplus for the ethnic majority.

8.3 Blind driver profiles and drivers forced into automatic mode

When drivers are forced into the automatic booking mode, and passengers do not have access to any information that identifies the gender and ethnic origin of the driver, both market sides become blind. In a sense, the market resembles that of online long-distance taxis: passengers cannot know the identity of the driver, and drivers cannot reject any passenger.

Table 10 contains the results of the third counterfactual exercise. In this counterfactual scenario, drivers need to balance two opposing forces. On the one hand, they must adapt prices to their new perceived identity. Thus, those drivers that are women or belong to the ethnic majority will have incentives to decrease prices to compensate passengers' expected utility loss, as happened in the second counterfactual exercise. On the other hand, given that all drivers lose their capacity to reject undesired requests, they will compensate themselves by increasing the only strategic variable whose control they have left, the price. As the proportion of less attractive passengers increases, the second force will dominate, and price levels will rise.

Table 10: Passenger and Driver Surplus. Driver Blind Profiles, Forced Automatic Mode

Panel A. All routes.

	Women Majority	Men Majority	Men Minority	Total
Manual mode	-100.00%	-100.00%	-100.00%	-100.00%
Price	-0.29%	+0.81%	+4.37%	+1.11%
Driver Surplus	-8.82%	-6.48%	+2.15%	-6.44%
Minority Passengers (over total pass.)	+16.70%	+12.67%	+5.47%	+12.91%
Participation (before driver choice)	-1.34%	-1.48%	-2.45%	-1.65%
Participation (after driver choice)	-1.21%	+3.07%	+17.04%	+4.07%
Passenger Surplus	-3.77%	-0.00%	+8.44%	-0.25%

Panel B. Top 10% routes with the highest proportion of ethnic minority drivers

	Women Majority	Men Majority	Men Minority	Total
Manual mode	-100.00%	-100.00%	-100.00%	-100.00%
Price	+0.97%	+1.26%	+2.91%	+1.70%
Driver Surplus	-13.84%	-11.36%	-6.02%	-10.96%
Minority Passengers (over total pass.)	+19.14%	+12.95%	+6.60%	+13.01%
Participation (before driver choice)	-3.47%	-3.41%	-5.45%	-3.92%
Participation (after driver choice)	-2.99%	+3.05%	+17.58%	+3.63%
Passenger Surplus	-8.20%	-2.20%	+6.01%	-3.01%

Notes: Panel A reports the differences between counterfactual and observed markets in all routes. Panel B the differences between counterfactual and observed markets only for those routes in the upper decile in terms of minority driver presence. The upper part of each panel reports driver-related decisions and surplus: manual mode usage, price choices, driver surplus, and the proportion of minority passengers with whom they share the trip (over the total number of accepted passengers). The lower part of each panel reports passenger-related decisions and surplus: request decisions (to participate in the platform before the driver choice), accepted requests (to participate in the platform after the driver choice), and passenger surplus.

On the other side of the market, three forces will determine passengers' welfare and effective participation. Given the sign of the price change, passengers will increase or reduce their requests. Nevertheless, the absence of information on the ethnic origin and gender of the drivers will directly and negatively impact their utility towards those riders that would be otherwise favored. Thirdly, the inability of the drivers to reject any request will benefit those passengers that suffer from a lower probability of acceptance.

Accounting for the various forces, the aggregate effect of a completely opaque marketplace is negative. Only drivers and passengers of the ethnic minority benefit from the change. Nonetheless, not even drivers of the ethnic minority benefit from this measure when the proportion of passengers and drivers of the ethnic minority is sufficiently high. Once again, female drivers and passengers are damaged the most. The situation of female passengers of the ethnic majority is especially prominent, as they experience the most profound surplus decrease, both compared to the rest of the population segments and counterfactual scenarios. The combination of the overall price increase, their stark preferences towards gender and ethnic origin, and their high baseline probabilities of acceptance make them the population segment, on the demand side, that is most negatively impacted by the alternative design.

9 Conclusions

To estimate the preferences of the different population segments towards gender and ethnic origin and to understand the implications of profile information in their surplus and participation decisions, I develop a two-sided model that accommodates the main strategic decisions taking place in BlaBlaCar. In this model, drivers decide whether they will be able to accept or reject the requests from prospective passengers, they choose the price at which transactions will take place, and, given their booking mode choice, they accept and reject the requests of incoming passengers. Passengers, anticipating the probability of being rejected, choose to make a request or not.

The model is estimated using a novel data set that contains information on all the drivers and passengers participating in *circa* one hundred thousand trips published in BlaBlaCar France, the leading car-sharing platform in the country. In particular, the data set identifies the gender and ethnic origin of all the users that operate through the application.

Model estimates show that women of the ethnic majority, especially passengers, prefer to travel with women counterparts. Moreover, a significant degree of ethnic-based homophily is shown to take place. The preference towards in-group members, in terms of ethnic origin, is even starker in the case of women of the ethnic majority.

Finally, counterfactual exercises assessing the implications of alternative marketplaces with varying degrees of opacity are performed. In these counterfactual scenarios, in contrast to the established perception, it is shown that blind profiles do not necessarily result in a higher surplus of ethnic minorities. Precisely, this paper shows that forcing the automatic mode does hinder the overall utility of drivers belonging to the ethnic minority. Nevertheless, overall, these counterfactual exercises prove that blind profiles have a remarkably harmful effect on women of the ethnic majority, as their participation and surplus decrease the most. These effects are especially prevalent amongst those routes that present the highest proportion of the minority population.

The identification of the mechanism behind women's preferences towards other women, the estimation of the model for women of the ethnic minority, and the role of the gender and ethnic identity of the reviewers remain to be solved. Regarding the mechanism, I hypothesize that safety concerns push women to prefer to travel with other women. To test this conjecture, I aim to use the setting of BlaBlaCar to assess the implications of sexual violence cases in women's participation. Concerning the role of the identity of reviewers, the current data set is being enhanced to cover the information of the specific rating and the identity attached to each review. Once this information is available, the main objective is to understand the value that users provide to the opinions of in-group and out-group members and the implications of a blind review system in which the reviewer's identity is no longer verifiable. Additionally, I will be able to analyze the relationship between the probability of leaving a review, the content and emotion embedded in the text, with the gender and ethnic origin of the user and the counterpart.

As a concluding remark, I would like to emphasize that this paper does not justify the fairness, or lack thereof, behind the motives backing the sorting choices of the different population segments. This paper has aimed to understand, in a purely descriptive way, firstly, the role that the gender and ethnicity of the deciding user and the counterpart play in their preferences and overall trust towards each other and, secondly, the asymmetric consequences, in terms of participation and utility, of alternative marketplaces with less information on these features. Of course, the motives behind the preferences of each population segment must inform those policies intended to influence the design of information transmission mechanisms. Assessing the underlying mechanism of the documented preference patterns is a research question that I can hopefully shed some light on in upcoming work.

References

Adachi, H. (2003). A search model of two-sided matching under nontransferable utility. *Journal of Economic Theory*, *113*(2), 182–198.

Becker, G. S. (1957). *The economics of discrimination*. Chicago: University of Chicago Press.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Brons, M., Pels, E., Nijkamp, P., & Rietveld, P. (2002). Price elasticities of demand for passenger air travel: a meta-analysis. *Journal of Air Transport Management*, *8*(3), 165–175.

Browning, M., Chiappori, P.-A., & Weiss, Y. (2014). *Economics of the family*. Cambridge University Press.

Burdett, K., & Coles, M. G. (1999). Long-term partnership formation: marriage and employment. *The Economic Journal*, *109*(456), F307–F334.

Burdett, K., & Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, *39*(2), 257–273.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of econometrics*, *34*(3), 305–334.

Choo, E., & Siow, A. (2006). Who marries whom and why. *Journal of political Economy*, *114*(1), 175–201.

Edelman, B., Luca, M., & 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., & Luca, M. (2014). Digital discrimination: The case of airbnb. com. *Harvard Business School NOM Unit Working Paper*(14-054).

Eeckhout, J., & Kircher, P. (2011). Identifying sorting—in theory. *Review of Economic Studies*, *78*(3), 872–906.

Farajallah, M., Hammond, R. G., & Pénard, T. (2019). What drives pricing behavior in peer-to-peer markets? evidence from the carsharing platform blablacar. *Information Economics and Policy*, *48*, 15–31.

Hitsch, G. J., Hortaçsu, A., & Ariely, D. (2010). Matching and sorting in online dating. *American Economic Review*, *100*(1), 130–63.

Ivaldi, M., & Palikot, E. (2020). Sharing when stranger equals danger: Ridesharing during covid-19 pandemic.

Kakar, V., Franco, J., Voelz, J., & Wu, J. (2016). Effects of host race information on airbnb listing prices in san francisco.

Lambin, X., & Palikot, E. (2018). Does reputation hinder entry? study of statistical discrimination on a platform.

Laouenan, M., & Rathelot, R. (2017). Ethnic discrimination on an online marketplace of vacation rental.

Laouénan, M., & Rathelot, R. (2022). Can information reduce ethnic discrimination? evidence from airbnb. *American Economic Journal: Applied Economics*, *14*(1), 107–32.

Pope, D. G., & Sydnor, J. R. (2011). What's in a picture? evidence of discrimination from prosper. com. *Journal of Human resources*, *46*(1), 53–92.

Postel-Vinay, F., & Robin, J.-M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, *70*(6), 2295–2350.

Reynaert, M., & Verboven, F. (2014). Improving the performance of random coefficients demand models: the role of optimal instruments. *Journal of Econometrics*, *179*(1), 83–98.

Shimer, R., & Smith, L. (2000). Assortative matching and search. *Econometrica*, *68*(2), 343–369.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, *8*, 321–340.

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

APPENDIX

A. The seat as the unit of analysis

This assumption eliminates any possibility of the driver choosing to accept or reject a certain passenger based on the utility provided by other potential or existing passengers. It simplifies the rejection decision of the driver, as each seat will entail a decision based exclusively on the utility provided by the specific requesting passenger and the specific utility associated with traveling with the empty seat. The alternative, allowing the driver to decide on the acceptance and rejection of all passengers making the request simultaneously and given the identity of all of them, is appealing: understanding the different car compositions as bundles from which the driver can choose provides the possibility of capturing synergies across the different passenger types. For example, if a majority driver has already accepted two majority passengers he may be more willing to accept a minority passenger in the third seat.

In a sense, following the previous line of reasoning, one would expect that majority drivers that publish one seat are less willing to accept a minority passenger than those who publish three seats or, at least, that they are willing to accept them with a different probability. Following an identical specification to that of (2), I analyze the role that the interaction between the booking mode, the gender and ethnic origin of the driver, and the number of published seats has in the proportion of traveling passengers.

Table 11 shows that minority passengers are not more prevalent in those cars driven by manual drivers with a larger number of published seats: it cannot be rejected that the interaction between the number of seats and the booking mode is different from zero. Along these lines, if the manual driver is a woman or belongs to an ethnic minority, a larger number of seats is not significantly correlated to a change in the proportion of accepted passengers of this or any other kind.

Thus, given that the proportion of passengers of any kind, and especially men and minorities, in cars driven by manual drivers, does not significantly vary with the number of published seats, it seems that manual drivers choose to accept and reject passengers without regard to the rest of potential passengers. Thus, it can be concluded that using the seat and not the car as the main unit observation does not entail any major concern in terms of the fitness of the data and the ability of the model to account for the real decisions taken by the users. Nevertheless, I include the number of originally published seats, in the vector of covariates referring to the mean utilities both of passengers and drivers, to appease any concern along these lines in a reduced form fashion.

B. Evidence on booking mode quality signals

The specification of the passengers' utility limits the difference in the requests received by manual and automatic drivers to the expected cost of rejection. In other words, the passenger cannot anticipate the quality of the driver by his booking mode choice.

It could be argued that the booking mode may signal certain quality aspects that may reflect a positive trait of the driver: drivers that choose a manual booking mode are more selective and caring, thus passengers are likely to be treated with more care; moreover,

Table 11: Results on the Seat as the Unit of Analysis

	Fema-Maj Pas. (1)	*Male-Maj Pas.* (2)	*Fema-Min Pas.* (3)	*Male-Min Pas.* (4)
Price	-0.0022*** (0.0003)	-0.0021*** (0.00003)	-0.0011*** (0.0001)	-0.0021*** (0.0002)
Female D.	0.0273** (0.0137)	-0.0129 (0.0128)	-0.0033 (0.0060)	-0.0178*** (0.0079)
Minority D.	-0.0700*** (0.0149)	-0.0476*** (0.0139)	0.0035 (0.0065)	0.0159** (0.0086)
Reputation D.	0.0279*** (0.0009)	0.0226*** (0.0009)	0.0049*** (0.0004)	0.0079*** (0.0005)
Experience D.	0.0001*** (0.0000)	0.0001*** (0.0000)	0.0000*** (0.0000)	0.0000*** (0.0000)
Manual	0.1088*** (0.0294)	0.0532** (0.0274)	-0.0234** (0.0128)	-0.0768*** (0.0172)
Man. x Female D.	-0.0327 (0.0201)	0.0265 (0.0187)	0.0138 (0.0088)	0.0119 (0.0122)
Man. x Minority D.	0.0365 (0.0244)	0.0396* (0.0227)	0.0213** (0.0106)	0.0480*** (0.0150)
Seat x Manual	-0.0055 (0.0035)	0.0007 (0.0033)	0.0013 (0.0015)	0.0016 (0.0020)
Seat x Man. x Fem. D.	0.0129** (0.0059)	-0.0047 (0.0033)	-0.0036 (0.0018)	-0.0049 (0.0034)
Seat x Man. x Min. D.	-0.0073 (0.0067)	0.0017 (0.0062)	-0.0044 (0.0029)	-0.0060 (0.0038)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.11	0.06	0.01	0.03
Observations	94,945	94,945	94,945	94,945

Notes: The unit of observation is the trip. The four dependent variables, in Columns (1)-(4), are the proportions of accepted passengers, belonging to each population segment, with respect to the initial number of published seats. The variable price is in euros, including the platform's commission. The variable female driver is a dummy that takes value one if the driver is identified as a female and zero otherwise. The variable minority driver is a dummy that takes value one if the driver belongs to the ethnic minority. The variable manual is a dummy that takes value one if the driver has chosen the manual booking mode and zero otherwise. The set of controls includes other information on the trip and the driver: driver's age, availability of personal picture, car value, comfort mode, the number of initially published seats, time until departure, distance from the city center, and others. The four regressions include route and time-fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

they are ensured that the rest of passengers in the trip have also been carefully chosen. In the opposite sense, manual drivers can be seen as more picky, therefore they may be more willing to establish stringent rules in the car or they may be less open and prone to conversation than automatic drivers.

Additionally, the booking mode of the driver may offer different information to passengers belonging to different population segments, or the offered information may be of more use for a specific gender or ethnic origin. Let us consider two examples. On the one hand, If female passengers are more sensitive in their interactions, they may value more those drivers that choose the manual mode as this booking alternative may signal a higher pickiness of the user. On the other hand, minority passengers may anticipate that manual majority drivers tend to have more prejudices towards them and thus, they could expect that, if making the trip with them, the driver may treat them worse than if they joined a comparable automatic driver.

I propose an exercise that shows that none of the previous concerns take place in reality. To tackle these issues, I make use of an alternative outcome variable, the grading left by the passengers to each driver, and I analyze how this variable is related to the booking mode of the driver and the interactions with the gender and ethnic origin of the passenger and the driver himself. The exact specification that I employ is,

$$R_{j,i,t} = X'_{j,i,t}\beta_1 + \epsilon_{j,i,t}$$

where $R_{j,i,t}$ is the rating left by passenger i to driver j in the trip taking place in route and time t, $X_{j,i,t}$ is a matrix containing, for each observation, the observable features of the driver (gender, ethnic origin, age, experience and previous average reputation), the observables of the passenger (gender, ethnic origin, experience and previous average reputation), the booking mode and their interaction.

Irrespective of the focus of the analysis, gender or ethnic origin, manual and automatic drivers receive the same average grading. In terms of the gender of the users, female drivers do not receive a significantly different grading when they choose the manual booking mode and female passengers do not provide significantly different ratings to drivers that use the manual booking mode and the automatic one. In terms of ethnic origin, the same result holds true, neither the ethnic origin of the driver nor that of the passenger presents a significant correlation with the grading provided to manual drivers. If there are no systematic differences between the grading of manual and automatic given the gender and ethnic origin of the users, it should be the case, all else equal, that they provide the same treatment to passengers. Consequently, passengers cannot extract from the booking mode of the driver any additional piece of information about the quality of the driver.

10 C. Chebyshev's Law of Large Numbers

The standard law of large numbers states that the average of independent and identically distributed random variables converges in probability to their expected value. This law is usually employed in logit discrete choice models to construct estimation moments that match the probabilities of consumption with the real observed market shares.

In this setting, each car and driver is different even if belonging to the same aggregation unit. Thus, the choice of the passengers to join any of the drivers on a route, even if they belong to the same aggregation unit, is a random variable with a different mean and variance. Let us denote $Y_{j,k,t}$ as the decision of driver j and passenger i of population segment k to travel together in route and time t

Table 12: Ethnic-Based Rating Results

	Rating (1)
Minority Pass.	0.038316* (0.0221)
Majority Driv.	0.0384742*** (0.0094)
Manual	-0.0022 (0.0122)
Min. Pass. x Maj. Driv.	-0.0686*** (0.0246)
Man x Min. Pass.	-0.0329 (0.0314)
Man x Maj. Driv.	-0.0038 (0.01295)
Man x Maj. Driv. x Min. Pass.	0.0506 (0.0346)
Driver average rating	0.6874*** (0.0118)
Driver experience	0.0001*** (0.0000)
Controls	Yes
Fixed Effects	Yes
R^2	0.04
Observations	82,939

Notes: The unit of analysis is the review. The table reports the coefficients of regressing the individual ratings over a series of explanatory variables. The explanatory variables include the ethnic belonging of the driver, the ethnic belonging of the passenger, a dummy variable for the manual booking mode, the interaction of the previous three variables, the previous rating and number of received reviews by the driver. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

$$E[Y_{j,i,t}] = \mu_{j,k,t} = pa_{j,k,t}pr_{k,j,t}.$$

When the number of total seats in the clusterization unit $\sum_{j\in c} S_{j,t}$ is sufficiently big, one can show that, in the application of a special case of the LLN,

$$\frac{1}{\sum_{j\in g} S_{j,t}} \sum_{j\in g} Y_{j,k,t} \to_p \frac{1}{\sum_{j\in g} S_{j,t}} \sum_{j\in g} pa_{j,k,t}pr_{k,j,t}.$$

This law of large numbers is based on independent random variables with different means and variances. Using Chebyshev's inequality, for any u,

Table 13: Gender-based Rating Results

	Rating (1)
Female Pass.	-0.03458*** (0.0070)
Female Driv.	0.0135 (0.0092)
Manual	-0.0038 (0.0068)
Fem. Pass. x Fem. Driv.	-0.0137 (0.0128)
Man. x Pass. Fem.	-0.0015 (0.0093)
Man. x Fem. Driv.	-0.0014 (0.0118)
Man. x Fem. Driv. x Fem. Pass.	0.0041 (0.0165)
Driver average rating	0.6871*** (0.0118)
Driver experience	0.0001*** (0.0000)
Controls	Yes
Fixed Effects	Yes
R^2	0.04
Observations	82,939

Notes: The unit of analysis is the review. The table reports the coefficients of regressing the individual ratings over a series of explanatory variables. The explanatory variables include the gender of the driver, the gender of the passenger, a dummy variable for the manual booking mode, the interaction of the previous three variables, the previous rating and number of received reviews by the driver. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

$$pr\left(\left|\frac{1}{\sum_{j\in g} S_{j,t}}\sum_{j\in g} Y_{j,k,t} - \frac{1}{\sum_{j\in g} S_{j,t}}\sum_{j\in g} pa_{j,k,t}pr_{k,j,t}\right| > u\right) = \frac{\sigma^2_{c,k}}{u^2\sum_{j\in g} S_j} \rightarrow_p 0,$$

where $\sigma^2_{c,k} = \frac{1}{\sum_{j\in g} S_{j,t}}\sum_j S_{j,t}E[(Y_{j,k,t} - \mu_{j,k,t})^2]$, because driver and passenger taste shocks are independent across drivers and passengers.

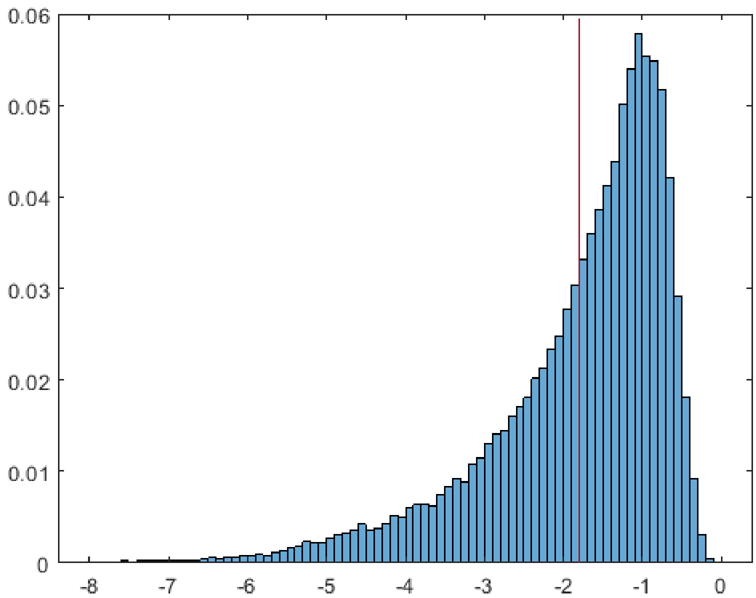
D. Detail of the estimation results

This appendix covers four different topics: the detail of the demand price elasticity; the description of the rest of the parameter estimates for passenger and driver trip utilities; the first stage regression results; the OLS estimates of driver and passenger trip parameters and the analysis of sensitivities.

D.1. Price elasticity of demand

The detail of the price elasticity estimates is shown in Figure 3. The mean price elasticity is -1.81 and the median is -1.47. The distribution of price elasticity estimates is tilted to the left.

Figure 3: Distribution of Price Elasticity Estimates



Notes: The Figure reports the distribution of price elasticity estimates for all the trips and passenger population segments.

These results are in line with other estimates found by the literature Brons et al. (2002). In particular, they are in line with the average elasticities of non-business class passengers.

D.2. Driver and passenger preferences

Passenger trip preferences for the main features are contained in Table 14. Driver trip preferences for the main features are contained in Table 15.

Table 14: Passenger GMM Preference Estimation for $c = 0$

	Passenger Preferences (1)		
	Male-Majority Pas.	*ΔFemale Pas.*	*ΔMinority Pas.*
Price	-0.1235*** (0.0062)	-0.0070*** (0.0021)	-0.0074** (0.0031)
Female Driv.	0.0227 (0.0722)	0.1541* (0.0896)	-0.0795 (0.1158)
Minority Driv.	-0.3110*** (0.0829)	-0.4777*** (0.0972)	0.8687*** (0.1491)
Reputation Driv.	0.3223*** (0.1131)	-0.0039 (0.0952)	-0.0419 (0.1632)
Experience Driv.	0.0014*** (0.0004)		
Profile picture Driv.	0.0680 (0.2404)		
Mail verific. Driv.	0.0777 (0.1541)		
Seats	0.1620 (0.1286)		
Comfort mode	0.3072 (0.2476)		
Car value	0.0000 (0.0000)		
Controls		Yes	
Fixed Effects		Yes	
R^2		0.89	
Observations		8,004	

Notes: The unit of analysis is the aggregation unit. The table reports the preferences of passengers towards the price of the trip, the gender and ethnic origin of the driver. Male-Majority Pas. reports the preferences of male passengers of the ethnic majority towards these features. ΔFemale Pas. and ΔMinority Pas. report the differences in the preferences between male passengers of the ethnic majority, female passengers of the ethnic majority and male passengers of the ethnic minority, respectively. The explanatory variable price is the final price paid by the passenger. The explanatory variable Female Driv. is a dummy that takes value one if the driver is a woman, and zero otherwise. The explanatory variable Minority Driv. is a dummy that takes value one if the driver belongs to the ethnic minority, and zero otherwise. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 15: Driver GMM Preference Estimation for $c = 0$

	Driver Preferences (1)		
	Male-Majority Driv.	*ΔFemale Driv.*	*ΔMinority Driv.*
Price	0.5826*** (0.0076)	-0.0224 (0.0091)	0.0260*** (0.0182)
Female Pass.	6.0738*** (0.0578)	0.2974*** (0.1119)	-1.0210*** (0.1382)
Minority Pass.	-2.1705*** (0.3117)	-0.1146 (0.1428)	0.6101** (0.2296)
Seats	0.0076 (0.2001)	0.0541 (0.2644)	0.0148 [illegible]

D.3. IV Construction

Now the question is, what is a valid instrument when the observation unit is the result of aggregating the decisions of various passengers and drivers? A natural first approach is to employ, for each aggregation unit, the weighted average of the instruments available at the seat level. Recall that the estimation process requires the aggregation of units that are identical in certain covariates (e.g., the ethnic origin of the driver) and different amongst others (e.g., the price of the driver). Moreover, the preferences of the driver are related to the econometric error in a non-linear fashion. These peculiarities make it advisable to understand how optimal instruments should be built.

With this purpose, a similar approach to that in Reynaert & Verboven (2014) is followed. Following Chamberlain (1987), the author develops optimal instruments in the context of a classical demand random coefficient model *a la Berry*. Optimal instruments are optimal combinations of exogenous regressors built to minimize the sum of squared errors. These instruments stem from the gradient of the econometric objective function and thus, their main component is the derivative of the econometric error with respect to the different types of covariates. In my setting, leaving the cost of rejection c aside, there exist three types of covariates and thus, three different types of instruments.

In particular, building on equations the moment conditions, the unobserved quality component can be transformed into a function of the utilities of the passengers and the drivers of each aggregation unit,

$$\begin{aligned}\xi^P_{k,k',t} &= log(\frac{\sum_{j\in g}\sum_{s\in S_{j,t}} a_{j,k,t}r_{k,j,t}}{q_{k,t}}) - \beta^{P,G}_k X^{P,G}_{i,j,t} - \rho^P_t \\ -log(\sum_{j\in g}\sum_{s\in S_{j,t}} &\frac{exp(\delta^D_{j,k,t})}{1+exp(\delta^D_{j,k,t})} exp(\alpha^P_{k_i} p_{j,t} + \beta^{P,NG}_{k_i} X^{P,NG}_{i,j,t} - \frac{c}{exp(\delta^D_{j,k,t})})(1-pr_{k,j,t})).\end{aligned} \tag{42}$$

The derivative of this equation with respect to the preference parameters of the driver is

$$\frac{\partial \xi^P_{k,k',t}}{\partial \beta^D} = -\frac{\sum_{j\in c}\sum_{s\in j} exp(X^{P,NG}_{i,j,t}\beta^{P,NG} - c(\frac{1}{pa_{j,i,t}}-1))(1-pr_{i,j,t})pa_{j,i,t}(1-pa_{j,i,t})X^D_{j,i,t}}{\sum_{j\in c}\sum_{s\in j} exp(X^D_{j,i,t}\beta^{P,NG} - c(\frac{1}{pa_{j,i,t}}-1))(1-pr_{i,j,t})pa_{j,i,t}}. \tag{43}$$

If the booking mode is automatic, the derivative, trivially, takes a value of zero, in line with a zero probability of rejection.

The derivative of the first equation with respect to the passenger nonlinear preference parameters is

$$\frac{\partial \xi^P_{k,k',t}}{\partial \beta^{P,NG}} = -\frac{\sum_{j\in c}\sum_{s\in j} exp(X^{P,NG}_{i,j,t}\beta^{P,NG} - c(\frac{1}{pa_{j,i,t}}-1))(1-pr_{i,j,t})pa_{j,i,t}X^{P,NG}_{i,j,t}}{\sum_{j\in c}\sum_{s\in j} exp(X^{P,NG}_{i,j,t}\beta^{P,NG} - c(\frac{1}{pa_{j,i,t}}-1))(1-pr_{i,j,t})pa_{j,i,t}} \tag{44}$$

The derivative with respect to the passenger linear preference parameters is

$$\frac{\partial \xi^{P}_{k,k',t}}{\partial \beta^{P,G}} = -X^{P,G}_{k_i,j,t} \tag{45}$$

In general terms, optimal instruments for the unit of analysis, the aggregation unit, should consist of a weighted average of the trip-level instruments. For linear passenger preference parameters, given that all drivers within the aggregation unit present the same covariates, instruments are also equal and any weights can be used to aggregate the trip-level instruments. For non-linear passenger preference parameters, weights depend on the overall utility provided by the non-linear part. The case of driver preference parameters is more complex, identification of the parameters relies on the differences between manual and automatic drivers. Thus, higher rates of rejection improve the identification of the parameters and are key for the construction of instruments. In fact, optimal instruments are proportional to the probability of rejection, which makes all automatic drivers have zero-valued instruments for the covariates influencing drivers' preferences.

D.4. First-stage IV Results

Instruments for prices are built using the average features of other drivers that travel in the same route in a 4-hour time span. The purpose of this instrument is to convey the unobserved cost of interaction that drivers face in a certain market. If multiple drivers join in a route and similar time frame, it must be the case that they have lower costs on average and thus, they will tend to establish lower prices.

Amongst the features of the rest of the drivers, the one that stands out the most in terms of significance and overall sense is the existence of other drivers in the same route and time frame (the average over the constant term).

Given that the gathered units of analysis (at the individual trip level) are different from those used by the model (at the aggregation unit level), I present first-stage results for trip-level prices and average aggregation unit prices. In the former, the first-stage regression is a simple OLS over all the features used in the estimation process and the, strictly speaking, instruments. In the latter, the same variables are employed but averaged over the members of the aggregation unit. Moreover, for aggregation-unit-level analysis, I employ the same weighting mechanism employed in the estimation process. Hence, instead of traditional OLS results shown in Table 16 are obtained *via* Weighted Least Squares.

D.5. Sensitivity Analysis

This appendix explores how different values of the rejection cost affect passenger and driver preference parameters and the average probability of acceptance.

Table 17 confirms the previous intuitions regarding the identification of the parameter c. As the calibrated value of c increases, the observed differences in the composition of

Table 16: IV. First Stage Results

	Individual price OLS (1)	Average price WLS (2)
Other driver presence	-0.6831*** (0.1460)	-1.1169*** (0.4923)
Other female driver presence	0.0868 (0.0440)	-0.3829 (0.5623)
Other minority driver presence	0.0948 (0.0748)	-0.0829 (0.5900)
Female driv.	0.1918 (0.1686)	0.1319 (0.1227)
Minority driv.	1.6315*** (0.0745)	1.4785*** (0.0830)
Controls	Yes	Yes
Fixed Effects	Yes	Yes
R^2	0.75	0.98
Observations	94,945	8,004

Notes: The unit of analysis is the trip in Column (1) and the aggregation unit in Column (2). The regression in Column 1 assigns equal values to all observations, while Column 2 weighs each observation according to the number of seats in the aggregation unit. The reported explanatory variables are the presence of other drivers in the route and a close time range to the trip, the presence other female drivers and minority drivers, a dummy variable that takes value one if the driver is a woman and a dummy variable that takes value one if the driver belongs to an ethnic minority. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

manual and automatic drivers are not only imputed to the choices of the drivers but also to the anticipation by passengers of the increasingly negative consequences of the rejection. Consequently, as the rejection cost increases, the average probability of acceptance that results from the estimation also increases. Regarding the values of the parameters, passenger preferences remain very similar both in terms of sign and magnitude. Driver preferences tend to scale, especially the price sensitivity parameter, likely due to the imputation of the differences in composition between manual and automatic drivers to the request decisions of the passengers.

D.6. Linear and non-linear decomposition of preferences

Equation (42) can be rewritten so as to explicitly convey the interactions between the unobserved component $\xi^P_{k,k',t}$ and the rest of the parameters of the model,

Table 17: Probability of Acceptance and Main Coefficients by Cost of Rejection

	c = 0.1 (1)	*c = 0.2* (2)	*c = 0.4* (3)	*c = 0.8* (4)
Average Probability of Acceptance	92.34%	93.98%	95.34%	97.46%
Price Sensitivity Pass. pref.	-0.1234	-0.1117	-0.1102	-0.1073
Woman-to-Woman Pass. pref.	0.1759	0.1818	0.1790	0.1718
Majority-to-Minority Pass. pref.	-0.3108	-0.3715	-0.3637	-0.3753
Minority-to-Minority Pass. pref.	0.5545	0.5238	0.5882	0.5704
Price Sensitivity Driv. pref.	0.5799	0.5479	0.7771	1.1762
Woman-to-Woman Driv. pref.	6.4197	11.1712	12.2753	14.0309
Majority-to-Minority Driv. pref.	-2.1656	-1.9623	-2.2574	-2.4241
Minority-to-Minority Driv. pref.	-1.5681	-1.2513	-1.9559	-2.4794

Notes: The upper part of the table reports the average probability of acceptance for different costs of rejection. The middle part of the table reports the parameters of passenger preferences towards price, women and minorities. The lower part of the table reports the parameters of driver preferences towards price, women and minorities.

$$\sum_{j\in g}\sum_{s\in S_{j,t}} a_{j,k,t}r_{k,j,t} = \tag{46}$$

$$= q_{k,t}\sum_{j\in g}\sum_{s\in S_{j,t}} \frac{exp(\delta^D_{j,k,t})}{1+exp(\delta^D_{j,k,t})} exp(\delta^P_{k,j,t} - \frac{c}{exp(\delta^D_{j,k,t})})(1-pr_{k,j,t}) \tag{47}$$

$$\propto q_{k,t}exp(X^{P,G\prime}_{i,j,t}\beta^{P,L}_k + \rho^P_t + \xi^P_{k,k',t})\sum_{j\in g}\sum_{s\in S_{j,t}} \frac{exp(\delta^D_{j,k,t})}{1+exp(\delta^D_{j,k,t})} exp(p_{j,t}\alpha^P_k + X^{P,NG\prime}_{i,j,t}\beta^{P,NG}_k - \frac{c}{exp(\delta^D_{j,k,t})}) \tag{48}$$

where k' is the population segment of the drivers that belong to group g. I factor the elements of passengers' utilities that are common to the aggregation unit $\{X^{P,G\prime}_{i,j,t}\beta^{P,L}_k, \rho^P_t\}$. Through this exercise, it can be seen that these components interact linearly with the econometric error, while the rest of the components are non-linearly related to the unobserved quality component in varying fashions. This condition is essential for the identification strategies of the different parameter types, as it will become clear from the following discussion. Accordingly, automatic mode moments can be rewritten,

$$\sum_{j\in g}\sum_{s\in S_{j,t}} r_{k,j,t} = \tag{49}$$

$$\bar{q}_{k,t}\sum_{j\in g}\sum_{s\in S_{j,t}} exp(\delta^P_{k,j,t})(1-pr_{k_p,j,t}) \tag{50}$$

$$\propto q_{k,t}exp(X^{P,G\prime}_{i,j,t}\beta^{P,G}_k + \rho^P_t + \xi^P_{k,k',t})\sum_{j\in g}\sum_{s\in S_{j,t}} exp(\alpha^P_k p_{j,t} + X^{P,NG\prime}_{i,j,t}\beta^{P,NG}_k). \tag{51}$$

The equivalences in (48) and (51) highlight how the different types of parameters interact with the econometric error term $\xi^P_{k,k',t}$. Although in different ways, driver preference parameters $\beta^D_{k'}$ and passenger preference parameters towards the features of the driver and the trip that differ across drivers in the same aggregation unit $\beta^{P,NG}_k$ are non-linearly related to the unobserved quality component. The average utility perceived by passengers for making the trip and the utility derived from making it with a driver that is a woman or that belongs to an ethnic minority $X^{P,G\prime}_{i,j,t}\beta^{P,G}_k$, and the utility of making the trip in a specific route and time ρ^P_t are linearly related to the error term.

E. The mechanism behind female preferences

This appendix aims to provide evidence supporting the conjecture that women's preference to travel with other women is influenced by safety concerns and, in particular, is influenced by concerns regarding sexual violence and abuse.

In BlaBlaCar, drivers and passengers share a trip with an *a priori* completely stranger. This situation exposes the user, especially in the case of passengers, to a prolonged interaction in an isolated environment. Unfortunately, these circumstances limit the possibility of defense, putting women in a highly vulnerable position.

Given that passengers are the side of the market most exposed to these negative consequences, I propose an analysis that tries to understand how is the prevalence of each population segment related to events that raise awareness of sexual abuse.

To do so, I exploit the legal procedures that took place in relation to, what was called by the press, the *Affaire Julie* and the following multitudinous manifestations taking place in the cities of Paris, Lyon, Marseille, Strasbourg, Metz and Avignon in support of Julie. Julie is the name of the woman that, between the ages of 13 and 15, was subject to sexual violence by 20 firemen in Paris. After the initial demand for rape against the 20 firemen, in July 2019, the court replaced the raping charges with sexual aggression charges. In November 2020, the reclassification of the charges was ratified by the Court d'Apellation, ratification which was appealed by Julie's mother in the days that follow the court decision. The first audience after the appeal was announced to take place on the tenth of February 2021. In response to this audience, manifestations were arranged in

Table 18: Results on Female Preference Mechanism

	Fema-Maj Pas. (1)	*Male-Maj Pas.* (2)	*Fema-Min Pas.* (3)	*Male-Min Pas.* (4)
Price	-0.0025*** (0.0003)	-0.0021*** (0.0003)	-0.0011*** (0.0001)	-0.0020*** (0.0002)
Woman D.	0.0181*** (0.0038)	-0.0004 (0.0034)	0.0028* (0.0016)	-0.0045** (0.0021)
Minority D.	-0.0484*** (0.0041)	-0.0400*** (0.0039)	0.0020 (0.0018)	0.0108*** (0.0024)
Reputation D.	0.0284*** (0.0009)	0.0226*** (0.0009)	0.0050*** (0.0004)	0.0079*** (0.0005)
Experience D.	0.0002*** (0.0000)	0.0002*** (0.0000)	0.0002*** (0.0000)	0.0000*** (0.0000)
Treat	-0.0285*** (0.0091)	-0.0111 (0.0084)	-0.0072* (0.0039)	-0.0045 (0.0052)
Treat x Woman D. x Manual	0.0441*** (0.0186)	0.0065 (0.0172)	0.0156* (0.0080)	-0.0001 (0.0105)
Treat x Minority D. x Manual	-0.0079 (0.0189)	0.0065 (0.0172)	0.0027 (0.0082)	-0.0134 (0.0108)
Controls	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.11	0.06	0.01	0.03
Observations	94,945	94,945	94,945	94,945

Notes: The unit of observation is the trip. The four dependent variables, in columns (1)-(4) are the proportions of accepted passengers, belonging to each population segment, with respect to the initial number of published seats. The variable price is in euros, including the platform's commission. The variable female driver is a dummy that takes value one if the driver is identified as a female and zero otherwise. The variable minority driver is a dummy that takes value one if the driver belongs to the ethnic minority. The variable manual is a dummy that takes value one if the driver has chosen the manual booking mode and zero otherwise. The set of controls includes other information on the trip and the driver: driver's age, availability of personal picture, car value, comfort mode, the number of initially published seats, time until departure, distance from the city center, and others. The four regressions include route and time-fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the aforementioned cities and, in the week preceding the dates, initiatives against sexual violence and abuse started to take place.

As with the empirical exercises proposed in paper 3, the objective is to understand the relation between the proportion of passengers of each kind with the features, gender and ethnic origin, of the driver. A new variable is added to capture how this relation was modulated in the week preceding the audience of the tenth of February in those cities in which the manifestations took place. This variable is denoted by *treat*, as these are the regions and times in which the publicity of Julie's case was most prominent, and is interacted with the gender and ethnic origin of the driver. As the case of Julie is a clear sign of the existence of sexual violence against women, exposure to events that support the victim raises awareness of this issue and makes it more visible.

From the results in Table 18, it is shown that women, both from the ethnic minority and majority, are less prevalent in those cars driven by men in the routes and times in which the events of analysis took place. The prevalence of men does not experience any alteration in these routes and times. These results appear to be in line with the conjecture that women's preference towards women is modulated by safety concerns.

In the previous analysis, I do not claim to have performed an empirical exercise based on an exogenous source of variation that allows to ensure the causal effect between sexual violence and women's preference, I just provide some evidence pointing towards this direction. As a future line of research, I aim to exploit the geographical and time variation of news of sexual aggression to understand women's preferences. Moreover, if news contains expressions and terms that may allow us to identify the ethnic origin of the aggressor, I may also exploit this additional source of variation to test the mechanism behind the differences in preferences across different ethnic groups.

Car-sharing resembles a wide variety of standard situations and encounters that women have to face in their daily life. Women's decisions, unlike those of men, are likely to be influenced by the possibility of suffering sexual violence. These possibilities shape their preferences towards the different kinds of interlocutors they face, limiting their actions and economic possibilities.

11 F. Covid restrictions

Firstly, I would like to briefly summarize the main restrictions taking place in France at the time in which the covered transactions took place. As a general note, restrictions during this period are a result of the different waves and variants of COVID-19. Given the similar size and nature of the analyzed cities, restrictions tend to affect them uniformly. On the 20th of July 2020, masks became compulsory in France in a wide array of public and private spaces, including BlaBlaCar. Although during the summer, no major lockdown was imposed by the French government, from the 17th of October to the 15th of December, new restrictions were established. In particular, a curfew prevented all citizens from staying out of their residences between 9 p.m. and 6 a.m. Other measures intended to limit non-essential mobility between departments were also established. On December 15th, all traveling restrictions were lifted, and the curfew was removed. No mobility restrictions were reestablished until the 20th of March, when non-essential travel was again prohibited in the most densely populated departments.

Until this day, transactions between the same routes have been extracted. The recorded comparable period goes from the 27th of October 2021 to the 24th of March 2022. Although this period was free from any mobility restraints, it was characterized by an outstanding increase in the price of gasoline and diesel. This price increase pronounced significantly with the entry of the new year. Gasoline prices and BlaBlaCar prices show a strong positive correlation. This feature hampers the direct comparison between the two periods of time.

In broad terms, the number of published trips increased remarkably. Especially when comparing the number of published trips between the end of 2020 and the end of 2021,

Table 19: Data Differences. 2020-2021 vs 2021-2022

	27.10.20-31.12.21	*01.01.20-24.03.21*	*27.10.21-31.12.22*[1]	*01.01.21-24.03.22*[2]
Total Number of trips	31,264	63,681	83,423	121,268
Proportion of majority woman D.	0.30	0.30	0.32	0.30
Proportion of majority men D.	0.52	0.52	0.52	0.55
Proportion of minority woman D.	0.02	0.02	0.04	0.03
Proportion of minority man D.	0.16	0.16	0.12	0.12
Proportion of majority woman P.	0.43	0.43	0.43	0.44
Proportion of majority men P.	0.36	0.36	0.35	0.36
Proportion of minority woman P.	0.07	0.07	0.09	0.08
Proportion of minority man P.	0.13	0.13	0.13	0.12
Manual booking mode	0.68	0.67	0.68	0.68
Seats	2.66 (0.76)	2.65 (0.76)	2.68 (0.77)	2.69 (0.77)
Passengers per seat	0.48 (0.42)	0.46 (0.44)	0.42 (0.40)	0.37 (0.39)
Price	24.56 (9.72)	24.82 (9.73)	26.64 (9.97)	28.09 (10.12)
Gasoline av. price[3]	1.26	1.33	1.55	1.73

Notes: (1) and (2) Passenger and driver composition from a sample of 100 individuals. (3) information from carbu.com. In each column, I report a series of key statistics for a specific period of time. I divide each data set into two time frames to avoid potential misreporting due to seasonal behaviors. The upper part of the table analyzes the number of published trips and the proportions of each driver population segment. The middle part of the table analyzes the proportion of accepted passengers, for each population segment. The lower part of the table reports the multiple features of the trip: the proportion of manual mode usage, the number of initially published seats per trip, the number of passengers per seat, the price of the trip and the price of gasoline.

it seems that mobility restrictions were indeed altering the regular traveling patterns of the French population.

The question is whether these mobility restrictions had a higher impact on a specific population segment, either as drivers or passengers. To assess this feature, I extract random samples of 100 drivers and passengers for the new periods of comparison, classifying each person according to the aforementioned gender and ethnic classifications. Composition-wise, the proportions of all population segments remain almost identical. If anything, the proportion of men minority drivers seems to have reduced, which is likely related to the fact that this population segment occupies job positions in which remote work is difficult to implement.

In terms of the main features of the app, all of them (including booking mode and the number of published seats) remain similar. The main difference is the price increase, which seems related to the overall pricing pattern of fuel. At the same time, the overall proportion of passengers per seat has decreased. Given the estimated price elasticities, the reduction in the proportions of passengers seems to be a direct result of the price increase.