Driving the Drivers:
Algorithmic Wage-Setting in Ride-Hailing

Yanyou Chen, Yao Luo, Zhe Yuan *

PRELIMINARY DO NOT CITE

This Version: *March 24, 2022*

Abstract

New algorithmic technologies, such as ride-hailing services, facilitate flexible work arrangements. However, we argue that ride-hailing companies may exercise algorithmic wage-setting and discriminate over certain work schedules, hence limiting drivers' real-time flexibility. Using rich data from China's leading ride-hailing company, we show that low-performing drivers earn 8 percent less hourly wages than high-performing drivers, despite comparable ride prices and take rates. Three main factors drive the wage differentials: high-performing drivers get more rides, less idle time, and riders with lower cancellation rate. Next, we rule out several alternative explanations, including drivers' strategic choice of service area, requests, and driving speed. Lastly, we construct and estimate a dynamic model of drivers' labor supply, accounting for time-varying reservation values. Combined with an estimated rider demand model, our counterfactual analysis shows that "fair" pay leads to about 5 percent gains in driver surplus and 0.78 percent to 1.37 percent loss in consumer surplus and platform revenue. Overall, social welfare increases by 1.71 percent to 2.10 percent.

Keywords: Two-Sided Market, Fair Pay, Work Schedule, Dynamic Discrete Choice, Market Power, Conditional Choice Probability

*Contact: Chen: University of Toronto, yanyou.chen@utoronto.ca; Luo: University of Toronto, yao.luo@utoronto.ca; Yuan: Zhejiang University, yyyuanzhe@gmail.com. We thank Duoyi Zhang, Anna Li, Mohaddeseh Heydari Nejad, Isaac Shiying Xi, and Weiyue Zhang for excellent research assistance. All errors are our own.

1 Introduction

Recent years have witnessed the rapid acceleration of algorithmic technologies. Online retailers now adopt algorithmic pricing and assortment planning to personalize customers' option sets and prices. In labor markets, algorithmic scheduling and wage-setting approaches have spread and changed the relationship between workers and employers. Employers collect a wide array of information on workers to help manage the workforce, direct tasks, and set wages. For example, a company can use an algorithm to screen job applicants or fire workers who fail to meet targets. Gig workers rely on assignments controlled by the algorithms of app-based platforms. Such algorithms aim to optimize the employers' benefits, this significantly changes how workers' wages and work schedules are determined. While advocates maintain that such technologies increase efficiency, critics argue the power asymmetry creates new sources of biases. Despite the difficulty in determining the evolution of algorithmic technologies, policymakers and academics are becoming increasingly concerned with the issues that arise in their applications. For example, while it is illegal to discriminate against specific demographics, an algorithm may penalize highly correlated factors. Thus, there is an urgent need to better understand the emerging challenges posed by algorithmic technologies in the labor market.

Prominent examples are ride-hailing markets. The proliferation of smartphones and mobile internet is driving the demand for ride-hailing services across the globe. A ride-hailing platform provides riders an economical mode of transportation, such as for daily work commutes, and allows drivers to create their own work schedules to best fit the job into their lives. Geolocation-based matching of drivers and riders creates substantial efficiency gains; the flexibility in work schedules increases driver utility. See Chen et al. (2019), which estimates drivers' valuation of flexible work schedules. However, not all work schedules are the same to the platform: some yield more revenue. Inevitably, an optimizing algorithm rewards "high-performing" drivers, who work long and consecutive hours, for the benefit of the platform.

This paper provides the first empirical study of algorithmic wage-setting and documents wage differentials due to work schedules in the ride-hailing market. While it is well-known that ride-hailing companies use dynamic pricing in rider fare that reflects demand and supply at different times, few papers study how wage rate is affected by hourly work schedule. Earnings in an hour depend not only on the particular hour, but also depend on whether the driver works in other hours. Conditional on customer ratings and quality of the driver, the algorithm may prioritize order assignment for drivers with long and stable working history on the platform.[1] For example, at 7AM of the day, the effective wage rate of a driver who also worked at 4–6AM may be higher than a driver who only works at 7AM, because the algorithm may prioritize the former driver in order assignment. To understand why platform's algorithm may discriminate over certain work schedules, and study how these factors affect drivers and consumers requires an equilibrium view of rider demand, drivers' labor supply, and the platform's algorithm. Our analysis consists of two parts. First, we document significant wage differentials due to work schedules using rich data from Platform X, the leading ride-hailing company in China. Second, we construct and estimate a dynamic labor supply model with elastic rider demand and endogenous platform pricing and wage-setting. We use the estimated model to quantify the effects of discriminatory wage-setting relative to a non-discriminatory setting — "equal pay for equal work".

First, we argue that ride-hailing companies exercise algorithmic discrimination in wage-setting and thereby limit driver utilization of real-time flexibility. Although tasks are standard and easily transferable in ride-sharing markets, each driver incurs a "warm-up" cost to start working, such as cleaning the car and filling up on gas. Expecting a continuous arrival of riders, having a driver work long hours is more efficient than multiple drivers working short hours. Moreover, because rider arrival and drivers' outside options fluctuate throughout the day, hiring only long-hour workers would lead to excess labor supply in off-peak hours. Thus, the platform combats this by mixing driver types and using time-varying ride prices and driver wage differentials to balance demand and supply.

Using high-frequency data for drivers of China's largest ride-hailing company, we first show that low-performing drivers earn 8 percent less in hourly wage from algorithmic discrimination. Three main factors drive the wage differential: high-performing drivers are given more ride requests per hour, wait fewer minutes for each request, and receive more requests from customers with lower cancellation rates. On the one hand, the fares riders pay are only time-varying, and the drivers receive similar portions of the received fare. Next, we rule out several alternative explanations, some of which are documented in the literature for other ride-hailing markets (see, e.g., Cook et al. (2021)). We find that drivers' strategic choice of service area, rider requests, and driving speed do not explain the wage differentials that we identify in our data. The large wage difference that we identify is mainly due to algorithmic wage-setting, which penalizes low-performing drivers.

Second, to quantify the effects of discriminatory wage-setting, we construct and estimate a dynamic model of drivers' hourly labor supply in a day. We propose a dynamic discrete choice

[1] For example, Uber has "Uber Pro" program which categorises drivers by how many rides drivers have finished. Drivers score higher scores by working long hours, and working at particular time of the day. Drivers with higher scores may get prioritized in matching with consumers.

model that is similar to Frechette et al. (2019), and we further incorporate decisions of the platform in the two-sided market. Our model accounts for riders' downward-sloping demand, drivers' dynamic labor supply, and the platform's fare and wage setting. There are two market power sources that drive the platform's pricing decisions: the driver faces alternative time-varying outside options, and the rider has alternative modes of transportation. Drivers first choose work schedule types and then hourly work schedules by solving finite-horizon dynamic discrete choice problems. While drivers are free to set their own work schedules, the platform rewards high-performing drivers by assigning them more frequent and rewarding trips, leading to wage differentials between work schedules.

In estimating rider preferences, we employ an IV approach to deal with unobserved factors that may affect demand and thus rider fare schedules. We obtain an estimated elasticity of about 1.2, indicating an elastic demand and the significance of alternative transportation modes. In estimating drivers' preferences, we account for driver heterogeneity. To this end, we propose a GMM estimator that integrates the CCP estimator of Hotz and Miller (1993) and accounts for moments from heterogeneous drivers and work schedules. The average reservation value for non-local drivers is 57.6 CNY and 51.3 CNY for those below age 40 and above age 40, respectively. In comparison, the corresponding estimated valuations for local drivers are 46.6 CNY and 49 CNY.

Combining the estimated labor supply model and the rider demand model, our counterfactual analysis shows how a non-discriminatory wage-setting scheme changes consumer surplus, driver surplus, and platform revenue. We recalculate the market outcomes when the platform is constrained to assign trips "fairly" and re-optimize the fare schedule. In this case, fare prices increase, especially during off-peak hours. Evidently, without targeting specific drivers, a "fair" algorithm needs to increase prices to motivate drivers on average. Overall, the new algorithm decreases consumer surplus by about 0.78 to 1.37 percent and platform revenue by similar percentages. On the other hand, drivers benefit by about 5 percent on average. In aggregate, total surplus could increase by 1.71 to 2.10 percent.

Related Literature

Our model builds on the literature on two-sided markets, which has been focusing on how the platform sets prices for both sides of the market. See Rysman (2009) for a recent survey. We take this view to the ride-hailing market, allowing for two sources of market power: driver and rider outside options. While Rysman (2004) proposes a general setting with oligopoly competition between platforms, we choose to focus on China's dominant platform, Platform X Express. This simplification approximates well the industry structure and allows us to incorporate important dynamics in drivers' labor supply. Castillo (2020) studies Uber's surge pricing using an empirical model of the two-sided market with riders, drivers, and the platform. Because Platform X effectively abandoned surge pricing, our focus is on the platform's wage-setting decisions.

Our results add to the labor literature on wage differentials. There has been extensive documentation of wage differentials based on demographics, such as gender and race. Altonji and Blank (1999) provides a great overview of this literature. Blau and Kahn (2017) surveys the literature on the gender pay gap. Another small literature studies part-time and full-time wage differentials. For example, Aaronson and French (2004) studies the joint determination of hours and wages exploiting variation in labor hours induced by social security rules. Our paper is the first to study algorithmic wage-setting and document wage differentials due to hourly work schedules. While algorithms can be designed to be neutral to specific demographics, they could evolve into evaluating highly correlated factors, such as work schedules. Therefore, our analysis sheds light on the discussion surrounding the impact of algorithmic technologies on efficiency and equality.

Our paper is closely related to Chen et al. (2020), which also estimates a dynamic model of labor supply with driver preferences over work schedules. Like us, they model each period as an hour and estimate Uber drivers' dynamic labor supply. Our paper complements their paper in several ways. Instead of treating each driver as living infinitely many periods, we model each driver's hourly work schedule in a day, leading to a finite horizon dynamic choice problem. While they model hourly wage as an exogenous process, our paper mainly focuses on endogenous wage differentials. As a result, we explicitly incorporate elastic rider demand and platform decisions on ride prices and wages. Relatedly, Chen et al. (2019) employs a static model to estimate drivers' hourly reservation wages using data from Uber. Taxi drivers' labor supply has been studied extensively in the literature. See, e.g., Camerer et al. (1997) and Farber (2015). Our setting fits more into the first paper's finding that cabdrivers make labor supply decisions "one day at a time."

Lastly, our empirical strategies follow the empirical industrial organization literature. In particular, to deal with potential unobserved heterogeneity, we employ an IV approach in estimating rider demand, similar to Kalouptsidi (2014). Moreover, in estimating the parameters that govern drivers' labor supply, we combine the GMM approach and the CCP estimator of Hotz and Miller (1993) from the dynamic discrete choice literature.

The remainder of the paper is organized as follows. Section 2 describes the ride-hailing industry in China and our data. Section 3 provides reduced-form evidence of algorithmic wage-setting that discriminates against low-performing drivers. Moreover, we conduct robustness analysis to exclude alternative explanations. Section 4 describes an equilibrium model with a dynamic model of drivers' labor supply. Section 5 discusses our estimation results and counterfactual experiments. Section 6 concludes. The Appendix contains all omitted details.

2 Industry Background and Data

In this section, we first describe the ride-hailing market in China and our data source, Platform X Express.[2] We then provide some summary statistics of our data and how we construct the working database.

2.1 Industry Background

With the development of China's residential travel demand, the number of ride-hailing users in China grew to 365 million by the end of the year 2020.[3] As the leading ride-hailing platform in China, Platform X captures over a 90 percent share of the mainland Chinese market. According to files submitted to SEC, Platform X has 15 million ride-hailing drivers and serves 493 million people globally, collecting an annual revenue of $21.6 billion USD in 2020.[4]

Matching Steps: Like Uber and Lyft, Platform X Express is a centralized two-sided platform that connects customers and drivers. Drivers and riders are matched in three steps: First, riders request a trip through their phone app. The request includes the origin and destination of their route, as well as a estimated fare. The fare is based on the distance of the proposed trip, predicted traffic status, and time of day. Second, the platform distributes the request to a nearby driver. Platform X usually distributes requests to drivers within three kilometers, but the area may be larger when there are few available drivers nearby. Unlike Uber, Platform X drivers cannot decline requests unless under extreme circumstances (e.g., the waiting time for the rider is too long). Third, the drivers, upon receiving the request, drive to the specified origin to pick up their rider and deliver them to their destination. The drivers cannot change the destination or distance of a trip and must follow the guidance of the platform. Most drivers strictly follow the route directed by the GPS. For each completed trip, drivers are paid a base fare plus a per-mile and per-minute rate according to the time of a day and distance of the hailed trip.

Fare Schedules: Drivers receive 79.1 percent of the rider fare, according to Platform X's

[2] Liu et al. (2019) studies Platform X Hitch, a small decentralized peer-to-peer platform, and quantifies the welfare loss due to decentralized matching.

[3] Page 47, China Internet Network Information Center (CNNIC) report.

[4] Platform X's SEC filings for its IPO on June 10, 2021.

annual report. The share is fairly constant according to the drivers that we interviewed. In the city we study (as of 2018), Platform X fare schedules divide a work day into six intervals: (1) peak-AM 7:00-10:00 (2) midday 10:00-16:00 (3) peak-PM 16:00-19:00 (4) afternoon off-peak 19:00-22:00 (5) late-night 22:00-00:00 (6) early-AM 00:00-6:59 (next day). Riders pay a 10 CNY base fare plus 0.38 CNY per minute and 1.9 CNY per mile for each Platform X Express trip during the off-peak periods. However, during the morning rush hours (7:00-10:00), the per-mile rate increases to 2.5 CNY, and during the evening rush hours (16:00-19:00), late-night (20:00-0:00), and early-morning (0:00-7:00), the per-mile rate is 2.4 CNY.

Driver Descriptions: Like most ride-hailing companies, Platform X touts its flexible work schedules. Compared to Uber and Lift, Platform X drivers work more hours and a greater proportion of drivers view this job as their main source of income. Based on anecdotal evidence and our interviews with Platform X drivers, Platform X categorizes drivers into two groups: (a) high-performing drivers, who work longer hours and/or during specific hours on the platform, and (b) low-performing drivers, whose work schedules are more casual. The platform's algorithm favors high-performing drivers by assigning more and better trip orders to them. While we do not observe such driver labels and orders in our data, we construct these variables using a combination of institutional knowledge and machine learning, both of which are validated by our reduced-form analysis.

2.2 Data Description

We obtain data from the Transportation Bureau of a major city in China. In these data, we observe all the completed transactions in December 2018, as well as all the created orders (including both finished transactions and canceled orders) from December 1st to December 10th, 2018.

Table 1 reports summary statistics at the order level. Panel I reports summary statistics for the first ten days, and Panel II reports for all 31 days. For the first ten days, 9 million customer orders were created, among which 58 percent were assigned to a driver and eventually completed.[5] An average order is 6.8 km with a drive time of 17 minutes. The average price per order is 24.8 CNY (about $4 USD). The average drive distance, drive time, and price of the entire 31 days are similar to those of the first 10 days.

The unit of observation in our raw database is driver-rider-order. However, it is challenging to conduct demand estimation, compare driver wages or simulate market outcomes

[5]There are two scenarios where an order is canceled: First, orders are automatically cancelled if no driver is available nearby and the order remains unassigned for more than 3 minutes. Second, drivers or customers may cancel the order after a match is made. In our data, 23.39 percent of canceled orders are initiated by drivers, 76.52 percent are initiated by customers, and the remainder is canceled by the platform.

Table 1: Summary Statistics of Orders and Transactions

Variable	Mean	Std. Dev.	Min	Max
Day 1 - 10, Both created orders and matched orders				
Price	24.82	26.18	0	2,194
DriveDistance (km)	6.84	6.89	0	1,377
Drivetime (minutes)	17.53	13.42	0	1,191
Probability of Order Matching		58%		
Number of Observations		9,192,983		
Day 1 - 31, only matched orders from day 11				
Price	25.31	26.44	0	3,387
DriveDistance (km)	6.92	6.85	0	727
Drivetime (minutes)	17.36	13.14	0	1,458
Number of Observations		14,471,573		

based on driver-rider-order level information because it is not standardized to the same measurement. We propose an algorithm and construct a driver-hour working dataset. For the working dataset, we are interested in driver operation and revenue information and construct several important variables for each driver-hour:

- **Earning time** is the trip duration, measured as the amount of time a driver spends with the rider. A driver can transport riders and collect revenue only during the earning time.

- **Drive Distance** measures the distance a driver serves a rider in an hour.

- **Driver's Hourly Wage** measures the revenue of a driver in an hour.[6] Given that the platform fee is around 20 percent of the revenue, the driver income is roughly 80 percent of the ride price.

- **Pickup Time** measures the time a driver is on the way to pick up riders.

- **Idle Time** is the time a driver spends waiting for orders. In a specific hour, we have the following relationship: Idle time = 60 - Work time - Pickup time.

- **Number of Orders** measures the number of orders a driver receives in an hour.[7]

[6] Our definition is different from Chen et al. (2019), which defines "wage rate" as a driver's total earnings in an hour, divided by minutes worked, multiplied by sixty. In other words, they study the wage rate when the driver is driving a rider, and we focus on the wage rate when the driver is active on the platform.

[7] In rare cases, an order may span several hours, which we attribute to the hour of departure.

Below, we discuss our algorithm of how we construct an hourly level dataset from the driver-rider-order level dataset:

- **Drop Outliers** We keep all orders with departure and arrival in the urban area (8 districts) within the city, drop orders with a price of zero, a price above 200, or that span over 4 hours. In total, we drop less than 0.5 percent of the observations.

- **Construct Work Schedules** Following Chen et al. (2019), we define a driver as working in hour t if he works at least ten minutes out of the hour. At night, when orders are sparse (from 22pm to 6am), we define a driver as working in hour t if he/she works at hour $t-1$ as well as hour $t+1$. All working hours of a driver comprise his/her work schedule.

- **Match Order to Hour** Suppose an order spans x hours. We divide this order into x sub-orders, with each sub-order corresponding to an hour. The hourly wage rate and driving distance are defined to be proportional to each hour. For instance, suppose an order starts at 8:50 and finishes at 9:20, yielding a revenue of 60 CNY. We say that $\frac{10}{10+20} = \frac{1}{3}$ of the order belongs to 8 am operations, and the rest contributes to 9 am operations. By doing so, we divide this order into two sub-order operations: The driver drives 10 minutes and earns 20 CNY at 8 am and drives 20 minutes (10 miles) and makes 40 CNY at 9 am. After matching orders to hours, we aggregate all sub-orders in an hour and obtain this driver's earning time, ride prices, pickup time, idle time, and number of orders in this hour.

With our algorithm, we obtain a driver-hour level database. Below, we first generate summary statistics for the driver-hour level database in Section 2.3. Then, we discuss potential discrimination in Section 3.

2.3 Driver Characteristics

To account for driver heterogeneity, we define four groups based on their age and birth city: (1) non-local young drivers; (2) non-local middle-aged drivers (3) local young drivers; (4) local middle-aged drivers. We define "non-local" drivers as those who were born outside the province we study. Drivers who were born before 1978 are labeled as "mid-aged" and the rest are "young". Because of the well-known and much debated discrimination surrounding China's "Hukou" system and age-related issues in more formal sectors, these groups face substantially different outside opportunities. In fact, around half of the drivers in our sample are non-local young drivers.

Table 2: Driver Characteristics

	Total (1)	Non-local Young (2)	Non-local Middle-aged (3)	Local Young (4)	Local Middle-aged (5)
Panel I: Performance (in a month)					
Work Days	12.0	12.7	13.1	10.1	11.3
Work Hours	104.3	114.1	124.3	74.8	89.9
# orders	196.8	218.1	230.1	142.2	167.3
Distance (km)	1,471.5	1,638.3	1,708.2	1,071.0	1,238.2
Panel II: Performance (in an hour)					
Earning Time (minutes)	30.6	30.6	31.0	30.0	30.5
Pickup time (minutes)	10.6	10.5	10.8	10.4	10.8
Idle Time (minutes)	18.8	18.9	18.1	19.6	18.7
# orders	1.9	1.9	1.9	1.9	1.9
Distance (km)	14.1	14.4	13.7	14.3	13.8
# obs	40,104	16,954	8,161	7,568	7,421
Share Obs	100%	42%	20%	19%	19%

Table 2 reports some main driver characteristics.[8] We report the characteristics of an average driver in Column (1) and the characteristics of the four groups in Column (2)-(5). Panel I reports driver performance in the month we study. Non-local drivers work longer in terms of both workdays and work hours. Though young and middle-aged drivers have similar workdays, middle-aged drivers work longer earning times. Drivers who work longer receive more orders and earn a higher monthly wage. Panel II summarizes the driver performance in an hour. For each hour worked, drivers spend only half of the time driving riders. On average, around 10 minutes are spent picking up customers, and another approximately 20 minutes are wasted waiting for orders. Non-local young drivers receive more orders and thus earn the highest hourly wage.

3 Reduced-Form Estimates

As discussed in Section 2.1, Platform X publishes a fare schedule for riders. While there is little scope of using the take rate to discriminate among drivers, Platform X can distribute higher quality orders to certain drivers, thereby affecting driver income. We have interviewed drivers about their experience of receiving low-/high-quality orders (Section 3.1). Drivers mention that their work schedules may affect their order quality. Section 3.2 motivates our choice of indices, based on which the algorithm discriminates drivers. We further define high-performing and low-performing drivers in Section 3.3. We find that high-performing drivers earn a higher hourly wage, and we attribute the wage differential to difference in

[8] More summary statistics can also be found in Online Appendix D.

order quality. Finally, we rule out alternative explanations in Section 3.4.

3.1 A Survey

To inform our empirical analysis, we hired a group of assistants to take a random set of real trips using Platform X Express and conduct interviews with their drivers. During our interviews with drivers, drivers' foremost concern is the quantity and quality of orders they receive, which ultimately determine their hourly wage. Low hourly wage drivers complain that they "wait too long for orders" and "spend most of their time picking up customers." High hourly wage drivers are fairly satisfied with the status quo, "always receiving a new order immediately after finishing an order." Thus, systematic differences in order assignment creates significant hourly wage differentials across drivers.

Our survey suggests that drivers are aware of the types of drivers the Platform X algorithm favors: i.e., those who work more total hours, preferably in consecutive hours, and those who work late at night, and work regularly (i.e. from 7 am - 10 pm every day), as well as how it rewards them (i.e. by assigning more trips and "better" riders, who have lower cancellation rates, are more polite, and are less far away). Drivers who earn a lower hourly wage receive significantly fewer orders than others.

3.2 Discrimination across Schedules

We explore the determinants of driver hourly wage differentials. Suppose drivers are identical except for their working schedules (i.e., some work longer hours, etc.). The platform algorithm may discriminate among drivers, distributing a different quantity and quality of orders to different drivers based on their work schedules, resulting in a higher hourly wage for some drivers but a lower hourly wage for others. We evaluate how a driver's hourly wage depends on their work schedule.

On any given day, the number of possible driver work schedules is 2^{24}, which is impossible to track. In our sample, the number of observed/realized schedules is more than 30 thousand. To reduce the dimensionality and explore discrimination across schedules, we divide the 24 hours in a day into four intervals: (1) peak-AM 7:00-9:59 (2) midday 10:00-15:59 (3) peak-PM 16:00-18:59 (4) night 19:00-6:59 (next day). This definition follows the Platform X fare schedules. Our peak hour definition also coincides with their definition of peak hour for the city we study.[9]

As a motivating fact, Table 3 shows that drivers who work more hours, especially off-peak hours, earn a higher hourly wage than those who work fewer hours. Column (2) in Table

[9]The city's peak hour is 7am-9am and 4:30pm-6:30pm, Mon-Fri.

Table 3: Hourly Wage Discrimination

Hourly Wage	(1)	(2)
# Work Hours in a month	0.003***	0.003***
	(0.000)	(0.000)
% Off-peak Hours		18.724***
		(0.170)
Constant	54.918***	39.201***
	(0.126)	(0.190)
Observations	4,182,331	4,182,331
R-squared	0.040	0.043
Day-Hour FE	Y	Y
Origin/Destination FE	Y	Y

Notes: Standard errors in parentheses. *** p<0.01.

3 shows that, when a driver works one more hour in a month, his hourly wage increases by 0.3 cents (CNY), controlling for the operation area, the day fixed effect, and hour fixed effect.[10] A more important feature for the wage gap is the fraction of off-peak hours. When a driver spends 1 percent more time working in off-peak hours, his/her hourly wage increases by 0.187 CNY.[11]

Remark: Appendix A contains additional regression tables on robustness checks and verifies driver demographics that may have explanatory power for wage differentials. Table A.1 reports how the driver's hourly wage depends on the fraction of different time intervals. We find that drivers' hourly wages are higher when they work more during midday and at night. Table A.2 shows that, given their work schedule, there is little evidence of wage differentials based on driver demographics. Indeed, including driver characteristics barely changes the R-squared. Moreover, one standard-error change in the fraction of off-peak hours changes wage rate by 14, which is an order of magnitude higher than the most important demographic variable, age. Despite the statistical significance of gender, it is economically insignificant in determining wage rate. In contrast, the coefficients on birth city and age reflect work schedule variation across driver groups conditioning on the first two variables. These results motivate controlling for age and birth city in our empirical analysis.

[10] Given that the 25th/75th percentile driver works 50/280 hours, their hourly wage gap is (280-50)*0.002 = 0.69 CNY.

[11] Given that the 25th/75th percentile driver spends 73 percent/84 percent of their time in off-peak hours, their hourly wage rate gap is (84%-73%)*18.7 = 2.06 CNY.

3.3 High-Performing and Low-Performing Drivers

While Platform X distinguishes between high-performing and low-performing drivers, our data is missing such labels. To tackle this problem, we first apply machine learning methods to identify key features that define these groups and then validate our definition via regression analysis. Online Appendix E reports the machine learning algorithm we use to cluster drivers based on their hourly wage. We find that drivers can be broadly divided into two groups. Drivers with relatively higher hourly wages tend to work longer hours and work more during midday and night time intervals. Therefore, we explore the determinants of driver hourly wage variation by regressing driver hourly wage on the number of working hours and fraction of time working in each interval. A driver is defined as a high-performing driver if he worked at least two consecutive hours in midday or night for at least 8 out of 21 workdays in the month we study and a low-performing driver otherwise.

We first show that there is a significant difference in hourly wage between high-performing and low-performing drivers.

Figure 1: Hourly Revenue

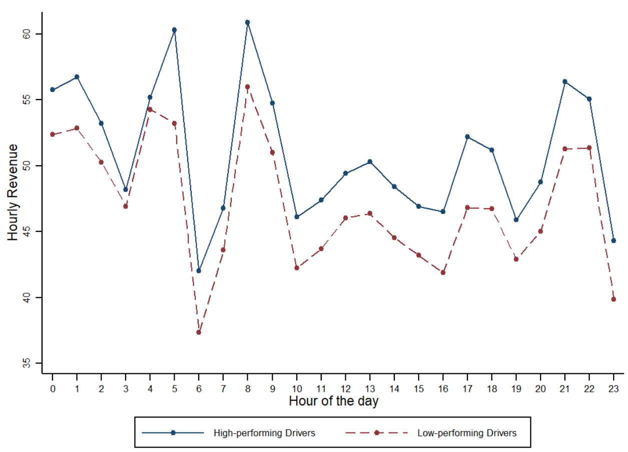


Table 4: High-performing v.s. Low-performing

Dependent Variables	Hourly Wage		
	(1)	(2)	(3)
High-performing	3.886***	3.794***	3.851***
	(0.0397)	(0.0393)	(0.0391)
Constant	46.49***	46.57***	47.24***
	(0.0376)	(0.0372)	(0.0701)
Day-Hour FE	N	Y	Y
Origin FE	N	N	Y
Destination FE	N	N	Y
Observations	4,182,328	4,182,328	4,182,328
R-squared	0.002	0.039	0.050

Notes: Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 1 illustrates the average hourly wage for high-performing and low-performing drivers at each time of the day. According to Figure 1, high-performing drivers earn a higher wage compared to low-performing drivers, and this difference persists across hours. We then conduct a more rigorous regression analysis on driver hourly wage. Table 4 shows that high-performing drivers earn 3.8 CNY (8.2 percent) more hourly compared to their low-performing counterparts. These results are robust when we control for day-hour fixed effects and origin/destination district dummies.

Given that high-performing drivers earn a significantly higher hourly wage, we then investigate what forces are driving the wage difference. We study the characteristics of orders and analyze the difference in characteristics between high-performing and low-performing drivers. In brief, results in Table 5 show that high-performing drivers receive more orders and orders of higher quality compared to low-performing drivers. On average, high-performing drivers receive 0.125 more orders (column 1) every hour. Moreover, orders assigned to high-performing drivers are 2.8 percent less likely to be cancelled by customers (column 2). Because high-performing drivers get assigned more orders every hour, they also drive 0.748 more kilometers (column 3) and 1.6 more minutes while carrying riders in an hour (column 4). High-performing drivers also spend 2.1 fewer minutes waiting for orders (column 5).[12]

Table 5: Driving Forces of Wage Difference

Dependent Variables	# Orders	Cancellation Rate (Consumer)	Drive Dist	Earning Time	Idle Time
	(1)	(2)	(3)	(4)	(5)
High-performing	0.125***	-0.0023***	0.748***	1.579***	-2.140***
	(0.0018)	(0.0004)	(0.0003)	(0.0187)	(0.0221)
Constant	1.468***	0.0894***	12.85***	32.35***	17.04***
	(0.00313)	(0.0005)	(0.0212)	(0.0334)	(0.0395)
Day-Hour FE	Y	Y	Y	Y	Y
Origin District FE	Y	Y	Y	Y	Y
Destination FE	Y	Y	Y	Y	Y
Low-performing Mean	1.76	0.082	13.4	29.3	20.4
Change	7.1%	-2.8%	5.6%	5.4%	-10.5%
Observations	4,182,318	4,815,026	4,182,318	4,182,318	4,182,318
R-squared	0.080	0.006	0.045	0.100	0.115

Notes: In columns (1) and (3)-(5), we use order-matched for the analysis. Order-matched is available from Dec. 1st, 2018 to Dec. 31st, 2018. In column (2), we use order-created to compute customer cancellation rates. Information on order-matched is available from Dec. 1st, 2018 to Dec. 10th, 2018. Standard errors are in parentheses. Column (1) reports the marginal effect from probit estimation.

In summary, the results show there are three main factors driving the wage differentials between high-performing and low-performing drivers. High-performing drivers are given more rides from the platform, waste less idle time waiting for orders, and receive more orders from customers of higher ratings (lower probability of cancellation from consumers). As assignment of orders is determined by the platform's algorithm, we hereafter term the systematic difference in the quantity and quality of order assignments based on work schedule (high-performing versus low-performing) as algorithmic discrimination.

Table D.1 zooms in and reports more detailed driver characteristics of high-/low-performing drivers. In our period of study, there are 23,712 high-performing drivers and 16,392 low-

[12] Gaineddenova (2021) shows that drivers prefer more expensive trips with a shorter pickup distance, using data from a decentralized ride-hailing platform.

performing drivers. High-performing drivers are more likely to be non-local (69 percent versus 53 percent) and younger (aged 37.2 versus 37.4). Because drivers with different demographics may have different outside options, they may self-select into high-performing versus low-performing work schedules based on their demographics. This phenomenon motivates us to estimate a separate set of parameters and reservation values for different type of drivers (local versus non-local and young versus middle-aged) in our analysis.

3.4 Alternative Explanations

This subsection discusses some alternative explanations that might drive the wage differentials we found. Rather than having the algorithm discriminate over different work schedules and assign orders differently, some may argue that drivers make decisions endogenously, resulting in the observed wage difference. For example, Cook et al. (2021) finds that the gender earnings gap amongst drivers can be entirely attributed to three factors: experience on the platform (learning-by-doing), preferences and constraints over where to work (driven largely by where drivers live and, to a lesser extent, safety), and preferences for driving speed. To provide a robustness check for our findings, we consider four alternative explanations: First, high-performing drivers may have better knowledge of more popular rider pickup areas and hence get more orders. Second, high-performing drivers may learn how to strategically reject and cancel rides, as found in Cook et al. (2021). Third, high-performing drivers may drive faster than others and hence earn a higher hourly rate. Fourth, high-performing drivers may know the streets better and hence choose better routes than low-performing drivers. We show that none of these alternative explanations are sufficient to explain the wage difference we find in our data.

First, let us explore whether high-performing drivers might be more likely to wait in areas where more orders, hence generating an earnings gap.[13] In our empirical analysis, we control for origin and destination fixed effects. For drivers working in the same area, we still find an 8.2 percent difference in hourly wage between high-performing and low-performing drivers (column 3 in Table 4). We also compare the operation districts for high-performing and low-performing drivers in Table 6. There are eight main districts in our studied city. Results in Table 6 show that there is no significant difference between high-performing and low-performing drivers in terms of where they work. Therefore, the wage differential found between high-performing and low-performing drivers is unlikely to be driven by drivers endogenously choosing where to work.

Second, Cook et al. (2021) shows that more experienced drivers learn how to strategically

[13] Haggag et al. (2017) finds that New York taxi drivers accumulate neighborhood-specific experience, which helps them to find riders.

Table 6: Active Area for High-performing and Low-performing Drivers

District	Origin		Destination	
	Low-performing	High-performing	Low-performing	High-performing
1	7%	7%	7%	7%
2	9%	8%	9%	8%
3	20%	22%	21%	23%
4	7%	7%	7%	7%
5	16%	15%	15%	14%
6	10%	11%	10%	11%
7	16%	15%	16%	15%
8	16%	15%	15%	13%
Total	100%	100%	100%	100%

reject and cancel rides, hence earning more. To examine whether such a mechanism exists in our data, we regress the probability of driver cancellations on driver type. Results in column (1) of Table 7 show that, in our data, high-performing drivers actually have lower cancellation rates than low-performing drivers. Therefore, it is unlikely to be true that the higher hourly wage of high-performing drivers is caused by drivers strategically rejecting and canceling rides in our case.

Third, some drivers may drive faster than others, and hence complete more trips and earn a higher hourly wage. Column (2) of Table 7 compares the speed of high-performing and low-performing drivers. We find that high-performing drivers do drive slightly (0.5 percent) faster than low-performing drivers. Because high-performing drivers have 5.4 percent more earning time, the difference in driving speed contributes an additional 0.53 percent hourly wage. This is consistent with Columns (3) and (4) of Table 5, which show the difference in drive distance is slightly higher than work time. Therefore, if anything, this 0.5 percent faster-driving speed only converts into an extra 0.26 CNY per hour, and thus explains very little of the large (8.2 percent or 3.8 CNY) wage difference between high-performing and low-performing drivers.

Last, some may argue that high-performing drivers may know the streets better, and hence high-performing drivers will use shortcuts or less congested routes. As our dataset contains only the origin and destination of an order, we thus do not observe the exact route chosen by the driver. Instead, according to our interviews with Platform X drivers and engineers, we document that Platform X drivers mostly follow the GPS recommended route on the Platform X app. Consumers may file complaints to Platform X if they notice that

Table 7: Driver Cancellation and Driver Speed

Dependent Variables	Probability of Cancellation (Driver)	Speed
	(1)	(2)
High-performing	-0.0062***	0.1313***
	(0.0002)	(0.0194)
Constant	0.0365***	0.410***
	(0.0003)	(0.0006)
Day-Hour FE	Y	Y
Origin/Destination FE	Y	Y
Low-performing Mean	0.034	24.63
Change compared to Low-type	-18.2%	0.5%
Observations	4,815,026	4,168,889
R-squared	0.004	0.089

Notes: In column (1), we use order-matched to compute driver cancellation rates . Information on order-created is available from Dec. 1st, 2018 to Dec. 10th, 2018. Column (1) reports the marginal effect from probit estimation.
In column (2), we use order-matched for the analysis. Order-matched is available from Dec. 1st, 2018 to Dec. 31st, 2018. Standard errors in parentheses.

drivers do not follow the suggested route. Therefore, drivers have less incentive to deviate from the recommended route. To sum up, these findings show that, in our data, none of the four alternative explanations are sufficient to explain the observed wage differential.

4 Model

We model the decisions of market participants for one day. At each hour of the day, there is a demand curve for rides, $D_t(P_t)$. Facing this demand curve, the platform makes two types of decisions. The platform first determines the price to charge consumers, P_t. We allow for dynamic pricing and thus allow price to vary across different times of the day. Second, the platform's algorithm allocates ride orders to each driver. The algorithm distinguishes between two types of drivers — high-performing drivers and low-performing drivers. High-performing drivers are committed to work consecutively for at least 2 hours during off-peak hours between 10AM-4PM and 7PM–6AM the next day. Low-performing drivers make no work schedule commitments. A driver i, who is characterized by $\boldsymbol{X_i}$, first decides whether to be a high-performing or low-performing driver $\tau \in \{H, L\}$. We assume that drivers choose their status (H or L) at the start of the day, and drivers cannot change their working status throughout the day. Conditioning on these choices, each driver chooses whether or not to work at each hour of the day. The problem is dynamic, because whenever a driver starts working or resumes working after a break, there is a fixed "warm-up" cost. If the driver chooses to be the high-performing type, the dynamic problem is under the constraint that working hours need to satisfy certain work schedules. Otherwise, the problem is unconstrained.

We use bold typeface to denote vectors containing values for each hour of the day. For example, $\boldsymbol{P}$ denotes all prices from $t = 1$ to $t = 24$. The sequence of wage rates is $\boldsymbol{W^{\tau}}$, which is determined by the platform's pricing decisions $\boldsymbol{P}$ and algorithm deciding who gets to drive customer $\boldsymbol{s}$.

4.1 Drivers' Dynamic Labor Supply

In the first stage, a driver chooses either to be a high-performing type or a low-performing type. Low-performing drivers solve an unconstrained dynamic discrete choice problem of when to work. To be a high-performing type, based on our discussion in Section 3.3, drivers need to work longer hours in off-peak hours. More specifically, we assume in our model that high-performing drivers are required to work consecutively for at least 2 hours during off-peak hours between 10AM–4PM and 7PM–6AM the next day. Besides fulfilling the required working hours, a high-performing driver makes an hourly choice of whether or not to work. We assume that drivers choose their status (H or L) at the start of the day, and drivers cannot change their working status during the day. If a driver chooses to be high-performing, the driver chooses which minimum requirement to satisfy in advance. For example, driver A may choose to be a high-performing driver by committing to work between 10AM and 12AM. Between 10AM and 12AM, driver A will be online with probability 1, and in any other time of the day, driver A can choose freely whether to work or not. A low-performing driver B does not commit to any work schedule. Ex post, even if driver B ends up working long hours, including from 10AM to 12AM, driver B would still be considered low-performing.

There are 16 possible work schedules that satisfy the high-performing requirement.[14] Work status is summarized by different work schedules, $\mathcal{L} \equiv \{0\}$ and $\mathcal{H} \equiv \{1, \cdots, 16\}$. The choice of work schedule is a simple logit model that motivates

$$N_j = N \cdot \frac{\exp(EV^j)}{\sum_{k=0}^{16} \exp(EV^k)},$$

where N is the total number of potential drivers, and EV^j represents the expected value

[14] For example, if a driver chooses to satisfy the high-performing requirement by working 10AM–12PM, then the driver is categorized as schedule 1. If a driver chooses to satisfy the high-performing requirement by working 11AM–1PM, then the driver is categorized as schedule 2, etc.

of choosing work schedule j.[15] Therefore, the total number of high-performing drivers is $N^H = \sum_{k=1}^{16} N_k$, and the total number of low-performing drivers is $N^L = N_0$.

In the second stage, drivers find the optimal solution to their dynamic discrete choice problem by choosing whether to work at each time t. Drivers observe the warm-up cost κ, sequence of hourly wages $\boldsymbol{W^\tau}$, and reservation values $\boldsymbol{O}$. Low-performing drivers, at each time t, compare the hourly wage plus the difference in expected future values to the value of their outside option. Then, the driver decides whether to work at time t. This is a dynamic problem, because if the driver chooses to work at time t and continues working at $t+1$, the driver would not need to pay an extra warm-up cost at $t+1$. Hence, the expected value for the future at time t is higher if the driver chooses to work than if the driver chooses not to work at time t. High-performing drivers have to work with probability 1 during committed hours. At any other time of day, high-performing drivers solve the same dynamic discrete choice problem by comparing the hourly wage plus the difference in expected future values to the value of their outside option, and decide whether to work at each time t.

Specifically, at the beginning of hour t, a driver receives a random draw from this wage distribution and another draw from the outside option:

$$\begin{aligned} U_{1t}^\tau &= \underbrace{W_t^\tau}_{\substack{\text{discriminatory} \\ \text{wage rate}}} + \sigma \cdot \epsilon_{1t}, \\ U_{0t}^\tau &= \underbrace{O_t(\boldsymbol{X_i})}_{\text{outside option value}} + \sigma \cdot \epsilon_{0t}, \end{aligned} \tag{1}$$

where O_t represents the reservation value from working on something else, and $\epsilon_{\cdot t}$ represents the driver preference that is Type-I extreme value distributed. There is a fixed warm-up cost $\kappa(\boldsymbol{X_i}) > 0$ to start working if the person took the outside option in the previous hour. This is to rationalize that drivers often drive in consecutive hours. For each set of parameters $\boldsymbol{\theta}(\boldsymbol{X_i}) \equiv (\boldsymbol{O}, \kappa, \sigma)$ and sequence of wages $\boldsymbol{W} \equiv \{\boldsymbol{W^H}, \boldsymbol{W^L}\}$, we can solve each driver's problem and obtain type-specific values.

Appendix C contains the details of the backward induction algorithm that we use to solve these dynamic discrete choice problems. The solutions generate conditional choice probabilities and then the number of drivers working in each hour by work schedule. Individual

[15] We use the total number of unique drivers in the past 30 days as the total number of potential drivers in our one-day model.

driver choices in turn generate the aggregate labor supply for each hour by driver type:

$$\begin{aligned} N_t^H &= \sum_{j=1}^{16} N_j \times \Pr(\text{work in hour t}|\text{work schedule j}), \\ N_t^L &= N_0 \times \Pr(\text{work in hour t}|\text{work schedule 0}), \end{aligned}$$

where the probabilities $\Pr(\cdot|\cdot)$ are the solutions to the above-mentioned DDC problems, i.e. the conditional choice probabilities. We denote the type-specific labor supply as

$$\begin{aligned} N_t^H &= \mathcal{N}_t^H(\boldsymbol{W^H}; \boldsymbol{\theta}) = \mathcal{N}_t^H(\boldsymbol{P}, \boldsymbol{s}; \boldsymbol{\theta}), \\ N_t^L &= \mathcal{N}_t^L(\boldsymbol{W^L}; \boldsymbol{\theta}) = \mathcal{N}_t^L(\boldsymbol{P}, \boldsymbol{s}; \boldsymbol{\theta}). \end{aligned}$$

4.2 Demand for Rides and the Platform's Problem

Customers only demand driver earning hours. The number of earning hours demanded is $D_t(P_t)$, where P_t is the hourly serving rate that the platform posts at hour t. For simplicity, we assume that demand for rides is downward sloping and iso-elastic:

$$Q_t = D_t(P_t) = \delta_t P_t^{-\epsilon},$$

where ϵ is the constant demand elasticity. The demand shifter δ_t includes daily weather indices, such as precipitation and temperature.

The platform takes demand shifter δ_t and demand elasticity ϵ as given, and chooses prices and assignments to balance the demand and supply of rides to maximize platform profit. The platform's choice of $(\boldsymbol{P}, \boldsymbol{s})$ maximizes its own payoff:

$$\begin{aligned} \max_{(\boldsymbol{P},\boldsymbol{s})} \quad & (1-\eta)\sum_t P_t D_t(P_t) \\ \text{s.t.} \quad & D_t(P_t)s_t \le \lambda_t^H \mathcal{N}_t^H(\boldsymbol{P}, \boldsymbol{s}; \boldsymbol{\theta}) \\ & D_t(P_t)(1-s_t) \le \lambda_t^L \mathcal{N}_t^L(\boldsymbol{P}, \boldsymbol{s}; \boldsymbol{\theta}). \end{aligned} \tag{2}$$

$D_t(P_t)$ is demand of rides measured in serving hours, and N_t^τ represents the total number of working hours (hours online) provided by the drivers. We have λ_t^τ as the technological constraint restricting the relationship between working hours and serving hours, and $\lambda_t^\tau \in (0, 1)$. For example, $\lambda_t^\tau = 0.5$ means that for every 15 minutes driving with a rider, a typical driver spends another 15 minutes on pick up, payment, etc. If $\lambda_t^\tau = 1$, there is no time spent on pick up. We have idle drivers waiting for trip requests when one of the two inequalities is unbinding. In our empirical analysis, we set the commission rate η equals to 20 percent.[16]

Given the choice of prices and assignments $(\boldsymbol{P}, \boldsymbol{s})$, the platform effectively determines the sequence of wages $(\boldsymbol{W^H}, \boldsymbol{W^L})$. Each high-performing and low-performing type expects to receive a wage rate:

$$\begin{aligned} W_t^H &= \eta P_t D_t(P_t) s_t \frac{1}{N_t^H}, \\ W_t^L &= \eta P_t D_t(P_t)(1 - s_t) \frac{1}{N_t^L}, \end{aligned} \tag{3}$$

where η is the revenue share that is received by the driver; s_t represents how the algorithm favors high-performing drivers, i.e. the proportion of orders assigned to high-performing drivers.

Remark: In general, $\boldsymbol{W^H} \neq \boldsymbol{W^L}$. The platform manages a large workforce of drivers who complete standardized and easily transferable tasks. However, each driver incurs a "warm-up" cost to start working, such as cleaning the car and filling up on gas. Because riders arrive continuously, having a driver work continuous hours could be more efficient than multiple drivers working in short segments. On the other hand, because rider arrival and drivers' outside option values fluctuate on an hourly basis, the platform cannot only hire longer working drivers, as that would lead to excess labor supply in off-peak hours. Therefore, the platform can mix driver types and use time-varying ride prices and wage differentials to balance demand and supply. We further elaborate on why the platform has incentive to discriminate among drivers in Appendix B.

5 Estimation

We first report demand estimation results in Section 5.1. Then, we estimate structural parameters in Section 5.2. Finally, we conduct a counterfactual experiment by setting the same hourly wage for both high-performing and low-performing drivers to evaluate market outcomes in Section 5.3.

5.1 Demand Estimation

We first estimate the demand for earning time in the city we study. We consider each hour to be a different market and aggregate our data to the day-hour level. We obtain the

[16] According to Platform X's IPO document, the nationwide average commission rate is 20.9 percent. In the survey we conducted, most drivers suggest that the commission rate is about 20 percent. Therefore, we use $\eta = 20$ percent in our empirical analysis.

logarithm of total earning time (Q_t) and the logarithm of average hourly wage (P_t). Demand parameters are estimated through:

$$\log Q_t = \log \delta_h - \epsilon \log P_t + e_h. \tag{4}$$

Our demand estimation suffers from classic supply-demand endogeneity. When there is a higher demand shock in the market, the platform may set a higher price. Therefore, our OLS estimates may be biased. Similar to Kalouptsidi (2014), we use the number of cars in competing ride-sharing companies in any day as our supply-side instrumental variable. This instrument is not correlated to hour-level demand shocks. Second, the number of cars each company operates in a day are correlated to the hourly wage. Therefore, it is a valid instrument.

Table 8: Demand Estimation

Dependent Variables	ln(Earning Hour)			
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	IV
ln(Hourly Wage)	-5.151***	-5.158***	-0.767***	-1.186**
	(0.0743)	(0.0737)	(0.152)	(0.553)
Rain		-0.0020	-0.0005	-0.0006
		(0.002)	(0.0007)	(0.0007)
Temperature		0.0127***	0.0094***	0.0098***
		(0.0033)	(0.0011)	(0.0012)
Constant	32.12***	32.06***	10.62***	12.69***
	(0.350)	(0.348)	(0.752)	(2.736)
Hour FE	N	N	Y	Y
Day of Week FE	N	N	Y	Y
Observations	744	744	744	744
R-squared	0.866	0.869	0.988	0.988

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8 reports demand estimates for the city of study. Column (1) reports the estimates without fixed effects. Column (2) reports estimates with weather as a demand shifter. Column (3) further includes day and hour fixed effects. Column (4) reports our IV estimates. All estimates of demand elasticities are negatively significant. After controlling for hour fixed effects and day of the week fixed effects, column (3) reports a demand elasticity of -0.767. Our IV estimates in column (4) show that the demand elasticity is -1.186, indicating a downward slope. When hourly price paid by customers increase by 1 percent, their total driving time decreases by 1.2 percent. Our estimated demand elasticity at -1.186 is comparable to the values estimated in the literature. Frechette et al. (2019) estimates an elasticity of -1.225 for New York City's taxi market. Cohen et al. (2016) relies on Uber's surge pricing algorithm and estimates a smaller price elasticity for UberX (between -0.4 and -0.6).

5.2 Estimation of Structural Parameters

We make use of observed conditional choice probabilities to estimate reservation values and parameters in drivers' choices in Section 4.1. We first derive the conditional choice probability of working for each type of driver at each time t. Given the utility of working for each hour in equation 1, the conditional probability of working for low-performing drivers is

$$\begin{aligned} P_L(a_t = 1|a_{t-1} = 1) &= \frac{\exp(W_t^L + \beta EV_{1t+1}^L)}{\exp(W_t^L + \beta EV_{1t+1}^L) + \exp(O_t + \beta EV_{0t+1}^L)}, \\ P_L(a_t = 1|a_{t-1} = 0) &= \frac{\exp(W_t^L - \kappa_t + \beta EV_{1t+1}^L)}{\exp(W_t^L - \kappa_t + \beta EV_{1t+1}^L) + \exp(O_t + \beta EV_{0t+1}^L)}, \end{aligned} \tag{5}$$

where $t \in [2, T-1]$. Throughout our model, EV's subscript 1 represents $a_{t-1} = 1$. In this case, EV_{1t}^L represents the expected value of low-performing drivers at time t if $a_{t-1} = 1$. Similarly, we derive the conditional probability for high-performing drivers for each schedule. At any time t, the conditional probability for high-performing drivers is

$$\begin{aligned} P_H(a_t = 1|a_{t-1} = 0) &= \sum_{j=1}^{16} \widetilde{P}_j \cdot P_j(a_t = 1|a_{t-1} = 0), \\ P_H(a_t = 1|a_{t-1} = 1) &= \sum_{j=1}^{16} \widetilde{P}_j \cdot P_j(a_t = 1|a_{t-1} = 1), \end{aligned} \tag{6}$$

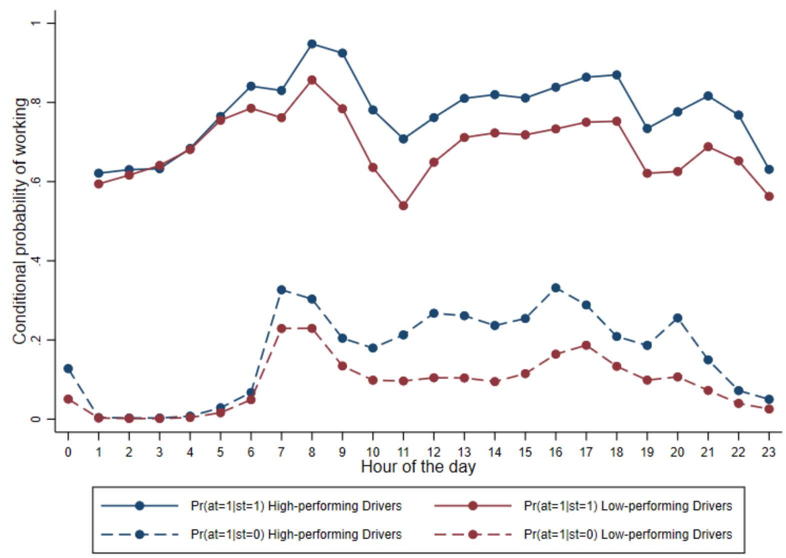
where $\widetilde{P}_j$ is the probability of choosing schedule j within high-performing drivers:

$$\widetilde{P}_j = \frac{\exp(EV^j)}{\sum_{k=1}^{16} \exp(EV^k)}.$$

Therefore, equations 5 and 6 show the model-predicted CCPs as a function of observed wage sequence $\{\boldsymbol{W^H}, \boldsymbol{W^L}\}$ and parameters $\boldsymbol{\theta}$. In Appendix C.3, we show the detailed derivation of $P_\tau(a_t = 1|a_{t-1} = 0)$ and $P_\tau(a_t = 1|a_{t-1} = 1)$ for $\tau \in \{L, H\}$ and $t \in [1, 24]$.

Then, from the data, we obtain the observed CCPs of high-performing and low-performing drivers at each time t. Figure 2 shows the CCPs obtained from the data. First, we can see that the conditional probability of working when the driver has already started working ($a_{t-1} = 1$) is much higher than the conditional probability of working when the driver was not working ($a_{t-1} = 0$) for both types of drivers. This indicates that there exists a significant warm-up cost. Second, the observed conditional probability of working for high-performing drivers is higher than low-performing drivers at all times of the day. However, the difference is smaller at night between 11PM to 6AM the next day. Third, The conditional probability of working is higher during morning peak hours from 7AM–10AM and afternoon peak hours from 4PM–7PM. This indicates that either drivers are responding to surge pricing, or drivers' reservation values vary across different times of the day.

Figure 2: Conditional Probability of Working (from data)



In our model, there are 26 parameters to be estimated. The warm-up cost is κ, the scale parameter σ, and the hourly reservation value is $O_t, \forall t \in [1, 24]$. The estimation of the dynamic parameters $\boldsymbol{\theta}_0$ is implemented according to the following procedures. First, denote the actual CCPs obtained from the data as $\boldsymbol{P^d_\tau}(\cdot)$, which is a 48-by-1 vector. Second, for a given set of parameters $\boldsymbol{\theta}$, the model-simulated CCPs are $\boldsymbol{P^S_\tau}(\cdot)$. The MSM estimate $\widehat{\boldsymbol{\theta}}$ minimizes the weighted distance between the data moments and the simulated moments:

$$L(\boldsymbol{\theta}) = \min_{\boldsymbol{\theta}} \quad [\boldsymbol{P^d_\tau} - \boldsymbol{P^S_\tau}(\boldsymbol{\theta})]' W [\boldsymbol{P^d_\tau} - \boldsymbol{P^S_\tau}(\boldsymbol{\theta})],$$

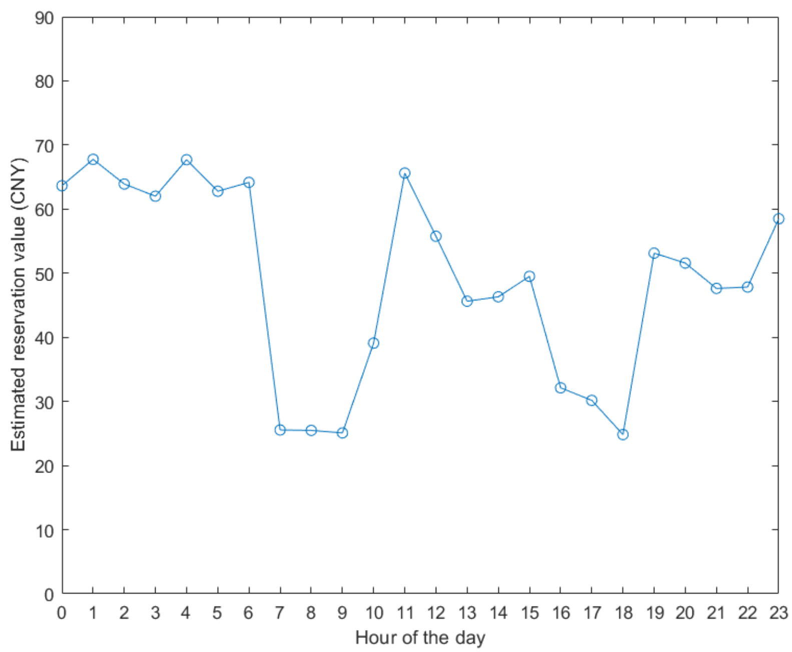
where W is a positive definite matrix. Figure C.1 shows the model's goodness of fit. Overall, the simulated values fit the observed CCPs well.

Figure 3 shows the estimated reservation values for a driver. The average estimated reservation value is 49 CNY ($7 USD).[17] The reservation value is lowest during morning peak hours, around 25.4 CNY ($3.7 USD), and highest at late night, around 64.5 CNY ($9.4

[17]The conversion rate between CNY and USD was 1:0.1459 for December 2018.

USD). For context, the minimum hourly wage in our city of study was 18.5 CNY ($2.7 USD) in 2018. From the estimated results, we can see that drivers have higher reservations values during off-peak hours between 10AM–2PM and 7PM–5AM. This helps explain why the ride-hailing platform wants to implement algorithmic discrimination to incentivize drivers to work more during off-peak hours. The estimated warm-up cost is 124.4 CNY ($18 USD), which is around 2.5 times the average hourly reservation value. The large warm-up cost helps explain why drivers usually choose to drive consecutive hours.

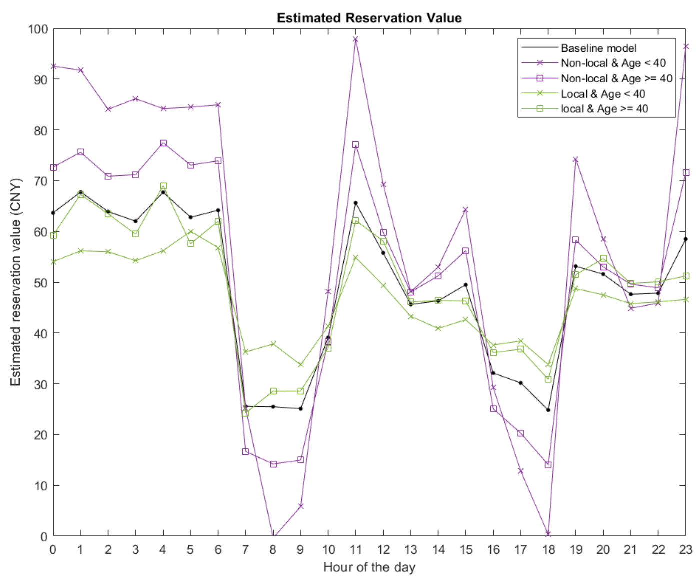
Figure 3: Estimated Reservation Values



In our benchmark model, we assume that the scale parameter σ and warm-up cost κ are constant, and that reservation values $\{O_t\}$ only vary across time. Next, we aim to capture the heterogeneity of drivers and allow for correlation between structural parameters and observed driver characteristics, $\boldsymbol{\theta}(\boldsymbol{X_i}) \equiv (\boldsymbol{O}(\boldsymbol{X_i}), \kappa(\boldsymbol{X_i}), \sigma(\boldsymbol{X_i}))$. We focus on two types of characteristics: whether the driver is a resident of the province of study, and whether the driver is less than 40 years of age. For each group of drivers, we obtain the observed CCPs from the data. Then, we use the same method to estimate structural parameters for each group of drivers. Figure 4 shows the estimated reservation values for each type of driver.

The average reservation value for non-local drivers is 57.6 CNY and 51.3 CNY for drivers younger than 40 and older than 40, respectively. In comparison, the corresponding estimated values for local drivers are 46.6 CNY and 49 CNY. From Figure 4, we can see that non-local drivers have higher reservation values than local drivers during off-peak hours between 10AM–4PM and 7PM–6AM. During morning peak hours from 7AM–10AM and afternoon peak hours from 4PM–7PM, non-local drivers have lower reservation values than local drivers. Among locals, drivers younger than 40 have lower reservation values at night than their older counterparts. Regarding the estimated warm-up costs, the value for non-local young drivers is 273.1 CNY. For non-local middle-aged drivers, the value is lower at 179.8 CNY. In comparison, the warm-up cost for local young drivers is only 69.5 CNY in contrast to 108.1 CNY for local middle-aged drivers. In general, the warm-up cost is much lower for local drivers. One explanation for the difference in warm-up cost is that non-locals live in the outer areas of the city. Table 2 shows that non-local drivers are more likely to be living in the two suburb areas. In comparison, local middle-aged drivers are the least likely to be living in the suburb areas.

Figure 4: Estimated Reservation Values with Driver Heterogeneity



5.3 Counterfactual Experiment

In the counterfactual analysis, we study changes in welfare and pricing if ride-hailing platforms are prevented from discriminating based on work schedule. In this case, orders would be randomly assigned to nearby active workers, and all drivers would have an equal chance to receive better orders. Effectively, the hourly wage each driver earns will become

$$\widetilde{W}_t = \frac{\eta P_t D_t(P_t)}{N_t}.$$

Given the new sequence of hourly wages, drivers solve the unconstrained dynamic discrete choice problem at each hour t:

$$
\begin{aligned}
U_{1t} &= \underbrace{\widetilde{W}_t}_{\substack{\text{non-discriminatory} \\ \text{wage rate}}} + \sigma \cdot \epsilon_{1t}, \\
U_{0t} &= \underbrace{O_t(\boldsymbol{X_i})}_{\text{outside option value}} + \sigma \cdot \epsilon_{0t},
\end{aligned}
$$

where we have replaced the discriminatory wage rate W_t^τ by the "fair" rate $\widetilde{W}_t$.

In the counterfactual analysis, the platform can only choose the hourly prices $\boldsymbol{P}$ to maximize its payoff:

$$
\begin{aligned}
\max_{\vec{P}} \quad & (1-\eta)\sum_t P_t D_t(P_t) \\
\text{s.t.} \quad & D_t(P_t) \leq \widetilde{\lambda}_t \mathcal{N}_t(\boldsymbol{W_t}; \boldsymbol{\theta})
\end{aligned}
\tag{7}
$$

Note that the two feasibility constraints in equation 2 becomes one, because all drivers have the same likelihood of receiving a task, and $\widetilde{\lambda}_t$ is the technology restriction without algorithmic discrimination.[18] Because the matching efficiency will be affected by how the platform allocates orders to drivers, we conduct the counterfactual analysis by imposing a lower bound and upper bound for the technology restriction $\widetilde{\lambda}_t$:[19]

$$
\begin{aligned}
\text{Lower bound:} \quad & \widetilde{\lambda}_t = \lambda_t^L, \\
\text{Upper bound:} \quad & \widetilde{\lambda}_t = \lambda_t^H.
\end{aligned}
$$

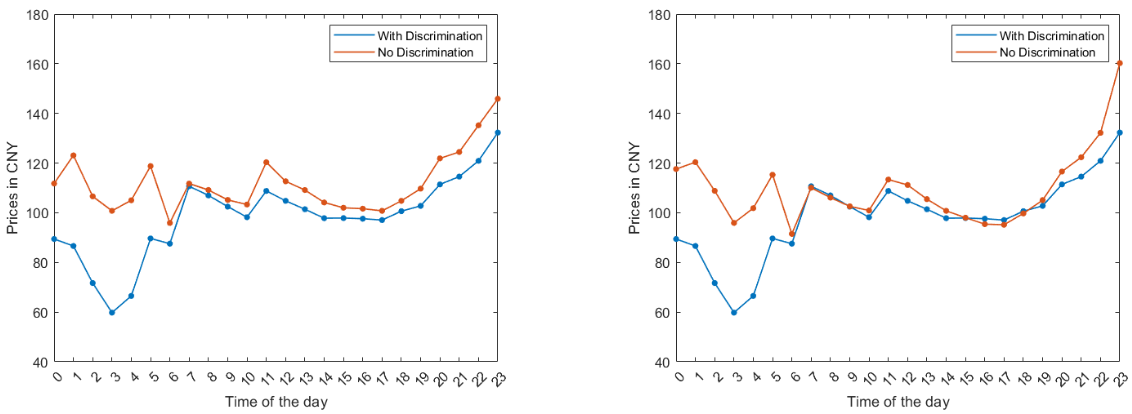
Using the estimated parameters, we solve for the new equilibrium outcome when platform is banned from algorithmic discrimination. Figure 5 shows the equilibrium prices with different assumptions on technological restriction. When $\widetilde{\lambda}_t = \lambda_t^L$, the matching efficiency without discrimination is assumed to be at its lower bound. Panel (a) shows that in this case, the prices will be higher at all hours of the day when discrimination is banned. The reason is that the platform now needs to offer drivers higher wages to incentivize them to work. The increment is larger during off-peak hours when the reservation values of drivers are high. Panel (b) shows the change in prices if there is improvement in matching efficiency without discrimination. In this case, pick-up time is minimized. Results show that prices will increase during off-peak hours, especially at night time. However, if the matching efficiency is set to the upper bound, hourly prices would not change during morning peak hours

[18] Under "fair" pay, there is only one group so that $s = 1$.

[19] We obtain technology restriction for high-performing and low-performing drivers (λ_t^H and λ_t^L) from the data.

and would slightly decrease in afternoon peak hours. This is because reservation values are low during peak hours. Drivers who used to be high-performing and work off-peak hours will now re-optimize their work schedules and shift to working more during peak hours. In conclusion, if algorithmic discrimination is banned and matching becomes relatively efficient, consumers would need to pay higher prices during off-peak hours and slightly lower prices during peak hours.

Figure 5: Hourly Price in Equilibrium



(a) $\widetilde{\lambda}_t$ equals lower bound

(b) $\widetilde{\lambda}_t$ equals upper bound

Next, we study how welfare changes if the platform can no longer discriminate based on work schedule. We calculate consumer welfare as $\sum_t \int_{P_t}^{\infty} \delta_t x^{-\epsilon} dx$ and driver surplus as

$$N \cdot EV_0 = N \cdot \left[\ln \Big(\exp((W_1 - \kappa + \beta EV_{12})/\sigma) + \exp((O_1 + \beta EV_{02})/\sigma) \Big) + \gamma \right],$$

where EV_0 represents the expected value before choosing the work schedule type.[20]

Table 9 shows the changes in platform revenue, consumer surplus, and driver surplus when the platform is banned from algorithmic discrimination. Under "fair" pay, consumer surplus, driver surplus, and platform revenue account for 56 percent, 42 percent, and 2 percent of total surplus, respectively. Columns 1 and 2 display the results from the benchmark model where reservation values vary only across time and not across driver types. The platform's total revenue decreases by 1.3 percent when matching efficiency is set to the lower bound and decreases by 0.7 percent if matching is efficient without discrimination. In comparison, drivers benefit by having a very large (5.1 percent) increase in surplus. The intuition is as follows. First, without algorithmic discrimination, platforms need to pay drivers higher wages to incentivize them to work, hence charging higher fees to consumers. Second, drivers

[20]Note that consumer surplus $\sum_t \int_{P_t}^{\infty} \delta_t x^{-\epsilon} dx = \sum_t \frac{\delta_t}{\epsilon - 1} (P_t)^{1-\epsilon} = \frac{1}{(\epsilon-1)(1-\eta)} \times$ platform revenue.

have more flexibility over choosing when to work and will not be punished by making such choices. Hence, driver surplus experiences a large increase when discrimination is banned. As a result, although consumer surplus decreases slightly if algorithmic discrimination is banned, the total surplus will increase.

Table 9: Changes in Platform Revenue, Consumer Surplus, and Driver Surplus

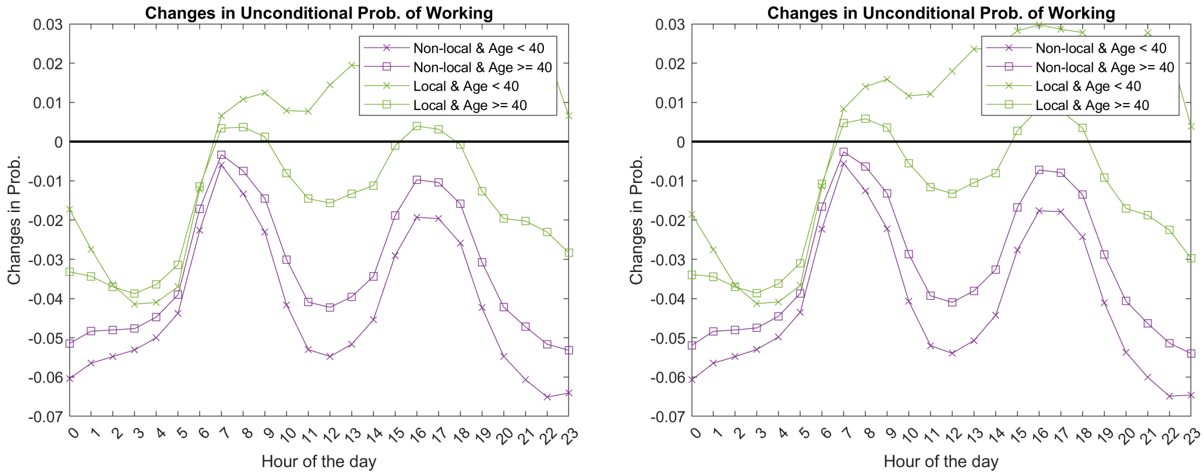
	(1) Benchmark		(2) Heterogeneous Drivers	
	Lower bound	Upper bound	Lower bound	Upper bound
Platform Revenue	-1.28%	-0.70%	-1.37%	-0.78%
Consumer Surplus	-1.28%	-0.70%	-1.37%	-0.78%
Driver Surplus	5.09%	5.13%		
Non-local × Young			5.97%	5.99%
Non-local × Middle-Aged			5.73%	5.77%
Local × Young			4.72%	4.77%
Local × Middle-Aged			4.75%	4.79%
Total Surplus	1.33%	1.74%	1.71%	2.10%

Columns 3 and 4 of Table 9 shows the welfare changes when we allow for reservation values to vary across different types of drivers. In aggregate, results show a similar degree of welfare change for consumers. Consumer surplus will decrease by 0.78 percent if matching efficiency is high, and decrease by 1.37 percent if matching efficiency is low. For drivers, change in driver surplus is similar for different age groups. However, non-local drivers will have a larger increase in driver surplus. The reason is because non-local drivers were more likely to select into the high-performing type in the face of algorithmic discrimination. Therefore, with algorithmic discrimination banned, high-performing drivers benefit more from the flexibility of choosing their own working hours.

To unpack the mechanism of how labor supply decisions are altered with or without discrimination, we show changes in the unconditional probability of driving for each type of driver in Figure 6. There are two effects that alter labor supply decision once algorithmic discrimination is banned: the wage effect, where hourly price is higher without discrimination, and the flexibility effect, where drivers can choose working hours freely without facing any punishment.

From the results, we can see that non-local drivers decrease their probability of working in every hour of the day, regardless of the matching efficiency. The decrease is larger during off-peak hours between 10AM–2PM and 8PM–4AM the next day. This is because non-local drivers were more likely to be high-performing drivers given their high warm-up cost and distribution of reservation values. Without extra incentives provided by the discriminatory

Figure 6: Changes in Unconditional Probability of Working



(a) Technological restriction $\widetilde{\lambda}_t$ equals lower bound

(b) $\widetilde{\lambda}_t$ equals upper bound

algorithm, non-local drivers will largely reduce their labor supply during off-peak hours. Local middle-aged drivers will also decrease their probability of working during off-peak hours. However, local middle-aged drivers increase their probability of working during peak hours. This is because the estimated reservation values are low during peak hours (Figure 4). When there is greater flexibility of choosing working hours, local middle-aged drivers will increase their probability of working during off-peak hours. Local young drivers will increase their probability of working for most times of the day from 6AM to 11PM. Given that local young drivers were less likely to be high-performing drivers, they thus benefit more from the wage effect when discrimination is banned.

Remark: People argue that the growth of the "gig" economy could help narrow the gender wage gap in the economy. For example, Hyperwallet (2017) documents that 86 percent of female gig workers believe gig work "offers the opportunity to make equal pay to their male counterparts". However, algorithmic discrimination over certain work schedules may impose extra job-flexibility penalties and indirectly discriminate against female drivers. Chen et al. (2019) shows that among the 1.87 million drivers in the U.S., about 512, 000 (27.3 percent) are female. In our study period, female drivers only account for 2.7 percent of all active drivers. Among the active female drivers, 47 percent are high-performing and 53 percent are low-performing. By comparison, 60 percent of the entire sample of drivers are high-performing and 40 percent are low-performing. This suggests that female drivers value the flexibility of work hours more than the average driver, and hence are less likely to commit to working long and consecutive hours. Given such a small number of female drivers, it is not plausible to estimate separate reservation values based on gender. Instead, in our benchmark model, we calculate the change in driver surplus for high-performing and low-performing drivers separately. Then, using the observed proportion of high-performing and low-performing types among female drivers, we can calculate the change in welfare for female drivers under the “fair” rate. Results indicate that female drivers will experience a 4.75 percent increase in surplus under the “fair” rate.

6 Conclusion

We provide the first empirical study on algorithmic wage-setting and document wage differentials arising from different hourly work schedules in the ride-hailing market. Using rich data from China’s leading ride-hailing company, we first show that low-performing drivers receive lower-paying and fewer trips than high-performing drivers. Next, we construct and estimate a structural model with elastic rider demand, dynamic labor supply, platform pricing and wage-setting. Finally, our counterfactual analysis demonstrates that “fair” pay leads to gains in driver wages at the cost of higher prices and consumer welfare. Overall, total welfare increases.

References

Aaronson, Daniel and Eric French, "The effect of part-time work on wages: Evidence from the social security rules," *Journal of Labor Economics*, 2004, *22* (2), 329–252.

Altonji, Joseph G and Rebecca M Blank, "Race and gender in the labor market," *Handbook of Labor Economics*, 1999, *3*, 3143–3259.

Blau, Francine D and Lawrence M Kahn, "The gender wage gap: Extent, trends, and explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler, "Labor supply of New York City cabdrivers: One day at a time," *The Quarterly Journal of Economics*, 1997, *112* (2), 407–441.

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

Chen, Kuan-Ming, Claire Ding, John A List, and Magne Mogstad, "Reservation wages and workers' valuation of job flexibility: Evidence from a natural field experiment," Technical Report, National Bureau of Economic Research 2020.

Chen, M Keith, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen, "The value of flexible work: Evidence from Uber drivers," *Journal of political economy*, 2019, *127* (6), 2735–2794.

Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe, "Using big data to estimate consumer surplus: The case of uber," Technical Report, National Bureau of Economic Research 2016.

Cook, Cody, Rebecca Diamond, Jonathan V Hall, John A List, and Paul Oyer, "The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers," *The Review of Economic Studies*, 2021, *88* (5), 2210–2238.

Farber, Henry S, "Why you can't find a taxi in the rain and other labor supply lessons from cab drivers," *The Quarterly Journal of Economics*, 2015, *130* (4), 1975–2026.

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.

Gaineddenova, Renata, "Pricing and Efficiency in a Decentralized Ride-Hailing Platform," 2021.

Haggag, Kareem, Brian McManus, and Giovanni Paci, "Learning by driving: Productivity improvements by new york city taxi drivers," *American Economic Journal: Applied Economics*, 2017, *9* (1), 70–95.

Hotz, V Joseph and Robert A Miller, "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 1993, *60* (3), 497–529.

Hyperwallet, "The Future of Gig Work is Female," 2017.

Kalouptsidi, Myrto, "Time to build and fluctuations in bulk shipping," *American Economic Review*, 2014, *104* (2), 564–608.

Liu, Tracy, Zhixi Wan, and Chenyu Yang, "The efficiency of a dynamic decentralized two-sided matching market," *Available at SSRN 3339394*, 2019.

Rysman, Marc, "Competition between networks: A study of the market for yellow pages," *The Review of Economic Studies*, 2004, *71* (2), 483–512.

_ , "The economics of two-sided markets," *Journal of Economic Perspectives*, 2009, *23* (3), 125–43.

A Additional Regression Tables on Wage Differentials

Table A.1: Ride-Prices by Different Schedules

Dependent Variables	Hourly Rate			
	(1)	(2)	(3)	(4)
# Work Hours in a month	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
% Peak-AM Hours	-15.895***			
	(0.185)			
% Midday Hours		0.753***		
		(0.148)		
% Peak-PM Hours			-14.407***	
			(0.284)	
% Night Hours				11.131***
				(0.119)
Constant	55.882***	54.823***	56.358***	51.137***
	(0.126)	(0.127)	(0.129)	(0.132)
Day FE	Y	Y	Y	Y
Hour FE	Y	Y	Y	Y
Origin FE	Y	Y	Y	Y
Destination FE	Y	Y	Y	Y
Observations	4,182,318	4,182,318	4,182,318	4,182,318
R-squared	0.042	0.040	0.041	0.042

Standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table A.2: Distinguish Sources of Discrimination

Dependent Variables	Hourly Rate				Coef × Std
	(1)	(2)	(3)	(4)	(5)
# Work Hours in a month	0.003***	0.003***	0.004***	0.003***	0.818
	(0.000)	(0.000)	(0.000)	(0.000)	
% Off-peak Hours		18.724***		18.216***	14.042
		(0.170)		(0.171)	
Nonlocal			-0.332***	-0.138***	-0.097
			(0.036)	(0.027)	
Age			-0.061***	-0.045***	-1.689
			(0.001)	(0.001)	
Female			-0.677***	-0.431***	-0.009
			(0.081)	(0.081)	
Constant	54.918***	39.201***	57.419***	41.342***	
	(0.126)	(0.190)	(0.141)	(0.203)	
Day FE	Y	Y	Y	Y	
Hour FE	Y	Y	Y	Y	
Origin FE	Y	Y	Y	Y	
Destination FE	Y	Y	Y	Y	
Observations	4,182,331	4,182,331	4,182,318	4,182,318	
R-squared	0.040	0.043	0.041	0.043	

Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

B Incentives to Wage Discriminate

Take $D_h(P) = \delta_h D(P)$ as an example. We discuss why the platform wants to implement wage discrimination. When there is imbalance between demand δ_h and supply O_h, the firm can change the price to clear the market. However, it might be more cost efficient to encourage drivers to become high-performing rather than recruit more low-performing drivers. One possibility is that fixed cost κ to start working is high and extending one's work hours is cheaper. Everything else equal, the platform can get an existing driver to complete an additional order with a small incentive. Enticing a driver to start working would be more costly. This could also explain China's infamous "9-9-6" work culture (i.e., from 9am to 9pm, 6 days a week) which, assuming overtime is paid, some firms prefer to hiring more workers to work flexible schedules.

Another reason is that, in the face of uncertainties, high-performing workers are more likely to work. We show this in a two-hour model. There are two decisions: work or not work during the day and evening. There is a representative consumer with willingness to pay u_1 during the day and u_2 during the night. The driver's utility from working is specified as in equation 1. Assume that the wage for the day and night are the same ($w_1 = w_2 = w$). Also assume that the reservation value is the same for the day and night, and normalize the reservation value to 0 ($o_1 = o_2 = 0$). The low-performing driver can choose whether to work in each period, but the high-performing driver needs to pre-commit to work in both periods if he/she chooses to work. A low-performing driver's probability of working in each period is

$$Pr(w + \epsilon_1 > o + \epsilon_0) = \frac{\exp(w)}{\exp(w) + \exp(0)}.$$

A high-performing driver's probability of working is

$$Pr(2w + \epsilon_1 > 2o + \epsilon_0) = \frac{\exp(2w)}{\exp(2w) + \exp(0)}.$$

Therefore, the expected revenue for the platform is

$$\begin{aligned} R_{part} &= \left[\frac{u\exp(w)}{\exp(w) + \exp(0)} - \frac{w\exp(w)}{\exp(w) + \exp(0)}\right] \times 2, \\ R_{full} &= \frac{2u\exp(2w)}{\exp(2w) + \exp(0)} - \frac{2w\exp(2w)}{\exp(2w) + \exp(0)}. \end{aligned}$$

We can see that hiring a high-performing driver yields a higher expected profit for the platform. The intuition is that for a driver to work both periods, we need to have the following:

- low-performing: $w_1 + \epsilon_1 > o_1 + \epsilon_0$ and $w_2 + \epsilon_1 > o_2 + \epsilon_0$

- high-performing: $w_1 + w_2 + \epsilon_1 > o_1 + o_2 + \epsilon_0$

We can see that for low-performing drivers to work both periods, the expected payoff needs to exceed the outside option in both periods. But for high-performing drivers to work, we only need the summation of the expected wages to be larger than the summation of the outside option value. It is easier to satisfy the inequality of the high-performing drivers.

The next question is why the platform has incentive to employ a mixture of high-performing and low-performing drivers. We conduct simulations to demonstrate that, in our full model, the platform indeed benefits from having both high-performing and low-performing drivers.

The platform solves the maximization problem:

$$\max_{\vec{P},\overrightarrow{w_f},\overrightarrow{w_c}} \sum_h P_h D_h(P_h) - w_{fh}N_{fh}(w_{fh}) - w_{ch}N_{ch}(w_{ch}).$$

For any given prices $\vec{P}^*$, the optimization problem can be re-written as

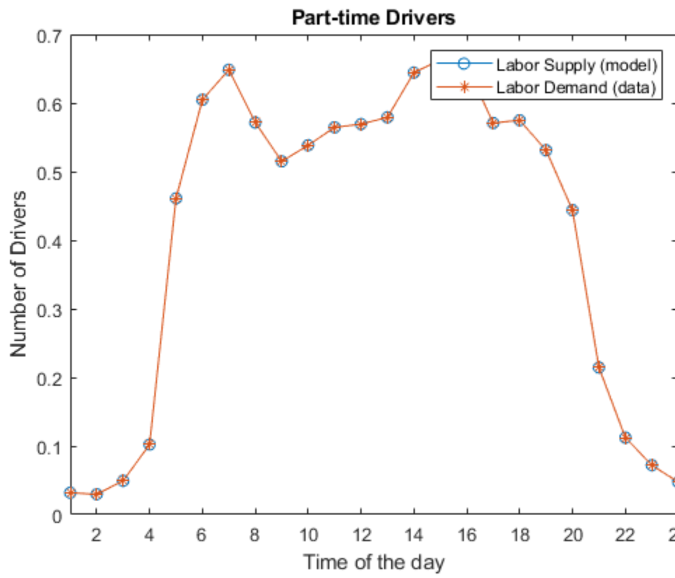
$$\begin{aligned} &\min_{\overrightarrow{w_f},\overrightarrow{w_c}} \sum_h w_{fh}N_{fh}(w_{fh}) + w_{ch}N_{ch}(w_{ch}) \\ &\quad s.t. \quad \forall h, N_{fh} + N_{ch} \geq D_h(P_h^*). \end{aligned}$$

The intuition of the cost minimization problem is that, given the prices, the total demand is determined. Then, the platform needs to find the right mixture of high-performing and low-performing drivers to 1) fulfil the total demand at each hour h and 2) minimize the total wage expenditure.

If the platform only uses low-performing drivers, the platform can find a vector of wages $\overrightarrow{w_c}$ that perfectly fit the labor demand $D_h(P_h^*)$ in any hour h. Figure B.1 shows the results of the platform's cost minimization problem if the platform only uses low-performing drivers. We can see that there is no excess supply or demand of drivers at the optimal wage rates.

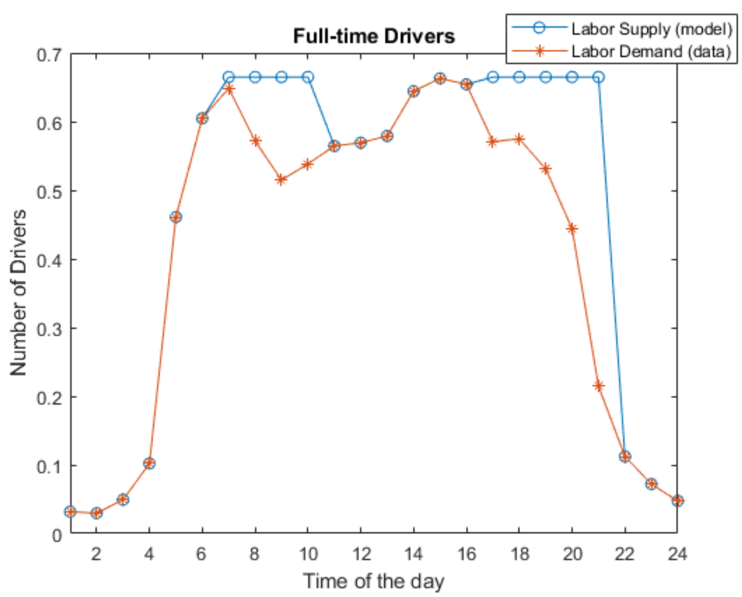
Figure B.2 shows the results of platform's cost minimization problem if the platform only uses high-performing drivers. First, the simulation results show that for the given amount of labor demand $D_h(P_h*)$ (as shown in the figure), using only high-performing drivers yields a lower total wage expenditure than using only low-performing drivers. Second, there is excess labor supply if only using high-performing drivers. The reason is that, in this example, the platform needs a sufficient number of drivers to satisfy high demand at 15 o'clock. However, because the high-performing drivers are required to work from 7-10 and 17-21, the number of drivers working at those other time periods will be no less than that at 15 o'clock. Thus, given

Figure B.1: Labor supply if only use low-performing drivers



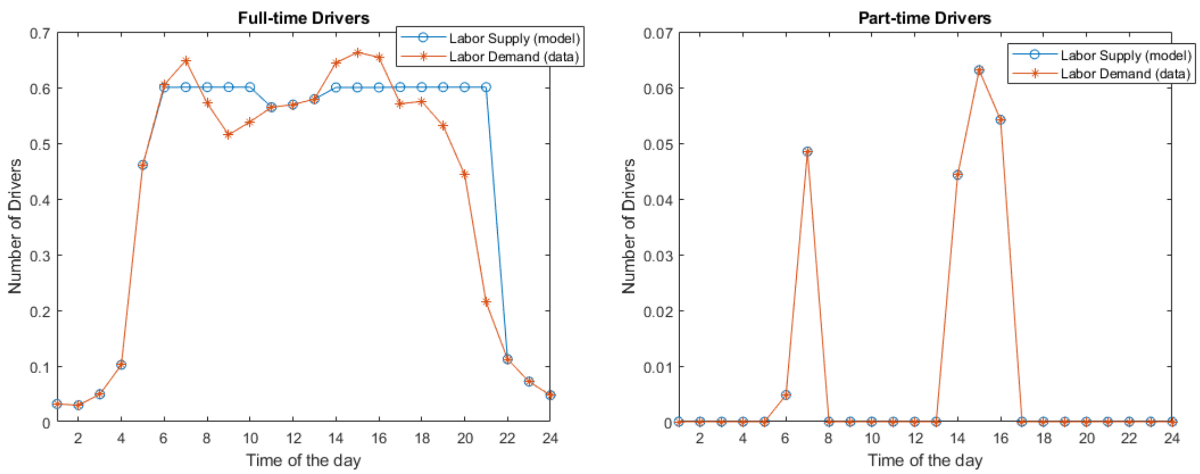
the labor demand curve of this example, there will be excessive labor supply in the required working hours from 7-10 and 17-21. Third, the platform's cost minimization problem has no unique solution if the platform only hires high-performing drivers. This is because, for the required working hours, only the summation of wages in those hours matter, assuming the discount factor is 1. Therefore, wage rates at those hours can be any combination, expect that the summation needs to be fixed.

Figure B.2: Labor supply if only use high-performing drivers



The above results show that hiring only high-performing drivers is more profitable for the platform than hiring only low-performing drivers. However, this is not the optimal solution that minimizes the platform's aggregate wage expenditure. The intuition is that, because hiring only high-performing drivers would lead to excess labor supply, the platform can lower the wage rate to mitigate the excess labor supply. The problem is that, for certain hours like 15 o'clock in the example, there would not be enough drivers to meet the demand. To solve this problem, the platform can use a small fraction of low-performing drivers to serve the excessive demand. Figure B.3 shows the labor supply if the platform uses a mixture of high-performing and low-performing drivers.

Figure B.3: Labor supply if using a mixture of high-performing and low-performing drivers



(a) Labor supply of high-performing drivers (b) Labor supply of low-performing drivers

C Drivers' Finite-Horizon Dynamic Problem

This appendix describes in detail drivers' finite-horizon dynamic choices and our CCP estimator. At each hour t, the utility of working and not working are specified as

$$U_{1t}^{\tau} = W_t^{\tau} + \sigma \cdot \epsilon_{1t},$$
$$U_{0t}^{\tau} = O_t + \sigma \cdot \epsilon_{0t}.$$

Drivers observe random shocks ϵ first, and then decide whether to work or not.

C.1 Low-Performing Drivers

At the final period T,

$$V_T^L = \begin{cases} W_T^L + \sigma \cdot \epsilon_{1T} & \text{if } a_T = 1 \ \& \ a_{T-1} = 1, \\ W_T^L - \kappa + \sigma \cdot \epsilon_{1T} & \text{if } a_T = 1 \ \& \ a_{T-1} = 0, \\ O_T + \sigma \cdot \epsilon_{0T} & \text{if } a_T = 0. \end{cases}$$

So, the expected utility of the last period T is given by

$$EV_{1T}^L = \ln\Big(\exp(W_T^L/\sigma) + \exp(O_T/\sigma)\Big) + \gamma,$$
$$EV_{0T}^L = \ln\Big(\exp((W_T^L - \kappa)/\sigma) + \exp(O_T/\sigma)\Big) + \gamma.$$

Throughout our model, EV's subscript 1 represents $a_{t-1} = 1$. In this case, EV_{1T}^L represents the expected value of a low-performing driver at time T if $a_{T-1} = 1$.

At any time $t \in [T-1, 2]$,

$$V_t^L = \begin{cases} W_t^L + \sigma \cdot \epsilon_{1t} + \beta EV_{1t+1}^L & \text{if } a_t = 1 \ \& \ a_{t-1} = 1, \\ W_t^L - \kappa + \sigma \cdot \epsilon_{1t} + \beta EV_{1t+1}^L & \text{if } a_t = 1 \ \& \ a_{t-1} = 0, \\ O_t + \sigma \cdot \epsilon_{0t} + \beta EV_{0t+1}^L & \text{if } a_t = 0. \end{cases}$$

So, the expected utility of period t is given by

$$EV_{1t}^L = \ln\Big(\exp((W_t^L + \beta EV_{1t+1}^L)/\sigma) + \exp((O_t + \beta EV_{0t+1}^L)/\sigma)\Big) + \gamma,$$
$$EV_{0t}^L = \ln\Big(\exp((W_t^L - \kappa + \beta EV_{1t+1}^L)/\sigma) + \exp((O_t + \beta EV_{0t+1}^L)/\sigma)\Big) + \gamma.$$

At the first period, $t = 1$,

$$V_1^L = \begin{cases} W_1^L - \kappa + \sigma \cdot \epsilon_{11} + \beta EV_{12}^L & \text{if } a_1 = 1, \\ O_1 + \sigma \cdot \epsilon_{01} + \beta EV_{02}^L & \text{if } a_1 = 0. \end{cases}$$

The expected value of being a low-performing driver is then derived as

$$EV^L = \ln\Big(\exp((W_1^L - \kappa + \beta EV_{12}^L)/\sigma) + \exp((O_1 + \beta EV_{02}^L)/\sigma)\Big) + \gamma. \tag{C.1}$$

C.2 High-Performing Drivers

High-performing drivers are required to work at T_0, and need to work consecutively for at least 2 hours. T_0 can be any hour between 10AM–2PM and 7PM–5AM. There are 16 possible work schedules to choose from. For schedule $j \in \{1, \cdots, 16\}$ with committed working hours $[T_0, T_0 + 1]$:

If $T_0 + 2 < T$. Then at last period T,

$$V_T^j = \begin{cases} W_T^H + \sigma \cdot \epsilon_{1T} & \text{if } a_T = 1 \text{ \& } a_{T-1} = 1, \\ W_T^H - \kappa + \sigma \cdot \epsilon_{1T} & \text{if } a_T = 1 \text{ \& } a_{T-1} = 0, \\ O_T + \sigma \cdot \epsilon_{0T} & \text{if } a_T = 0. \end{cases}$$

The expected utility of period T is given by:

$$\begin{aligned} EV_{1T}^j &= \ln\Big(\exp(W_T^H/\sigma) + \exp(O_T/\sigma)\Big) + \gamma, \\ EV_{0T}^j &= \ln\Big(\exp((W_T^H - \kappa)/\sigma) + \exp(O_T/\sigma)\Big) + \gamma. \end{aligned}$$

At $t \in [T - 1, T_0 + 3]$,

$$V_t^j = \begin{cases} W_t^H + \sigma \cdot \epsilon_{1t} + \beta EV_{1t+1}^j & \text{if } a_t = 1 \text{ \& } a_{t-1} = 1, \\ W_t^H - \kappa + \sigma \cdot \epsilon_{1t} + \beta EV_{1t+1}^j & \text{if } a_t = 1 \text{ \& } a_{t-1} = 0, \\ O_t + \sigma \cdot \epsilon_{0t} + \beta EV_{0t+1}^j & \text{if } a_t = 0. \end{cases}$$

The expected utility of period $t \in [T-1, T_0+3]$ is given by:

$$EV^j_{1t} = \ln\Big(\exp((W^H_t + \beta EV^j_{1t+1})/\sigma) + \exp((O_t + \beta EV^j_{0t+1})/\sigma)\Big) + \gamma,$$
$$EV^j_{0t} = \ln\Big(\exp((W^H_t - \kappa + \beta EV^j_{1t+1})/\sigma) + \exp((O_t + \beta EV^j_{0t+1})/\sigma)\Big) + \gamma.$$

At time T_0+2, because the driver commits to work at T_0 and T_0+1, $a_{T0+1} = 1$ with probability 1:

$$V^j_{T_0+2} = \begin{cases} W^H_{T_0+2} + \sigma \cdot \epsilon_{1T_0+2} + \beta EV^j_{1T_0+3} & \text{if } a_{T_0+2} = 1, \\ O_{T_0+2} + \sigma \cdot \epsilon_{0T_0+2} + \beta EV^j_{0T_0+3} & \text{if } a_{T_0+2} = 0. \end{cases}$$

At T_0+1, the high-performing worker has to work. The expected value at any T_0+1 is given by

$$EV^j_{1T_0+1} = W^H_{T_0+1} + \beta EV^j_{1T_0+2}.$$

At period T_0, the expected value is

$$EV^j_{1T_0} = W^H_{T_0} + \beta EV^j_{1T_0+1},$$
$$EV^j_{0T_0} = W^H_{T_0} - \kappa + \beta EV^j_{1T_0+1}.$$

At any time before T_0, $t \in [T_0-1, 2]$, the expected utility is given by

$$EV^j_{1t} = \ln\Big(\exp((W^H_t + \beta EV^j_{1t+1})/\sigma) + \exp((O_t + \beta EV^j_{0t+1})/\sigma)\Big) + \gamma,$$
$$EV^j_{0t} = \ln\Big(\exp((W^H_t - \kappa + \beta EV^j_{1t+1})/\sigma) + \exp((O_t + \beta EV^j_{0t+1})/\sigma)\Big) + \gamma.$$

At period 1,

$$V^j_1 = \begin{cases} W^H_1 - \kappa + \sigma \cdot \epsilon_{11} + \beta EV^j_{12} & \text{if } a_1 = 1, \\ O_1 + \sigma \cdot \epsilon_{01} + \beta EV^j_{02} & \text{if } a_1 = 0. \end{cases}$$

The expected value of being a high-performing driver is then derived as

$$EV^j = \ln\Big(\exp((W^H_1 - \kappa + \beta EV^j_{12})/\sigma) + \exp((O_1 + \beta EV^j_{02})/\sigma)\Big) + \gamma. \tag{C.2}$$

C.3 Estimation

We now describe how we adapt the CCP estimator to our setting.

C.3.1 Low-performing Drivers

At the final period T, the conditional probability of working is

$$P_L(a_T = 1|a_{T-1} = 1) = \frac{\exp(W_t^L)}{\exp(W_t^L) + \exp(O_T^L)},$$
$$P_L(a_T = 1|a_{T-1} = 0) = \frac{\exp(W_t^L - \kappa_T)}{\exp(W_t^L - \kappa_T) + \exp(O_T^L)}.$$

For any $t \in [2, T-1]$,

$$P_L(a_t = 1|a_{t-1} = 1) = \frac{\exp(W_t^L + \beta EV_{1t+1}^L)}{\exp(W_t^L + \beta EV_{1t+1}^L) + \exp(O_t^L + \beta EV_{0t+1}^L)},$$
$$P_L(a_t = 1|a_{t-1} = 0) = \frac{\exp(W_t^L - \kappa_t + \beta EV_{1t+1}^L)}{\exp(W_t^L - \kappa_t + \beta EV_{1t+1}^L) + \exp(O_t^L + \beta EV_{0t+1}^L)}.$$

At $t = 1$,

$$P_L(a_1 = 1) = \frac{\exp(W_1^L - \kappa_1 + \beta EV_{12}^L)}{\exp(W_1^L - \kappa_1 + \beta EV_{12}^L) + \exp(O_1^L + \beta EV_{02}^L)}. \tag{C.3}$$

C.3.2 High-performing Drivers

For any schedule $j \in \{1, \cdots, 16\}$, the conditional probability of working in final period T is

$$P_j(a_T = 1|a_{T-1} = 1) = \frac{\exp(W_t^H)}{\exp(W_t^H) + \exp(O_T)},$$
$$P_j(a_T = 1|a_{T-1} = 0) = \frac{\exp(W_t^H - \kappa_T)}{\exp(W_t^H - \kappa_T) + \exp(O_T)}.$$

For any $t \in [T_0 + 3, T-1]$, we have

$$P_j(a_t = 1|a_{t-1} = 1) = \frac{\exp(W_t^H + \beta EV_{1t+1}^j)}{exp(W_t^H + \beta EV_{1t+1}^j) + \exp(O_t + \beta EV_{0t+1}^j)},$$
$$P_j(a_t = 1|a_{t-1} = 0) = \frac{\exp(W_t^H - \kappa_t + \beta EV_{1t+1}^j)}{\exp(W_t^H - \kappa_t + \beta EV_{1t+1}^j) + \exp(O_t + \beta EV_{0t+1}^j)}.$$

At $t = T_0 + 2$, we have

$$P_j(a_t = 1|a_{t-1} = 1) = \frac{\exp(W_t^H + \beta EV_{1t+1}^j)}{\exp(W_t^H + \beta EV_{1t+1}^j) + \exp(O_t + \beta EV_{0t+1}^j)}.$$

At $t = T_0 + 1$, we have

$$P_j(a_t = 1|a_{t-1} = 1) = 1.$$

At $t = T_0$,

$$\begin{aligned} P_j(a_t = 1|a_{t-1} = 1) &= 1, \\ P_j(a_t = 1|a_{t-1} = 0) &= 1. \end{aligned}$$

For any $t \in [2, T_0 - 1]$, we have

$$\begin{aligned} P_j(a_t = 1|a_{t-1} = 1) &= \frac{\exp(W_t^H + \beta EV_{1t+1}^j)}{\exp(W_t^H + \beta EV_{1t+1}^j) + \exp(O_t + \beta EV_{0t+1}^j)}, \\ P_j(a_t = 1|a_{t-1} = 0) &= \frac{\exp(W_t^H - \kappa_t + \beta EV_{1t+1}^j)}{\exp(W_t^H - \kappa_t + \beta EV_{1t+1}^j) + \exp(O_t + \beta EV_{0t+1}^j)}. \end{aligned}$$

At $t = 1$, we have

$$P_j(a_1 = 1) = \frac{\exp(W_1^H - \kappa_1 + \beta EV_{12}^j)}{\exp(W_1^H - \kappa_1 + \beta EV_{12}^j) + \exp(O_1 + \beta EV_{02}^j)}.$$

Therefore, at any t, the conditional probability for high-performing drivers is

$$\begin{aligned} P_H(a_t = 1|a_{t-1} = 0) &= \sum_S \widetilde{P}_j \cdot P_j(a_t = 1|a_{t-1} = 0), \\ P_H(a_t = 1|a_{t-1} = 1) &= \sum_S \widetilde{P}_j \cdot P_j(a_t = 1|a_{t-1} = 1), \end{aligned}$$

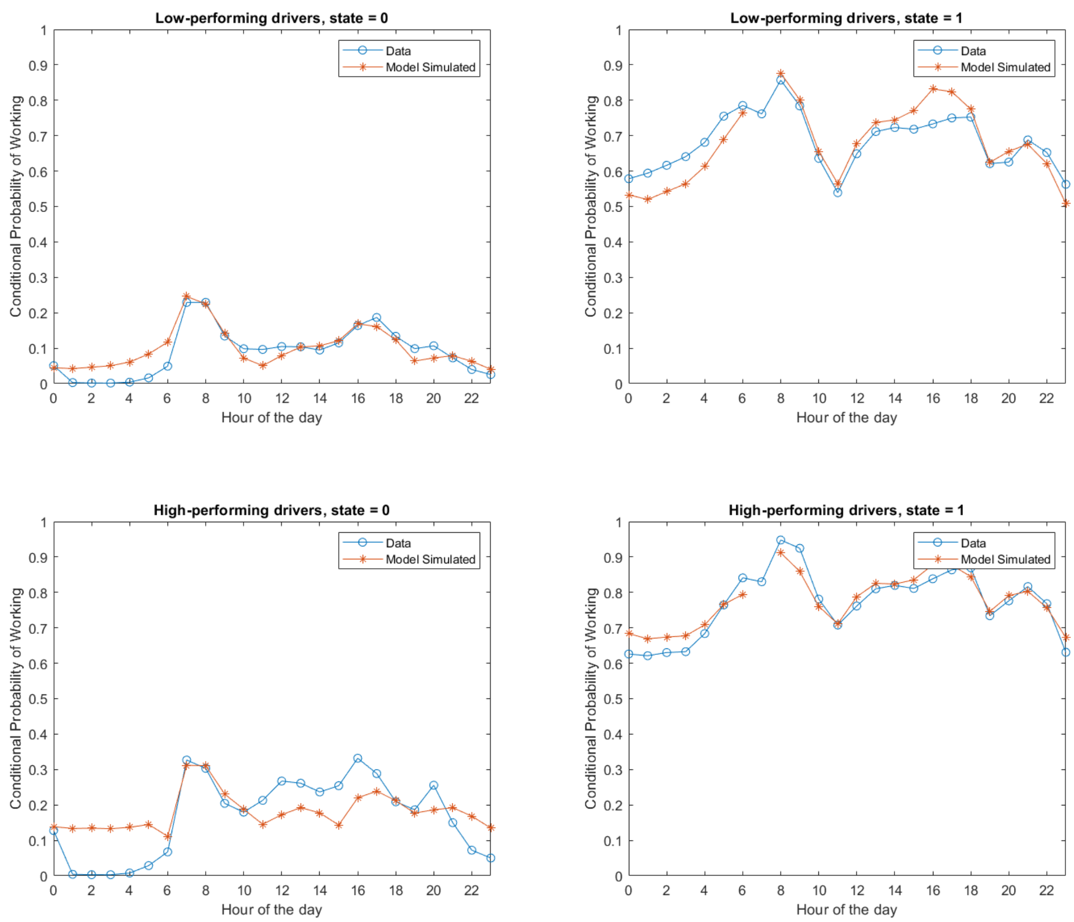
where $\widetilde{P}_j$ is the probability of choosing each minimum requirement within high-performing drivers:

$$\widetilde{P}_j = \frac{\exp(EV^j)}{\sum_{k=1}^{16} \exp(EV^k)}.$$

C.4 Model Validation

Figure C.1 shows the model's goodness of fit. First, for both high-performing and low-performing drivers, the simulated CCPs fit the observed CCPs well when $a_{t-1} = 1$. The simulated CCPs fit well for low-performing drivers when $a_{t-1} = 0$. However, the fit of high-performing drivers at $a_{t-1} = 0$ is a little off. This might be because the observed CCPs of high-performing drivers and low-performing drivers are quite different when $a_{t-1} = 0$. Our model is not able to capture this difference, because we assume that the reservation values are the same among low-performing and high-performing drivers in our benchmark model. As a robustness check in a later section, we allow the reservation values to be different by observable characteristics of consumers.

Figure C.1: Model Goodness of Fit



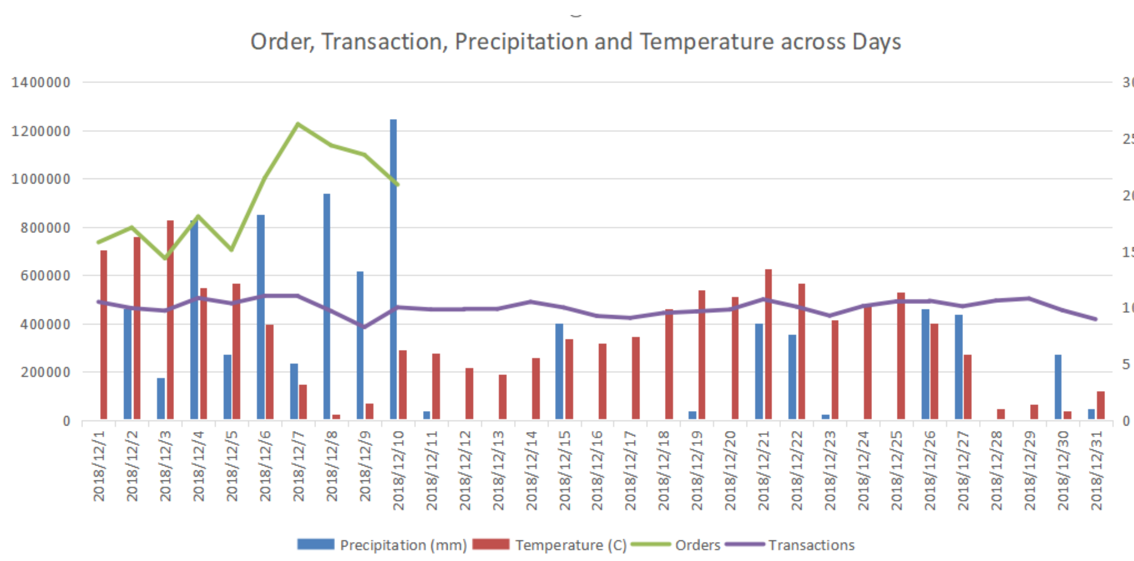
Note: Figure C.1 shows the model's simulated values against the empirically observed CCPs.

Online Appendices: Not For Publication

D Additional Summary Statistics

Order, Transaction, Precipitation and Temperature. Figure D.1 reports the daily number of orders and transactions during our sample period (10 days of order data and 31 days of transaction data). We compare them with daily precipitation and average temperatures. From Dec. 6 to Dec. 10, the precipitation increases and temperature decreases, resulting in more orders (customer demand). However, the number of completed transactions across days remain the same throughout our sample period. Information about precipitation and temperature is used in our demand estimation.

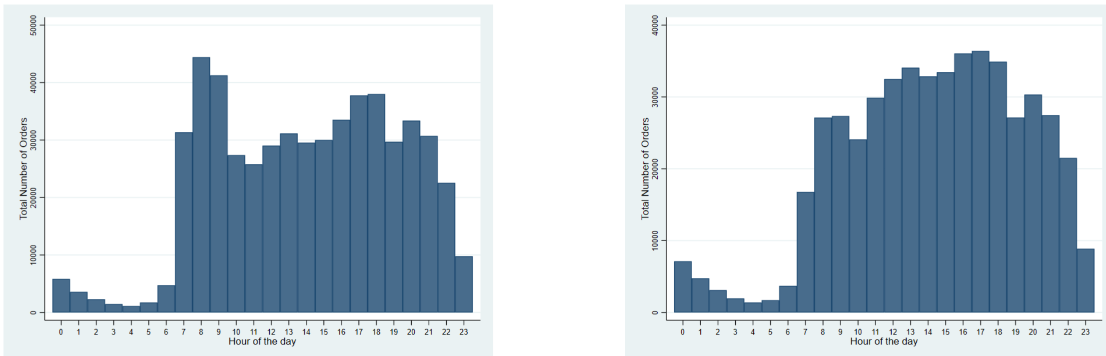
Figure D.1: Orders, Transactions, Precipitation, and Temperature across Days



Order Information. Figure D.2 shows the total number of orders for each hour of the day, and the distribution for weekdays versus weekends. The number of orders is significantly lower at night and early morning from 11PM to 6AM. For weekdays, there is a clear pattern of a morning peak from 7AM to 9AM and an afternoon peak from 5PM to 7PM. Such patterns of peak hours do not exist on weekends.

Figure D.3 shows the average travel distance and average price per transaction. Regarding average travel distance on weekends, the average distance is stable between 6AM and 9PM at about 9KM per transaction. The travel distance per transaction is slightly higher at

Figure D.2: Total Number of Orders



(a) Number of orders on weekdays (b) Number of orders on weekends

night at about 10KM. There is a sharp increase in average travel distance between 4AM and 6AM. This might be caused by early transit to the airport. The general pattern of average travel distance on weekdays is similar to that of weekends. However, on weekdays we observe significantly lower travel distances at peak hours (7AM-9AM and 5PM-7PM). This indicates that riders travel in shorter distances during peak hours. Regarding prices, the average price is about 5.5 CNY per KM. Prices are higher during late night and early morning, because the platforms charge extra fees for late night travels. On weekdays, prices are higher during peak hours than off-peak hours. On weekends, there is no significant difference between prices during peak hours and off-peak hours.

Driver Characteristics of High/Low-performing Drivers Table D.1 summarizes the characteristics of high-performing and low-performing drivers. There are $23,712$ high-performing drivers and $16,392$ low-performing drivers. Panel I reports the driver/vehicle characteristics. There are few female drivers in the sample (2.2% for high-performing drivers and 3.5% for low-performing drivers). High-performing drivers are more likely to be non-local (69% v.s. 53%) and younger (aged 37.2 v.s. 37.4). Panel II reports driver performance in a month. On average, high-performing drivers work longer (17 v.s. 5 days and 159 v.s. 26 hours), receive more orders (301 v.s. 46) and earn 6.6 times more than an average low-performing driver. Panel III reports the average driver performance in an hour. High-performing drivers have more time serving passengers (30.7 minutes v.s. 29.3 minutes) and spend less time waiting for orders (18.6 minutes v.s. 20.4 minutes). High-performing drivers have more orders (1.9 v.s. 1.74 orders) and earn more (50.4 v.s. 46.5 RMB every hour).

Multi-Homing versus Single-Homing. In our main analysis, we focus on DiDi because DiDi accounts for more than 90% of China's mainland ride-hailing market share. Nonethe-

Figure D.3: Average Travel Distance and Price Per Transaction

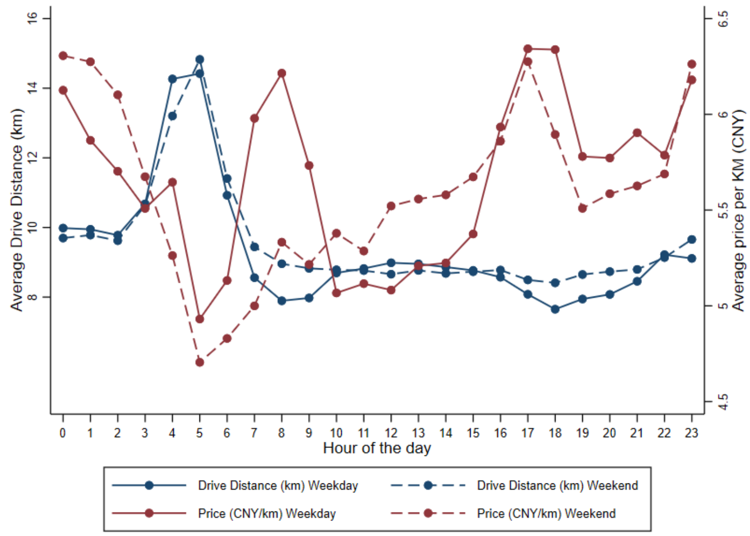


Table D.1: High/Low-performing Driver Characteristics

	High-performing (1)	Low-performing (2)
Panel I: Driver/Vehicle Characteristics		
% femal	2.2%	3.5%
% non-local	69%	53%
Age	37.2	37.4
MSRP Mean	12.2	13.0
MSRP Median	11.2	11.6
% New Energy	42%	28%
Panel II: Performance (in a month)		
Work Days	17	5
Work Hours	159	26
# orders	301	46
Monthly Revenue	7985	1202
Panel III: Performance (in an hour)		
Work Time	30.7	29.3
Pickup time	10.7	10.2
Idle Time	18.6	20.4
# orders	1.90	1.76
Hourly Revenue	50.4	46.5
# drivers	23712	16392
Share Drivers	59.1%	40.9%
# driver-hours	3758519	423799

Notes: MSRP data comes from 17,347 drivers for whom vehicle type information is available. Measured in 10,000 CNY.

less, there is a concern that drivers may switch between working for different platforms if they pay different hourly wages. To address such concerns, we document the number of vehicles/drivers that are multi-homed versus single-homed in our data. First, we look at the number of vehicles that are multi-homed from registration data. Panel (a) of Table D.2 shows that 85% of vehicles are registered to only one platform, and only 1.8% of vehicles are registered to more than two platforms. Therefore, multi-homing is not very common based on vehicle registration information. Then, we look at how common multi-homing is directly from actual transactions. Panel (b) of Table D.2 shows that among all the vehicles that conducted business in December 2018, 92.5% used a single platform and never switched to another platform within the month. Only 0.3% of vehicles used more than two platforms in the given month. The evidence shows that the majority of vehicles/drivers are single-homed.

Table D.2: Multi-Homing versus Single-Homing

Number of Registered Platforms	Number of Vehicles	Percent
1	86,422	84.6%
2	13,838	13.5%
3	1,866	1.8%

(a) Based on Vehicle Registration Data

Number of Used Platforms	Number of Vehicles	Percent
1	49,213	92.5%
2	3,836	7.2%
3	141	0.3%

(b) Based on Transactional Data

Among the multi-homed drivers, we further study how these drivers switch between different ride-sharing platforms. We also calculate the number of multi-homed and single-homed drivers within a whole day based on actual transactions. In any given day of December, only about 1% of drivers used more than one platform within a day. This shows that it is very rare for drivers to switch between platforms within a given day. Therefore, the results suggest that drivers in our data are mostly single-homed and rarely switch between platforms.

E Cluster Schedules and Hourly Rates

This online appendix describes how we cluster drivers using their work schedule and hourly rate data. We use the data that contains driver schedules and hourly revenue for all drivers on Dec. 3rd, 2018. Our sample includes 23,689 drivers (observations).

E.1 Data Cleaning

We use a 24×1 vector for each driver to describe their working schedule and hourly rate. The n^{th} element represents the hourly revenue at n o'clock. If the driver does not work at this hour, we denote the element value to be 0. For instance, suppose the driver worked at 7 am and earned 18 CNY, the 7th element is 18 for this vector.

In addition, we construct the following variables to measure the driver's working schedule and include them in our study:

- **earlymorning**: driver's working hour during early morning (0 - 7)
- **morningpeak**: driver's working hour during morning peak (7 - 10)
- **miday**: driver's working hour during mid-day (10 - 16)
- **peakaf**: driver's working hour during afternoon peak (16 - 19)
- **offpeakaf**: driver's working hour during afternoon off-peak (19 - 22)
- **latenight**: driver's working hour during late night (22 - 24)
- **workhour**: driver's working hour in one day
- **start**: the starting working hour
- **end**: the end working hour

- **consecutive**: the consecutive working hours in a day

- **consecutive1/2/3**: We divide 24 hours into 3 parts. Consecutive1/2/3 indicates the consecutive working hours in each part of the day.

- **consecutive4**: the consecutive working hours during afternoon off-peak (19 - 22)

- **morningPeakCon/peakAfCon**: the consecutive working hours during morning peak and afternoon peak

- **HourlyRate**: the average hourly rate in a day

E.2 K-means clustering

We apply the k-means method and cluster the drivers in our working database. The purpose of this clustering is to explore how different work schedules can affect the driver's hourly revenue. The k-means clustering method divides observations into a certain number of (k) groups according to their similarity. We do not know ex ante the number of groups we define. Therefore, we have tried $k = 2, 3, 4$ different clusters. The set of independent variables includes all 40 variables described above. We use the average hourly revenue for the driver in one day as the dependent variable.

Table 1 illustrates the results when $k = 2$. Drivers are divided into low hourly rates (cluster 1) and high hourly rates (cluster 2). High hourly rate drivers are more likely to work longer and consecutive hours.

Tables 2 and 3 report cluster results for $k = 3$ and $k = 4$, respectively. Though we pre-set more clusters, drivers can be separated into two groups. When $k = 3$, we have low hourly rate drivers (cluster 1) and high hourly rate drivers (clusters 2 and 3). When $k = 4$, we have low hourly rate drivers (cluster 1) and high hourly rate drivers (clusters 2, 3, and 4). Moreover, no matter which k we choose, the characteristics of the lower-income schedules are similar: they work shorter and fewer consecutive hours.

Table E.1: Clustering results for k=2

	cluster1	cluster2
Count	11,226	12,472
earlymorning	0.59	0.29
morningpeak	1.2	2.15
miday	1.61	4.86
peakaf	0.93	2.71
offpeakaf	0.87	2.16
latenight	0.7	1.46
workhour	5.62	13.01
start	9.21	7.4
end	16.59	20.92
consecutive	4.56	11.59
consecutive1	1.3	1.61
consecutive2	2.16	6.24
consecutive3	1.91	4.75
consecutive4	1.29	2.98
morningPeakCon	1.56	2.95
peakAfCon	1.22	3.52
HourlyRate	40.03	46.91

Table E.2: Clustering results for k=3

	cluster1	cluster2	cluster3
Count	8,912	7,745	7,041
earlymorning	0.51	0.52	0.27
morningpeak	1.21	1.43	2.51
miday	1.46	3.42	5.38
peakaf	0.76	2.58	2.48
offpeakaf	0.6	2.89	1.43
latenight	0.43	2.51	0.59
workhour	4.79	12.37	12.34
start	9.36	8.07	7.14
end	15.57	22.51	19.35
consecutive	3.92	10.36	11.35
consecutive1	1.24	1.34	1.83
consecutive2	1.98	4.4	6.91
consecutive3	1.34	6.16	3.27
consecutive4	0.84	4.41	1.7
morningPeakCon	1.56	1.98	3.42
peakAfCon	0.98	3.52	3.11
HourlyRate	38.28	47.73	46.12

Table E.3: Clustering results for k=4

	cluster1	cluster2	cluster3	cluster4
Count	7,921	6,246	5,328	4,203
earlymorning	0.48	0.24	0.27	0.82
morningpeak	1.32	2.59	2.36	0.3
miday	1.41	4.73	5.37	2.11
peakaf	0.72	2.79	2.33	2.16
offpeakaf	0.46	2.9	0.98	2.75
latenight	0.28	2.26	0.23	2.46
workhour	4.55	14.55	11.4	9.67
start	9.12	6.87	7.32	9.77
end	14.93	22.27	18.54	22.47
consecutive	3.71	12.44	10.53	8.16
consecutive1	1.29	1.82	1.72	0.96
consecutive2	1.94	6.09	6.85	2.76
consecutive3	1.09	6.04	2.51	5.74
consecutive4	0.62	4.19	1.06	4.26
morningPeakCon	1.69	3.51	3.25	0.47
peakAfCon	0.91	3.74	2.85	3.02
HourlyRate	37.54	46.78	45.77	48.05