The Efficiency of A Dynamic Decentralized Two-sided Matching Market*

Tracy Xiao Liu
Tsinghua University

Zhixi Wan
University of Hong Kong

Chenyu Yang
University of Maryland

April 20, 2021

Abstract

This paper studies a decentralized dynamic matching market by using data from a Chinese ride-sharing platform to estimate a model of search and matching between drivers and passengers. We measure passenger valuations of trips, driver preferences, distributions of planned search length, and waiting costs. To estimate driver preferences, we develop a new inequality approach based on an incomplete dynamic model. We assess whether centralized algorithms that require different information sets can improve efficiency, and we show that information on agent preferences and planned search length can enable the platform to increase revenue and the total surplus of drivers and passengers.

1 Introduction

A key question online platforms face is how to organize matching between participants. Uber and Lyft are two prominent examples of centralized platforms that use algorithms to direct drivers to passengers. On other platforms such as Airbnb, eBay, and most online marketplaces, a user has substantial autonomy in choosing a partner. While a centralized algorithm could maximize certain objectives—such as revenue or the number of matches—by design, the algorithm may require (sometimes extensive) user information as input and may be itself an expensive engineering product. In contrast, a decentralized market may be less costly to operate. Yet due to constraints such as an individual user's limited

*We thank discussants and participants at many seminars and conferences as well as the referees and editor for their helpful comments. Correspondence to: Tracy Xiao Liu: liuxiao@sem.tsinghua.edu.cn; Zhixi Wan: zhixiwan@hku.hk; Chenyu Yang: chny.yang@gmail.com

information and search frictions, the outcome could fall short of social-planning or profit-maximizing goals. In this paper, we study the extent to which centralization can improve the welfare of a dynamic decentralized market. Our empirical context is a ride-sharing platform in China. We exploit data on driver and passenger choices to estimate passenger valuations of trips, driver preferences for passengers, distributions of planned search length, and costs of waiting on the platform. Using the estimates, we compare the outcomes of the decentralized market with the performance of a series of centralized algorithms that require increasingly more information.

The platform we study is DiDi Hitch (herein "Hitch"), which is a platform operated by DiDi, the largest ride-sharing company in the world. Based in China, DiDi operates multiple differentiated ride–sharing platforms. In 2020, an average of 60 million rides occurred on DiDi every day, more than Uber and Lyft combined. The majority of the rides came from the company's main operation, DiDi Express (herein "Express"), a centralized platform similar to Uber or Lyft. This paper focuses on the smaller decentralized peer-to-peer platform, Hitch. During our data period in 2018, Hitch facilitated one million rides per day.

Hitch has several features similar to other decentralized platforms. First, rather than relying on a dispatch system to allocate passenger demands to waiting drivers, the platform lets drivers view and choose from a list of passenger requests that specify departure time as well as pickup and drop-off locations. Second, preference heterogeneity on one side of the market is important in the matching process. The platform features sufficiently low distance-based fares that most drivers on this platform are commuters (rather than professional drivers) who travel from one particular location to another, such as from home to work. A driver thus prefers a passenger who requests a similar route to reduce the detour. Third, due to the nature of commuting, both drivers and passengers have limited time to search on the platform; if they cannot find suitable partners, drivers will drive alone and passengers will use alternative transportation to ensure an on–time departure.[1]

We aim to measure how centralization can improve the efficiency of the matching market. We consider three sets of information necessary for centralization: (a) preferences of drivers and passengers currently in the market, (b) how long these drivers and passengers plan to search, and (c) the arrival time of future drivers and passengers (plus their preferences and how long they will search). The information in (a) helps an algorithm to identify matches that maximize social surplus, and the information in (b) and (c) allows

[1]The latter two features are shared to various degrees by a number of decentralized platforms. Similar ride-sharing services are offered by BlaBlaCar in Europe and Grab in Southeast Asia. On Airbnb, a prospective customer searches for hosts near the planned vacation location before starting the trip. On dating websites and apps, people search for partners with certain attributes.

an algorithm to correctly trade off the social values of making matches now versus later. In addition to measuring the welfare gap between the decentralized market and the welfare-maximizing planner with information in (a), (b), and (c), we also evaluate the performance of algorithms that use a subset of the information.

The first step of our analysis is to recover passenger valuations of trips on Hitch, driver preferences for passengers, distributions of planned search length, and costs of waiting on the platform. In our model, drivers and passengers make decisions over finite and privately known horizons. On the passenger side, we assume that passengers are forward-looking and choose between waiting on the Hitch platform, requesting on Express to get rides immediately, and using alternative transportation. We assume that the passenger valuation of a trip does not depend on the driver identity, and we characterize the optimal passenger behavior. The characterization allows us to exploit passenger choices and their timing to estimate the passenger parameters. On the driver side, we use a new inequality approach to estimate drivers' heterogeneous preferences for passengers. We assume that a driver faces a finite horizon, and the option value of waiting is 0 when a driver reaches the end of her horizon and nonnegative before. We do not additionally specify how the driver computes this option value. We then derive bounds of a driver's choice probabilities, which lead to conditional moment inequalities that set identify driver preferences. To estimate the discount factor, we fully specify the driver's dynamic choice problem and exploit data on the timing of driver decisions. Our moment inequality method uses the estimator in Chernozhukov et al. (2019).

In our welfare evaluation, we first consider a greedy algorithm, which requires preference information in (a) but not (b) or (c). The algorithm matches a new driver (passenger) to one of the waiting passengers (drivers) upon the new agent's arrival, provided the match is individually rational. In our context, because we assume that passenger valuations for Hitch rides are higher than passengers' outside option and do not depend on driver identities, individual rationality means that the utility of the driver in a match is higher than her outside option. Compared with the decentralized market, we find that the algorithm is able to achieve higher match rates (percentages of matched drivers and passengers) at the cost of worse match qualities (measured by the length of the detour of realized matches). The total surplus of drivers and passengers decreases by 5%–10% (95% confidence interval). We next consider what Akbarpour et al. (2020) call a "patient" algorithm, which uses information in (a) and (b). In this algorithm, when any driver (passenger) is about to leave, the platform matches the leaving driver (passenger) with one of the waiting passengers (drivers) if an individually rational match exists. This algorithm produces a thicker market (a higher average number of passengers and drivers waiting at an instant) and achieves higher match rates and platform revenue than the greedy algorithm or the decentralized market. However, the total surplus is 6%–10% lower than the decentralized market due to the longer waiting time and lower match qualities. We next consider two variants of the patient algorithm. The first variant emphasizes the qualities of matches and the second variant emphasizes the number of matches. Compared with the patient algorithm, the total surplus is 10%–16% higher and the platform revenue is 0.5%–5% lower under the first variant, while the total surplus is 11%–18% lower and the platform revenue is 1%–3% higher under the second variant. Finally, we simulate the solution of a planner with information in (a), (b), and (c) with the objective of maximizing the total surplus of drivers and passengers. We find that this solution achieves similar match rates as the patient algorithm but it obtains large and statistically significant improvements in welfare and platform revenue over the decentralized market and other centralized algorithms. The total surplus is 22%–30% higher and the platform revenue is 18%–31% higher than the decentralized market.

Contributions and Related Work

First, we contribute to the literature of dynamic matching. The market design literature studying the optimality of dynamic matching markets has identified the trade–offs between market thickness, waiting time, and match quality in various frameworks (Loertscher et al. (2018); Ashlagi et al. (2019); Baccara et al. (2020); Akbarpour et al. (2020)). In particular, Akbarpour et al. (2020) consider an environment similar to our empirical setting. We generalize the theoretical framework by considering agents with ex ante heterogeneous preferences and measuring the qualities of the matches and agent surplus.

Second, we make a methodological contribution. Our estimation of driver preferences uses a set of weak assumptions about how drivers compute the option value of waiting. A number of papers have explored the implications of incomplete static models in auctions (Haile and Tamer (2003)) and discrete games (Ciliberto and Tamer (2009); Ho (2009); Eizenberg (2014); Pakes et al. (2015); Berry et al. (2016); Wollmann (2018); Ciliberto et al. (2020); Fan and Yang (2020)). With dynamic models, a typical approach fully specifies the optimization problem or a notion of dynamic equilibrium. In comparison, the estimates of our preference parameters rely on sufficiently weak conditions for us to accommodate situations, for example, where drivers condition their strategies on subsets of a complex state space or use heterogeneous strategies. A paper related to our inequality approach is Holmes (2011), which estimates a fully specified spatial dynamic optimization problem facing Walmart and constructs inequalities implied by the optimality conditions.

There have been recent interests in understanding the efficiency of decentralized dynamic matching markets in empirical contexts. Dynamic models featuring optimizing agents have been used to analyze the mechanism in the allocation of kidneys from deceased donors (Agarwal et al. (2021)), public housing assistance (Waldinger (2018)), and bear-hunting licenses (Reeling and Verdier (2020)).[2] We focus on the information necessary to achieve centralization and the extent to which centralization improves welfare.[3]

In the transportation industry, several papers use dynamic models to study the taxi market (Lagos (2003); Frechette et al. (2019); Buchholz (2021))[4] and ride-sharing platforms (Shapiro (2018); Bian (2020); Castillo (2020); Rosaia (2020)). Another transportation context is dry bulk shipping (Brancaccio et al. (2020a,b)). These papers are unified in considering carriers that make optimal labor supply decisions (and in most cases, the decisions on which locations to search) as opposed to choosing customers.

Finally, with the rise of the sharing economy, a number of empirical papers have studied peer-to-peer decentralized matching markets. Some examples include estimating the importance of preferences in explaining assortative matching on an online dating website (Hitsch et al. (2010)), the heterogeneous competitive effects of Airbnb (Farronato and Fradkin (2018)), how search strategies respond to experimental variations in information on market thickness (Bimpikis et al. (2020); Fong (2020); Li and Netessine (2020)), and the determinants of a platform's growth (Cullen and Farronato (2020)).[5]

Road Map In the rest of the paper, we discuss the empirical context and the data in Section 2. We describe our model and estimate passenger preferences in Section 3. Section 4 presents our driver model and inequality approach. We conduct the welfare evaluation by simulating the outcomes of a decentralized market and a series of centralized algorithms in Section 5, and briefly conclude in Section 6.

[2]We direct readers to Agarwal and Somaini (2020) for a review of the empirical models of nontransferable utility matching. For transferable utility-matching games, we refer readers to Choo and Seitz (2013); Chiappori and Salanié (2016); Fox (2017). In particular, Choo (2015) studies a frictionless dynamic marriage market and Fox (2008) considers a repeated game in a labor market setting.

[3]A theoretical literature on kidney exchanges (e.g., Roth et al. (2005, 2007); Ünver (2010)) also examines properties of centralized dynamic matching algorithms. We also note a large empirical literature on trade frictions, the effects on allocative efficiency, and the roles of intermediaries (e.g., Allen, Clark and Houde (2014); Gavazza (2011, 2016); Salz (2020) and references therein), and in particular dynamic auctions (Adachi (2016); Hendricks and Sorensen (2016); Bimpikis, Elmaghraby, Moon and Zhang (2020); Bodoh-Creed, Boehnke and Hickman (2021)).

[4]We note that the methodology in these papers is fundamentally related to the search models in the labor literature, where the dynamic search and match problems have been studied both theoretically and empirically in a large number of papers (surveyed in, e.g., Mortensen (1986); Pissarides (2000); Shimer and Smith (2001); Canals and Stern (2002); Rogerson, Shimer and Wright (2005); Eckstein and Van den Berg (2007)).

[5]We also refer readers to the review article Einav et al. (2016).

2 Empirical Context

In the first subsection, we describe the mechanism and interfaces of the Hitch platform. The next subsection presents the data and summary statistics. To further understand driver behavior and motivate the structural model, we conducted interviews with drivers. The last subsection summarizes our findings.

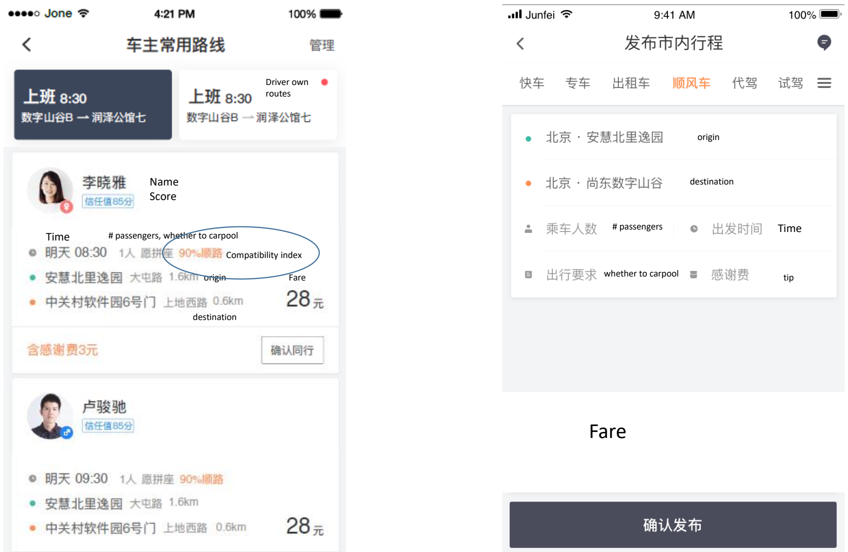
2.1 Institutional Details

The Hitch platform operates without a centralized dispatch system. On Hitch, a prospective passenger sends a ride request to the platform, where waiting drivers decide whether to answer it. The passenger request specifies the pickup location, drop-off location, and the preferred departure time. A prospective driver can view all requests via the platform's app, and the first driver who answers a request "wins" the trip.[6]

Figure 1 shows the primary interfaces.[7] The left panel is the driver view. The top highlighted portion contains the driver's departure time and route (consisting of the starting location and destination) entered by the driver. The app requires the driver to input at least the route information to start the search. If the departure time is not set, the system assumes that the driver wishes to "leave now" and sets the departure time to the search time. The passenger information viewed by the driver includes the passenger pseudonym, departure time, a compatibility index, and the pickup and drop-off locations. The compatibility index is approximately defined as 1 minus the detour distance divided by the total trip distance. Specifically, for a driver who intends to travel from A to B, and a passenger who wishes to travel from a to b, the total trip distance is defined as the distance of $A \to a \to b \to B$, and the detour distance is defined as the total trip distance minus the distance of $A \to B$. If a driver answers the request, this driver must travel to the pickup location by the passenger departure time, then deliver the passenger to the drop-off location. The driver can sort the list in several ways; the default is based on the compatibility index and departure time. The fare is calculated from the distance of a system-chosen route consistent with the locations in the passenger request, and it is fixed once the driver answers the request (therefore, there are no idling charges, tolls, or adjustments due to deviations from the pre-calculated route). Our data are from one city in China. During the time of the sample, the fare consisted of an initial charge and 1 RMB

[6]Our data are from 2018. As of March 2020, the Hitch platform has slightly modified the pickup process to allow passengers to choose whether to accept a driver's offer.

[7]Due to safety concerns, actual passenger photos are generic and do not display gender. The figures are demos provided by DiDi.



Passenger list for a searching driver

Launch page for a passenger

Figure 1: Driver and Passenger Interfaces

Yuan per km charge (we use $\mathbb{1}(\cdot)$ to denote indicator functions):

$$\text{fare} = 5 \text{ Yuan} + (\text{passenger distance} - 2 \text{ km}) \cdot \mathbb{1}(\text{passenger distance} > 2 \text{ km}).$$

The platform commissions were 10% of the fares. In comparison, Express (and taxis) in our sample city charged 3 Yuan for the first 3 km and 2 Yuan per additional km.

The right panel is the passenger view. The passenger fills out the pickup and drop-off locations next to the dots. The passenger also needs to indicate the number of people in the group, the departure time, whether the passenger is willing to carpool,[8] and whether the passenger is willing to offer a tip that will be added to the fare shown to the driver.[9] The bottom of the page displays the fare before the passenger confirms the request.

2.2 Data Summary

Our data cover a short time window for 20 weekdays in the summer of 2018 in a city in southern China. On the driver side, our data are left–truncated. We observe drivers who arrived between 4:40 p.m. and 5:00 p.m., including their driver IDs, routes, proposed

[8]In our data, we find carpooling to be rare: of the 6,813 answered passenger requests, two pairs of passengers were able to carpool. We focus on one-to-one matches throughout the empirical analysis.

[9]About 4% of passengers offered tips. We include the price variations due to the tips in our empirical analysis.

driver_id	driver_route	refresh_time_stamp	proposed_departure_time
1	$r_1 \in \mathbb{R}^4$	16:40:03,16:42:12;16:53:01	17:10
2	$r_2 \in \mathbb{R}^4$	16:40:11,16:55:11,16:59:05;17:11:01	17:25
$\vdots$			

passenger_id	passenger_route	create_time	proposed_departure_time	answer_time	cancel_time	driver_id	status
1	$r_1 \in \mathbb{R}^4$	16:22:02	17:15	16:40:02		5	Success
2	$r_2 \in \mathbb{R}^4$	16:30:11	17:00		16:40:05		Fail
$\vdots$							

Table 1: Data Sample

departure time, and the time stamps each time they refreshed the list of passengers from 4:40 p.m. to 5:30 p.m. On the passenger side, we observe passengers *present* between 4:40 p.m. and 5:00 p.m. (This includes passengers who arrived before 4:40 p.m. but the duration of their requests partially overlapped with the time window of 4:40 p.m. to 5:00 p.m.). We observe the passenger ID; time of arrival; time of proposed departure; time of cancellation;[10] and time of being answered (even when it was canceled or answered after 5:00 p.m., or it was answered by a driver who arrived before 4:40 p.m. or after 5:00 p.m.); the ID of the answering driver; and the passenger routes. We also observe the same passenger data on the Express platform. In Table 1, we give a (hypothetical) sample of our Hitch data. We use r to denote a "route," which consists of four real numbers that indicate the two sets of coordinates representing the starting location and the destination of a driver, or the pickup and drop-off locations of a passenger. For the driver, "refresh_time_stamp" shows the time stamps each time a driver viewed the list of passengers. On the passenger side, "create_time" means the time the request was created, "proposed_departure_time" indicates the time the passenger wanted to be picked up, and "answer_time" is the time a driver answered the request. The fields of "answer_time" and "driver_id" are filled only if a driver answered the corresponding passenger's request.

Table 2 presents summary statistics. The left panel is based on driver data. The average distance of the driver routes is 21 km. To examine the quality of the match, we link the driver and passenger data through driver IDs. The 2,170 common driver IDs account for 33% of the drivers and 18% of the passenger requests in our data. Based on these linked

[10]The passenger may cancel a failed search, or if a request is not answered by the specified departure time, the system cancels the request. In our data, the system canceled 0.8% of the failed searches. A passenger may send another request after cancellation; about 8% of all canceled searches were "reentered." When calculating the total waiting time of a passenger, we group the requests on the same day by the same passenger, and use the first arrival as the time of arrival and the last cancellation or the time of being answered as the time the search ends.

driver IDs, we find that the average detour distance is less than four km. We note that the share of 33% understates the driver match rate, because the drivers in our data could have answered a passenger who arrived after 5:00 p.m. For the waiting time, we focus on when a driver started and stopped her search on the platform (as opposed to the precise time the driver started the trip or picked up a passenger). For drivers who answered one of the requests in our passenger data, we calculate the average waiting time as the difference between the driver's first search (the first time point in "refresh_time_stamp") and the time the driver answered the request. The average is less than four minutes. The "Time in Market" of 403 seconds is the unconditional average waiting time across all drivers. For those who did not answer any request in the data, the average is calculated as the difference between the first and last time points in "refresh_time_stamp."

The right panel is based on the passenger data. The requested trips are slightly shorter than the drivers' routes. The match rate for passengers is 52% (including matches with drivers who arrived before 4:40 p.m. or after 5:00 p.m.). We further link the passenger IDs to transactions on the Express platform. We find that 9% of the passengers canceled their Hitch requests and completed the same trips by requesting rides from the Express platform. The Express platform can match a passenger with a driver much more quickly (usually within seconds). Similar to the driver side, we focus on how long passengers waited on the Hitch platform, and we calculate the "Time Till Being Answered" as the average difference between the time a request was created and when it was answered. The "Time in the Market" is the average difference between the time a request was created and the time it was either answered or canceled. We provide more details on our data in Appendix A.

	Driver			Passenger	
	Mean	Std		Mean	Std
Trip Length (km)	20.78	11.36	Trip Length (km)	18.67	11.66
Detour Length (km)	3.56	2.94			
			Match Rate	0.52	
			Canceled & Requested Express	0.09	
Time Till Answering (Sec)	224.96	276.58	Time Till Being Answered (Sec)	222.92	253.46
Time in the Market (Sec)	403.76	663.56	Time in the Market (Sec)	286.2	308.55
# New Requests/10 Sec	2.69	1.74	New Requests/10 Sec	3.64	2.55
Number of Observations	6,454			13,013	

Table 2: Summary Statistics

2.3 Driver Behavior

To understand how drivers search on the Hitch platform, we interviewed 12 DiDi drivers from November 2020 to December 2020.[11] Three main findings motivate our specification of a driver search model that features a finite horizon and heterogeneous preferences for passengers.

1. A majority of drivers are strict about departing before a self-imposed deadline. None of the 12 drivers we interviewed was a full-time driver on any DiDi platform, and all reported using Hitch for commuting purposes. Seven of the drivers said they would absolutely depart before their stated departure time, and five drivers said that on some days they could wait past the stated departure time for 10 to 30 minutes. The main reasons for setting a deadline included (1) the need to pick up a family member (e.g., spouse from work, children from school) and (2) the desire to leave their workplace by a certain time.[12]

2. Drivers sometimes adjusted their actual departure time to accommodate passengers. Five drivers mentioned negotiating with passengers to find a mutually agreeable pickup time instead of simply adhering to the pickup time on the passenger request.

3. The two most important factors to the drivers (11 out of 12) for whether to answer a request were (1) the compatibility of the routes and (2) whether the driver and the passenger could find a mutually agreeable departure time. When measuring the compatibility of the routes, some drivers emphasized using the compatibility index—which essentially takes into account both the detour distance and the fares (the distance of the passenger trip)—while others focused on the length of the detour. Other factors include the passenger's DiDi credit score[13] (4/12) and whether the passenger offered a tip (3/12).

Our data produce some evidence for the first finding. In the data, 39.5% of the drivers asked to "leave now" (i.e., the departure time is set to the time of the search), and 60.5% of the drivers specified a departure time later than the time of the search. We find that 90% of the drivers of the first type ended their search within 7.5 minutes, and 91% of the second type ended their search before the stated departure time.

[11] The interviews were conducted in Beijing by students who posted 4:00 p.m. requests on weekdays. Based on conversations with DiDi staff, we believe the driver profiles are comparable to those in our sample city.

[12] One driver stated, "I do not count on this (Hitch) to make money. I have a desk job and there is no point sitting around at the company after clocking out."

[13] This score is specific to the overall DiDi platform and reflects whether the passenger had been late for rides in the past.

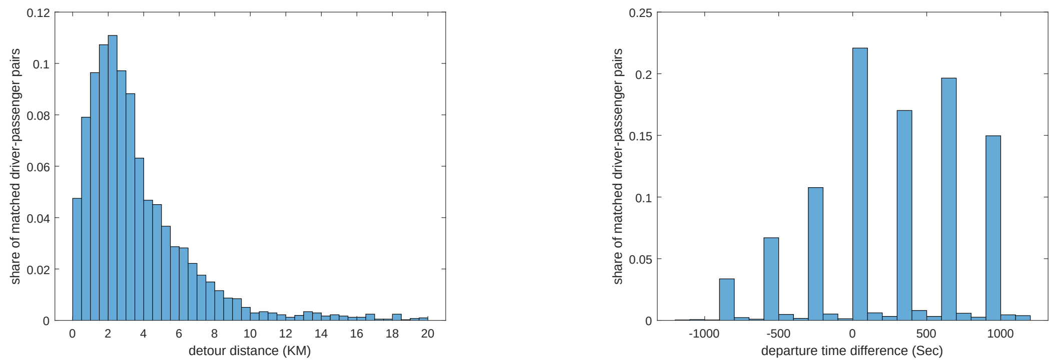


Figure 2: Distributions of Detour Length and Difference in Departure Time

Our data also yield evidence for the third finding. Specifically, we find that the length of the detour is a far better indicator for match quality than the difference in the stated passenger and driver departure time. On the left panel of Figure 2, we plot the distribution of the detour for the subsample of passengers answered by drivers in our driver data. On the right, we plot the distribution of the driver departure time minus the passengers' departure time for the same subsample. The latest passenger departure time is 5:30 p.m., and the latest driver departure time is 5:30 p.m. The distribution of the detour is heavily skewed towards 0, with a mean detour distance of 3.56 km, which is about 17% of the average driver route distance. However, the right panel shows a more spread-out distribution, where the absolute difference between the driver's and passenger's departure time is more than 600 seconds for half of this subsample. In comparison, the average driver waiting time for this subsample is 225 seconds.[14]

Overall, we think that it is best to view the stated departure time as an upper bound for how long drivers plan to *search on the platform*. Anecdotally, most drivers stop looking for passengers sometime before the stated departure time as drivers need time to prepare to leave work or walk to their cars. In the data, the average time between a driver's stated departure time and her last search is on average 384 seconds. This difference is too large to stem from the difference between the search time of the drivers who answered a request and the average search time (Table 2).

[14]The spikes of the distribution of the departure time difference are the result of departure time often being set at five-minute marks, such as 16:45, 16:50, etc.

3 Passenger

At a high level, the key passenger parameters are (1) the valuations for the Hitch rides, (2) the discount factor, and (3) how long a passenger plans to search. Our data detail how long each passenger's request lasted on the platform (which may be shorter than the planned search span due to drivers answering the requests), when and which passengers canceled requests, whether such passengers requested cars from Express after the cancellation, and fares for the rides on Hitch and Express. Our model shows that with a higher discount factor, more passengers are likely to stop waiting for a Hitch driver early in the search process, opting instead for Express. We then use the share of such passengers and the variations in Hitch fares relative to Express fares to estimate the discount factor and the distribution of the valuations. We use an inequality approach to set estimate the distribution of planned passenger search length.

3.1 Model

We use a finite-horizon dynamic model to describe passenger behavior. At time $\underline{t_j}$, a passenger j sends a request that specifies a route $r_j \in \mathbb{R}^4$ (consisting of coordinates of pickup and drop-off locations). The passenger plans to wait for a maximum of T_j seconds on the platform. The passenger is forward-looking and continuously discounts the future payoff at rate ϑ. We use ϑ to capture the cost of time while searching on the platform. We note that the time a driver picked up a passenger and started a trip is later than the time the request was answered. Our model does not assess how passengers valued time after a driver answered the request and before the actual pickup. Between $\underline{t_j}$ and $\underline{t_j} + T_j$, the passenger can cancel the Hitch request and make one on the Express platform, cancel the Hitch request and choose the outside transportation option (such as public transit), or wait for a Hitch driver to answer her request. If the passenger cancels the Hitch request or a Hitch driver answers the request, the passenger's search ends and she does not return to the platform. If a passenger does not cancel the request and is not answered by a Hitch driver before $\underline{t_j} + T_j$, the passenger waits till $\underline{t_j} + T_j$ and cancels the request. In either case, we use $\overline{t_j}$ to denote the time the search ends. In the above, the starting and ending time points of the search $\underline{t_j}$ and $\overline{t_j}$ are clock time (e.g., 4:30:00 p.m. and 4:45:00 p.m.), and T_j is a duration (e.g., 1200 seconds). The duration T_j represents how long the passenger plans to search, which we alternatively refer to as a passenger's "horizon." We also note that $\overline{t_j}$ is not the time a passenger physically departs for a trip but the time her search on Hitch ends—either when she cancels the request or when it is answered. We assume that passenger valuations of trips are heterogeneous across passengers but do not depend on driver identities, because passengers do not know ex ante who will answer the requests. We use w_j^H and w_j^E to denote the values of the trips through the Hitch and Express platforms. Among these variables, $\left(\underline{t_j}, \overline{t_j}, r_j\right)$ and the outcome of j's request are data, and $\left(\vartheta, w_j^H, w_j^E, T_j\right)$ are key model primitives that are not observed and need to be estimated from data.

We now use a set of assumptions to formalize the model above. Our first assumption is a normalization to ensure that sending a request on Hitch is not a dominated strategy for prospective passengers relative to their outside option whose value is normalized to 0.

Assumption 1. $w_j^H > 0$.

We next assume continuous discounting.

Assumption 2. *A passenger discounts future payoffs at rate* ϑ.

The third assumption states that passengers have multiple opportunities to decide whether to cancel the search. Because Assumptions 1 and 2 together imply that a passenger would not have any incentive to cancel the search by taking the outside option valued at 0 as long as the probability of receiving a ride on the Hitch platform is positive, we assume below that at each opportunity of move, the passenger either chooses to wait or to take Express. Consistent with our setting where passengers do not actively search for drivers or observe them, we assume that the option value of waiting is a function of time.

Assumption 3. *A passenger j on the platform can take actions at privately known random time points $\mathcal{T}_j = \left(t_j^1, t_j^2, \ldots, t_j^{N_j}\right)$, where $\underline{t_j} = t_j^1 < t_j^2 < \ldots < t_j^{N_j} = \underline{t_j} + T_j$. The passenger sends a request at t_j^1. At the time point t_j^k, $k = 1, \ldots, N_j$, the passenger forms the option value of waiting $v_j\left(t_j^k\right)$. If $w_j^E > v_j\left(t_j^k\right)$, the passenger j cancels the search in order to take Express at t_j^k. The passenger receives an instantaneous utility of w_j^E. The passenger's search ends at $\overline{t_j} = t_j^k$, and the passenger leaves the platform. Otherwise the passenger continues to wait. At a rate $h_j(t)$, a Hitch driver answers j's request at a time t, $t_j^k < t < t_j^{k+1}$. If this occurs, the passenger receives an instantaneous utility of w_j^H and leaves the platform at t. The search ends at $\overline{t_j} = t$ and the passenger leaves the platform. If the passenger does not cancel at t_j^k and is not answered between t_j^k and t_j^{k+1}, then the passenger reaches t_j^{k+1}, the passenger forms a new option value of waiting $v_j\left(t_j^{k+1}\right)$, and the same decision process repeats. If the passenger neither cancels the search at any t_j^k nor is answered by the time $t_j^{N_j}$, the passenger cancels the search at $\underline{t_j} + T_j$ and receives an instantaneous utility of $\max\left\{0, w_j^E\right\}$. The search ends at $\overline{t_j} = \underline{t_j} + T_j$, and the passenger leaves the platform.*

In the above, the time points in $\mathcal{T}_j$ are clock time, and the time points in $t_j^2, \ldots, t_j^{N_j}$ are not observed by the researchers. We will show that our characterization of passenger behavior is robust to the distribution of $\mathcal{T}_j$ so long as $\underline{t_j} = t_j^1$ and $t_j^{N_j} = \underline{t_j} + T_j$. As a result, our empirical approach does not rely on particular distributional assumptions on $\mathcal{T}_j$. Next, we make a simplifying assumption about a passenger's belief of $h_j(t)$, where we assume that passengers believe that the event of being answered occurs at rate $\varkappa$.

Assumption 4. *Passengers believe that $h_j(t) = \varkappa$.*

The assumption is equivalent to assuming that the probability of being answered by a Hitch driver between any t and t', $t < t'$, is $1 - e^{-\varkappa(t'-t)}$. We now define the optimal passenger strategy consistent with our assumption above on passenger belief.

Assumption 5. *(Definition of $v_j(t)$ by backward induction) The set of action time points $\mathcal{T}_j$ is known to the passenger j. The option value of waiting is:*

$$
\begin{aligned}
v_j\left(t_j^{N_j}\right) &= 0, \\
&\vdots \\
v_j\left(t_j^{k}\right) &= e^{-(\vartheta+\varkappa)\left(t_j^{k+1}-t_j^{k}\right)} \max\left\{w_j^E, v_j\left(t_j^{k+1}\right)\right\} \\
&\quad + \int_0^{t_j^{k+1}-t_j^{k}} e^{-\vartheta\tau} w_j^H \varkappa e^{-\varkappa\tau} d\tau, \\
&\vdots
\end{aligned}
$$

The following theorem characterizes passenger behavior.

Theorem 1. *Given Assumptions 1 to 5,*

1. *If $w_j^E < 0$, the passenger j does not choose Express; the passenger either is answered by a Hitch driver or chooses the outside option at $\underline{t_j} + T_j$.*
2. *If $0 < w_j^E < \dfrac{\varkappa}{\varkappa+\vartheta} w_j^H$, the passenger either is answered by a Hitch driver at a time between $\underline{t_j}$ and $\underline{t_j} + T_j$ or chooses Express at $\underline{t_j} + T_j$.*
3. *If $0 < \dfrac{\varkappa}{\varkappa+\vartheta} w_j^H < w_j^E$, the passenger chooses Express at $\underline{t_j}$.*

We provide the proof in Appendix B.1. The theorem states that if a passenger chooses Express, she does so either at $\underline{t_j}$ or $\underline{t_j} + T_j$, and a passenger is more likely to choose Express at $\underline{t_j}$ if ϑ is large. Therefore the share of passengers choosing Express at $\underline{t_j}$ helps inform the discount factor ϑ. We discuss how we define such passengers in data in the next section.

We next derive bounds for the distribution of T_j. Our approach requires one independence assumption and does not rely on the previous assumption on passenger belief

(Assumption 4) or that passengers are forward-looking (Assumption 5). We use F_T^{psg} to denote the distribution of the T_j and use $D_{\underline{t_j}}$ to denote the set of waiting drivers. The independence assumption we need is:

Assumption 6. *T_j and r_j are distributed i.i.d. across j and independent of $D_{\underline{t_j}}$.*

One justification for the assumption above is that, during our sample period, passengers could not directly view the set of waiting drivers. The next theorem bounds the distribution of T_j. We use "single" to mean that a passenger is not answered by a Hitch driver when she ends the search.

Theorem 2. *Given Assumptions 3 and 6, for any fixed duration of time $\tau > 0$, the following holds:*

$$\Pr\left(\text{single}, \overline{t_j} - \underline{t_j} < \tau \,\middle|\, r_j, D_{\underline{t_j}}\right) < F_T^{\text{psg}}(\tau) < \Pr\left(\overline{t_j} - \underline{t_j} < \tau \,\middle|\, r_j, D_{\underline{t_j}}\right) \tag{1}$$

Proof. For the left-hand side, we note that Bayes' rule and Assumption 3 imply:

$$\begin{aligned}
\Pr\left(\text{single}, \overline{t_j} - \underline{t_j} < \tau \,\middle|\, r_j, D_{\underline{t_j}}\right) &= \Pr\left(\overline{t_j} - \underline{t_j} < \tau \,\middle|\, \text{single, } r_j, D_{\underline{t_j}}\right) \cdot \Pr\left(\text{single} \,\middle|\, r_j, D_{\underline{t_j}}\right) \\
&= \Pr\left(T_j < \tau \,\middle|\, \text{single, } r_j, D_{\underline{t_j}}\right) \cdot \Pr\left(\text{single} \,\middle|\, r_j, D_{\underline{t_j}}\right) \\
&= \Pr\left(T_j < \tau \,\middle|\, r_j, D_{\underline{t_j}}\right) \cdot \Pr\left(\text{single} \,\middle|\, T_j < \tau, r_j, D_{\underline{t_j}}\right) \\
&\leq \Pr\left(T_j < \tau \,\middle|\, r_j, D_{\underline{t_j}}\right).
\end{aligned}$$

It follows from Assumption 6 that $\Pr\left(T_j < \tau \,\middle|\, r_j, D_{\underline{t_j}}\right) = F_T^{\text{psg}}(\tau)$.

For the right-hand side, given Assumption 3, if $T_j < \tau$, then a passenger's search must end within t seconds, and therefore Assumption 6 implies that:

$$F_T^{\text{psg}}(\tau) = \Pr\left(T_j < \tau \,\middle|\, r_j, D_{\underline{t_j}}\right) < \Pr\left(\overline{t_j} - \underline{t_j} < \tau \,\middle|\, r_j, D_{\underline{t_j}}\right).$$

□

3.2 Identification and Estimation of the Passenger Model

Theorem 2 provides the upper and lower bounds of F_T^{psg}, the distribution of planned passenger search length. In principle, it is straightforward to calculate the bounds of the CDF F_T^{psg}, based on the unconditional probabilities:

$$\Pr\left(\text{single}, \overline{t_j} - \underline{t_j} < \tau\right) < F_T^{\text{psg}}(\tau) < \Pr\left(\overline{t_j} - \underline{t_j} < \tau\right),$$

where both sides can be readily calculated from data. To obtain more precise estimates, we incorporate the conditioning variables in Theorem 2. The intuition is that the two-sided bounds are close when the probability of being answered is low, such as when there are few drivers waiting whose routes are similar to j's. In the extreme case, the distribution is identified point-wise when there are no waiting drivers.

Specifically, we model F_T^{psg} as an exponential distribution with mean $\frac{1}{\lambda^{\text{psg}}}$. We combine the following conditional moment inequalities (where "single" means the passenger leaves the platform without being answered) across multiple values of the duration τ (such as 5 minutes, 10 minutes, ...):

$$E\left[\mathbb{1}\left(\text{single}, \overline{t_j} - \underline{t_j} < \tau\right) - F_T^{\text{psg}}(\tau) \,\middle|\, r_j, D_{\underline{t_j}}\right] \leq 0$$
$$E\left[F_T^{\text{psg}}(\tau) - \mathbb{1}\left(\overline{t_j} - \underline{t_j} < \tau\right) \,\middle|\, r_j, D_{\underline{t_j}}\right] \leq 0.$$

Several difficulties arise when using these moment inequalities. The number of moments we use can be large because of the high dimensionality of the conditioning variables. In addition, moments based on multiple ts are correlated in different ways than the unconditional moments obtained by interacting the inequalities with positive functions of the conditioning variables. The estimator in Chernozhukov et al. (2019), which is based on the maximum of Studentized inequality-specific statistics, is well-suited to handle this case for both computational and statistical reasons. We provide the implementation details in Appendix C. We obtain a 95% confidence interval of $[0.0016, 0.0020]$ for λ^{psg}.

The passenger utility function is estimated separately and based on the behavioral characterization in Theorem 1, which does not involve F_T^{psg}. The key to estimating the discount factor ϑ is the share of passengers canceling the search on Hitch and requesting on Express shortly after arriving at the Hitch platform.[15] In the following, we define the indicator function $\mathbb{1}\left(j \text{ chooses Express at } \underline{t_j}\right) = 1$ on day d if j:

1. requested Express with route r_j on day d, (2) did not send a Hitch request with route r_j on day d, and (3) sent a Hitch request with route r_j on any day other than d; or

2. canceled a Hitch request with route r_j within 10 seconds, and (2) requested the same route r_j on Express afterwards on the same day.

Sixteen such trips in our data, or 0.12% of all requested passenger trips, satisfy either

[15] Passengers may also opt for taxis. We expect this case to be less common, given that the passengers who know how to use the DiDi app can arrange for an Express car to come to the pickup locations for the same fares as a taxi. Among all ride-sharing apps, DiDi commanded 91% of the market share in 2018. However, we recognize that the options of alternative transportation modes comparable to Express or Hitch may cause us to underestimate this share and thus underestimate the discount factor.

criterion. If the threshold in the second criterion is set to 20 seconds, there are 18 such trips, which lead to similar estimates. The small number of trips indicates a small discount rate ϑ.

We model the distribution of w_j^H as a truncated normal distribution $N(\mu, \sigma^2)$ on the support of $(0, +\infty)$, and we estimate different parameter μs for passengers who requested trips shorter or longer than 20 km. Because both Express and Hitch drivers use personal vehicles to transport passengers, we assume that the passenger experiences are similar with either service, and the value of taking Express is $w_j^H - \left(f_j^E - f_j^H\right)$. We also assume that the event of being answered is independent of w_j^H conditional on the trip distance. Based on Theorem 1, we estimate the parameters using GMM with the following moment conditions:

$$E\left(\left(\begin{array}{c} \mathbb{1}\left(j \text{ chooses Express at } \underline{t_j}\right) - \Pr\left(\frac{\varkappa}{\varkappa+\vartheta} w_j^H < w_j^H - \left(f_j^E - f_j^H\right)\right), \\ \mathbb{1}\left(j \text{ chooses Express after } \underline{t_j}\right) - \Pr\left(0 < w_j^H - \left(f_j^E - f_j^H\right) < \frac{\varkappa}{\varkappa+\vartheta} w_j^H\right) \cdot \left(1 - \mathbb{1}\left(j \text{ is answered on Hitch}\right)\right), \\ \mathbb{1}\left(j \text{ exits without a match}\right) - \Pr\left(w_j^H - \left(f_j^E - f_j^H\right) < 0\right) \cdot \left(1 - \mathbb{1}\left(j \text{ is answered on Hitch}\right)\right) \end{array}\right) \otimes \boldsymbol{d}_j\right) = 0, \tag{2}$$

where $\boldsymbol{d}_j$ is a vector of indicator variables for whether the length of the trip $|r_j| < 5, 6, \ldots, 40$ km, and "j exits without a match" means that the passenger was not answered by a Hitch driver and did not hail Express. In the above, as a behavioral assumption, we use $\varkappa = 0.0013$ for the hazard rate of being answered by a Hitch driver. The parameter is estimated as the hazard rate of being answered in a constant hazard model with random right censoring (due to passenger cancellation) via MLE (Klein and Moeschberger (2006)). This $\varkappa$ implies that passengers would perceive the probability of being answered within 10 minutes to be 54%.

Table 3 reports the results. The units are in Yuan for μ and σ, and the estimates indicate that the distribution of valuation is more right-skewed for passengers going on longer trips. We find that allowing for two types of μ significantly improves the fit over just one type, while the effect of adding more types is minimal. The model predicts 0.10% of the passengers requesting Express upon arrival (versus 0.12% in data) and 8.9% of the passengers canceling and requesting Express (which corresponds with the second moment in (2) with a sample average of 9%). To interpret the estimate of the discount factor, we note that from the simulation in Section 5, the 95% confidence interval of the average passenger surplus in a decentralized market is [3.77 Yuan, 4.22 Yuan], implying an average waiting cost of 0.042–0.047 Yuan per minute at the moment a passenger sends a request.[16]

[16] Buchholz et al. (2020) estimate the value of time to be about 1 Yuan per minute for taxi passengers in Prague. The city in our sample has a similar per capita income. We think the main discrepancy is due to selection: passengers (and drivers) who have low valuations for the trips and are not pressed for time use the Hitch platform, while others opt for a readily available service such as taxi or request cars from one of the centralized platforms.

Table 3: Estimates of Passenger Preferences

		Est	Se
Valuation of Hitch Rides (Yuan)	μ ($\lvert r_j \rvert < 20$ km)	-16.90	0.96
	μ ($\lvert r_j \rvert > 20$ km)	19.31	0.48
	σ	13.34	0.24
Discount Factor (10^{-4})	ϑ	1.86	0.08

4 Driver

We now model how a driver searches over a finite horizon. In the first subsection, we work through the main idea of our approach in a two-period model. In the second subsection, we formally state a set of behavioral assumptions that are consistent with an empirical model of dynamically optimizing agents. In particular, we do not assume how the option value of waiting is calculated. Next, we show that these assumptions lead to bounds on a driver's choice probability, which help to set identify the driver's preferences for passengers. We then discuss how additional assumptions help to estimate the discount factor. In the last subsection, we construct moment inequalities, discuss their roles in identification, and present the estimation results.

4.1 A Two-Period Model

To illustrate our approach, we first consider the estimation of a two-period dynamic discrete choice model. A driver i starts in period 1; she can choose a passenger from a set of waiting passengers S_1 and leave, or wait to enter the second period. If the driver chooses to wait and enter the second period, she faces a potentially different choice set S_2, and she can either choose a passenger and leave, or leave without a match. For simplicity, we assume that there are no other drivers. The instantaneous payoff for matching with a passenger j is $u_{ij}(x_{ij}, \varepsilon_{ij})$, where x_{ij} is a vector of covariates observable to the researchers, and ε_{ij} is an unobservable independent across j and independent of x_{ij}. Both x_{ij} and ε_{ij} are time invariant. The driver receives a utility of 0 if she chooses to leave without a match in the second period. Finally, the driver discounts the payoff in the second period by $\frac{1}{1+\delta}$, where $\delta > 0$.

A typical approach is to assume that a forward-looking driver is aware of the transition dynamics $\Pr(S_2|S_1)$ and calculate the option value of waiting in period 1 as

$$v(S_1) = \frac{1}{1+\delta} E_{S_2|S_1} \max\left\{0, \max_{j \in S_2} u_{ij}\right\},$$

and the probability of choosing j in period 1 as $\Pr\left(u_{ij} > \max\left\{v\left(S_1\right), \max_{\tilde{j}\in S_1\setminus j} u_{i\tilde{j}}\right\} | S_1\right)$. These probabilities can be taken to data either via a moment approach or MLE to estimate the preferences.

In our approach, we assume that data consist of the choices of many drivers where each driver faces large choice sets (S_1 and S_2 are large). Therefore, it may not be reasonable to assume that a driver has perfect knowledge of the conditional probability $\Pr\left(S_2|S_1\right)$. To address this concern, we construct inequalities to identify preferences without the assumption about how $v\left(S_1\right)$ is computed. Instead, we merely require that $v\left(S_1\right) \geq 0$, and we do not assume whether the driver knows the distribution $\Pr\left(S_2|S_1\right)$. Our inequalities focus on the probability that a driver chooses a particular type of passenger and the probability that a driver chooses any passenger. For example, if we assume that x_{ij} is a one -dimensional characteristic, we can define $J^\star = \{j : x_{ij} = x^\star\}$. We then consider the probability that a driver chooses a passenger in $J^\star$ in period 1 or period 2. For the upper bound, we note that a necessary condition for a passenger $j \in J^\star$ to be chosen (in either period) is that there exists a $j \in J^\star$ in either period 1 or 2 such that $u_{ij}\left(x_{ij}, \varepsilon_{ij}\right) > 0$, and we have:

$$\Pr\left(\text{driver } i \text{ chooses any passenger } j \in J^\star \mid S_1\right) \leq \Pr\left(\max_{j\in S_1\cup S_2\cap J^\star} u_{ij} > 0 | \, S_1\right), \tag{3}$$

where the max operator is 0 if the maximum is taken over an empty set. For the lower bound, we note that a sufficient condition for a $j \in J^\star$ to be chosen in either period is that such a j is chosen in the second period. Therefore:

$$\begin{aligned} &\Pr\left(\max\left\{0, \max_{j\in S_1\cap J^\star} u_{ij}\right\} > \max_{j'\in S_1\setminus J^\star} u_{ij'}, \max_{j\in S_2\cap J^\star} u_{ij} > \max\left\{0, \max_{j'\in S_2\setminus J^\star} u_{ij'}\right\} \mid S_1\right) \\ \leq\, &\Pr\left(\text{driver } i \text{ picks up any passenger } j \in J^\star \mid S_1\right) \end{aligned} \tag{4}$$

We can also construct bounds for the probability that a driver i chooses any passenger in either period:

$$\begin{aligned} &\Pr\left(\max_{j\in S_2} u_{ij} > 0 | \, S_1\right) \\ \leq\, &\Pr\left(\text{driver } i \text{ chooses any passenger} \mid S_1\right) \\ \leq\, &\Pr\left(\max_{j\in S_1\cup S_2} u_{ij} > 0 | \, S_1\right). \end{aligned} \tag{5}$$

These inequalities are useful for set identifying the payoff u_{ij} without involving the discount factor δ. Consider the following example: $u_{ij} = x_{ij} + \varepsilon_{ij}$, $S_1 = \{j_1, j_2\}$, and

$\Pr\left(S_2\,|S_1\right)$ is given by:

$$\begin{aligned}\Pr\left(S_2 = S_1\,|S_1\right) &= \frac{1}{2}\\ \Pr\left(S_2 = \{j_1\}\,|S_1\right) &= \frac{1}{6}\\ \Pr\left(S_2 = \{j_2\}\,|S_1\right) &= \frac{1}{6}\\ \Pr\left(S_2 = \emptyset\,|S_1\right) &= \frac{1}{6}.\end{aligned}$$

The objective is to set identify the distribution of ε_{ij}, F_ε. If we assume that x_{ij} has the support of $\mathbb{R}$, we can bound $F_\varepsilon\left(x^\star\right)$ for an arbitrary fixed number $x^\star$: conditioning on S_1 such that $x_{ij_1} = x_{ij_2} = -x^\star$, the lower bound of the inequality (5) becomes

$$\frac{1}{2}\left(1 - \left(F_\varepsilon\left(x^\star\right)\right)^2\right) + \frac{1}{3}\left(1 - F_\varepsilon\left(x^\star\right)\right)$$

and the upper bound becomes:

$$1 - \left(F_\varepsilon\left(x^\star\right)\right)^2$$

Therefore we obtain the following bounds for the CDF of ε_{ij}:

$$\begin{aligned}&\sqrt{\frac{16}{9} - 2\Pr\left(\text{driver } i \text{ chooses any passenger}\,|\,S_1\right)} - \frac{1}{3}\\ \leq& F_\varepsilon\left(x^\star\right)\\ \leq& \sqrt{1 - \Pr\left(\text{driver } i \text{ chooses any passenger}\,|\,S_1\right)},\end{aligned}$$

whenever the upper bound is defined for the reals.

A key feature of these inequalities is that they are agnostic about the information of when a choice is made, and they exploit the information on what that choice is to partially identify driver preferences. We contrast this result with Magnac and Thesmar (2002), who show that one is generally not able to separately identify the discount factor from payoffs in an infinite-horizon dynamic discrete-choice model. A discrete time version of such a model features an agent choosing from K alternatives over time, where the instantaneous utility of each alternative in period t takes on the form of $u_{kt} + \varepsilon_{kt}$. The unobservable ε_{kt} specific to time and each alternative is redrawn at the beginning of a period and typically assumed to be independent across alternatives and time conditional on some observables. Our setup differs in two aspects. First, we assume a finite horizon. Second, we also assume that the choice sets change across time, but the payoff (including the unobserved component) does not vary for the same choice. Then intuitively, we can construct inequalities that "collapse" the different choice sets across time. The usefulness of these inequalities depends on how fast the choice sets evolve over time and the length of the horizon. In the example above, the bounds become wider if the number of periods increases or the probability of the choice set staying the same in the next period ($\Pr(S_2 = S_1 | S_1)$) decreases.

For our purpose, it is also desirable to obtain an estimate of the discount factor δ to understand the cost of waiting, which necessarily requires specifying how drivers compute v. We bring back the assumption that:

$$v(S_1) = \frac{1}{1+\delta} E_{S_2|S_1} \max \left\{ 0, \max_{j \in S_2} u_{ij} \right\}.$$

Given an identified set of the preferences, the probability of whether i chooses any passenger in S_1, $\Pr(\max_{j \in S_1} u_{ij} > v(S_1) | S_1)$, pins down a set of δ.

Below, we generalize this two-period model to fit our empirical setting, but the intuition of identification is similar.

4.2 Model Assumptions

We start by specifying the preferences of drivers for passengers. We use r_i and r_j to denote the routes of driver i and passenger j. Driver i's utility for answering a passenger j's request is:

$$u_{ij} = u(r_i, r_j, \beta_i, \varepsilon_{ij}), \tag{6}$$

where β_i is a vector of driver-specific random coefficients, and ε_{ij} is a time-invariant, driver-passenger specific random variable. The random variables β_i and ε_{ij} are unobserved by the researchers. The random coefficient β_i could capture drivers' heterogeneous willingness to detour, and the variable ε_{ij} could capture—for example—local traffic conditions if i were to travel to j's location, whether i finds j's stated departure time agreeable, and its effect on i's utility. In our interviews, while some drivers pointed to passenger-specific characteristics in their decisions, the most important ones are specific to driver-passenger pairs (Section 2.3). In particular, given that we do not observe the stated departure time for about 40% of the drivers, and drivers and passengers sometimes negotiate to find a mutually agreeable departure time, we use ε_{ij} to account for this dimension of compatibility in addition to other match specific unobservables. We also do not expect traffic conditions to drastically change while drivers search, which lasts on average fewer than seven minutes in our data. We show in Appendix A that traffic conditions are stable during the time window we consider. Therefore we develop our approach under the assumption that the unobserved component of the utility function is match specific and time invariant.

Our model assumes that drivers search for passengers over a finite amount of time. Formally, each driver is associated with an exogenous vector of preference parameters β_i, how long the driver plans to search on the platform T_i, a time of arrival $\underline{t_i}$, and an endogenous time $\overline{t_i}$ when the driver ends the search. We alternatively refer to T_i as the driver's horizon. A driver ends her search either when she answers no passenger's request and leaves the platform at $\underline{t_i} + T_i$, or when she answers a request at a time between $\underline{t_i}$ and $\underline{t_i} + T_i$. As a reminder, $\overline{t_i}$ is the time a driver stops the search on the platform, and it is not the actual pickup time or a driver's departure time. Like in the passenger model, $\underline{t_i}$ and $\overline{t_i}$ are clock time, and T_i is a duration. We formalize the meanings of $\underline{t_i}, \overline{t_i}$, and T_i in subsequent assumptions. We use F_T^{drv} to denote the distribution of T_i.

Based on a set of behavioral assumptions given below, we use data on $\underline{t_i}$, r_i, F_T^{drv}, and driver outcomes to construct inequalities on choice probabilities that will help us to learn the distributions of $(\beta_i, \varepsilon_{ij})$. We discuss how we calibrate F_T^{drv} in Section 4.5. We use S_t to denote the set of passengers waiting at t, and use $\tilde{S}(t_1, t_2)$ to denote the set of new passengers arriving at the platform between time t_1 and t_2, inclusive of t_1 and t_2.

Assumption 7. *Attributes $\underline{t_i}, r_i, \beta_i, \varepsilon_{ij}$, and T_i are distributed i.i.d. across i and j.*

Assumption 8. *Attributes $\underline{t_i}, r_i, \beta_i$, and T_i are independent of the passenger identities in $S_{\underline{t_i}}$ and $\tilde{S}\left(\underline{t_i}, t\right)$ for any $t > \underline{t_i}$.*

Assumption 9. *The distribution of $\underline{t_i}$ and T_i are continuous, and T_i has a density function.*

Assumption 10. $S_t \subseteq S_{\underline{t_i}} \cup \tilde{S}\left(\underline{t_i}, t\right)$ *for any $t > \underline{t_i}$.*

The first assumption says that each dimension of a driver's attributes is independent of each other. We note that our approach relies on the independence of ε_{ij} across drivers (the i dimension), otherwise the distribution of ε_{ij}s of the waiting passengers would be an equilibrium outcome and differ from the distribution of ε_{ij}s of the new passengers. In Assumption 8, we assume that the identities of the waiting passengers $S_{\underline{t_i}}$ and the new arrivals $\tilde{S}\left(\underline{t_i}, t\right)$ are independent of the attributes of a driver who arrives at $\underline{t_i}$. Assumption 9 is a weak technical condition. Assumption 10 is a weak assumption on the evolution of the set of waiting passengers.

We next state assumptions on driver behaviors. Similar to the passenger model, we assume that a driver i makes choices at a set of N_i finite time points. We use $\mathcal{T}_i = \left\{t_i^1, \ldots, t_i^{N_i}\right\}$ to denote the clock time of these time points. The information set of a driver at t is denoted as pairs of a time point in $\mathcal{T}_i$ and the set of waiting passengers at the time for all such time points before t:

$$\mathcal{S}_{i,t} = \left\{ \left(t_i^k, S_{t_i^k} \right) \right\}_{t_i^k \in \mathcal{T}_i, t_i^k \leq t}$$

Assumption 11. *A driver i takes actions at random time points $\mathcal{T}_i = \left(t_i^1, t_i^2, \ldots, t_i^{N_i} \right)$, where $\underline{t_i} = t_i^1 < t_i^2 < \ldots < t_i^{N_i} = \underline{t_i} + T_i$. Conditional on N_i, $t_i^{k+1} - t_i^k$ has a continuous distribution with positive support. At the time point t_i^k, where $k = 1, \ldots, N_i$, the driver forms the option value of waiting $v_i \left(\mathcal{S}_{i,t_i^k} \right)$. If there exists a $j \in S_{t_i^k}$ such that:*

$$u_{ij} > \max \left\{ v_i \left(\mathcal{S}_{i,t_i^k} \right), \max_{j' \in S_{t_i^k} \setminus j} u_{ij'} \right\},$$

the driver answers j's request, receives an instantaneous utility of u_{ij} and ends the search at $\overline{t_i} = t_i^k$. The driver i then takes no further action. Otherwise the driver waits until t_i^{k+1} and the same decision process repeats. If a driver i chooses no passenger at any t_i^k, the driver ends the search at $\overline{t_i} = \underline{t_i} + T_i$, receives a utility of 0, and takes no further action.

Assumption 12. $v_i \left(\mathcal{S}_{i,t} \right) \geq 0$ *and* $v_i \left(\mathcal{S}_{i,t_i^{N_i}} \right) = 0$.

Assumption 11 states that a driver chooses the best passenger whose match utility is above the option value of waiting. Assuming discrete move opportunities at continuously distributed action time points helps to break ties. At any instant, multiple drivers may want to choose to answer one passenger request, but with Assumption 9, the probability of any two drivers taking action simultaneously is 0. The assumption also states that once a driver chooses a passenger or reaches the end of her horizon without a match, the driver leaves the platform and no longer makes any choice. Similar assumptions are used in continuous-time dynamic models (Doraszelski and Judd (2012); Arcidiacono et al. (2016)). Like passenger models, we do not assume the distribution of the time points in $\mathcal{T}_i$ other than $\underline{t_i} = t_i^1$ and $t_i^{N_i} = \underline{t_i} + T_i$. Assumption 12 states that the option value is nonnegative until the search ends, at which point the option value of waiting is 0 and the driver solves a static choice problem before permanently leaving the platform.

The driver choice rules assumed above are consistent with dynamic models with rational expectations and finite horizons, but we do not specify assumptions on how drivers compute the option value of waiting before $\underline{t_i} + T_i$, or how drivers form expectations about future passengers. We also do not assume the distribution of the action time points in $\mathcal{T}_i$ (except for continuity) or its correlation with driver preferences. In the next section, we derive bounds based on these assumptions given above, and these bounds will hold with heterogeneous driver strategies that flexibly depend on the information sets.

4.3 Bounds on Choice Probabilities

Our main result is a set of inequalities bounding the probability that a driver i chooses a type of passenger or any passenger by a time t. We define a passenger to be of type $J^\star$ if j is an element of a nonrandom set $J^\star$. In the empirical implementation, an example of $J^\star$ could be the set of passengers such that the fare is below a fixed number $f^\star$. Our first inequality bounds the probability of choosing a particular type of passengers, corresponding to the bounds in (3) and (4) in the two-period model in Section 4.1. The probability of a driver i choosing a $j \in J^\star$ is bounded from above by the probability that there exists a passenger j from the intersection of $J^\star$ and the superset that contains the passengers waiting at $\underline{t_i}$ as well as the new passengers arriving between $\underline{t_i}$ and t such that $u_{ij} > 0$. For the lower bound, the choice probability is bounded by the probability that the driver i does not choose a $j \notin J^\star$ before $\underline{t_i} + T_i$, and a $j \in J^\star \cap S_{\underline{t_i}+T_i}$ at $\underline{t_i} + T_i$ generates a nonnegative driver utility higher than other alternatives. We note that we can directly apply the inequalities to our right-truncated data by focusing on bounds of the probability of making a choice before t. Our second set of bounds on whether a driver chooses any passenger corresponds to the inequalities in (5) in the two-period model.

To state the theorem, we use $J^\star$ to denote a nonrandom set of passengers. We use $\mathcal{R}(t_1, t_2)$ to denote the time points at which any passenger arrives at the platform or ends her search (either because her request is answered or she leaves without a match) between the time points (clock time) t_1 and t_2, inclusive of t_1 but not t_2. In other words, $\mathcal{R}(t_1, t_2)$ represents the set of instances at which the set of waiting passengers changes. $\mathcal{R}(t_1, t_2)$ is a set of random time points that depends on the set of passengers waiting at t_1 and the arrival and exit processes of the passengers. We provide the proof in Appendix B.2.

Theorem 3. *For fixed $t > \underline{t_i}$, Assumptions 7 to 12 imply that*

$$\begin{aligned}
&\int_0^{t-\underline{t_i}} E\left[\prod_{\tilde{t}\in\mathcal{R}\left(\underline{t_i},\underline{t_i}+\tau\right)} \mathbb{1}\left(\max\left\{0, \max_{j\in S_{\tilde{t}}\cap J^\star} u_{ij}\right\} > \max_{j\in S_{\tilde{t}}\setminus J^\star} u_{ij'}\right)\right.\\
&\left.\cdot\mathbb{1}\left(\max_{j\in S_{\underline{t_i}+\tau}\cap J^\star} u_{ij} > \max\left\{0, \max_{j'\in S_{\underline{t_i}+\tau}\setminus J^\star} u_{ij'}\right\}\right) \middle| r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right] dF_T^{drv}(\tau)\\
\le &\Pr\left(\text{driver } i \text{ chooses any passenger } j\in J^\star \text{ by } t \middle| r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right)\\
\le &\Pr\left(\max_{j\in S_{\underline{t_i}}\cup\tilde{S}\left(\underline{t_i},t\right)\cap J^\star} u_{ij} > 0 \middle| r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right),
\end{aligned} \tag{7}$$

and

$$\begin{aligned}
&\int_0^{t-\underline{t_i}} \Pr\left(\max_{j\in S_{\underline{t_i}+\tau}} u_{ij} > 0 \;\middle|\; r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right) dF_T^{drv}(\tau) \\
\leq &\Pr\left(\textit{driver } i \textit{ chooses any passenger by } t \;\middle|\; r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right) \\
\leq &\Pr\left(\max_{j\in S_{\underline{t_i}}\cup\tilde{S}\left(\underline{t_i},t\right)} u_{ij} > 0 \;\middle|\; r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right). \qquad (8)
\end{aligned}$$

By varying $J^\star$ and with variations in $S_{\underline{t_i}}$, we can bound the probability of a driver choosing a passenger of certain characteristics. This allows us to use the responses of choices to variations in choice sets to identify driver preferences.

4.4 Strategic Delays and the Discount Factor

To estimate the discount factor, we exploit the possible presence of strategic delays: a forward-looking driver i may strategically forgo a passenger j and continue to wait, even if $u_{ij} > 0$. We do so by specifying how drivers compute $v_i\left(\mathcal{S}_{i,t_i^1}\right)$, which helps us to compute the probability that a driver i chooses any passenger at $\underline{t_i}$.

We start with assumptions on how drivers form the option value of waiting. Given that drivers can potentially view the set of all waiting passengers and drivers, the state variables are complex and high-dimensional. We use a dimension-reduction assumption inspired by the "inclusive value" approach in the dynamic demand literature (Melnikov (2013); Gowrisankaran and Rysman (2012)). Specifically, we assume that the information set of a driver consists of the utilities from the best and second-best matches.

Assumption 13. $v_i\left(\mathcal{S}_{i,t}\right) = v_i\left(\varsigma_i\left(\mathcal{S}_{i,t}\right)\right),$ *where*

$$\varsigma_i\left(\mathcal{S}_{i,t}\right) = \left\{t, u_{ij_1}, u_{ij_2} : \{j_1, j_2\} \subset S_t, u_{ij_1} \geq u_{ij}, u_{ij_2} \geq u_{ij}, \forall j \in S_t, j \neq j_1, j_2\right\}.$$

We next model how drivers perceive the evolution of S_t. This step requires assumptions about how drivers perceive the arrival and exit of the passengers. We assume that drivers use constant rates to approximate these processes. For the rate of arrival, we simulate a large set J of $|J|$ potential passenger routes from the empirical distribution of the passenger routes. Given the rate of passenger arrival $\gamma^{\text{psg}} = 0.36$ (Section 2.2), the rate of arrival of a passenger with the route r_j is $\gamma^{\text{psg}}/|J|$. We also assume that drivers believe any existing passenger request to disappear at the rate of $\varkappa^{\text{psg}}$, the inverse of the average duration of a passenger's request in the market ($\varkappa^{\text{psg}} = \dfrac{1}{286.2 \text{ sec}} = 0.0035$).[17]

[17]We note that this is a behavioral assumption: a request may disappear because another driver answers

Assumption 14. *Drivers believe that a new request with route r_j arrives at rate $\gamma^{psg}/|J|$, and that any existing request disappears at rate $\varkappa^{psg}$.*

Finally, we define the option value of waiting. We assume that the arrival of a passenger $j^\star$ allows the driver to update $\varsigma_{i,t}$ by choosing the best two from the new set $S_t \cup \{j^\star\}$. The driver treats the exits of j_1 or j_2 as a "partial stockout," updating ς_i to $\{t, -\infty, u_{ij_2}\}$ or $\{t, u_{ij_1}, -\infty\}$. We also assume that the action time points in $\mathcal{T}_i$ arrive at rate λ_i. A large choice of λ_i reduces the chance that a driver randomly misses opportunities to match with a passenger. In practice, we choose $\lambda_i = 0.5$, which implies that a driver checks once every two seconds on average, so that drivers are highly unlikely to randomly miss some passengers.. The change of value functions from a higher λ_i is negligible.

Given Assumption 13, we use (u_1, u_2, t) to represent the state variables of a driver, and $\varDelta$ is an infinitesimal amount of time. The assumptions above lead to the following value functions: for any $\underline{t_i} \le t < \underline{t_i} + T_i$,

$$\begin{aligned}
v_i\left(u_1, u_2, t\right) = \frac{1}{1+\delta\varDelta} \Bigg[& \left(1 - 2\varDelta\varkappa^{\text{psg}} - \varDelta\gamma^{\text{psg}} - \varDelta\lambda_i\right) v_i\left(u_1, u_2, t+\varDelta\right) \\
& + \varDelta\lambda_i \max\left\{u_1, u_2, v_i\left(u_1, u_2, t+\varDelta\right)\right\} \\
& + \varDelta\varkappa^{\text{psg}} \left(\max\left\{u_1, v_i\left(u_1, -\infty, t+\varDelta\right)\right\} + \max\left\{u_2, v_i\left(-\infty, u_2, t+\varDelta\right)\right\}\right) \\
& + \frac{\varDelta\gamma^{\text{psg}}}{|J|} \sum_{j=1}^{|J|} \max\left\{u_{ij}, u_1, u_2, v_i\left(\max\left\{u_{ij}, u_1, u_2\right\}, \max\left\{\left\{u_{ij}, u_1, u_2\right\} \setminus \max\left\{u_{ij}, u_1, u_2\right\}\right\}, t+\varDelta\right)\right\} \Bigg],
\end{aligned}$$

with the boundary condition

$$v_i\left(u_1, u_2, \underline{t_i} + T_i\right) = 0.$$

The probability of a driver choosing a passenger when the driver first arrives to the market is

$$\Pr\left(\max_{j \in S_{\underline{t_i}}} u_{ij} > v_i\left(\varsigma_i\left(\mathcal{S}_{i,\underline{t_i}}\right)\right) \middle| r_i, S_{i,\underline{t_i}}\right).$$

With identified preference parameters and a distribution F_T^{drv}, a higher δ increases this probability, and we can compare the model-predicted choice probability with the data counterpart to learn δ. In principle, we can construct the probability of waiting at $t_i^2, \ldots$, but the calculation is more involved given that the actual time of choice is unobserved.

it, and therefore the actual rate of exit is a complicated function that in part depends on the set of waiting passengers. Our assumption should be interpreted as drivers using the average rate of exits in computing the expected value of waiting.

4.5 Identification and Estimation of Driver Preferences

Empirical Model and Inequality Moments We start by specifying the utility of driver i for matching with passenger j. We use d_{ij} to denote the detour distance if i answers j, f_j to denote the fare of the request, and $|r_i|$ to denote the distance of the driver's route. The utility function is then given by

$$u_{ij} = u_0 - (\beta_d + \sigma_d \nu_i)\, d_{ij} + f_j + \beta_r\, |r_i| + \sigma_\varepsilon \varepsilon_{ij}. \tag{9}$$

The parsimonious specification allows for driver-specific heterogeneous willingness to pay for the detour $(\beta_d + \sigma_d \nu_i)$. We also allow the overall willingness to answer a request to depend on driver distance $|r_i|$. Both ν_i and ε_{ij} are distributed i.i.d. standard normal. As discussed in Section 2.3, we do not observe the stated departure time for about 40% of the drivers, and drivers and passengers may be flexible about adjusting their actual departure time. We therefore subsume this dimension of compatibility (and other unobserved factors such as the local traffic condition) in the error term ε_{ij}.

In the following, we define our moment inequalities based on Theorem 3. We first define the constraints on covariates $J^\star$ to be one of the two forms:

$$J^\star = \{j : d_{ij} < d^\star\} \text{ or } J^\star = \{j : f_j < f^\star\}$$

to bound the probabilities of choosing passengers such that the detour or the fare is below a certain level. The conditional inequality moments are

$$\begin{aligned} E\Bigg[\int_0^{t-\underline{t_i}} E\Bigg[\prod_{\tilde{t}\in\mathcal{R}\left(\underline{t_i},\underline{t_i}+\tau\right)} \mathbb{1}\left(\max\left\{0, \max_{j\in S_{\tilde{t}}\cap J^\star} u_{ij}\right\} > \max_{j\in S_{\tilde{t}}\setminus J^\star} u_{ij'}\right) \\ \cdot \mathbb{1}\left(\max_{j\in S_{\underline{t_i}+\tau}\cap J^\star} u_{ij} > \max\left\{0, \max_{j'\in S_{\underline{t_i}+\tau}\setminus J^\star} u_{ij'}\right\}\right) \,\Big|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\Bigg] dF_T^{\mathrm{drv}}(\tau) \\ - \mathbb{1}\left(\text{driver } i \text{ picks up any passenger } j \text{ by } t, j \in J^\star\right) \Big| r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\Bigg] \le 0, \end{aligned} \tag{10}$$

$$\begin{aligned} E\Big[\mathbb{1}\left(\text{driver } i \text{ picks up any passenger } j \text{ by } t, j \in J^\star\right) \\ - \Pr\left(\max_{j\in S_{\underline{t_i}}\cup\tilde{S}\left(\underline{t_i},t\right)\cap J^\star} u_{ij} > 0 \,\Big|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right) \Big| r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\Big] \le 0. \end{aligned} \tag{11}$$

$$E\left[\int_0^{t-\underline{t_i}} \Pr\left(\max_{j\in S_{\underline{t_i}+\tau}} u_{ij} > 0 \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right) dF_T^{\mathrm{drv}}(\tau)\right.$$
$$\left. -\mathbb{1}\left(\text{driver } i \text{ picks up any passenger by } t\right) \middle| r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right] \leq 0, \tag{12}$$
$$E\left[\mathbb{1}\left(\text{driver } i \text{ picks up any passenger by } t\right)\right.$$
$$\left. - \Pr\left(\max_{j\in S_{\underline{t_i}}\cup\tilde{S}\left(\underline{t_i},t\right)} u_{ij} > 0 \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right) \middle| r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right] \leq 0. \tag{13}$$

As discussed in Section 2.2, we observe whether drivers answer a passenger from those present between 4:40 p.m. and 5:00 p.m., but we do not observe in our data whether a driver answered a request sent after 5:00 p.m. Our moments handle the truncation well: we construct bounds for the probability of a driver answering a passenger before $t =$4:42 p.m., 4:44 p.m., ..., 5:00 p.m., which correspond with the collection of the inequalities (10) through (13) across those ts.

The identification argument of the preference parameters is similar to those in Section 4.1. As the two-period model demonstrates, inequalities (10) through (13) alone are capable of set identifying the preference parameters. The intuition for identifying the parameters in the utility function is thus similar to the idea of special regressors in static choice models, where the exogenous variations of d_{ij} and f_j in the choice sets $S_{\tau+\underline{t_i}}$ and $S_{\underline{t_i}} \cup \tilde{S}\left(\underline{t_i}, t\right)$ help to pin down the distribution of driver preferences (Manski (1988); Lewbel (2019)). As a reminder, our model differs from the framework in Magnac and Thesmar (2002) in two aspects. First, we consider agents that optimize over a finite horizon. Second, we assume that unobservables are specific to each choice (passenger), but not to a choice-instance. The first assumption is motivated by the institutional setting discussed in Section 2.3, and the second assumption is motivated by the nature of the unobservables, which capture the local traffic conditions on the route to pick up a passenger, and whether a driver and a passenger can find a mutually agreeable pickup time. The two restrictions help us to learn useful information about preferences from the data on who matches with whom. Intuitively, in the extreme case where the planning horizon of a driver (T_i) is close to 0, the role of dynamics in driver strategies is limited. One can fully learn the distribution of preferences from the choices and choice-specific covariates by applying the identification argument in a static choice model. On the other hand, if the length of the horizon is ∞, the lower bound of the choice probability in inequality (10) would approach 0 for any finite t and not be useful for identification.

To implement the moments defined in (10), we calibrate the distribution F_T^{drv} based on the stated departure time of the drivers. In principle, this distribution can be estimated with bounds similar to Theorem 2 (and jointly estimated with the inequalities above), but our driver data are truncated and we do not observe whether a driver answered a passenger who arrived after 5:00 p.m., which means we do not have the equivalent lower bound (namely, the probability of a driver exiting the market without a passenger) in the theorem. We instead use the stated departure time to calibrate F_T^{drv}. As discussed in Section 2.3, we consider the stated departure time a proxy for the upper bound of when a driver plans to stop searching on the platform. Conditional on specifying a departure time (which accounts for 60.5% of drivers in data), the average time between arrival and the stated departure time is 28 minutes. Assuming that the horizon of the other 39.5% of drivers is 7.5 minutes (90% of these drivers searched for fewer than 7.5 minutes), we compute an upper bound for the average driver's horizon: $28\text{min} \cdot 0.605 + 7.5\text{min} \cdot 0.395 =$19.7 minutes (1,135 seconds). We thus model the driver's horizon as following an exponential distribution with a mean of $\frac{1}{\lambda^{\text{drv}}}$, $\lambda^{\text{drv}} = \frac{1}{1135 \text{ sec}} = 8.5 \times 10^{-4}$. This calibration may overstate how long on average a driver plans to search, and given the discussion above on identification, we likely would obtain a conservative estimate of the confidence set.

The estimation of the discount factor δ is based on:

$$E\left(\Pr\left(\max_{j\in S_{t_i^1}} u_{ij} > v_i\left(\mathcal{S}_{i,t_i^1}, t_i^1\right)\right) - \mathbb{1}\left(\text{driver } i \text{ answers any passenger at } t_i^1\right) \middle| r_i, S_{t_i^1}\right) = 0,$$

following the discussion in Section 4.4. To use this moment with the other inequality moments above in a unified inequality framework, we consider the implied inequality moments:

$$E\left(\Pr\left(\max_{j\in S_{t_i^1}} u_{ij} > v_i\left(\mathcal{S}_{i,t_i^1}, t_i^1\right)\right) - \mathbb{1}\left(\text{driver } i \text{ answers a passenger at } t_i^1\right) \middle| r_i, S_{t_i^1}\right) \geq 0 \tag{14}$$

$$E\left(\Pr\left(\max_{j\in S_{t_i^1}} u_{ij} > v_i\left(\mathcal{S}_{i,t_i^1}, t_i^1\right)\right) - \mathbb{1}\left(\text{driver } i \text{ answers a passenger at } t_i^1\right) \middle| r_i, S_{t_i^1}\right) \leq 0. \tag{15}$$

The size of the identified set of the preference parameters pinned down by inequalities (10) through (13) determines the set of identified δ through (14) and (15).

Finally, similar to how we estimate F_T^{psg}, our inference is based on Chernozhukov et al. (2019) to accommodate the large number of moments and allow for arbitrary correlations across the moments. We jointly estimate the preference parameters and the discount factor based on the inequalities in (10) through (15) across multiple ts and $J^{\star}$s. We discuss the empirical implementation in detail and the model fit in Appendix D. We simulate the market outcomes using the estimated parameters in Section 5.2. We show that the

Table 4: Driver Parameter Estimates

			95% CI
Mean Coefficients	Intercept (Yuan)	u_0	[12.36, 21.05]
	Detour (Yuan/km)	β_d	[3.28, 4.67]
	Driver Trip Length (Yuan/km)	β_r	[-0.81, -0.57]
Unobserved Heterogeneity	Detour (Yuan/km)	σ_d	[0.11, 0.28]
	Intercept (Yuan)	σ_ε	[2.39, 12.25]
Discount Factor (10^{-4})		δ	[1.93, 4.10]

simulated passenger match rate, the average length of detour conditional on the matches, and the average waiting time align well with their data counterparts in the summary statistics.

Results Table 4 reports the projected 95% confidence intervals of the parameters. For the parameters in the driver utility function (9), the units are in Yuan or Yuan per km. We find that 1 km of detour reduces the willingness to accept by 3.28–4.67 Yuan, and that the overall willingness to answer a request decreases in the length of a driver's trip. The unobserved heterogeneity for the disutility of detours is limited (σ_d), and the magnitude of the driver-passenger specific unobservables is larger (but the estimate is noisier). To understand the magnitude of these estimates, the average fare of answered requests is 22 Yuan, and the disutility from a 5 km detour would nearly cancel out the incentives of the fares for most drivers. Furthermore, the identification of these parameters is almost entirely through the moments that do not involve the discount factor δ and do not rely on the assumptions on how the option values of waiting are calculated. Preference parameters estimated with moments corresponding with (10) through (13) are nearly identical to the results in Table 4.

The estimates of the discount factor imply that the maximum per-minute cost of waiting is approximately 1.2%–2.4% of the option value of waiting. As we show in the next section, the 95% confidence interval of the average driver surplus is $[6.55 \text{ Yuan}, 10.33 \text{ Yuan}]$, which implies an average driver's waiting cost of 0.10–0.25 Yuan per minute when a driver first arrives at the market.

5 Welfare Evaluation

We first discuss a series of centralized matching mechanisms. The algorithms have different information requirements and prioritize different components of the welfare (such as the number of matches, the qualities of the match, and the costs of waiting). Next, in Section 4.4, we compare the simulated outcomes of a decentralized market where driver strategies are defined with the outcomes of the centralized algorithms. Finally, we reflect on the implementation of a centralized algorithm.

5.1 Centralized Matching Algorithms

We consider six types of centralized matching algorithms related to the proposals in Akbarpour et al. (2020). A platform can harness three sets of information to implement a centralized algorithm: (a) preferences of drivers and passengers currently in the market, (b) how long these drivers and passengers plan to search, and (c) the arrival time of future drivers and passengers (as well as their preferences and how long they will search). For (a), we observe the routes of drivers and passengers in the data, but we (or the platform) do not observe a specific driver's willingness to accept a ride, her tolerance of a detour, or a specific passenger's valuation of the trip. We also do not observe the match-specific shock ε_{ij}. Finally, market participants are unlikely to have high-quality predictions for the information in (c). We focus on what a platform can do if it can predict, infer, or solicit the information in (a), (b), and (c)—but not how it obtains such information. We discuss the latter in Section 5.3. We also note that the information in (a) and (b) is privately known by agents in the decentralized market but not used to coordinate the matching process. We organize our discussions in terms of which algorithm a platform may implement if it is given access to successively richer sets of information on drivers and passengers.

Greedy Algorithm The greedy algorithm uses just the preference information in (a) and works in the following way: when a driver arrives at the platform, the platform checks the set of waiting passengers to see if any match generates a positive driver surplus. If more than one such match exists, the platform breaks ties by choosing the passenger who maximizes the ex ante sum of driver and passenger surplus. If the arrival time of the driver i is t, the platform chooses passenger j maximizing $u_{ij} + e^{-\vartheta\left(t-\underline{t_j}\right)} w_j^H$ among the set of js such that the match is individually rational ($u_{ij} \geq 0$). If no such match exists, the platform lets the driver wait. When a passenger arrives, the platform makes a similar attempt to match the passenger to the set of waiting drivers.

We see the greedy algorithm as approximating the optimum algorithm in the class of all centralized matching algorithms that (1) maximize the match rate and (2) have information in (a) but not (b) or (c). Following the notation in Akbarpour et al. (2020), we use OPT to denote this optimum algorithm. Intuitively, OPT cannot identify leaving agents, and any agent who reaches the end of her horizon will leave without a match. Therefore, OPT has a strong incentive to maintain a small pool of waiting agents, which counteracts the incentive to maintain a thick market to improve the chance of matching a new agent. In a model where any two agents can be matched with the same probability, Theorem 4 of Akbarpour et al. (2020) formally shows that the match rates under OPT and the greedy algorithm are close.[18] Furthermore, when market participants incur waiting costs, the greedy algorithm improves welfare by economizing on the waiting time.

Patient Algorithm We next examine the performance of what Akbarpour et al. (2020) call a "patient algorithm." The algorithm uses the information in (a) and (b) as input. To be precise, we note that the additional information required by the version of the patient algorithm considered here is in fact a subset of (b). It consists of (1) the time when the next agent would leave the market without a match, and (2) the identity of this agent. The algorithm works in the following way: when a driver (passenger) arrives at the market, the platform does not take any action and lets the driver (passenger) wait. If any passenger (driver) is about to leave the market, the platform attempts to match the passenger (driver) with the set of waiting drivers (passengers). We use the same tie-breaking rule as in the greedy algorithm.

The patient algorithm can substantially improve the match rate through the increased market thickness, ensuring that all departing agents are given a chance to match before perishing. In fact, Theorem 1 of Akbarpour et al. (2020) shows that the performance of the patient algorithm is nearly indistinguishable from an "omniscient" planner that maximizes the match rate and is aware of the information in (a), (b), and (c), provided that the arrival rate is sufficiently high and the probability of an agent being able to find a match is not overly low.[19]

However, there are two reasons to think that the patient algorithm can be improved in

[18] The authors show that in a matching market where any two agents can be matched with the probability $\frac{d}{m}$ where m is the arrival rate, the additional gain in the match rate from an optimal (and potentially more complex) algorithm using the same amount of information as the greedy algorithm decreases at the rate of d^{-1} when the arrival rate $m \to +\infty$. While the formal theoretical result is obtained in a more abstract model, we have reason to believe that using the greedy algorithm to approximate the performance of OPT is sensible in our context. In particular, we experimented with an approach inspired by the batch auctions, making matches between the sets of waiting drivers and passengers every Δt seconds, and found the match rates and total surplus decreasing in Δt and achieving their maxima when the algorithm coincided with the greedy algorithm at Δt close to 0.

[19] Theorem 1 of Akbarpour et al. (2020) shows that the additional gains in the match rate from the

our context. First, the waiting costs (due to keeping agents in the market potentially longer than in the decentralized market or under the greedy algorithm) may offset the welfare gains. More importantly, the patient algorithm provides no guarantee that a potential high-value match will materialize, because match opportunities are triggered by any agent leaving the market. Therefore even if there exists a high-value ij match between a driver i and a passenger j who both currently wait in the market, a low-value match ij' could instead occur as long as the passenger j' departs before i and j, and the ij' match is a tie-breaking choice. The occurrence of the ij' match would preclude the ij match as the matching is 1 to 1.

As a result, below we consider two variants of the patient algorithm that may improve the patient algorithm. The first one, which we call a hybrid algorithm, aims to increase the match quality by reducing the waiting time of select agents. The second, which is called a "Departure-Aware Greedy" algorithm in Akbarpour et al. (2020), further exploits the information on how long an agent will wait; it may result in even higher match rates than the patient algorithm but also longer waiting time.

Patient Algorithm Variant 1: Hybrid Algorithm We propose a hybrid algorithm that blends the greedy and patient algorithms and prioritizes high-quality matches. The information required for this algorithm is the same as for the patient algorithm. Upon the arrival of a driver or a passenger, the hybrid algorithm first identifies the driver utility of each individually rational match. The algorithm then executes the match immediately if the driver utility u_{ij} in the match exceeds a certain "acceptance threshold." The acceptance threshold is a parameter of the algorithm, set by the platform and fixed across matches. We use the tie-breaking rule in the greedy algorithm for all matches where the driver utility exceeds the threshold. If there is no such match, the new arrival waits with others. Immediately before a driver or a passenger reaches the end of her horizon, the algorithm looks for matches for the leaving agent and uses the tie-breaking rule in the patient algorithm. Choosing a threshold of 0 leads to the greedy algorithm, whereas a choice of $+\infty$ leads to the patient algorithm. The hybrid algorithm reduces the likelihood of a potential high-value match being broken up by a third party, while preserving match opportunities for the leaving agents. It also economizes on the waiting time to allow high-quality matches to happen quickly. We use a threshold of 11.5 Yuan in the results we present in the next section.

omniscient algorithm decrease at the rate of $e^{-d/2}$ when the arrival rate $m \to +\infty$, where $\frac{d}{m}$ is the probability that any two agents can be matched (also see footnote 18).

Noisy Hybrid Algorithm In this case, we relax the assumption on the information needed to implement the hybrid algorithm. Specifically, we assume the platform's information on the horizon of agents is noisy: for 95% of all agents, the platform-predicted horizon is a uniform random drawn between 80% and 100% of the true horizon. For the other 5%, the predicted horizon is a uniform random drawn between 100% and 120% of the true horizon. The probability of being an under/overpredicted agent is i.i.d. across drivers and passengers. In the case where the predicted horizon is longer than the actual horizon (overprediction), if the platform is unable to match the driver (passenger) upon her arrival, the platform lets the driver (passenger) wait and tries to match her again if any new passenger (driver) arrives or an existing passenger (driver) is about to reach the end of her horizon. The agent leaves the market without a match when she reaches the end of her actual horizon. In the case where the predicted horizon is shorter than the actual horizon (underprediction), the process is similar until the agent reaches the platform's predicted horizon, and the platform then attempts to match the agent with her waiting counterparts. If the platform succeeds in matching the agent, the matched agents leave; otherwise the platform predicts the waiting time to be $+\infty$ and the process in the case of overprediction applies.

Patient Algorithm Variant 2: Departure-Aware Greedy (DAG) Algorithm This algorithm uses the information in (a) and (b) as input. The DAG algorithm we consider is identical to the patient algorithm with the exception of the tie-breaking rule. If there are multiple agents who are compatible with the leaving agent (the match generates a positive driver surplus), DAG chooses the one who is the first to leave. For example, if driver i is to stop searching on the platform, we choose passenger $\tilde{j} \in S_t$ such that:

$$\tilde{j} = \arg \min_{j \in S_t, u_{ij} > 0} \underline{t_j} + T_j$$

to match with i. The case with a leaving passenger is similar. A motivation of this algorithm is that, if any agent is equally likely to be matched within a unit of time, DAG prioritizes matching the agents most at risk of leaving without a match.

Omniscient Algorithm This algorithm assumes that the platform knows the information in (a), (b), and (c). The planner's objective is to maximize the sum of driver and passenger surpluses (discounted to the respective time of arrival). The algorithm represents the upper bound of the welfare given the preferences, arrival, and exit processes of all agents. Because information about future agents is available, the algorithm can quickly match agents who will not have a good match in the future. It only allows select agents to wait, trading off the cost of time and the arrival of a good match opportunity. We discuss the computation of the algorithm in Appendix E.

5.2 Results

We simulate the outcomes of the decentralized market and the centralized algorithms.[20] In each simulation, we assume that passengers and drivers arrive at constant rates as estimated in Table 2. We report the projected 95% confidence interval of each outcome in Table 5.[21] In the last two rows, we report two measures of market thickness, which are the average numbers of waiting passengers and drivers. We calculate these measures by randomly choosing 10,000 instances from the simulation and averaging the number of waiting drivers and passengers across those instances. We also report the projected 95% confidence interval of the differences of key welfare measures (driver surplus, passenger surplus, total surplus, and platform revenue) across these scenarios in Table 6, where cell (k, ℓ) reports the 95% confidence interval of column ℓ minus column k in Table 5.

Column (i) of Table 5 reports outcomes from the decentralized market. We first note that the fit is quite reasonable: the confidence intervals of the percentage of matched passengers (row (2)), the duration of a driver's or a passenger's time in the market (rows (7) and (8)), and the length of the detour conditional on matched agents (row (11)) all include the values observed in the data (52%, 403 seconds, 286 seconds, and 3.56 km).

The results under the greedy algorithm (Table 5, column (ii)) show a trade-off between the match rates and match quality: the passenger match rate (Table 5, row (2)) is statistically significantly higher than the observed match rate in the data, but the average match quality is lower, measured by the higher average detour length (Table 5, row (11)), leading to an overall decrease of driver utility and total surplus compared with the decentralized market (Table 6, Driver Surplus and Total Surplus, cell (i, iii)). The greedy algorithm increases the platform revenue by 3%–8% (Table 6, Platform Revenue, cell (i, iii)) and reduces market thickness (Table 5, rows (10) and (11)).

The patient algorithm (column (iii) in Table 5), on the other hand, produces a thicker market and further improves the match rates upon the greedy algorithm. However, there is no significant improvement in the match quality (detour) for the drivers. Overall, there

[20]To focus on the stationary states, we simulate the outcomes for 21,000 agents and report the results based on the 1,500th to 12,499th agent.

[21]Our confidence set consists of parameter vectors that are from a large grid and correspond with a test statistic below the critical value (see Appendices C and D). We use a stratified sampling procedure to include parameters on the boundary as well as those in the interior of the set. Our simulation is based on 40 parameter vectors.

Table 5: Simulated Market Outcomes

		(i) Decentralized	(ii) Greedy	(iii) Patient	(iv) Hybrid	(v) Noisy Hybrid	(vi) DAG	(vii) Omniscient
	Match Rate							
(1)	Driver	[0.72, 0.76]	[0.76, 0.80]	[0.80, 0.84]	[0.77, 0.81]	[0.75, 0.79]	[0.83, 0.87]	[0.80, 0.84]
(2)	Passenger	[0.52, 0.56]	[0.56, 0.58]	[0.58, 0.61]	[0.55, 0.59]	[0.54, 0.58]	[0.60, 0.62]	[0.58, 0.61]
	Surplus (Yuan)							
(3)	Driver*	[6.55, 10.33]	[5.73, 8.81]	[5.54, 9.07]	[6.71, 10.84]	[6.59, 10.73]	[4.47, 6.94]	[7.78, 12.94]
(4)	Passenger†	[3.77, 4.22]	[3.94, 4.45]	[3.84, 4.29]	[3.89, 4.35]	[3.83, 4.29]	[3.84, 4.29]	[4.88, 5.94]
(5)	Total‡	[10.76, 14.13]	[10.18, 12.93]	[9.83, 13.00]	[11.03, 14.79]	[10.89, 14.69]	[8.74, 10.87]	[13.28, 18.16]
(6)	Revenue (Yuan)†	[1.02, 1.15]	[1.09, 1.19]	[1.11, 1.23]	[1.08, 1.21]	[1.06, 1.18]	[1.14, 1.24]	[1.25, 1.46]
	Time in the Market (Sec)							
(7)	Driver	[389.68, 490.02]	[239.24, 306.28]	[486.02, 534.91]	[362.73, 428.52]	[359.76, 421.09]	[549.75, 604.87]	[393.32, 595.64]
(8)	Passenger	[242.65, 324.57]	[218.91, 293.19]	[423.57, 546.61]	[294.20, 428.47]	[287.05, 419.13]	[461.04, 599.45]	[265.60, 360.85]
	Number of Waiting Agents							
(9)	Drivers	[102.67, 128.21]	[62.29, 79.53]	[126.06, 140.05]	[94.56, 110.90]	[93.83, 110.34]	[142.43, 157.21]	[105.04, 159.41]
(10)	Passengers	[87.92, 117.30]	[79.34, 106.25]	[153.37, 197.80]	[106.42, 154.78]	[103.70, 151.72]	[167.02, 216.97]	[98.17, 135.67]
(11)	Detour (km)	[3.54, 4.09]	[4.24, 4.59]	[4.03, 4.47]	[3.72, 4.26]	[3.70, 4.27]	[4.47, 4.86]	[3.71, 4.14]

* : on a per–driver basis; †: on a per–passenger basis; ‡ : weighted average driver and passenger surplus by the respective entry rates.

Table 6: Welfare Changes Across Scenarios

Driver Surplus (Yuan)*	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
(i)	[-1.60, -0.76]	[-1.45, -0.88]	[0.13, 0.51]	[0.03, 0.40]	[-3.53, -2.09]	[1.17, 2.78]
(ii)		[-0.25, 0.59]	[0.98, 2.05]	[0.86, 1.92]	[-2.01, -1.26]	[1.93, 4.19]
(iii)			[1.09, 1.86]	[1.00, 1.66]	[-2.27, -1.07]	[2.05, 4.01]
(iv)				[-0.37, 0.00]	[-4.04, -2.24]	[0.83, 2.35]
(v)					[-3.93, -2.13]	[0.94, 2.40]
(vi)						[3.31, 6.08]
Passenger Surplus (Yuan)†						
(i)	[0.17, 0.38]	[-0.01, 0.20]	[0.06, 0.22]	[-0.01, 0.17]	[-0.05, 0.22]	[0.96, 1.74]
(ii)		[-0.22, -0.04]	[-0.19, 0.02]	[-0.21, -0.09]	[-0.27, -0.05]	[0.65, 1.50]
(iii)			[-0.02, 0.13]	[-0.07, 0.08]	[-0.08, 0.09]	[0.82, 1.67]
(iv)				[-0.11, 0.01]	[-0.18, 0.04]	[0.83, 1.62]
(v)					[-0.11, 0.09]	[0.85, 1.65]
(vi)						[0.80, 1.67]
Total Surplus (Yuan)‡						
(i)	[-1.36, -0.53]	[-1.41, -0.77]	[0.21, 0.67]	[0.04, 0.57]	[-3.42, -2.01]	[2.40, 4.20]
(ii)		[-0.39, 0.38]	[0.86, 1.91]	[0.71, 1.77]	[-2.22, -1.40]	[3.01, 5.44]
(iii)			[1.12, 1.95]	[0.98, 1.72]	[-2.30, -1.08]	[3.22, 5.30]
(iv)				[-0.43, -0.06]	[-4.09, -2.29]	[2.10, 3.65]
(v)					[-3.99, -2.15]	[2.25, 3.75]
(vi)						[4.41, 7.45]
Platform Revenue (Yuan)†						
(i)	[0.04, 0.09]	[0.08, 0.11]	[0.05, 0.08]	[0.03, 0.06]	[0.09, 0.14]	[0.20, 0.34]
(ii)		[0.02, 0.06]	[-0.02, 0.02]	[-0.03, -0.00]	[0.03, 0.07]	[0.12, 0.28]
(iii)			[-0.06, -0.01]	[-0.08, -0.04]	[0.01, 0.04]	[0.10, 0.26]
(iv)				[-0.04, -0.00]	[0.03, 0.07]	[0.13, 0.28]
(v)					[0.05, 0.09]	[0.15, 0.30]
(vi)						[0.07, 0.24]

* : on a per–driver basis; †: on a per–passenger basis; ‡ :weighted average driver and passenger surplus by the respective entry rates.

is no statistically significant improvement in the total surplus compared with the greedy algorithm, which the longer waiting time and the ensuing higher costs of time could explain: wen discount factors are set to 0, the driver, passenger, and total surpluses would be higher under the patient algorithm than under the greedy algorithm. We also note that, due to the poorer match qualities and longer waiting time, the algorithm does not improve efficiency over the decentralized market (in terms of total surplus), but it does increase the platform revenue (7%–12%).

For the hybrid algorithm (column (iv) in Table 5), we use a threshold of 11.5 Yuan; that is, an agent who just arrives at the market and whose match with a waiting agent produces a driver surplus u_{ij} greater than 11.5 Yuan, will be matched immediately. This feature reduces the welfare costs of the waiting time and improves the match quality by "locking in" high-quality matches as soon as they occur. Through experimentation, we find that at the threshold of 11.5 Yuan (and in its $1-$ Yuan neighborhood), the total surplus under the hybrid algorithm is higher than the decentralized market (Table 6, Total Surplus, cell (i, iv)). The hybrid algorithm achieves match rates and waiting time between the greedy and patient algorithms but higher-quality matches.

Next, column (v) of Table 5 shows that adding noise to the information on how long agents plan to search reduces the hybrid algorithm's match rates and surpluses, but it still delivers similar surpluses as the decentralized market while generating a higher platform revenue. We systematically investigate the two dimensions of the noise (the proportion with an underpredicted horizon and degree of under/overprediction) by simulating the market outcomes where the share of agents with an underpredicted horizon varies from 50% to 95% (so the share of overpredicted horizons varies from 50% to 5%), and the noise level varies from 5% to 50%.[22] Figure 3 reports the midpoints between the lower and upper bounds of the 95% confidence intervals of the passenger match rate and total surplus for each combination of the two noise parameters. On the left, the passenger match rate increases markedly from about 0.5 to 0.56 as the share of underpredicted agents increases from 0.5 to 0.95 for any noise level, but the reduction of the noise level from 50% to 5% increases the midpoints of the match rates by 0.004. The intuition for the result is as follows: under the hybrid algorithm, a waiting driver (passenger) may be matched either because a passenger (driver) is about to leave or because the waiting driver (passenger) herself will leave. In the latter case, the platform attempts to match the driver (passenger) with the set of all

[22]As in Section 5.1, if the share of underpredicted horizons is ρ, the share of overpredicted horizons is $1-\rho$; if the noise level is w, the platform prediction of the horizon of the underpredicted agents uniformly distributes between $[(1-w)\,T_i, T_i]$, where T_i is the true horizon, and the prediction of the overpredicted agents' horizons uniformly distributes between $[T_i, (1+w)\,T_i]$. The probability of being an under/overpredicted agent is i.i.d. for both drivers and passengers.

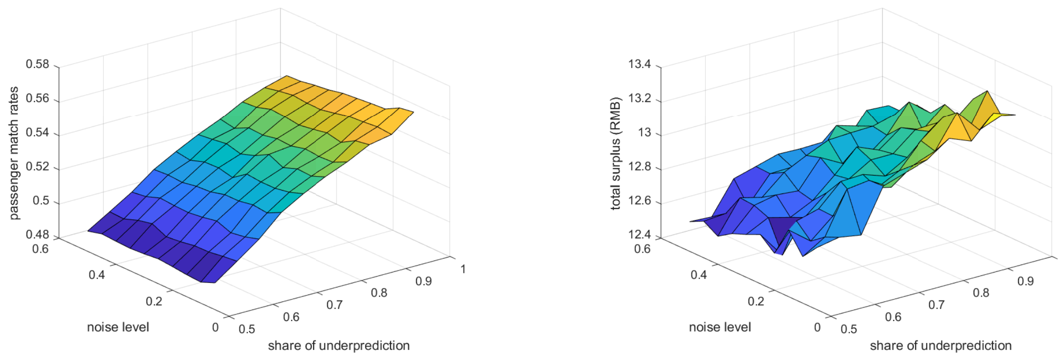


Figure 3: Noisy Hybrid: Passenger Match Rate and Total Surplus

waiting passengers (drivers) and stands a greater chance of success. Therefore, to maintain a high match rate, it is particularly important for every driver and passenger to get a chance to choose from all of their waiting counterparts. The right panel shows the change of total welfare. Although the noise level has little effect on the passenger match rate, the reduction of the noise level can increase the match quality, especially when the share of underprediction is high.

DAG (column (vi) of Table 5), another variant of the patient algorithm, achieves the highest match rates but the lowest surpluses among all scenarios due to the low qualities of the matches and long waiting time. On the other hand, the algorithm also achieves higher revenues than the aforementioned scenarios.

Finally, we compare the results above with the omniscient planner solution (column (vii)). The solution does not necessarily produce the highest match rate (it could be lower than DAG's for some parameters in the confidence set). Compared with the decentralized market, the solution produces statistically significant welfare improvement for both drivers and passengers, and 17%–30% increases in platform revenues.

5.3 Implementation

Firms today have access to consumer information from different sources and the ability to integrate such information. For example, panel data—which can be combined with field experiments and analyzed with machine learning tools—allow marketers to recover unobserved heterogeneity of consumers and significantly improve the efficacy of targeting (Rossi et al. (1996)) or even to use a personalized pricing policy (Dubé and Misra (2019)). The information required for the centralized algorithms is similar in scope; repeated observations of the same agent across days could potentially allow the platform to infer private information such as her preferences, arrival time, or the planned length of her search. We also show that even if the platform's prediction is noisy, the scope to raise revenues while maintaining the agent surplus at the level of the decentralized market may still be sizable. Furthermore, we note that a platform may be able to design the matching rule appropriately to solicit the information required for the matching algorithm from agents on the platform. For example, Theorem 5 of Akbarpour et al. (2020) suggests a "patient mechanism" to induce agents to inform the platform when they reach the end of their horizons. We examine the feasibility of implementing this mechanism in our context in Appendix F.

6 Conclusion

In this paper, we empirically quantify the gains from centralizing the matching process on a decentralized ride-sharing platform where drivers have heterogeneous preferences over passengers. The outcomes of the centralization depend on the objectives of the algorithms and the information available. We show that the welfare gap between the decentralized market and an omniscient planner's solution is substantial. To improve welfare over the decentralized market, we find that an algorithm needs to explicitly take into account match qualities and how long agents plan to search. Such an algorithm need not be omniscient or maximize the match rates, and it can improve the welfare of drivers and passengers even if the information on agents' horizons has a moderate amount of noise. In addition to understanding the merits of each specific *algorithm*, we view the simulations as also providing a range of expected outcomes given that the platform has access to a certain level of *information*, either by means of data-collecting technology or market design.

References

Adachi, Anna, "Competition in a Dynamic Auction Market: Identification, Structural Estimation, and Market Efficiency," *The Journal of Industrial Economics*, 2016, *64* (4), 621–655.

Agarwal, Nikhil and Paulo Somaini, "Empirical Models of Non-Transferable Utility Matching," 2020.

_ **, Itai Ashlagi, Michael A Rees, Paulo Somaini, and Daniel Waldinger**, "Equilibrium allocations under alternative waitlist designs: Evidence from deceased donor kidneys," *Econometrica*, 2021, *89* (1), 37–76.

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan, "Thickness and information in dynamic matching markets," *Journal of Political Economy*, 2020, *128* (3), 783–815.

Allen, Jason, Robert Clark, and Jean-François Houde, "The effect of mergers in search markets: Evidence from the Canadian mortgage industry," *The American Economic Review*, 2014, *104* (10), 3365–3396.

Arcidiacono, Peter, Patrick Bayer, Jason R Blevins, and Paul B Ellickson, "Estimation of dynamic discrete choice models in continuous time with an application to retail competition," *The Review of Economic Studies*, 2016, *83* (3), 889–931.

Ashlagi, Itai, Maximilien Burq, Patrick Jaillet, and Vahideh Manshadi, "On matching and thickness in heterogeneous dynamic markets," *Operations Research*, 2019, *67* (4), 927–949.

Baccara, Mariagiovanna, SangMok Lee, and Leeat Yariv, "Optimal dynamic matching," *Theoretical Economics*, 2020, *15* (3), 1221–1278.

Berry, Steven, Alon Eizenberg, and Joel Waldfogel, "Optimal product variety in radio markets," *The RAND Journal of Economics*, 2016, *47* (3), 463–497.

Bian, Bo, "Search Frictions, Network Effects and Spatial Competition: Taxis versus Uber," Technical Report, mimeo, Penn State University 2020.

Bimpikis, Kostas, Wedad J Elmaghraby, Ken Moon, and Wenchang Zhang, "Managing Market Thickness in Online Business-to-Business Markets," *Management Science*, 2020, *66* (12), 5783–5822.

Bodoh-Creed, Aaron L, Joern Boehnke, and Brent Hickman, "How efficient are decentralized auction platforms?," *The Review of Economic Studies*, 2021, *88* (1), 91–125.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou, "Geography, transportation, and endogenous trade costs," *Econometrica*, 2020, *88* (2), 657–691.

_ , _ , _ , and Nicola Rosaia, "Search Frictions and Efficiency in Decentralized Transportation Markets," Technical Report, National Bureau of Economic Research 2020.

Buchholz, Nicholas, "Spatial equilibrium, search frictions and efficient regulation in the taxi industry," Technical Report 2021.

— , **Laura Doval, Jakub Kastl, Filip Matějka, and Tobias Salz**, "The value of time: Evidence from auctioned cab rides," Technical Report, National Bureau of Economic Research 2020.

Canals, José J and Steven Stern, "Empirical search models," in "Search Theory and Unemployment," Springer, 2002, pp. 93–129.

Castillo, Juan Camilo, "Who Benefits from Surge Pricing?," *Available at SSRN 3245533*, 2020.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato, "Inference on causal and structural parameters using many moment inequalities," *The Review of Economic Studies*, 2019, *86* (5), 1867–1900.

Chiappori, Pierre-André and Bernard Salanié, "The econometrics of matching models," *Journal of Economic Literature*, 2016, *54* (3), 832–861.

Choo, Eugene, "Dynamic marriage matching: An empirical framework," *Econometrica*, 2015, *83* (4), 1373–1423.

— **and Shannon Seitz**, "The Collective Marriage Matching Model: Identification, Estimation, and Testing," in "Structural Econometric Models," Emerald Group Publishing Limited, 2013, pp. 291–336.

Ciliberto, Federico and Elie Tamer, "Market structure and multiple equilibria in airline markets," *Econometrica*, 2009, *77* (6), 1791–1828.

— , **Charles Murry, and Elie T Tamer**, "Market structure and competition in airline markets," *Available at SSRN 2777820*, 2020.

Cullen, Zoë and Chiara Farronato, "Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms," *Management Science*, 2020.

Doraszelski, Ulrich and Kenneth L Judd, "Avoiding the curse of dimensionality in dynamic stochastic games," *Quantitative Economics*, 2012, *3* (1), 53–93.

Dubé, Jean-Pierre and Sanjog Misra, "Personalized pricing and customer welfare," *Available at SSRN 2992257*, 2019.

Eckstein, Zvi and Gerard J Van den Berg, "Empirical labor search: A survey," *Journal of Econometrics*, 2007, *136* (2), 531–564.

Einav, Liran, Chiara Farronato, and Jonathan Levin, "Peer-to-peer markets," *Annual Review of Economics*, 2016, *8*, 615–635.

Eizenberg, Alon, "Upstream innovation and product variety in the us home pc market," *Review of Economic Studies*, 2014, *81* (3), 1003–1045.

Fan, Ying and Chenyu Yang, "Merger, product repositioning and firm entry: The retail craft beer market in california," *Available at SSRN 3681556*, 2020.

Farronato, Chiara and Andrey Fradkin, "The welfare effects of peer entry in the accommodation market: The case of airbnb," Technical Report, National Bureau of Economic Research 2018.

Fong, Jessica, "Search, Selectivity, and Market Thickness in Two-Sided Markets: Evidence from Online Dating," *Selectivity, and Market Thickness in Two-Sided Markets: Evidence from Online Dating (June 12, 2020)*, 2020.

Fox, Jeremy T, "An Empirical, Repeated Matching Game Applied to Market Thickness and Switching," 2008.

_ , "Specifying a Structural Matching Game of Trading Networks with Transferable Utility," *American Economic Review*, 2017, *107* (5), 256–260.

Frechette, Guillaume R, Alessandro Lizzeri, and Tobias Salz, "Frictions in a competitive, regulated market: Evidence from taxis," *American Economic Review*, 2019, *109* (8), 2954–92.

Gavazza, Alessandro, "The role of trading frictions in real asset markets," *The American Economic Review*, 2011, *101* (4), 1106–1143.

_ , "An empirical equilibrium model of a decentralized asset market," *Econometrica*, 2016, *84* (5), 1755–1798.

Gowrisankaran, Gautam and Marc Rysman, "Dynamics of consumer demand for new durable goods," *Journal of political Economy*, 2012, *120* (6), 1173–1219.

Haile, Philip A and Elie Tamer, "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 2003, *111* (1), 1–51.

Hendricks, Kenneth and Alan Sorensen, "The Value of Intermediaries in Dynamic Auction Markets," 2016.

Hitsch, Gunter J, Ali Hortaçsu, and Dan Ariely, "Matching and sorting in online dating," *American Economic Review*, 2010, *100* (1), 130–63.

Ho, Katherine, "Insurer-provider networks in the medical care market," *American Economic Review*, 2009, *99* (1), 393–430.

Holmes, Thomas J, "The diffusion of Wal-Mart and economies of density," *Econometrica*, 2011, *79* (1), 253–302.

Klein, John P and Melvin L Moeschberger, *Survival analysis: techniques for censored and truncated data*, Springer Science & Business Media, 2006.

Lagos, Ricardo, "An analysis of the market for taxicab rides in New York City," *International Economic Review*, 2003, *44* (2), 423–434.

Lewbel, Arthur, "The identification zoo: Meanings of identification in econometrics," *Journal of Economic Literature*, 2019, *57* (4), 835–903.

Li, Jun and Serguei Netessine, "Higher market thickness reduces matching rate in online platforms: Evidence from a quasiexperiment," *Management Science*, 2020, *66* (1), 271–289.

Loertscher, Simon, Ellen V Muir, and Peter G Taylor, "Optimal market thickness and clearing," *Unpublished paper, Department of Economics, University of Melbourne.[1224]*, 2018.

Magnac, Thierry and David Thesmar, "Identifying dynamic discrete decision processes," *Econometrica*, 2002, *70* (2), 801–816.

Manski, Charles F, "Identification of binary response models," *Journal of the American statistical Association*, 1988, *83* (403), 729–738.

Melnikov, Oleg, "Demand for differentiated durable products: The case of the us computer printer market," *Economic Inquiry*, 2013, *51* (2), 1277–1298.

Mortensen, Dale T, "Job search and labor market analysis," *Handbook of labor economics*, 1986, *2*, 849–919.

Pakes, Ariel, Jack Porter, Kate Ho, and Joy Ishii, "Moment inequalities and their application," *Econometrica*, 2015, *83* (1), 315–334.

Pissarides, Christopher A, *Equilibrium unemployment theory*, MIT press, 2000.

Reeling, Carson and Valentin Verdier, "Welfare effects of dynamic matching: An empirical analysis," 2020.

Rogerson, Richard, Robert Shimer, and Randall Wright, "Search-theoretic models of the labor market: A survey," *Journal of economic literature*, 2005, *43* (4), 959–988.

Rosaia, Nicola, "Competing Platforms and Transport Equilibrium: Evidence from New York City," Technical Report, mimeo, Harvard University 2020.

Rossi, Peter E, Robert E McCulloch, and Greg M Allenby, "The value of purchase history data in target marketing," *Marketing Science*, 1996, *15* (4), 321–340.

Roth, Alvin E. and Marilda A. Oliveira Sotomayor, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis* Econometric Society Monographs, Cambridge University Press, 1990.

Roth, Alvin E, Tayfun Sönmez, and M Utku Ünver, "Pairwise kidney exchange," *Journal of Economic theory*, 2005, *125* (2), 151–188.

_ , _ , **and** _ , "Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences," *American Economic Review*, 2007, *97* (3), 828–851.

Salz, Tobias, "Intermediation and competition in search markets: An empirical case study," Technical Report, National Bureau of Economic Research 2020.

Shapiro, Matthew H, "Density of Demand and the Benefit of Uber," 2018.

Shimer, Robert and Lones Smith, "Matching, search, and heterogeneity," *Advances in Macroeconomics*, 2001, *1* (1).

Ünver, M Utku, "Dynamic kidney exchange," *The Review of Economic Studies*, 2010, *77* (1), 372–414.

Waldinger, Daniel, "Targeting in-kind transfers through market design: A revealed preference analysis of public housing allocation," Technical Report, Working paper 2018.

Wollmann, Thomas G, "Trucks without bailouts: Equilibrium product characteristics for commercial vehicles," *American Economic Review*, 2018, *108* (6), 1364–1406.

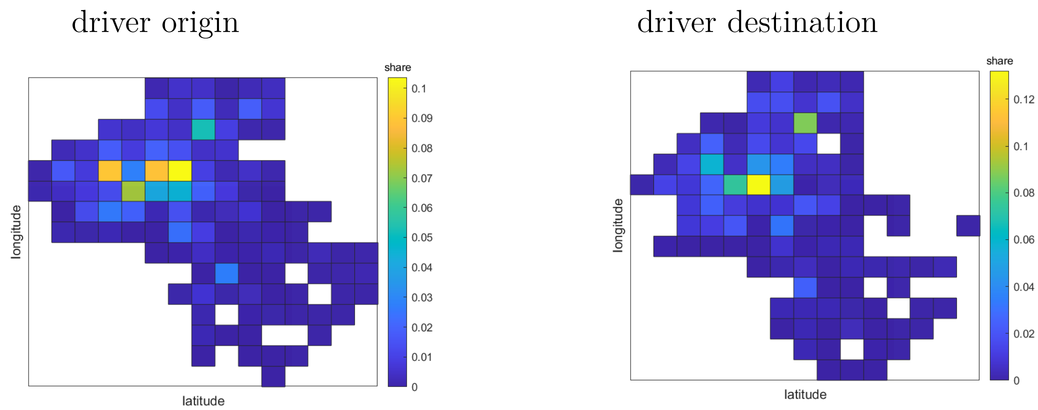


Figure 4: Driver

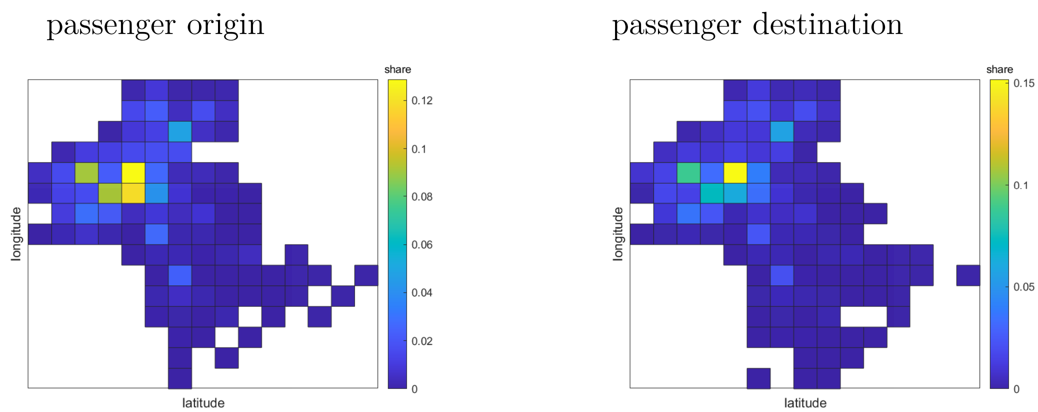


Figure 5: Passenger

A Additional Summary Statistics

In Figures 4 and 5, we plot the shares of trip origins and destinations by drivers and passengers on heat maps. The vertical bar indicates the share of the trips.

In Figure 6, we examine to what extent a passenger's compatibility score affects the actual choice of a driver. Conditional on linked driver–passenger pairs, we plot the maximum score (subject to the truncation) in the set of waiting passengers facing the driver at the moment of the choice and the score of the passenger actually chosen. A point on diagonal indicates that the driver chooses the passenger with the maximum score, which account for 69.5% of the matched driver–passenger pairs.

Lastly, we examine the traffic conditions in Figures 7 and 8 and show that that the traffic condition is quite stable for our drivers.. We select 42 routes most frequently requested by passengers in the data. We first examine the traffic condition across time after aggregating

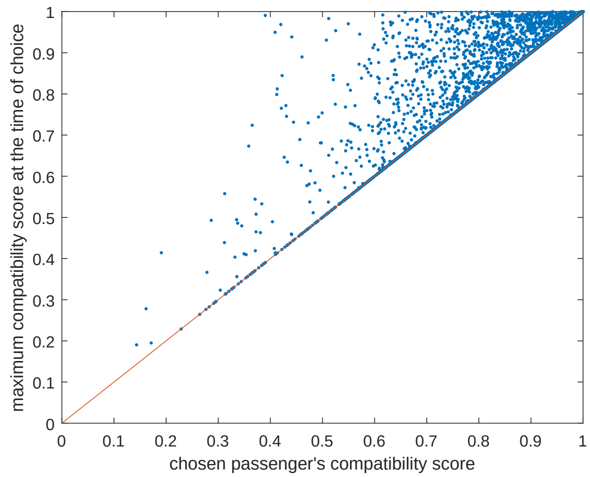


Figure 6: Chosen Passenger's Score and the Maximum Score

over routes: we plot the average speed in Figure 7 for traveling at 4:30 p.m., 4:34 p.m., ..., 6:30 p.m. across days. The speed slightly declines from 39KM/h to 36KM/h over the 2–hour period. We next examine the traffic condition across routes, after aggregating over time: in Figure 8, we plot the maximum and minimum travel time across 2 hours against the respective route distance.

B Proofs

B.1 Proof of Theorem 1

The first statement follows from Assumption 5, which states that the option value of waiting is non–negative. We prove the second statement by showing $v_j\left(t_j^k\right) > w_j^E$ for $k = 1, \ldots N_j - 1$ through induction. We first note that

$$v_j\left(t_j^k\right) = e^{-(\vartheta+\varkappa)\left(t_j^{k+1}-t_j^k\right)} \max\left\{w_j^E, v_j\left(t_j^{k+1}\right)\right\} + \frac{\varkappa}{\varkappa+\vartheta}\left(1 - e^{-(\vartheta+\varkappa)\left(t_j^{k+1}-t_j^k\right)}\right) w_j^H.$$

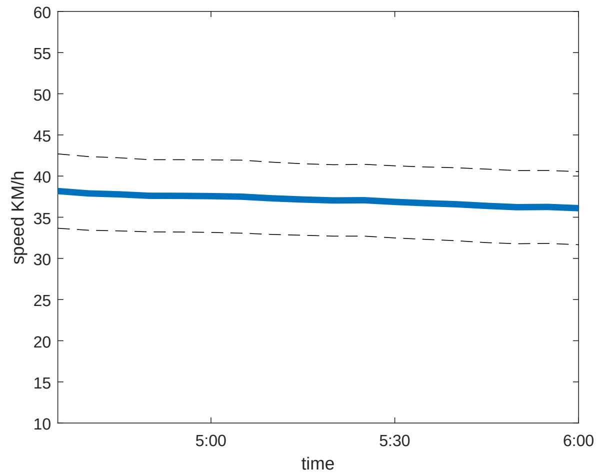


Figure 7: Average Speed, 4:30 p.m.-6:30 p.m.

The dotted lines represent the 95% confidence interval for the average speed across 42 routes.

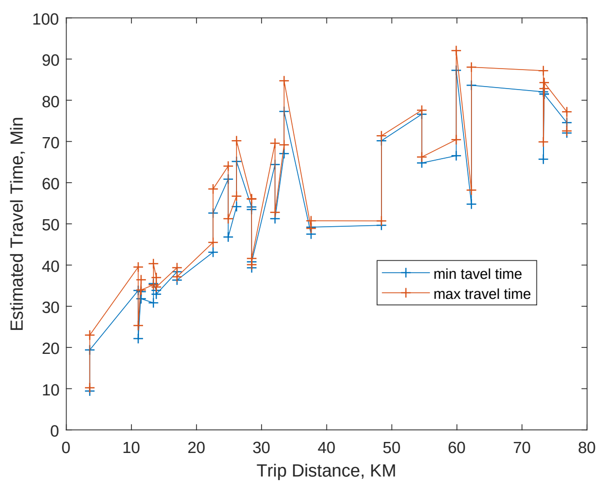


Figure 8: Travel Time on Popular Routes

The maximum time and minimum travel time on the 42 routes we select between 4:30 p.m. and 6:30 p.m. The min and max are taken within a route across the travel time measured at 4:30 p.m., 4:34 p.m., ..., 6:30 p.m.

Therefore if $0 < w_j^E < \frac{\varkappa}{\varkappa+\vartheta} w_j^H$, then

$$\begin{aligned} v_j\left(t_j^{N_j-1}\right) =& e^{-(\vartheta+\varkappa)\left(t_j^{N_j}-t_j^{N_j-1}\right)} w_j^E \\ &+\left(1-e^{-(\vartheta+\varkappa)\left(t_j^{N_j}-t_j^{N_j-1}\right)}\right)\frac{\varkappa}{\varkappa+\vartheta} w_j^H \\ >& e^{-(\vartheta+\varkappa)\left(t_j^{N_j}-t_j^{N_j-1}\right)} w_j^E + \left(1-e^{-(\vartheta+\varkappa)\left(t_j^{N_j}-t_j^{N_j-1}\right)}\right) w_j^E \\ =& w_j^E. \end{aligned}$$

Then for any $k+1 \leq N_j - 1$, if $v_j\left(t_j^{k+1}\right) > w_j^E$, then

$$\begin{aligned} v_j\left(t_j^{k}\right) =& e^{-(\vartheta+\varkappa)\left(t_j^{k+1}-t_j^{k}\right)} \max\left\{w_j^E, v_j\left(t_j^{k+1}\right)\right\} \\ &+\left(1-e^{-(\vartheta+\varkappa)\left(t_j^{k+1}-t_j^{k}\right)}\right)\frac{\varkappa}{\varkappa+\vartheta} w_j^H \\ >& e^{-(\vartheta+\varkappa)\left(t_j^{k+1}-t_j^{k}\right)} w_j^E + \left(1-e^{-(\vartheta+\varkappa)\left(t_j^{k+1}-t_j^{k}\right)}\right) w_j^E \\ =& w_j^E. \end{aligned}$$

To prove the third statement, we note that the same calculation shows that $v_j\left(t_j^k\right) < w_j^E$ for $k = 1, \ldots N_j - 1$, if $\frac{\varkappa}{\varkappa+\vartheta} w_j^H < w_j^E$.

B.2 Proof of Theorem 3

For the lower bound, by Assumption 11, for any duration $\tau > 0$, the condition

$$\prod_{\tilde{t}\in\mathcal{R}\left(\underline{t_i},\underline{t_i}+\tau\right)} \mathbb{1}\left(\max\left\{0, \max_{j\in S_{\tilde{t}}\cap J^\star} u_{ij}\right\} > \max_{j\in S_{\tilde{t}}\setminus J^\star} u_{ij'}\right) > 0$$

ensures that the driver i does not answer any passenger not in $J^\star$ at any $t_i^k < \underline{t_i} + \tau$. The condition

$$\max_{j\in S_{\underline{t_i}+\tau}\cap J^\star} u_{ij} > \max\left\{0, \max_{j'\in S_{\underline{t_i}+\tau}\setminus J^\star} u_{ij'}\right\}$$

ensures that the driver answers a passenger in $J^\star$ at $\underline{t_i} + \tau$ if her horizon is of length τ. Therefore combining the two, we have a lower bound of the probability of a driver answering a passenger $J^{\star}$ by $\underline{t_i}+\tau$ conditional on $T_i=\tau$:

$$p_{i,\underline{t_i},\tau}=E\left[\prod_{\tilde{t}\in\mathcal{R}\left(\underline{t_i},\underline{t_i}+\tau\right)}\mathbb{1}\left(\max\left\{0,\max_{j\in S_{\tilde{t}}\cap J^{\star}}u_{ij}\right\}>\max_{j\in S_{\tilde{t}}\setminus J^{\star}}u_{ij'}\right)\right.$$
$$\left.\cdot\mathbb{1}\left(\max_{j\in S_{\underline{t_i}+\tau}\cap J^{\star}}u_{ij}>\max\left\{0,\max_{j'\in S_{\underline{t_i}+\tau}\setminus J^{\star}}u_{ij'}\right\}\right)\,\middle|\,r_i,S_{\underline{t_i}},\tilde{S}\left(\underline{t_i},t\right),\underline{t_i},T_i=\tau\right].$$

Therefore for any time $t>\underline{t_i}$ (we use E_X to denote the expectation with respect to X),

$$\begin{aligned}
&\Pr\left(\text{driver } i \text{ picks up a passenger } j\in J^{\star} \text{ by } \underline{t_i}+T_i\,\middle|\,r_i,S_{\underline{t_i}},\tilde{S}\left(\underline{t_i},t\right),\underline{t_i}\right)\\
=&E_{T_i}\left(\text{driver } i \text{ picks up a passenger } j\in J^{\star} \text{ by } \underline{t_i}+T_i\,\middle|\,r_i,S_{\underline{t_i}},\tilde{S}\left(\underline{t_i},t\right),\underline{t_i},T_i\right)\\
=&E_{T_i}\left[\mathbb{1}\left(t<\underline{t_i}+T_i\right)\cdot\Pr\left(\text{driver } i \text{ picks up a passenger } j\in J^{\star} \text{ by } t\,\middle|\,r_i,S_{\underline{t_i}},\tilde{S}\left(\underline{t_i},t\right),\underline{t_i},T_i\right)\right.\\
&\left.+\mathbb{1}\left(\underline{t_i}+T_i<t\right)\cdot\Pr\left(\text{driver } i \text{ picks up a passenger } j\in J^{\star} \text{ by } \underline{t_i}+T_i\,\middle|\,r_i,S_{\underline{t_i}},\tilde{S}\left(\underline{t_i},t\right),\underline{t_i},T_i\right)\right]\\
\geq&E_{T_i}\left[\mathbb{1}\left(\underline{t_i}+T_i<t\right)\cdot\Pr\left(\text{driver } i \text{ picks up a passenger } j\in J^{\star} \text{ by } \underline{t_i}+T_i\,\middle|\,r_i,S_{\underline{t_i}},\tilde{S}\left(\underline{t_i},t\right),\underline{t_i},T_i\right)\right]\\
\geq&E_{T_i}\left[\mathbb{1}\left(\underline{t_i}+T_i<t\right)\cdot p_{i,\underline{t_i},T_i}\right]\\
=&\int_0^{t-\underline{t_i}}E\left[\prod_{\tilde{t}\in\mathcal{R}\left(\underline{t_i},\underline{t_i}+\tau\right)}\mathbb{1}\left(\max\left\{0,\max_{j\in S_{\tilde{t}}\cap J^{\star}}u_{ij}\right\}>\max_{j\in S_{\tilde{t}}\setminus J^{\star}}u_{ij'}\right)\right.\\
&\left.\cdot\mathbb{1}\left(\max_{j\in S_{\underline{t_i}+\tau}\cap J^{\star}}u_{ij}>\max\left\{0,\max_{j'\in S_{\underline{t_i}+\tau}\setminus J^{\star}}u_{ij'}\right\}\right)\,\middle|\,r_i,S_{\underline{t_i}},\tilde{S}\left(\underline{t_i},t\right),\underline{t_i}\right]dF_T^{\text{drv}}\left(\tau\right).
\end{aligned}$$

For the upper bound, by Assumptions 8, 10, 11 and 12, we have the following

$$
\begin{aligned}
&\Pr\left(\text{driver } i \text{ picks up a passenger } j \text{ by } t, j \in J^{\star} \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right) \\
=&E_{\mathcal{T}_i}\left[\mathbb{1}\left(\sum_{k, t_i^k \leq t} \sum_{j \in S_{t_i^k} \cap J^{\star}} \mathbb{1}\left(u_{ij} > v_i\left(\left\{S_{t_i^{\kappa}}\right\}_{\kappa \leq k}, t_i^k\right)\right) \cdot \mathbb{1}\left(u_{ij} > \max\left\{0, \max_{j' \in S_{t_i^k}, j' \neq j} u_{ij'}\right\}\right) > 0\right) \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}, \mathcal{T}_i\right] \\
\leq&E_{\mathcal{T}_i}\left[\mathbb{1}\left(\sum_{k, t_i^k \leq t} \sum_{j \in S_{t_i^k} \cap J^{\star}} \mathbb{1}\left(u_{ij} > v_i\left(\left\{S_{t_i^{\kappa}}\right\}_{\kappa \leq k}, t_i^k\right)\right) > 0\right) \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}, \mathcal{T}_i\right] \\
=&E_{\mathcal{T}_i}\left[1 - \prod_{k, t_i^k \leq t}\left(1 - \mathbb{1}\left(\max_{j \in S_{t_i^k} \cap J^{\star}} u_{ij} > v_i\left(\left\{S_{t_i^{\kappa}}\right\}_{\kappa \leq k}, t_i^k\right)\right)\right) \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}, \mathcal{T}_i\right] \\
\leq&E_{\mathcal{T}_i}\left[1 - \prod_{k, t_i^k \leq t}\left(1 - \mathbb{1}\left(\max_{j \in S_{\underline{t_i}} \cup \tilde{S}\left(\underline{t_i}, t\right) \cap J^{\star}} u_{ij} > 0\right)\right) \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}, \mathcal{T}_i\right] \text{ (Assumptions 10 and 12)} \\
=&E_{\mathcal{T}_i}\left[1 - \left(1 - \mathbb{1}\left(\max_{j \in S_{\underline{t_i}} \cup \tilde{S}\left(\underline{t_i}, t\right) \cap J^{\star}} u_{ij} > 0\right)\right) \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}, \mathcal{T}_i\right] \\
=&\Pr\left[\max_{j \in S_{\underline{t_i}} \cup \tilde{S}\left(\underline{t_i}, t\right) \cap J^{\star}} u_{ij} > 0 \,\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right), \underline{t_i}\right].
\end{aligned}
$$

The proof of the second set of inequalities is similar.

C Empirical Implementation: Estimating F_T^{psg}

We specify F_T^{psg} as an exponential distribution with a mean parameter to be estimated. The estimating equations are given by the conditional moment inequalities,

$$
\begin{aligned}
E\left[\mathbb{1}\left(\text{single}, \overline{t_j} - \underline{t_j} < \tau\right) - F_T^{\text{psg}}(\tau) \,\middle|\, r_j, S_{\underline{t_j}}\right] &\leq 0 \\
E\left[F_T^{\text{psg}}(\tau) - \mathbb{1}\left(\overline{t_j} - \underline{t_j} < \tau\right) \,\middle|\, r_j, S_{\underline{t_j}}\right] &\leq 0.
\end{aligned}
$$

for τ =300, 600, 900, ...1800 seconds. To capture the information in the conditioning variables, we compute the number of drivers who need to detour less than 5KM to answer the request of passenger j, and we construct indicator variables for whether this number is 0, between 1 and 5, 6 and 10, ..., 26 and 30, or greater than 30. Given a value of τ, we interact the indicator functions with the inequalities and average the interactions to obtain the sample analog. We eventually have K =96 moments.

We next construct the test statistic in Chernozhukov et al. (2019). To simplify the notation, we use $\hat{m}_k$ and $\hat{s}_k$ to denote the sample average and standard deviation of the

kth moment, and the test statistic is

$$\max_{1\le k\le K} \frac{\sqrt{n}\hat{m}_k}{\hat{s}_k},$$

where n is the sample size. For a test of size α, the critical value is[23]

$$\frac{\Phi^{-1}\left(1-\alpha/K\right)}{\sqrt{1-\frac{1}{n}\left(\Phi^{-1}\left(1-\alpha/K\right)\right)^2}}.$$

D Empirical Implementation: Estimating Driver Preferences

D.1 Driver Preferences

In this subsection, we discuss how to compute the sample analog of moment inequalities in Section 3. Give the conditioning variables, it is straightforward to compute the probability

$$E\left[\max_{j\in S_{\underline{t_i}}\cup\tilde{S}\left(\underline{t_i},\tau\right)\cap J^\star} u_{ij}>0\;\middle|r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i},t\right), \underline{t_i}\right].$$

We simulate 50 draws from the standard normal distribution to compute the unobserved driver types, $\beta_d+\sigma_d\nu_i$ and $\sigma_\varepsilon\varepsilon_{ij}$. For each type and a given value of t, we determine whether the utility of the best passenger in the set $S_{\underline{t_i}}\cup\tilde{S}\left(\underline{t_i},t\right)\cap J^\star$ is positive and average the indicators to obtain the simulated probability.

The calculations of

$$\int_0^{t-\underline{t_i}} E\left[\prod_{\tilde{t}\in\mathcal{R}\left(\underline{t_i},\underline{t_i}+\tau\right)} \mathbb{1}\left(\max\left\{0,\max_{j\in S_{\tilde{t}}\cap J^\star} u_{ij}\right\}>\max_{j\in S_{\tilde{t}}\setminus J^\star} u_{ij'}\right)\right.$$
$$\left.\cdot\mathbb{1}\left(\max_{j\in S_{\underline{t_i}+\tau}\cap J^\star} u_{ij}>\max\left\{0,\max_{j'\in S_{\underline{t_i}+\tau}\setminus J^\star} u_{ij'}\right\}\right)\;\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i},t\right), \underline{t_i}\right] dF_T^{\mathrm{drv}}\left(\tau\right)$$

and

$$\int_0^{t-\underline{t_i}} \Pr\left(\max_{j\in S_{\underline{t_i}+\tau}} u_{ij}>0\;\middle|\, r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i},t\right), \underline{t_i}\right) dF_T^{\mathrm{drv}}\left(\tau\right)$$

require simulating from the distribution of F_T^{drv} and the distribution of $S_{\tau+\underline{t_i}}$ given the

[23]The paper also offers a bootstrap version of the test. The results from either test statistic are nearly identical and do not affect the results reported in the tables.

conditioning variables. For F_T^{drv}, we use the calibrated distribution in Section (4.5). If $t < \underline{t_i}$, we set the integral to 0. To calculate the quantities inside the integrals, we sample from the empirical distribution of $\left\{S_{\tau+\underline{t_i}}\right\}_{0\leq\tau\leq t-t_i}$. We first define $N(r,S)$ as the number of passengers in S such that a driver with route r detours less than 5KM to pick up and that $\frac{\text{fare}}{|r|} > 0.75$. The function measures the number of passengers requesting similar routes to r. We then create bins for paths of the set of waiting passengers within a duration of $t-\underline{t_i}$ based on $N(r,S)$, where a path of the set of waiting passengers starting at the clock time t_m with a duration of $t-\underline{t_i}$ is $\{S_{\tau+t_m}\}_{0\leq\tau\leq t-\underline{t_i}}$. For pairs of integers

$$\left(n_1^S, n_2^S\right) \in \{(0,1),(1,2),(2,5),(5,10)\},$$

we define

$$\mathcal{N}_i\left(n_1^S, n_2^S\right) = \left\{\{S_{\tau+t_m}\}_{0\leq\tau\leq t-\underline{t_i}} : N\left(r_i, \cup_{0\leq\tau\leq t-\underline{t_i}}\{S_{\tau+t_m}\}\right) \in \left[n_1^S, n_2^S\right)\right\}.$$

We then randomly sample 50 paths of waiting passengers, $\{S_{\tau+t_m}\}_{0\leq\tau\leq t-\underline{t_i}}$, where $m = 1,\ldots,50$, such that each path falls in the same bin $\mathcal{N}_i$ as $S_{\underline{t_i}} \cup \tilde{S}\left(\underline{t_i}, t\right)$. For each path, we calculate the integral with respect to F_T^{drv} using the quadrature rule with 20 equally spaced nodes. We then average over the paths to obtain the simulated probabilities. The simulation error is likely small, because further increasing the number of simulations has little effects on the results.

Finally, for each i, we transform the conditional moments into unconditional ones. For pairs of distances and integers

$$(d_1, d_2) \in \{(0,22.5),(22.5,25),(25,27.5),(27.5,30),(30,35),(35,40),(40,+\infty)\},$$

and

$$\left(n_1^S, n_2^S\right) \in \{(0,1),(1,2),(2,5),(5,10)\},$$

we define

$$\begin{aligned}\mathcal{C}_S\left(d_1, d_2, n_1^S, n_2^S\right) = \Big\{r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right) : |r_i| \in [d_1, d_2),\\ N\left(r_i, S_{\underline{t_i}} \cup \tilde{S}\left(\underline{t_i}, t\right)\right) \in \left[n_1^S, n_2^S\right)\Big\}.\end{aligned}$$

We then interact the inequalities with indicator variables

$$\mathbb{1}\left(\left(r_i, S_{\underline{t_i}}, \tilde{S}\left(\underline{t_i}, t\right)\right) \in \mathcal{C}_S\left(d_1, d_2, n_1^S, n_2^S\right)\right)$$

for each combination of (d_1, d_2) and $\left(n_1^S, n_2^S\right)$ above.

We average across observations to obtain sample analogs of conditional moment inequalities. For the test statistic, we collect the moments for all combinations of t and $J^\star$, where t takes on 4:42 p.m., 4:44 p.m., ..., 5:00 p.m., and $J^\star$ takes on the forms of

$$\{j : d_{ij} < d^\star\}, d^\star = 1, 2.5, 5, 7.5, 10, 12.5$$

and

$$\{j : f_j < f^\star\}, f^\star = 15, 20, ..., 35.$$

We have a total of 6720 moments.

D.2 Driver Discount Factor

We define the driver to have answered a passenger's request at $\underline{t_i}$ if the driver answered the request within 1 minute of the arrival.[24] On the sample of drivers arriving between 4:40 p.m. and 5:00 p.m., 270 out of 6454 drivers accepted passenger requests upon the drivers' arrival. Similar to Appendix D.1, we define

$$\begin{aligned}\tilde{\mathcal{C}}_S\left(d_1, d_2, n_1^S, n_2^S\right) = \Big\{ & r_i, S_{\underline{t_i}} : |r_i| \in [d_1, d_2), \\ & N\left(r_i, S_{\underline{t_i}}\right) \in \left[n_1^S, n_2^S\right)\Big\},\end{aligned}$$

where the $N(\cdot)$ function is defined as in the previous section. We use $\mathbb{1}\left(\left(r_i, S_{\underline{t_i}}\right) \in \tilde{\mathcal{C}}_S\left(d_1, d_2, n_1^S, n_2^S\right)\right)$ to construct indicator variables and interact them with the inequalities. There are a total of 56 moments.

D.3 Inference

Our sample consists of the choices of 6454 drivers, and the test statistic is based on the moments in Appendix D.1 and D.2, a total of 6776 moments. The inference procedure is similar to Appendix C, where we use the maximum of the Studentized moment as the test statistic.

We next offer a brief discussion on the model fit. First, the model is not rejected at the 95% or 90% confidence level. At the parameter value minimizing the test statistic, 312 out of the 6776 moments are violated. Using this parameter, we plot the bounds of the probability of a driver accepting a passenger's request t seconds after the driver arrived to

[24] Changing the definition to within 45 seconds of arrival or 75 seconds of arrival generates similar estimates and does not qualitatively change the welfare results in Section 5

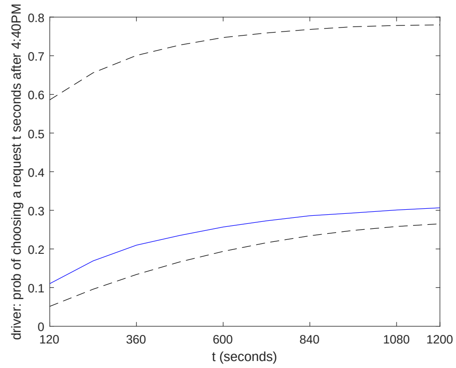


Figure 9: Model Fit: Probability of Choice

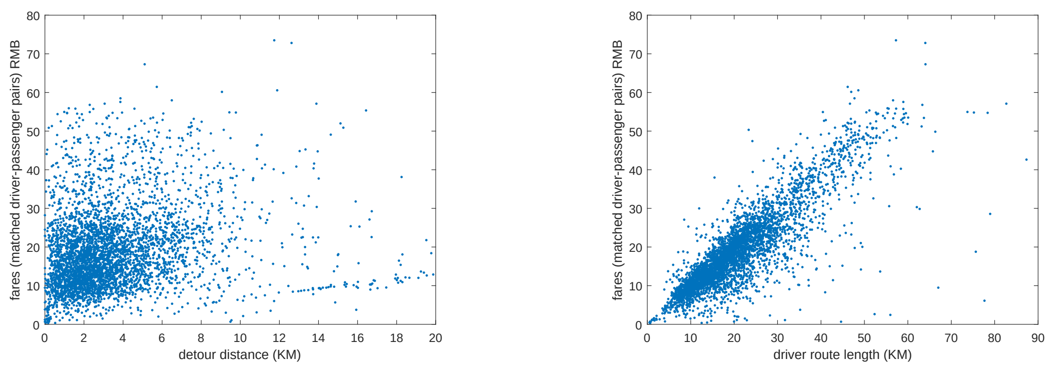


Figure 10: Data Patterns and Identification

the market in Figure 9. The blue line (solid) is based on data and the black ones (dash) are the bounds implied by the model.

Finally, we provide some data patterns that underpin the estimates in Table 4. As discussed in 4.5, identification hinges on the variation of attributes in the choice set (the set of waiting passengers). In the left panel of Figure 10, we show that there exist a positive correlation (0.2) between the detour length and fare conditional on the matched passengers and drivers, consistent with the interpretation that drivers need to be additionally compensated for extra detours. On the right, we show an even stronger positive correlation between the driver's trip length and the matched passenger fare. This pattern could be explained by (a) that the willingness of accepting a request decreases in the driver's trip length or (b) that there are fewer requests with short detours for these drivers. Our estimates, which account for the effects of detours and choice sets in (b) explicitly, still find evidence of the effects of (a).

E Omniscient Planner Solution

We discuss the computation of the planner solution. The planner solution is complex given that it needs to account for cross sectional and inter–temporal tradeoffs among a large number of agents. Our method leverages a result in calculating the equilibrium assignment in a transferable utility game, where the (integer) equilibrium allocation (1) maximizes the total surplus and (2) coincides with the solution to a continuous linear programming problem (Chapter 8, Roth and Sotomayor (1990)). Specifically, we define the payoff of a match as the following:

$$U_{ij} = \begin{cases} -\infty & \text{if } \underline{t_i} + T_i < \underline{t_j} \text{ or if } \underline{t_j} + T_j < \underline{t_i} \text{ or } u_{ij} < 0 \\ e^{-\delta\left(\underline{t_i} - \underline{t_j}\right)} u_{ij} + w_j^H & \text{if } \underline{t_j} < \underline{t_i} < \underline{t_j} + T_j, u_{ij} > 0 \\ u_{ij} + e^{-\vartheta\left(\underline{t_j} - \underline{t_i}\right)} w_j^H & \text{if } \underline{t_i} < \underline{t_j} < \underline{t_i} + T_i, u_{ij} > 0 \end{cases}$$

In the above, the specification takes into account whether a match is physically feasible (the driver and the passenger can physically meet in the market) and whether the match is acceptable to the driver (the driver utility $u_{ij} > 0$). If two agents are compatible on both dimensions, the platform matches them at the earliest possible moment (at the arrival of whoever comes later). Next, we define the utility of being unmatched:

$$\begin{aligned} U_{i0} &= 0 \\ U_{0j} &= \max\left\{0, w_j^E\right\}. \end{aligned}$$

The specification assumes that if an agent ends up being single, the platform moves it to its outside option as soon as the agent arrives. We use a_{ij} to denote whether driver i and passenger j are matched. The match outcomes obey the following constraints:

$$\begin{aligned} \sum_j a_{ij} &\leq 1, \forall i > 0 \\ \sum_i a_{ij} &\leq 1, \forall j > 0 \\ 0 \leq a_{ij} &\leq 1. \end{aligned}$$

The first two constraints say that the match is 1–1. The last constraint requires a_{ij} to be between 0 and 1. By not specifying the constraints on a_{i0} or a_{0j}, we assume that the platform can set an arbitrary number of agents to be single. Finally, the objective function is

$$\max_{a_{ij}} \sum_{i,j} a_{ij} U_{ij}.$$

The linear programming problem subject to the constraints above produces integer solutions that coincide with the optimal assignment.

We note that, although we use this linear programming problem to solve for the planner solution, the matching process in our context is still represented by a non–transferable utility game. We are able to monetize the welfare of both sides of the game and thus compute a notion of "total surplus", and our constraints respect the assumption that agents do not adjust fares to maximize match–level surplus.

F Implementability

In this section, we examine whether a mechanism suggested in Akbarpour et al. (2020) can induce agents to reveal how long they plan to search (T_i and T_j for drivers and passengers). In this mechanism, an agent has only one chance to announce that she is about to leave the market. If the agent chooses to make the announcement, she will be given the opportunity to match. If the match is unsuccessful, the platform will treat the agent as if she has left. While the agent waits, the platform tries to match the agent whenever there is an announcement from another agent. In our model, the passenger has no incentive to prematurely leave the market as long as the algorithm generates an acceptance rate weakly greater than $\varkappa$ in the decentralized market. On the driver side, we use our estimates to provide suggestive evidence that some but not all drivers would truthfully reveal whether they reach the end of their horizons in a steady state under the patient algorithm.

Our approach is to compare the driver's expected value of making the announcement now with the expected value of making the announcement a short time later. We assume that in this mechanism, a driver does not observe the set of waiting passengers until she announces she will leave, at which point the driver can choose one passenger from the set of waiting passengers. We use D_t and S_t to represent the sets of waiting drivers and passengers. Formally, the expected value of an announcement for driver i at t is:

$$V_t = E_{S_t}\left(\max\left(\max_{j \in S_t} u_{ij}, 0\right)\right).$$

To calculate the expected value of waiting for Δt seconds, we assume that between t and $t + \Delta t$, the driver participates in each potential match due to the exit of each passenger. If the driver i is not matched at $t + \Delta t$, the driver announces that she will leave and gets the chance to choose from the set of waiting passengers. The value of waiting, denoted as $V_{t+\Delta t}$, is the average driver utility from this process (with realized match values discounted to t) given the stationary distribution of $\{D_\tau, S_\tau, t \leq \tau < t + \Delta t\}$ under the patient algorithm.[25] We note that in a stationary environment, $V_{t+\Delta t}$ represents a lower bound of the option value of waiting because committing to waiting for a duration of Δt is likely a sub-optimal dynamic strategy.

We simulate the comparison for 10,000 drivers for parameters in the 95% confidence set. To simulate V_t and $V_{t+\Delta t}$, we randomly choose 50 ts and the associated sets of drivers and passengers $\left\{D_{\tilde{t}}, S_{\tilde{t}}, t \leq \tilde{t} < t + \Delta t\right\}$ for $\Delta t = 60$ seconds. We calculate V_t as the average of $\max\left(\max_{j \in S_t} u_{ij}, 0\right)$ across the 50 draws of S_t. For each draw of $\left\{D_{\tilde{t}}, S_{\tilde{t}}, t \leq \tilde{t} < t + \Delta t\right\}$, we calculate i's utility for waiting between t and $t+\Delta t$; if i remains unmatched at $t+\Delta t$, we calculate the driver utility as $e^{-\Delta t \cdot \delta} \max\left(\max_{j \in S_{t+\Delta t}} u_{ij}, 0\right)$; if i is matched with a leaving passenger at $\tilde{t}$ between t and $t + \Delta t$, we discount the utility to the moment of t by $e^{-\delta\left(\tilde{t}-t\right)}$. We take the average across the 50 draws from stationary distribution to obtain $V_{t+\Delta t}$.

Our results indicate that 37%–59% of drivers prefer to wait rather than announce now. The result is not driven by the discount factor δ but by the fact that matches triggered by leaving passengers are of lower qualities than the expected value of the match a driver can find when the driver makes the announcement. In particular, we find that the drivers who are more likely to find good matches are less likely to wait. Specifically, for each simulated driver, we calculate the share of passengers in J (the set of all observed passengers in the data) such that the detour associated with picking up the passenger is less than 5 km, and we use this share as our measure of ease to match. The median of this measure is 1% and the 75% quantile is 1.7%. In our simulation, we find that no driver with ease to match above 1.7% prefers to wait. These results suggest that the platform may need to offer incentives to induce these drivers to wait.

[25] The set of waiting drivers D_τ matters due to our tie-breaking rule.