Journal Pre-proofs

The Impact of COVID-19 on the Ride-Sharing Industry and Its Recovery: Causal Evidence from China

Wei Wang, Wei Miao, Yongdong Liu, Yiting Deng, Yunfei Cao

PII: S0965-8564(21)00260-3
DOI: https://doi.org/10.1016/j.tra.2021.10.005
Reference: TRA 3274

To appear in: *Transportation Research Part A*

Received Date: 15 December 2020
Revised Date: 29 September 2021
Accepted Date: 11 October 2021

The Impact of COVID-19 on the Ride-Sharing Industry and Its Recovery: Causal Evidence from China

Wei Wang
Address:
- 10 Huixin East Street, School of International Trade and Economics, University of International Business and Economics, Beijing 100029, China.

Affiliation:
- School of International Trade and Economics, University of International Business and Economics.

Wei Miao
Address:
- Level 38, One Canada Square, Canary Wharf, London E14 5AA, United Kingdom

Affiliation:
- UCL School of Management, University College London

Yongdong Liu
Address:
- Level 38, One Canada Square, Canary Wharf, London E14 5AA, United Kingdom

Affiliation:
- UCL School of Management, University College London

Yiting Deng
Address:
- Level 38, One Canada Square, Canary Wharf, London E14 5AA, United Kingdom

Affiliation:
- UCL School of Management, University College London

Yunfei Cao

Corresponding author
Address: No 5 South Zhongguancun Street, Haidian District, Beijing 100081, China
Email: caoyunfei1986@bit.edu.cn
tel: (+86)10-68918401
Affiliations:
- Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing, 100081, China
- School of Management and Economics, Beijing Institute of Technology, Beijing, 100081, China

Funding Information:

Wei Wang acknowledges the financial support from the National Natural Science Foundation of China (Grant No. 72173019 and 71703017), the Beijing Social Science Foundation (Grant No. 20JJA004), the Chinese Ministry of Education Research Funds on Humanities and Social

Sciences (Grant No. 21YJA790056). Yunfei Cao acknowledges the financial support from the National Natural Science Foundation of China (Grant No. 71704011).

Abstract

The COVID-19 pandemic has brought unprecedented disruptions to many industries, and the transportation industry is among the most disrupted ones. We seek to address, in the context of a ride-sharing platform, the response of drivers to the pandemic and the post-pandemic recovery. We collected comprehensive trip data from one of the leading ride-sharing companies in China from September 2019 to August 2020, which cover pre-, during-, and post-pandemic phases in three major Chinese cities, and investigate the causal effect of the COVID-19 pandemic on driver behavior. We find that drivers only slightly reduced their shift decision in response to increased COVID-19 cases, likely because they have to make a living from providing ride-sharing services. Nevertheless, conditional on working, drivers exhibit strong risk aversion: As the number of new cases increases, drivers strategically adjust the scope of search for passengers, complete fewer trips, and as a result, make lower daily earnings. Finally, our heterogeneity analyses indicate that the effects appear to vary both across drivers and over time, with generally stronger effects on drivers who are older, more experienced, more active before the pandemic, and with higher status within the firm. Our findings have strong policy implications: These drivers tend to contribute more to the focal company, and also rely more on providing ride-sharing services to make a living. Therefore, they should be prioritized in stimulus plans offered by the government or the ride-sharing company.

Keywords: COVID-19, Ride-sharing, Recovery, Instrumental Variables, Causal Inference

1. Introduction

The sharing economy has been booming in recent years, leading to a rapid increase in jobs in the "gig" economy. According to Hossain (2020), in the US alone, the sharing economy sector has created 6.23 million jobs with 78 million service providers, and 800 million people engage with it. The transportation sector is one of the most salient beneficiaries of the burgeoning sharing economy. For instance, commuting to work by shared bicycle (e.g., Citi Bike) has become an increasingly popular transportation option (Ford et al., 2019; Chen et al., 2020). The ride-sharing service (e.g., Uber) allows drivers to enjoy more flexibility in work, which is proven valuable to drivers (Chen et al., 2019) and has improved capacity utilization (Cramer and Krueger, 2016).

However, the COVID-19 pandemic has brought unprecedented disruptions to many industries, and the transportation industry is among the most disrupted ones. Further, the COVID-19 has raised concerns about the survivability of the sharing economy in general. It is reported that gross bookings on Uber rides were down by 75% in the three months through June 2020, and that Lyft's April ridership was down by 75% from April 2019.[1] One important question is whether and how the sharing economy could recover in the post-pandemic economy. This is of paramount importance both to the overall economic recovery, and to the companies within the sector.

To answer this question, it is crucial to understand the response of platform users to the pandemic and the post-pandemic recovery. China is the first major economy to recover from the pandemic with most cities returning to normal by the second or third quarter of 2020, which provides an ideal context to study the effect of the pandemic on the industry and the post-pandemic recovery of the industry. Therefore, we collected comprehensive hailing request and trip data for a random sample of drivers from one of the leading ride-sharing companies in

[1] https://www.washingtonpost.com/technology/2020/08/10/uber-coronavirus-lockdowns/

China from September 2019 to August 2020, which cover pre-, during-, and post-pandemic phases in three major Chinese cities. The data allow us to empirically quantify the effect of the COVID-19 on the ride-sharing industry and the recovery of the industry in the post-pandemic phase. Because the data are collected based on a random sample of drivers, we do not have complete panel data for riders, and therefore our focus is the impact of COVID-19 on driver behavior; however, we note that insights from the demand side are also extracted from our analyses.

Unlike the traditional taxi market, where taxi drivers rent vehicles from taxi companies and then directly provide transportation services to consumers, modern ride-sharing platforms, including the focal company under study, typically serve as the matching intermediary between drivers and passengers. Due to such two-sided market nature, the profitability of modern ride-sharing platforms (and sharing economy in general) highly depends on the interdependence or externality between the two sides of economic agents (Rysman, 2009).[2] Therefore, a ride-sharing platform would benefit from the network effect if more drivers work for them.[3] It is thus managerially important for the ride-sharing platform to understand whether COVID-19 has affected drivers' labor supply patterns and if yes, the magnitude of the effect across drivers and over time.

We first analyze how the number of daily confirmed COVID-19 new cases in the city affects a driver's decision on whether or not to work on a day, and conditional on working, number of trip requests, number of completed trips, average trip distance, average search distance, and earnings. Because of the potential effect of mobility on the COVID-19 case growth rate (Badr et al., 2020; Wielechowski et al., 2020), there is simultaneity-induced

[2] Most current ride-haling platforms make profits by charging commissions or service fees on successfully fulfilled trips by drivers. Therefore, these platforms naturally benefit from a larger network effect.

[3] In the context of a two-sided market, network effect is the phenomenon by which the utility of a user on one side of the platform depends on the number of users on the other side of the platform, i.e., the platform becomes more valuable when more users use its service on the other side (Jang et al., 2018).

endogeneity concern when estimating the effect of COVID-19 severity on drivers' labor supply. We apply instrumental variable estimation to deal with this endogeneity concern and identify the causal impact of COVID-19 and recovery on mobility. We select two instrumental variables: 1) number of confirmed cases in other cities in the same province; and 2) number of confirmed "imported cases" (i.e., people who were tested positive after arriving from other countries). These instrumental variables meet the two criteria for instrumental variables: (1) **Relevance condition**, i.e., the instrumental variables are correlated with the endogenous variable, daily confirmed COVID-19 new cases. The two selected instrumental variables are relevant because they affect platform users' perception of the severity of the pandemic. We also conduct the under-identification test and weak instrumental variable test to empirically confirm that the relevance condition holds with our selected instrumental variables. (2) **Exclusion restriction**, i.e., the instrumental variables do not directly affect drivers' labor supply decisions and the effect only takes place through the endogenous variable. These two selected instrumental variables also meet the exclusion restriction because these new cases originated from outside of the focal city and are unlikely caused by transportation usage in the focal city.

Our analyses indicate that, among all labor supply measures, drivers seem relatively less flexible in adjusting their shift decision in response to new COVID-19 cases: A one standard deviation increase in new confirmed cases, which equals 4.39 cases, only decrease drivers' probability of working by 1.3%, which suggests that drivers' decision of whether or not to work is only marginally affected by the pandemic. However, conditional on working, we find that as the number of COVID-19 cases increases, drivers tend to receive more requests from the booking aggregator channel, where passengers can select from a list of service providers.[4] At the same time, we see a significant decrease in requests received via the focal company's own booking app channel. Such patterns imply that during the pandemic, perhaps

[4] We discuss the different booking channels in more detail in Section 3.

due to the shrinkage in overall supplies, passengers tend to switch from a single ride-sharing service provider to booking aggregators, so that they can have a wider range of choices and hence a higher chance of getting a ride. Despite the increased aggregator channel requests, the total number of completed trips from both channels decreases, implying more intensive competition from other ride-sharing companies during the period we study. We also uncover the structural change in passengers' transportation need: When the COVID-19 is more severe, passengers refrain from taking long-distance trips and therefore drive down the average trip distance for drivers; meanwhile, drivers also strategically adjust their scope of search for passengers. As a result, we see a salient decrease in drivers' daily earnings.

We continue to investigate whether there are heterogeneous effects across drivers and over time. We find that for drivers who have worked longer at the company, the negative effects of new cases on the shift decision and the number of completed trips are smaller, but the negative effect on daily earnings is larger. The negative effects of new case on the shift decision, on the number of requests and the number of completed trips, and on daily earnings are stronger for drivers who are certified as "gold drivers" within the company, who have received more requests before the COVID-19 breakout, and for older drivers.

Finally, we also uncover pronounced intertemporal heterogeneity of the impact of COVID-19—Drivers are less sensitive to new cases during the post-COVID period and when the government emergency level is lower. Our heterogeneity analyses shed light on whom the government and company should prioritize in the stimulus or subsidy plans.

The rest of the paper is organized as follows. In section 2, we review the literature on the impact of COVID-19 on the transportation industry. In section 3, we introduce the industry background and our dataset. Section 4 discusses our empirical strategy and the instrumental variables used. Section 5 reports the estimation results and our findings. Section 6 concludes.

2. Literature

This paper relates to three streams of literature. First, it relates to the burgeoning literature on the economic impact of the COVID-19 pandemic. Studies have demonstrated the impact of the pandemic on mobility (Badr et al., 2020; Warren and Skillman, 2020; Wielechowski et al., 2020), taxi usage (Nian et al., 2020; Zheng et al., 2020), and small businesses (Bartik et al., 2020; Gourinchas et al., 2020; Kim et al., 2020), education (Sen and Tucker, 2020), mental health (Witteveen and Velthorst, 2020), among others. For example, Warren and Skillman (2020) find that the pandemic has led to a large reduction in mobility both in the US and globally. Badr et al. (2020) find a strong relationship between mobility pattern and decreased COVID-19 case growth rates for the most affected counties in the US. Zheng et al. (2020) conducted a spatiotemporal analysis of taxi demand and supply in Shenzhen, China, using data in the first three months of 2020. Their findings indicate that, after re-opening, the recovery of taxi travel was largely behind the recovery of overall vehicle travel, most drivers significantly cut back work hours and many adjusted work schedule to focus on peak-time. Nian et al. (2020) analyze the impact of Covid-19 on taxi usage in Chongqing, China. They find significant effects of the epidemic on the number of taxi trips, travel speed and time, and spatial distribution of taxi trips. Gourinchas et al. (2020) estimate the impact of the COVID-19 on business failures among small and medium size enterprises (SMEs) in seventeen countries. Bankruptcy rate of transport & storage sector is calculated to increase from 7.64% to 13.28% due to COVID-19. Kim et al. (2020) find in the early phases of the pandemic, small businesses and their owners experienced unprecedented disruptions of up to 40% drop in weekly revenues, expenses, and consumption. Our focus is the ride-sharing industry, an increasingly important market within the transportation sector, in which drivers can also be considered small business owners managing their own work schedule facing external factors. More importantly, we shed light on how the industry recovered in the post-pandemic phase, which is one of the earliest empirical research on market recovery from COVID-19.

Second, by analyzing driver behavior in response to COVID-19, we add to the literature on labor supply in the taxi industry and more recently, in the ride-sharing industry. In the taxi industry, it has been demonstrated that drivers exhibit reference-dependent preferences (Camerer et al., 1997; Crawford and Meng, 2011), respond positively to both unanticipated and anticipated increases in earnings opportunities (Farber, 2015), strategically respond to surge pricing (Cachon et al., 2017; Castillo, 2020; Miao et al., 2021), overtreat passengers under a usage-based pricing scheme (Miao and Chu, 2020), and are strategic in location choices (Buchholz, 2017). In the ride-sharing industry, prior research has demonstrated flexibility in driver schedule and labor supply (Hall and Krueger, 2016; Chen et al., 2019), and the effect of external incentives (Allon et al., 2018; Brodeur and Nield, 2018). For instance, Chen et al. (2019) find that Uber drivers earn more than twice the surplus they would in less flexible arrangements. Cramer and Krueger (2016) find that UberX drivers have higher capacity utilization rate compared with taxi drivers. We find drivers' labor supply and location choices are affected by confirmed COVID-19 cases after accounting for endogeneity, and the effect is stronger when the city is under the top-tier emergency level than under the lower-tier emergency levels.

Third, this paper also adds to the recent research on COVID-19 induced inequality. For example, Polyakova et al. (2020) show that while the COVID-19 induced excess mortality is higher for the older population, the economic damage is higher for younger adults – those between 25 to 44 years old experienced employment displacement of 11.6 percentage points, compared to 1.8 percentage points for those who are over 85 years old. Weill et al. (2020) find that social distancing responses to COVID-19 emergency declarations vary by income level of the region: wealthier areas decreased mobility significantly more than poorer areas. Witteveen and Velthorst (2020) find that the COVID-19 has exacerbated socioeconomic inequalities, and the economic hardship experienced by people in lower occupational ranking has led to much higher prevalence of expressing adverse mental health. Blundell et al. (2020) find that the COVID-19 has exacerbated many existing inequalities in terms of income, age, gender and ethnicity. For instance, younger workers and those on low incomes have a higher probability of losing jobs during the lockdown. Our results indicate that older drivers were affected to a larger extent.

To summarize, we see the contribution of this project to be three-fold. First, we broaden the research on the economic impact of COVID-19 by causally assessing its impact on the ride-sharing industry, and assessing the recovery of the industry during the post-pandemic re-opening phase. Second, we complement existing studies on drivers' labor supply by exploring their strategic behavioral responses to the unexpected pandemic during different phases. Third, we demonstrate asymmetric effects of COVID-19 on drivers with different characteristics, adding to the existing studies on COVID-19 induced inequality.

3. Industry Background and Data

To empirically examine the impact of COVID-19 on drivers' labor supply behavior, we first collect detailed trip-level data from one of the largest ride-sharing platforms in China. The company offers a wide range of booking services and its flagship service is the basic ride-sharing service similar to UberX, where a consumer sends her ride-sharing request and the company then matches the consumer with a nearby available driver. Besides the ride-sharing service, the company also provides other booking services such as advanced booking, airport pickup, chartered service, etc.

The company provides its booking services via two channels. First, the company runs its own booking app, as displayed in Figure 1, for both iOS and Android users. We denote booking requests made through the mobile app as app requests. Second, in 2017, the company started collaborating with Amap, a leading navigation company in China, to allow Amap users to hail the focal company's cars along with those of other competing ride-sharing service providers. When hailing a car from Amap, as illustrated in Figure 2, a consumer can select from a list of ride-sharing companies to be included in her booking request. As long as the focal company of interest is selected by the consumer, we can observe the booking request in our data. We denote booking requests received via Amap as aggregator requests. As a ride-sharing aggregator, Amap collaborates with all major ride-sharing companies in China (Didi, Shouqi, Caocao, etc.). Therefore, only a fraction of channel requests are eventually completed by the focal company. The difference between requested and completed trips through the aggregator can be regarded as a proxy for competition intensity: a larger difference would imply more intensive competition.

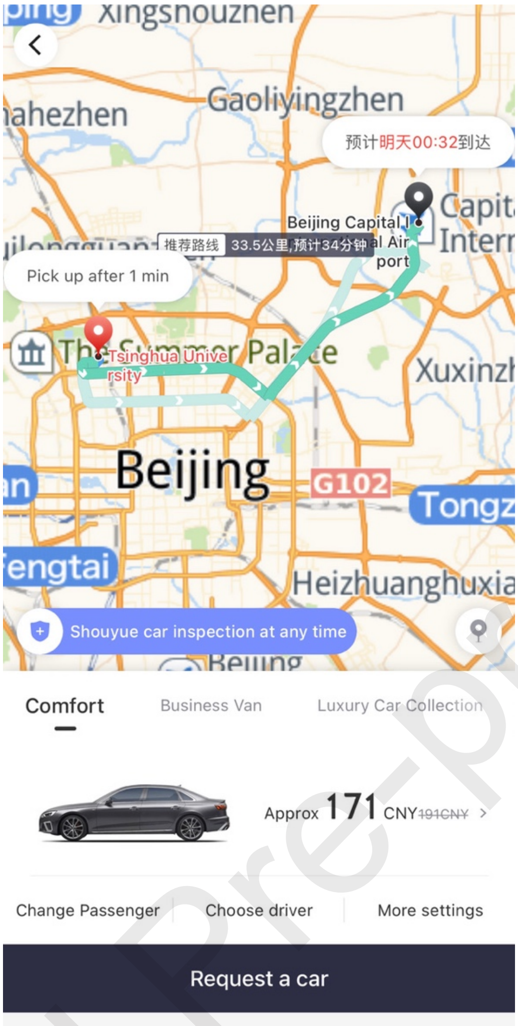

Figure 1: Screenshot of an App Hailing Request

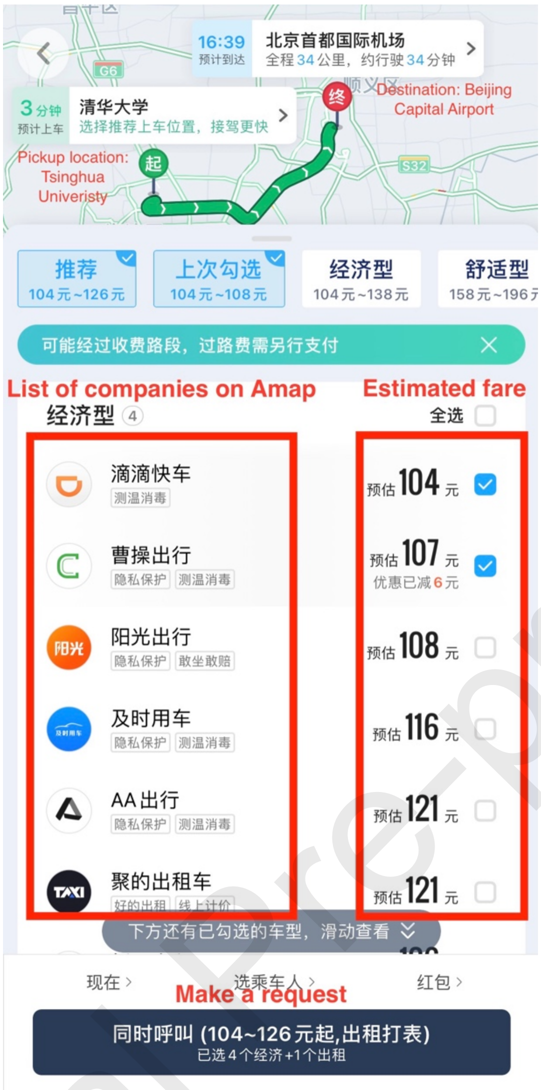

Figure 2: Screenshot of an Aggregator Hailing Request

Our trip-level data include all booking requests for a random sample of drivers in three major cities in China, Chengdu (in Sichuan Province), Shenzhen (in Guangdong Province) and Guangzhou (in Guangdong Province) from September 2019 to August 2020. Each booking request records driver ID, rider ID, job type (ride-sharing request, advanced booking, airport pickup, etc.), order source (app request or aggregator request), request time, driver confirmation time, trip start and end time, pickup location and destination geocoordinates, distance traveled in km, original price before discount, and final price after discount. Table 1, Panel A reports the summary statistics at the trip level: An average trip is 10 km and costs 35 Chinese Yuan.

Table 1: Summary Statistics

VARIABLES	Unit	N	Mean	S.D.	Q1	Median	Q3
A. Driver-trip level							
Trip distance	km	4,134,049	10.16	11.18	4.09	7.13	12.5
Trip fare	CN¥	4,133,786	34.84	49.68	14.4	22.75	38.91
B. COVID-19 city-day level							
New cases	count	678	2.01	5.27	0	0	2
Imported cases	count	678	0.62	1.57	0	0	0
Other city new cases	count	678	4.08	11.28	0	0	2
C. Driver-day level							
Whether to work	indicator	3,162,240	0.27	0.446	0	0	1
Number of requests	count	868,628	9.41	10.64	3	6	12
Number of completed trips	count	868,628	4.76	5.15	1	3	7
Average trip distance	km	868,628	10.15	12.84	5.32	8.33	12.17
Average search distance	km	868,628	5.98	5.74	2.75	4.8	7.67
Earnings	CN¥	868,628	165.78	196.3	35.32	107.05	234.86

To proxy for the severity of COVID-19, we further collect daily number of new cases (excluding imported cases) and imported cases (i.e., people who were tested positive after arriving from other countries) at the city level (i.e., Chengdu, Shenzhen, and Guangzhou) and at the province level (i.e., Sichuan Province and Guangdong Province) from the official websites of Health Commissions of the two provinces.[5] These data are available from January 18, 2020. The detailed statistics of new cases and imported cases on a given day were typically disclosed in the late evening on the same day or in the early morning (usually before 9 am) on the subsequent day. We discuss the details of how we account for such delays in new case disclosures in Section 4.2.

Table 1, Panel B reports the summary statistics of COVID-19 measures at the city-day level from January 18, 2020 to August 31, 2020: The average daily number of new cases across

[5] Data for Guangdong Province and Sichuan Province are respectively collected from the official websites of Health Commission of Guangdong Province (http://wsjkw.gd.gov.cn/zwyw_yqxx/) and Health Commission of Sichuan Province (http://wsjkw.sc.gov.cn/scwsjkw/gzbd01/ztwzlmgl.shtml).

the three focal cities is 2.01, with a standard deviation of 5.27; the average daily number of imported cases has a relatively smaller magnitude, with a mean of 0.62 and a standard deviation of 1.57. The large coefficients of variation, 2.62 (5.27/2.01) and 2.53 (1.57/0.62) respectively, suggest a considerable spatiotemporal dispersion of COVID-19 measures across cities and time. Indeed, this can be verified by the patterns presented in Figure 3, which plots the time trend of number of new cases by day in each city. Starting from January 18, 2020, the number of new cases in all cities increased drastically and quickly reached the peak around February 1, 2020. During this early outbreak, Chengdu had the largest number of new cases, followed by Guangzhou, with Shenzhen the least affected by the COVID-19. Since the beginning of February, the number of new cases decreased sharply and by March 1, 2020, almost all cities had close to zero new case. Afterward, there were two salient waves of surge in new cases respectively in April and June. The large spatiotemporal variation of COVID-19 measures provides us with sufficient statistical power to identify the causal effect of COVID-19 on driver behavior.

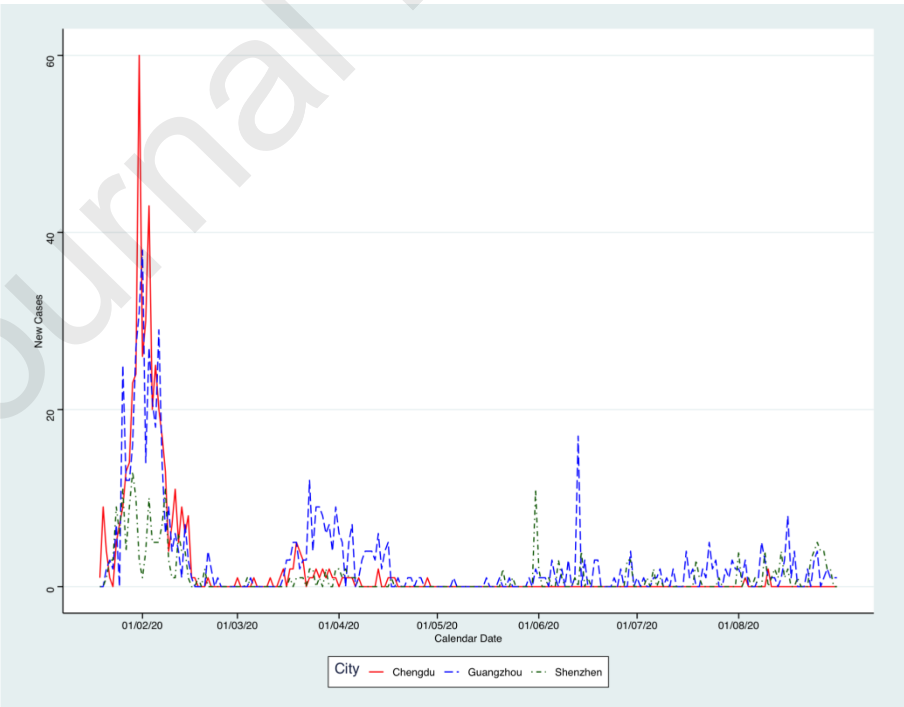

Figure 3: Number of New Cases in Each City by Day

4. Empirical Settings

4.1 Driver Behavior Measures

To facilitate our empirical analysis of drivers' responses to COVID-19, we follow the literature (e.g., Farber, 2008) and further aggregate trips into a higher level for each driver so that we can measure both extensive margin (i.e., whether to work) and intensive margin (i.e., how much to work) of drivers' labor supply. As our COVID-19 measures vary at the daily level, we aggregate the trip level data into driver-day level. Specifically, we construct the following driver-day level measures which serve as the dependent variables in our empirical analysis.

Whether or not to work. We construct the variable as a binary indicator, which equals 1 if a driver has at least one ride request and 0 otherwise. Following the economics literature (e.g., Farber, 2015), we use this variable to measure drivers' shift decision, i.e., willingness to work on a day, which proxies for the extensive margin of drivers' labor supply. It is ambiguous ex-ante how the number of new cases affects a driver's shift decision. On the one hand, more new cases may increase the risk of infection, which decease drivers' expected wellbeing, and therefore discourage drivers from working on a specific day; on the other hand, fewer drivers on the street suggest less competition among drivers and therefore higher chances of getting a passenger and potentially higher hourly earnings, which may motivate drivers to work (Camerer et al., 1997). It is important for the ride-sharing company to understand how the severity of COVID-19 affects drivers' willingness to work, so that the company can adjust their stimulus plans for drivers accordingly.

Total number of requests and number of completed orders. Both variables contain three aspects of information which are of policy and managerial interest. First, both variables can proxy for the length of drivers' daily labor supply. Conditional on working, if a driver decides to work for longer hours, then we expect the driver to have a larger number of requests/orders. Second, both variables contain information on consumer demand. We expect the total number of requests/orders to decrease if there is a lower demand for ride-sharing service from consumers due to the COVID-19 outbreak. Finally, both variables can measure the intensity of competition among drivers. Keeping the level of demand fixed, the total number of requests/orders would be larger when there are fewer drivers working on the day. Due to the complexity of information contained, ex-ante, it is not straightforward how the COVID-19 measures affect the total number of requests/orders for individual drivers.

Average trip distance. In our empirical context, drivers cannot reject a booking request once being matched with a passenger, therefore, the trip distance is largely determined by passengers. Since passengers may be reluctant to take long distance trips during the pandemic, we expect a negative impact of the number of new cases on the *average trip distance*.

Average search distance. Our trip data record the detailed geocoordinates of pickup location and destination for each trip. We then compute the line distance from a previous trip's destination to current trip's pickup location as a proxy for drivers' search diameter. The impact of new cases on drivers' search scope is unclear ex-ante: On the one hand, the significant decrease in consumer demand for ride-sharing service may motivate drivers to prolong and extend their search for passengers and therefore increase the *average search distance* by drivers; on the other hand, risk-averse drivers may choose to limit the search scope in order to reduce the infection risk.

Earnings. Earnings measure the driver's income from providing ride-sharing services, which is highly correlated with the number of completed orders and trip distance. It allows us to directly assess the impact of the COVID-19 on drivers' financial wellbeing.

The summary statistics for the above driver-day level labor supply measures are reported in Panel C, Table 1. On average, a driver receives 9.41 booking request, completes 4.76 trips, and make 165.78 Chinese Yuan on a typical day during our data period. We also plot the mean of each labor supply measure in each city by day in Figure 4.

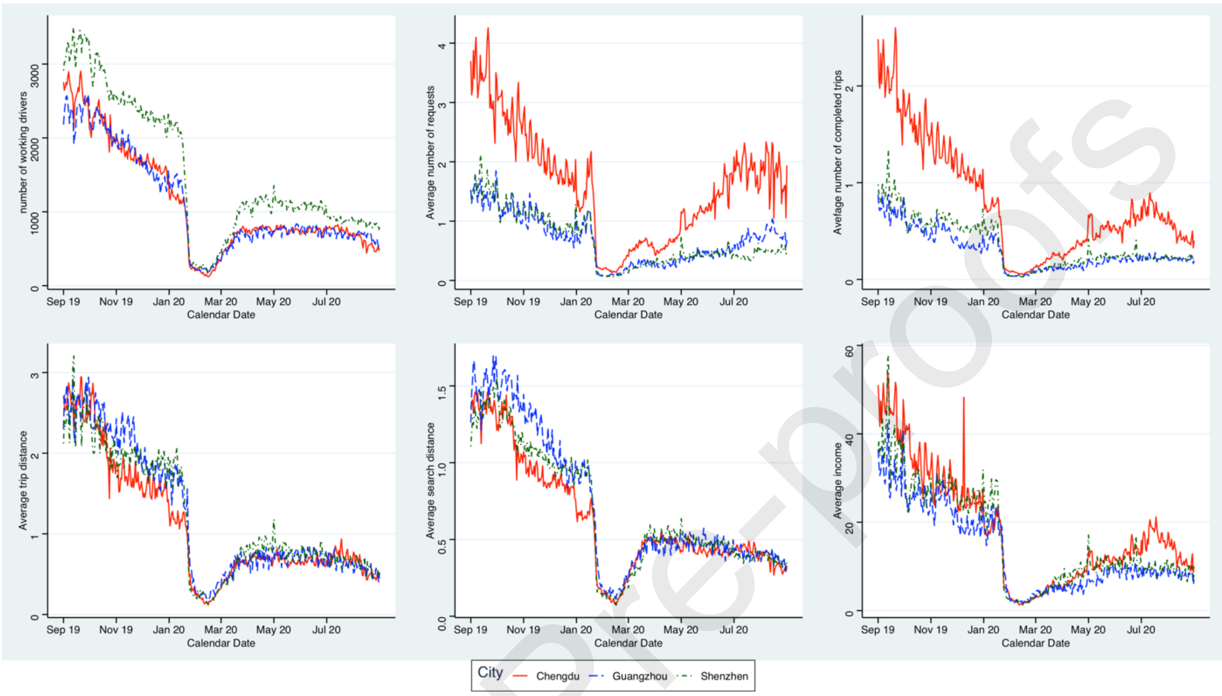

Figure 4: Average Labor Supply Measures Across Drivers in Each City by Day

4.2 Empirical Setup

To empirically investigate the causal impact of COVID-19 cases on driver behavior, we regress the labor supply measures of driver *i*, in city *j*, on day *t* on the COVID-19 measure and other covariates as follows:

(1) $$LS_{ijt} = \beta_0 + \alpha NewCases_{j,t-1} + \eta_i + \gamma_{jw(t)} + \varepsilon_{ijt},$$

where LS_{ijt} is the full set of drivers' daily labor supply measures discussed in Section 4.1, including *whether or not to work*, *total number of requests*, *total number of completed orders*, *average trip distance*, *average search distance, and earnings*. $NewCases_{j,t-1}$ is the variable of interest, which is the number of newly confirmed COVID-19 cases in city *j* on day *t-1* disclosed by the Health Commissions; and α captures the impact of lagged new cases on driver behavior, which is the key parameter of interest. We estimate the effect of new cases on the previous day instead of new cases on the focal day because drivers make decisions on whether and how long to work based on their perceived risk of working. Given the fact that the government typically discloses information on new case of a specific day in the late evenings of that day or early mornings of the next day, the number of new cases on day *t-1* should represent the best information for the drivers' decisions on day *t*.[6] η_i is a driver fixed effect and $\gamma_{jw(t)}$ is a city-week fixed effect where $w(t)$ stands for the week that day t falls into. ε_{ijt} is an idiosyncratic error term.

We first include driver fixed effects to control for driver-specific characteristics that may affect drivers' labor supply patterns. Such characteristics include, but are not limited to, the driver's sociodemographic characteristics (e.g., gender and age), the driver's degree of risk aversion, whether a driver is driving full-time or part-time, and the driver's innate abilities to search for passengers, etc. For instance, less risk-averse drivers may prefer to work on days when there are more new cases because they expect less competition from peer drivers and potentially higher profitability on such days. Another example is that, full-time drivers can be more subject to the impact of new cases compared to part-time drivers, because full-time drivers' income largely comes from providing ride-sharing services via the focal company. Driver fixed effects can mitigate such driver-specific time-invariant confounding effects and help us obtain more accurate estimates for our focal explanatory variable $NewCases_{j,t-1}$.

In addition to driver fixed effects that remove cross-sectional confounding effects across drivers, we also include time fixed effects in Equation (1) to mitigate the intertemporal confounding effects. Since our focal explanatory variable $NewCases_{j,t-1}$ varies at day level, to avoid multicollinearity problem, we consider time fixed effects at the week level. Moreover, given that the local government in each city may have enacted different policies on fighting

[6] We note that our main results are also robust to using $NewCases_{j,t}$, i.e., the non-lagged new cases on day *t*. The results are available upon request.

COVID-19 and/or stimulating economy (e.g., subsidizing drivers) during our data period, we further allow the week fixed effects to be city-specific, that is, we include city-week fixed effects in Equation (1) to control for any city-specific intertemporal confounding effects.

4.3 Endogeneity and Instrumental Variable Method

After including the driver fixed effects and city-week fixed effects in Equation (1), the only challenge to obtaining causal inference is the potential endogeneity of $NewCases_{j,t-1}$. Equation (1) could be subject to simultaneity issues because drivers' labor supply decisions and number of new cases may be interdependent. On the one hand, drivers may adjust their labor supply accordingly to the number of new cases. On the other hand, prior research has demonstrated the potential effect of mobility on the COVID-19 case growth rate (Badr et al., 2020; Wielechowski et al., 2020). If a city has a higher volume of private transportation through ride-sharing services, given the highly contagious nature of COVID-19, the city may have a higher number of new cases. Although drivers' labor supply would not affect the number of new cases on the previous day, dynamics among the error term ϵ_{ijt} may still lead to incorrect inferences (Bellemare et al, 2017), and should be accounted for in the empirical analysis.

To tackle the potential endogeneity issue, we use the instrumental variable (IV) method, leveraging exogenous sources of variation in the explanatory variable that are uncorrelated with the error term in Equation (1) using two-stage least squares (2SLS). We select two instrumental variables. The first instrumental variable is *imported new cases*, which measures the number of infected travelers from overseas in each city as disclosed by Health Commissions. Because the imported cases relate to travelers from overseas, it should be exogenous to local confirmed cases and meet the exclusion restriction. The second instrumental variable is *other city new cases*, which is the number of new cases confirmed in other cities within the same province. Take Chengdu as an example: On January 31, 2020, Chengdu had 3 new COVID-19 cases, and Sichuan Province, which Chengdu belongs to, had 33 new cases; therefore, *other city new cases* takes the value of 30 for drivers in Chengdu on January 31, 2020. Since confirmed cases in other cities within Sichuan province should not directly affect Chengdu's ride-sharing market, the variable *other city new cases* should also satisfy the exclusion restriction.

The first-stage regression is specified below in Equation (2), where the definitions of variables are the same as in Equation (1).

$$Newcases_{ijt} = \theta_0 + \theta_1(imported\ new\ cases)_{ijt} + \theta_2(other\ city\ new\ cases)_{ijt} + \cdots + \gamma_{jw(t)} + \varepsilon_{ijt}. \tag{2}$$

As discussed above, both our instrumental variable candidates satisfy the exclusion restriction as they do not directly affect the ride-sharing market of cities under study, so that they are valid instruments. At the same time, they also meet the relevance requirement: First, during the period from January 18, 2020 to August 31, 2020, the Pearson's correlation coefficient between *NewCases* and *imported new cases* is 0.199 (p-value < 0.0001), and the Pearson's correlation coefficient between *NewCases* and *other city new cases* is 0.835 (p-value < 0.0001). Both correlation coefficients suggest a significant and positive correlation between the endogenous variable and IVs. Second, the first-stage regression results (Table A1 in the Online Appendix) show that the estimated coefficients for both IVs are statistically and economically significant, with a large adjusted R-squared. Finally, following the standard procedure of IV method (Greene, 2012), we perform the under-identification test and weak-IV test and report the test statistics in Table A1. Both the under-identification test statistics (Kleibergen-Paap rk LM statistics) and the weak-IV test statistics (Kleibergen-Paap Wald rk F statistics) reject the null hypothesis of under-identification or weak IVs, which further confirms the relevance condition of our selected IVs.

From the above analysis, we conclude that our instrumental variables can solve the endogeneity issue in Equation (1) using 2SLS. For ease of interpretation, in all 2SLS regressions, we normalize all continuous independent variables to have a mean of zero and standard deviation of one. Also, we compute and report the robust standard errors clustered at the driver level to deal with potential serial correlation in driver behavior.

5. Results

In this section, we report our estimation results and present a comprehensive picture of how COVID-19 has affected driver behavior during the pandemic. We first present in Section 5.1 the 2SLS results for Equation (1), which depict the average treatment effects of new cases on both the extensive and intensive margins of drivers' labor supply. In Section 5.2, we then take advantage of the rich set of sociodemographic information we have to further explore the heterogeneous treatment effects of new COVID-19 cases across drivers and across different phases of the pandemic.

5.1 Main effect of COVID-19 on drivers' labor supply and earnings

Unlike the conventional taxi market, where taxi drivers rent vehicles from taxi companies and then directly provide transportation service to consumers, modern ride-sharing platforms, including the focal company under study, typically serve as the matching intermediary between drivers and passengers and there exists significant network effect. Therefore, drivers' labor supply is crucial to the survival and growth, and it is managerially important for ride-sharing platforms to understand whether COVID-19 has a significant impact on drivers' labor supply patterns and if yes, the direction and magnitude of the effect.

We first examine drivers' shift decision. As discussed in Section 4.1, two competing forces may drive our results. On the one hand, when there are more new COVID-19 cases, drivers may choose not to work so as to reduce the infection risk; on the other hand, less risk-averse drivers may increase their labor supply in response to lower expected competition from peer drivers. In the end, the impact of COVID-19 on drivers' shift decisions is unclear. We look for answers from our empirical analysis using Equation (1).

The 2SLS results for Equation (1) are reported in Table 2. To conserve space, we report the OLS results without the use of instrumental variables in Tables A2 and A3 in the Online Appendix. In Column (1), Table 2, we present the estimated coefficient for *new cases* with the dependent variable being drivers' shift decisions *whether or not to work*. The coefficient is -0.013 (s.e. 0.000) and statistically significant at 1% level. Since Column (1) is essentially a linear probability model,[7] the interpretation for the coefficient is that, a one standard deviation increase in new cases (i.e., 4.39 new cases) decreases drivers' probability of working on a day by 1.3 percentage point (i.e., a 4% decrease relative to the average probability of working on a day). The negative sign indicates that drivers' aversion to the infection risk outweighs the potentially higher profitability due to lower competition. However, despite the statistical significance, the economic magnitude of the coefficient seems moderate. One plausible reason is that the majority of drivers in our sample are full-time drivers at the company and hence do not have much flexibility in terms of shift decisions. This result can partially relieve ride-sharing platform's concern that drivers may greatly reduce their number of shifts during the pandemic. More importantly, our finding has a strong policy implication for the government and the ride-sharing companies: if they were to enact stimulus or recovery plans under limited budget, then they should focus more on drivers' working time than on their willingness to work, as drivers tend to be relatively less flexible in adjusting their shift decision.

[7] Although the dependent variable is a binary indicator, we estimate the linear probability model instead of a logistic model because the large number of fixed effects renders it computationally burdensome to estimate a logistic model.

Table 2: Impact of New COVID-19 Cases on Drivers' Labor Supply Measures

VARIABLES	(1) Whether to work	(2) Number of requests	(3) Number of completed trips	(4) Average trip distance	(5) Average search distance	(6) Earnings
Lagged New Cases	-0.013***	0.206***	-0.108***	-0.069**	-0.009	-3.540***
	(0.000)	(0.035)	(0.011)	(0.035)	(0.019)	(0.446)
Driver fixed effects	Y	Y	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y	Y	Y
Mean dependent variable	0.275	9.408	4.759	10.155	5.978	165.781
Coef. / Mean DV	-4%	2.2%	-2.3%	-0.6%	-0.15%	-2.1%
Hansen *J* statistic	0.155	173.4	9.391	2.756	0.00757	31.08
No. observations	3,162,240	868,628	868,628	868,628	868,628	868,628
R-squared	0.358	0.455	0.504	0.271	0.120	0.406

Notes. This table reports the second-stage estimation results in 2SLS for the impact of lagged number of new cases on driver behavior specified in Equation (1). The independent variable is the one-day-lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level. ** Significant at the 5% level.

Next, we examine labor supply measures that are directly related to driver earnings. Note that on days when drivers do not work, our labor supply measures have the value of 0. To avoid the inflation of zeros and capture of impact of COVID-19 more precisely, we condition our remaining analyses on days when drivers work. In Column (2), we report the estimation result with *total number of requests* as the dependent variable. Surprisingly, we discover a positive coefficient for *new cases*, 0.206 (s.e. 0.035), which is statistically significant at 1%—When there is a one standard deviation increase in new cases (i.e., 2.25 new cases in the conditional sample) total booking requests received by drivers do not decrease but increase by 0.206 (i.e., a 2.2% increase relative to the average number of requests). What is more intriguing is that, the coefficient in Column (3), where the dependent variable is *total number of completed trips*, has a negative and statistically significant coefficient of -0.108 (s.e. 0.011), which is opposite to the sign of coefficient in Column (2). It seems that when the COVID-19 situation is more severe, despite that each driver receives more booking requests on a day, the number of fulfilled trips actually decreases.

To better understand the mechanisms underlying the seemingly contradictory results above, we further decompose the *total number of requests* and *completed trips* based on the source of request. As introduced in Section 3, the focal company provides its booking service via two channels, the aggregator channel and the app channel. Therefore, we decompose the *total number of requests/completed trips* into aggregator channel and app channel and report the results in Table 3. The decomposition indeed reveals a clearer picture of how the pandemic has affected the ride-sharing business: We observe a significant increase (decrease) in requests received through the aggregator (app) channel. One plausible explanation is that when the COVID-19 is more severe, expecting fewer available drivers from a single service provider, passengers may have switched to aggregator channels so that they can select from a wider range of ride-sharing companies. Such switching behavior eventually leads to an increase in the aggregator channel requests. Nevertheless, despite the increase in requests, the actual number of completed trips decreases for drivers in our data, which suggests a stronger competition among different ride-sharing service providers during the pandemic. The coefficients for requests and completed trips via the app channel, reported in Columns (3) and (4) have consistent signs: The increase in new COVID-19 cases has a negative impact on both requests and completed trips in the app channel. The decrease could be due to a lower transportation demand during the pandemic or passengers' switching to aggregator channels or competing companies, or both.

Table 3: Decomposition by Source of Request

VARIABLES	(1) Number of aggregator requests	(2) Number of completed aggregator trips	(3) Number of app requests	(4) Number of completed app trips
Lagged New Cases	0.224***	-0.092***	-0.018***	-0.016***
	(0.034)	(0.010)	(0.004)	(0.003)
Driver fixed effects	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y
Mean dependent variable	8.368	3.955	1.040	0.804
Coef. / Mean DV	2.7%	2.3%	1.7%	1.7%
No. observations	868,628	868,628	868,628	868,628
R-squared	0.454	0.497	0.332	0.331

Notes. This table reports the 2SLS results for the decomposition of number of requests and completed trips by source of trip request. The independent variable is the one-day-lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level.

Another measure that indirectly affects drivers' earnings is *average trip distance*. In our empirical setting, drivers can only passively accept booking requests and hence cannot select into trips based on the trip distance. In other words, *average trip distance* is largely determined by the demand side and can reflect on the structural change in passengers' perceived mobility needs (i.e., shorter or longer trips). Previous studies show that commuters' perceived mobility needs are sensitive to many factors, such as the complexity of their daily activity patterns (Thorhauge et al., 2020); we expect that the pandemic should have a pronounced impact on commuters' mobility needs as well. At the same time, *average trip distance* can affect drivers' earnings in that long-distance trips can help drivers reduce idle time and may eventually increase their shift earnings. In Column (4), we report the estimation result with *average trip distance* as the dependent variable: A one standard deviation increase in new cases (2.25 cases) decreases the average trip distance by 0.069 km (s.e. 0.035). As the risk of infection increases with the time people stay in the same confined space (e.g., in the same vehicle), passengers seem to avoid long-distance trips when the COVID-19 situation is more severe.[8]

We next examine how the pandemic has affected drivers' search behavior, which serves as a proxy for the intensive margin of drivers' labor supply. Ex-ante, it is unclear how the severity of COVID-19 may affect drivers' scope of search for passengers, as two competing forces may play a role: On the one hand, expecting a lower demand, drivers may extend their search to more locations in order to reach their earnings target; on the other hand, more COVID-19 cases may discourage drivers from driving across too many regions in order to reduce the infection risk. Which effect dominates is an empirical question. Column (5), Table 2 indicates that the average effect of lagged new cases on drivers' search scope is statistically insignificant. However, as we will show in the subsample analyses in Section 5.2, the overall statistical insignificance is due to the heterogeneous, opposite effects across cities, which offset each other in the pooled sample.[9] Such salient heterogeneity is likely due to the difference in the local government policies at the city level. For instance, the Shenzhen government and Shenzhen taxi companies began to proactively subsidize taxi drivers starting from late January 2020,[10] which encourages drivers to actively search for passengers.

Finally, we examine the impact of new cases on the direct measure of drivers' daily earnings. The estimation result is presented in Column (6), Table 2. Unsurprisingly, as the consequence of decreased number of completed orders, lower average trip distance, and shrinking scope of search, we find a significant and negative impact of COVID-19 on drivers'

[8] We acknowledge that the average treatment effect on average trip distance seems small in relative terms—0.069 km is approximately 0.6% of the mean of average trip distance (10 km); however, our later heterogeneity analysis in Section 5.3 suggests that this is a result of salient intertemporal heterogeneity of the treatment effect: In the early phase of the pandemic (before March 1st, 2020), the treatment effect is as large as 2% of the mean of average trip distance.

[9] In the subsample analysis with Chengdu and Guangzhou (Column (5), Table A6), we find a negative and significant effect of new cases on drivers' average search distance. In contrast, in the subsample analysis with Chengdu and Shenzhen (Column (5), Table A8), we find a positive effect.

[10] See: https://finance.sina.com.cn/china/dfjj/2020-02-06/doc-iimxxste9213222.shtml

daily earnings, with a magnitude of 3.540 (s.e. 0.446). Given that the average daily earnings in the sample are 165.78 Chinese Yuan, a one standard deviation increase in new cases decreases drivers' daily earnings by 2.1%.

5.2 Robustness checks

To this end, we have fully investigated the causal impact of COVID-19 on both the extensive and intensive margins of drivers' labor supply. Before discussing the heterogeneity analysis, we test the robustness of our main findings to alternative IV specifications and subsample analyses. First, within the two selected IVs, *imported cases* is in theory more likely to satisfy the exclusion restriction than *other city new cases*, thus we test the robustness of our findings to using *imported cases* as the only IV. The results are reported in Table A4 and A5 in the Online Appendix. All results are consistent with our main findings.

Second, since the number of IVs is larger than the number of endogenous variables, we report the Hansen's *J*-statistics in Table 2 to test the over-identifying restrictions. The joint null hypothesis is that all IVs are uncorrelated with the error term and are correctly excluded from the estimated equation. We find that the *J*-statistics are overall small in magnitude and statistically insignificant except for Columns (2) and (6), further supporting the validity of both *other city new cases* and *imported cases*.

Finally, because Guangzhou and Shenzhen are in the same province, a potential concern is that new cases in one city will be counted towards *other city new cases* for the other city. We rule out such a concern by running two subsample analyses, replicating our main analyses excluding either Guangzhou or Shenzhen. The results which are reported in Tables A6 to A9 in the Online Appendix indicate that our main findings remain unchanged.

Based on the above empirical evidence, we are confident to conclude that our chosen IVs are valid instruments, and that our main results reported in Section 5.1 are robust.

5.3 Heterogeneity analysis

Since its outbreak, the COVID-19 pandemic has undoubtedly brought catastrophic impact on numerous industries, especially the transportation industry. One crucial question for both policymakers and ride-sharing platform managers is how to promulgate appropriate policies to speed up the recovery process during the post-COVID era. For this purpose, it is important for stakeholders to fully understand how the severity of COVID-19 may have heterogeneous effects across drivers of different characteristics and across different phases of the pandemic, so that they can enact more targeted policies. We provide answers to this question in this section.

We first look at whether drivers' sociodemographic characteristics moderate the impact of COVID-19 using Equation (3):

$$LS_{ijt} = \beta_0 + \alpha NewCases_{j,t-1} + \gamma NewCases_{j,t-1} * Char_i + \eta_i + \gamma_{jw(t)} + \varepsilon_{ijt}, \quad (3)$$

where similar to Equation (1), LS_{ijt} is the driver's labor supply measure for driver i, in city j, on day t. $NewCases_{j,t-1}$ is the lagged number of new confirmed COVID-19 cases in city j on day t. $Char_i$ is a measure of driver characteristics, including years of working experience at the focal company, premium status at the focal company (i.e., whether the driver is certified as a "gold driver"[11]), number of requests prior to the COVID-19 outbreak, and years of driving experience. γ is the coefficient for the interaction term between $NewCases_{j,t-1}$ and $Char_i$, which captures the moderation effect of driver characteristics on the impact of new cases. We also include driver and city-week fixed effects as in Equation (1). We estimate Equation (3) for each driver characteristic. Similar to Section 5.1, we deal with endogeneity by implementing 2SLS estimation, using *imported new cases* and *other city new cases* to instrument the number of new cases in the city.

[11] As an incentive, the company rewards drivers with the "gold status" for meeting a set of performance goals. As the gold status is rewarded based on periodical reviews, a driver can also lose the gold status. We construct a binary indicator for gold drivers, which takes the value of 1 if the driver had the gold status before the pandemic started.

Table 4, panels A, B, C and D report the 2SLS estimation results when $Char_i$ measures working experience, premium status, pre-COVID requests, and driving experience, respectively. Columns (1) to (6) correspond to the same set of dependent variables as in Table 2.

Table 4: Impact of New COVID-19 Cases on Drivers' Labor Supply Measures (with Moderating Effects)

	(1)	(2)	(3)	(4)	(5)	(6)
	Whether to work	Number of requests	Number of completed trips	Average trip distance	Average search distance	Earnings
	Panel A: Working experience					
Lagged New Cases	-0.013***	0.177***	-0.116***	-0.175**	0.003	-2.753***
	(0.000)	(0.035)	(0.012)	(0.073)	(0.020)	(0.575)
*Lagged New Cases * Exp*	0.002***	0.060***	0.016*	0.192	-0.027	-1.653*
	(0.001)	(0.017)	(0.009)	(0.126)	(0.017)	(0.853)
	Panel B: Gold driver status					
Lagged New Cases	-0.011***	0.312***	-0.016	-0.143***	-0.040**	0.941*
	(0.001)	(0.035)	(0.012)	(0.055)	(0.020)	(0.534)
*Lagged New Cases * Gold driver*	-0.008***	-0.394***	-0.340***	0.261	0.110***	-16.582***
	(0.002)	(0.039)	(0.021)	(0.167)	(0.032)	(1.428)
	Panel C: Pre-COVID requests					
Lagged New Cases	-0.017***	0.176***	-0.134***	-0.063*	-0.001	-4.491***
	(0.000)	(0.037)	(0.013)	(0.035)	(0.019)	(0.508)
*Lagged New Cases * Pre-Covid requests*	-0.021***	-0.320***	-0.276***	0.067	0.081***	-10.207***
	(0.001)	(0.023)	(0.016)	(0.052)	(0.012)	(0.714)
	Panel D: Driving experience					
Lagged New Cases	-0.019***	0.159***	-0.147***	-0.068*	0.005	-4.887***
	(0.001)	(0.038)	(0.013)	(0.035)	(0.019)	(0.528)
*Lagged New Cases * Driving exp*	-0.037***	-0.435***	-0.347***	0.001	0.114***	-12.197***
	(0.001)	(0.026)	(0.013)	(0.033)	(0.013)	(0.635)
	Panel E: Intertemporal effect					
Lagged New Cases	-0.025***	-0.120***	-0.163***	-0.203**	-0.054	-7.850***
	(0.001)	(0.028)	(0.014)	(0.080)	(0.042)	(1.010)
*Lagged New Cases * Post March 1*	0.029***	0.460***	0.086***	0.196**	0.045	6.521***
	(0.001)	(0.048)	(0.019)	(0.088)	(0.047)	(1.154)
	Panel F: Emergency response level					
Lagged New Cases	0.000	-0.088*	-0.065***	-0.775***	-0.386***	-14.017***
	(0.001)	(0.048)	(0.025)	(0.206)	(0.099)	(2.193)
*Lagged New Cases * Level 2*	0.007***	-0.193**	-0.062	0.846***	0.373**	9.254***
	(0.002)	(0.076)	(0.039)	(0.274)	(0.159)	(2.938)
*Lagged New Cases * Level 3*	-0.008	-0.939***	0.026	1.816***	0.543*	12.115**
	(0.006)	(0.260)	(0.129)	(0.490)	(0.326)	(5.579)
Driver fixed effects	Y	Y	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y	Y	Y

Notes. This table reports the results for heterogeneity effect of lagged new COVID-19 cases on driver behavior. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Panel A focuses on drivers' years of working experience (standardized) at the focal company. Column (1) indicates that drivers' working experience positively moderates the impact of COVID-19 on drivers' shift decision, possibly because more experienced drivers rely on providing ride-sharing services to make a living and therefore have less flexibility in terms of their shift decisions. Columns (2) and (3) suggest that working experience helps drivers cope with the bad situations: More experienced drivers were able to receive more requests and orders during the pandemic. However, Column (6) indicates a larger decrease in earnings for more experienced drivers, possibly because their earnings level was higher prior to the COVID-19 outbreak.

Panel B focuses on drivers' premium status at the focal company. Results indicate that compared to drivers without the gold status, drivers with gold status experienced larger decreases in number of requests, number of completed trips, and earnings. This is likely because gold drivers are more likely to be full-time drivers who rely heavily on the focal company to make a living and have higher pre-pandemic requests, trips and earnings. Our speculation is further corroborated by the result in Columns (5)—The effect of COVID-19 on average search distance is positive for drivers with gold status, suggesting that these drivers are more willing to increase their scope of search for passengers despite the higher infection risks, probably because working for the focal company is their major source of income.

Panel C focuses on drivers' number of requests prior to the COVID-19 outbreak (*pre-COVID requests)*, which is measured during the baseline period of Sep 1, 2019 to Jan 1, 2020 and is standardized. *pre-COVID requests* can be considered as the active level of the driver in the baseline period, which not only proxies for drivers' contribution to the focal company, but also measures how heavily drivers rely on providing ride-sharing services to make a living. Results are largely similar to those in Panel B.

Panel D focuses on drivers' (standardized) years of driving experience, which can be used to proxy driver age. Again, results are largely similar to those in Panels B and C.

Next, we look at whether the effect varies across different phases of the pandemic by estimating the following equation:

$$LS_{ijt} = \beta_0 + \alpha NewCases_{j,t-1} + \gamma NewCases_{j,t-1} * Post_t + \eta_i + \gamma_{jw(t)} + \varepsilon_{ijt}, \quad (4)$$

where $Post_t$ takes the value of 1 if day t is after March 1. Other variables are the same as in Equation (1). Results in Panel E of Table 4 indicate that in the post-COVID phase, each newly confirmed case has a smaller effect on shift decision, completed trips and earnings, possibly because drivers are better prepared for living and working with the pandemic in the post-COVID period.

Since the COVID-19 outbreak, the Chinese government has been adjusting the level of public health emergency response for each city, with first-level emergency response being the highest level, followed by second- and third-levels. Such emergency levels can affect drivers' perception on COVID-19 risk, and we report in Panel F of Table 4 the moderating effects of emergency levels which vary by city and day. Results indicate that compared with the most severe level (Level 1), when there are more new cases, at Level 2, drivers are more likely to increase shifts, drive longer trips, increase average search distance and receive more earnings; at Level 3, drivers are more likely to drive longer trips, increase average search distance and receive more earnings. As discussed earlier, new cases have two countervailing effects: the infection risk effect (i.e., more new cases may increase the risk of infection, which discourages drivers from working), and the competition effect (i.e., fewer drivers on the street suggest less competition among drivers and therefore higher chances of getting a passenger and potentially higher hourly earnings, which may motivate drivers to work). Our results indicate that when the COVID-19 situation is the most severe, the infection risk effect outweighs the competition effect. However, when the pandemic eases, the competition effect takes over, motivating drivers to increase labor supply.

Overall, the results in Panels E and F, Table 4 suggest that in the post-COVID phase, the impact of each newly confirmed case has a smaller effect on search scope and earnings, which implies a recovery of the ride-sharing industry alongside the recovery of the economy after the pandemic.

6. Discussion and Conclusion

This paper investigates the causal effect of the COVID-19 pandemic on driver behavior, especially their labor supply patterns, in one of the leading ride-sharing companies in China. Because China is one of the earliest major economies to recover from the pandemic and the data cover pre-, during-, and post-pandemic phases, we are able to empirically assess and quantify the effect of the COVID-19 on several key dimensions of driver behavior. In addition, we take advantage of the rich set of sociodemographic information in the data to explore the heterogeneous effects of new COVID-19 cases across drivers and across different phases of the pandemic.

We find that on average, drivers are risk averse: as the number of new cases increase, drivers are less likely to work on a given day, and conditional on working, reduce search scope, complete fewer trips and the trips are shorter on average, and as a result, receive lower earnings. In addition, there appear to be more intensive competition from other ride-sharing companies during the period we study.

Moreover, the effects appear to vary both across drivers and over time. We find the effects to be generally more prominent for older drivers, drivers who are more experienced and more active, and those who have higher status within the firm. These drivers tend to contribute more to the focal company, and also rely more on providing ride-sharing services to make a living. Therefore, they should be prioritized in stimulus plans offered by the government or the ride-sharing company.

Finally, we find that in the post-COVID phase, each newly confirmed case has a smaller effect on driver shift, completed trips and earnings, implying the recovery of the ride-sharing industry alongside the recovery of the economy after the pandemic.

There are several limitations to the present study that represent opportunities for future research. First, we measure driver earnings directly with the total trip fare on a day, and have ignored possible subsidies from the ride-sharing company or from the government. Should subsidy data be available, it would yield more precise results by incorporating subsidy into the analyses. Second, absent real-time GPS trajectory data, we are unable to directly measure drivers' scope of search but have to use the map distance between last trip's destination and current trip's pickup location as a proxy for drivers' activeness of search. Hence, further research with detailed GPS trajectory data can be used to corroborate our result on driver search behavior. Third, due to data limitation, we have mainly focused on the impact of the COVID-19 pandemic on the ride-sharing supply, and our analysis on the impact of COVID-19 on the ride-sharing demand has been limited on the average trip distance. It would be interesting to collect passenger data and directly explore the impact of the pandemic on the demand side, especially when passengers act as behavioral commuters (Lien et al., 2020). Finally, our data are from one ride-sharing company, so that we have abstracted away from competition between ride-sharing companies and drivers' multi-homing behavior. Measuring the effect of COVID-19 on competition among ride-sharing companies is a topic that could be interesting for future research.

REFERENCES

Allon, G., Cohen, M., Sinchaisri, W.P., 2018. The impact of behavioral and economic drivers on gig economy workers. *Available at SSRN 3274628*.

Badr, H.S., Du, H., Marshall, M., Dong, E., Squire, M.M., Gardner, L.M., 2020. Association between mobility patterns and covid-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases* 20, 1247-1254.

Bartik, A.W., Bertrand, M., Cullen, Z., Glaeser, E.L., Luca, M., Stanton, C., 2020. The impact of covid-19 on small business outcomes and expectations. *Proceedings of the National Academy of Sciences* 117, 17656-17666.

Bellemare, M. F., Masaki, T., & Pepinsky, T. B. (2017). Lagged explanatory variables and the estimation of causal effect. *The Journal of Politics*, *79*(3), 949–963.

Blundell, R., Costa Dias, M., Joyce, R., Xu, X., 2020. Covid-19 and inequalities. *Fiscal Studies* 41, 291-319.

Brodeur, A., Nield, K., 2018. An empirical analysis of taxi, lyft and uber rides: Evidence from weather shocks in nyc. *Journal of Economic Behavior & Organization* 152, 1-16.

Buchholz, N., 2017. Spatial equilibrium, search frictions and efficient regulation in the taxi industry.

Cachon, G. P., Daniels, K. M., & Lobel, R. (2017). The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity. *Manufacturing & Service Operations Management*, *19*(3), 368–384. https://doi.org/10.1287/msom.2017.0618

Castillo, J. C. (2018). Who Benefits from Surge Pricing? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3245533

Camerer, C., Babcock, L., Loewenstein, G., Thaler, R., 1997. Labor supply of new york city cabdrivers: One day at a time. *Quarterly Journal of Economics* 112, 407-441.

Chen, M.K., Rossi, P.E., Chevalier, J.A., Oehlsen, E., 2019. The value of flexible work: Evidence from uber drivers. *Journal of Political Economy* 127, 2735-2794.

Chen, Z., van Lierop, D., Ettema, D., 2020. Exploring dockless bikeshare usage: A case study of beijing, china. *Sustainability* 12, 1238.

Cramer, J., Krueger, A.B., 2016. Disruptive change in the taxi business: The case of uber. *American Economic Review* 106, 177-182.

Crawford, V.P., Meng, J., 2011. New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. *American Economic Review* 101, 1912-1932.

Farber, H.S., 2008. Reference-dependent preferences and labor supply: The case of new york city taxi drivers. *American Economic Review* 98, 1069-1082.

Farber, H.S., 2015. Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *Quarterly Journal of Economics* 130, 1975-2026.

Ford, W., Lien, J.W., Mazalov, V.V., Zheng, J., 2019. Riding to wall street: Determinants of commute time using citi bike. *International Journal of Logistics Research and Applications* 22, 473-490.

Gourinchas, P.-O., Kalemli-Özcan, Ṣ., Penciakova, V., Sander, N., 2020. Covid-19 and sme failures. National Bureau of Economic Research.

Greene, William H. 2012. Econometric Analysis. Book, Whole. Boston: Pearson.

Hall, J.V., Krueger, A.B., 2016. An analysis of the labor market for uber's driver-partners in the united states.

Hossain, M., 2020. The effect of the covid-19 on sharing economy activities. *Journal of Cleaner Production*, 124782.

Jang, D.-C., Kim, B., & Lee, S.-Y. (2018). A two-sided market platform analysis for the electric vehicle adoption: Firm strategies and policy design. *Transportation Research Part D: Transport and Environment*, *62*, 646–658. https://doi.org/10.1016/j.trd.2018.02.002

Kim, O., Parker, J.A., Schoar, A., 2020. Revenue collapses and the consumption of small business owners in the early stages of the covid-19 pandemic. *NBER Working Paper*.

Lien, J.W., Mazalov, V.V., Zheng, J., 2020. Pricing equilibrium of transportation systems with behavioral commuters. *Journal of Dynamics & Games* 7, 335.

Miao, W., Chu, J., 2020. Flat-rate pricing, incentive compatibility, and efficiency in the taxi market.

Miao, W., Liu, Y., Wang, W., Deng, Y., & Tang, C. S. (2021). The Effects of Surge Pricing on Driver Behavior in the Ride-Sharing Market: Evidence from a Quasi-Experiment. *Available at SSRN 3909409*.

Nian, G., Peng, B., Sun, D.J., Ma, W., Peng, B., Huang, T., 2020. Impact of covid-19 on urban mobility during post-epidemic period in megacities: From the perspectives of taxi travel and social vitality. *Sustainability* 12, 7954.

Polyakova, M., Kocks, G., Udalova, V., Finkelstein, A., 2020. Initial economic damage from the covid-19 pandemic in the united states is more widespread across ages and geographies than initial mortality impacts. *Proceedings of the National Academy of Sciences* 117, 27934-27939.

Rysman, M., 2009. The economics of two-sided markets. *Journal of Economic Perspectives* 23, 125-143.

Sen, A., Tucker, C.E., 2020. Social distancing and school closures: Documenting disparity in internet access among school children. *Available at SSRN 3572922*.

Thorhauge, M., Kassahun, H.T., Cherchi, E., Haustein, S., 2020. Mobility needs, activity patterns and activity flexibility: How subjective and objective constraints influence mode choice. *Transportation Research Part A: Policy and Practice* 139, 255-272.

Warren, M.S., Skillman, S.W., 2020. Mobility changes in response to covid-19. *arXiv preprint arXiv:2003.14228*.

Weill, J.A., Stigler, M., Deschenes, O., Springborn, M.R., 2020. Social distancing responses to covid-19 emergency declarations strongly differentiated by income. *Proceedings of the National Academy of Sciences* 117, 19658-19660.

Wielechowski, M., Czech, K., Grzęda, Ł., 2020. Decline in mobility: Public transport in poland in the time of the covid-19 pandemic. *Economies* 8, 78.

Witteveen, D., Velthorst, E., 2020. Economic hardship and mental health complaints during covid-19. *Proceedings of the National Academy of Sciences* 117, 27277-27284.

Zheng, H., Zhang, K., Nie, M., 2020. The fall and rise of the taxi industry in the covid-19 pandemic: A case study. *Available at SSRN 3674241*.

Online Appendix

Table A1: Results for First-stage Regression

	(1)	(2)
	DV: lagged new cases	DV: lagged new cases
Lagged Other City New Cases	0.639***	0.489***
	(0.004)	(0.013)
Lagged Imported Case	0.262***	0.408***
	(0.000)	(0.000)
Driver fixed effect	Y	Y
City-week fixed effect	Y	Y
Kleibergen-Paap rk LM statistic	6987.09***	1601***
Kleibergen-Paap Wald rk F statistic	1.54e+07***	6.99e+05***
Number of observations	3,162,240	868,510
Adjusted R-squared	0.842	0.852

Notes. This table reports the first-stage regression result in the 2SLS method. The dependent variable is *lagged new cases*. Column (1) corresponds to using whether to work being the dependent variable in the second-stage regression; Column (2) corresponds to using the remaining dependent variables in the second-stage regression, which are conditional on drivers working on a day. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level.

Table A2: OLS Results for Equation (1)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Whether to work	Number of requests	Number of completed trips	Average trip distance	Average search distance	Earnings
Lagged New Cases	-0.004***	0.072***	-0.049***	-0.024	-0.011	-1.524***
	(0.000)	(0.016)	(0.006)	(0.036)	(0.015)	(0.308)
Driver fixed effects	Y	Y	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y	Y	Y
Mean dependent variable	0.275	9.408	4.759	10.155	5.978	165.781
No. observations	3,162,240	868,628	868,628	868,628	868,628	868,628
R-squared	0.454	0.497	0.332	0.331	0.454	0.497

Notes. This table reports the OLS estimation results for the impact of number of lagged new cases on driver behavior specified in Equation (1). The independent variable is the lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level.

Table A3: Decomposition by Source of Request (OLS)

VARIABLES	(1) Number of aggregator requests	(2) Number of completed aggregator trips	(3) Number of app requests	(4) Number of completed app trips
lagged New Cases	0.081***	-0.042***	-0.009***	-0.007***
	(0.016)	(0.005)	(0.002)	(0.002)
Driver fixed effects	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y
Mean Dependent Variable	8.368	3.955	1.040	0.804
No. observations	868,628	868,628	868,628	868,628
R-squared	0.454	0.497	0.332	0.331

Notes. This table reports the OLS estimation results for the impact of lagged number of new cases on driver behavior specified in Equation (1). The independent variable is the lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level.

Table A4: Impact of New COVID-19 Cases on Drivers' Labor Supply Measures
(with only *imported cases* as IV)

VARIABLES	(1) Whether to work	(2) Number of requests	(3) Number of completed trips	(4) Average trip distance	(5) Average search distance	(6) Earnings
Lagged New Cases	-0.013***	0.413***	-0.091***	-0.028	-0.010	-1.724***
	(0.001)	(0.047)	(0.014)	(0.035)	(0.020)	(0.502)
Driver fixed effects	Y	Y	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y	Y	Y
Mean dependent variable	0.275	9.408	4.759	10.155	5.978	165.781
No. observations	3,162,240	868,628	868,628	868,628	868,628	868,628
R-squared	0.358	0.455	0.504	0.271	0.120	0.406

Notes. This table reports the second-stage estimation results in 2SLS for the robustness check with only *imported cases* as the IV. The independent variable is the one-day-lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level.

Table A5: Decomposition by Source of Request

(with only *imported cases* as IV)

VARIABLES	(1) Number of aggregator requests	(2) Number of completed aggregator trips	(3) Number of app requests	(4) Number of completed app trips
Lagged New Cases	0.421***	-0.082***	-0.008*	-0.009***
	(0.045)	(0.013)	(0.004)	(0.003)
Driver fixed effects	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y
Mean dependent variable	8.368	3.955	1.040	0.804
No. observations	868,628	868,628	868,628	868,628
R-squared	0.453	0.497	0.332	0.331

Notes. This table reports the second-stage estimation results in 2SLS for the robustness check with only *imported cases* as the IV. The independent variable is the one-day-lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level. * Significant at the 10% level.

Table A6: Impact of New COVID-19 Cases on Drivers' Labor Supply Measures
(Subsample Analysis with Chengdu and Guangzhou)

VARIABLES	(1) Whether to work	(2) Number of requests	(3) Number of completed trips	(4) Average trip distance	(5) Average search distance	(6) Earnings
Lagged New Cases	-0.013***	0.339***	-0.090***	-0.092***	-0.038**	-2.341***
	(0.001)	(0.042)	(0.012)	(0.033)	(0.019)	(0.454)
Driver fixed effects	Y	Y	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y	Y	Y
Mean dependent variable	0.269	11.133	5.409	9.812	5.822	161.262
No. observations	1,850,862	498,173	498,173	498,173	498,173	498,173
R-squared	0.351	0.449	0.514	0.171	0.137	0.376

Notes. This table reports the second-stage estimation results in 2SLS for the robustness check on the subsample analysis excluding Shenzhen. The independent variable is the one-day-lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level. ** Significant at the 5% level.

Table A7: Decomposition by Source of Request
(Subsample Analysis with Chengdu and Guangzhou)

VARIABLES	(1)	(2)	(3)	(4)
	Number of aggregator requests	Number of completed aggregator trips	Number of app requests	Number of completed app trips
Lagged New Cases	0.349***	-0.081***	-0.010***	-0.008***
	(0.041)	(0.011)	(0.004)	(0.003)
Driver fixed effects	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y
Mean dependent variable	10.171	4.673	0.962	0.736
No. observations	498,173	498,173	498,173	498,173
R-squared	0.447	0.508	0.300	0.300

Notes. This table reports the second-stage estimation results in 2SLS for the robustness check on the subsample analysis excluding Shenzhen. The independent variable is the one-day-lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level.

Table A8: Impact of New COVID-19 Cases on Drivers' Labor Supply Measures
(Subsample Analysis with Chengdu and Shenzhen)

VARIABLES	(1) Whether to work	(2) Number of requests	(3) Number of completed trips	(4) Average trip distance	(5) Average search distance	(6) Earnings
Lagged New Cases	-0.015***	0.379***	-0.174***	0.064	0.092***	-3.771***
	(0.000)	(0.061)	(0.018)	(0.040)	(0.024)	(0.640)
Driver fixed effects	Y	Y	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y	Y	Y
Mean dependent variable	0.273	10.361	5.436	9.846	5.684	173.703
No. observations	2,260,416	616,036	616,036	616,036	616,036	616,036
R-squared	0.364	0.459	0.494	0.314	0.117	0.416

Notes. This table reports the second-stage estimation results in 2SLS for the robustness check on the subsample analysis excluding Guangzhou. The independent variable is the one-day-lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level.

Table A9: Decomposition by Source of Request
(Subsample Analysis with Chengdu and Shenzhen)

VARIABLES	(1) Number of aggregator requests	(2) Number of completed aggregator trips	(3) Number of app requests	(4) Number of completed app trips
Lagged New Cases	0.411***	-0.142***	-0.033***	-0.032***
	(0.059)	(0.017)	(0.005)	(0.004)
Driver fixed effects	Y	Y	Y	Y
City-week fixed effects	Y	Y	Y	Y
Mean dependent variable	9.202	4.523	1.159	0.913
No. observations	616,036	616,036	616,036	616,036
R-squared	0.460	0.492	0.325	0.321

Notes. This table reports the second-stage estimation results in 2SLS for the robustness check on the subsample analysis excluding Guangzhou. The independent variable is the one-day-lagged number of new COVID-19 cases in the city. All columns control for fixed effects for drivers and city-week pairs. Robust standard errors clustered at driver level in parentheses. *** Significant at the 1% level.

Author Statement

The individual contributions to the paper using the relevant CRediT roles are as follows:

Wei Wang: Conceptualization; Data Curation

Wei Miao: Formal Analysis; Writing – Original Draft

Yongdong Liu: Methodology; Formal Analysis

Yiting Deng: Methodology; Writing – Review & Editing

Yunfei Cao: Validation; project administration