Driving the Drivers:

Preferential Assignment in Ride-Hailing

Yanyou Chen†
Yao Luo†
Zhe Yuan‡

† University of Toronto
‡ Zhejiang University

Motivations

- Recent years have witnessed the rapid acceleration of algorithmic technologies.
- Assignment Algorithms: preferential assignments or bonuses for completing many trips, which may result in a flexibility penalty.
  - Ride-hailing platforms: (US) Uber; (Europe) Bolt; (China) Didi
  - Parcel-delivery platforms: (US) TaskRabbit, Instacart; (Europe) Glovo
  - Food-delivery platforms: (US) DoorDash, Uber Eats; (Europe) Deliveroo; (China) Meituan, Ele.me
- Provide the first empirical study of a preferential assignment algorithm and its impact on worker behavior and welfare.

Research Questions

- Would assignment algorithms favor some workers? If yes, why and how?
- How would the platform revenue and driver surplus change if the platform cannot use preferential assignment algorithms?

Preview of Findings

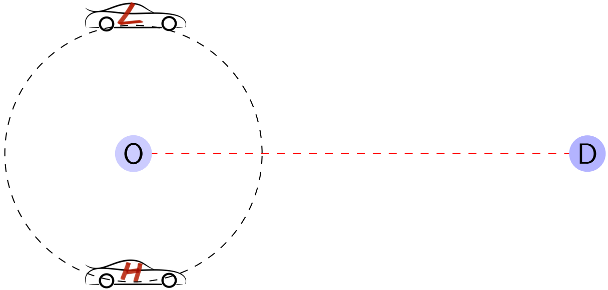
- Reduced-from Evidence
  - Preferential assignment algorithm is based on hourly work schedule.
  - Drivers favored by the algorithm earn 8 percent more hourly than the other drivers.
- Structural Model
  - Platform revenue decreases by 12 percent.
  - Drivers, especially young and local ones, have higher surplus without the preferential algorithm.

Literature

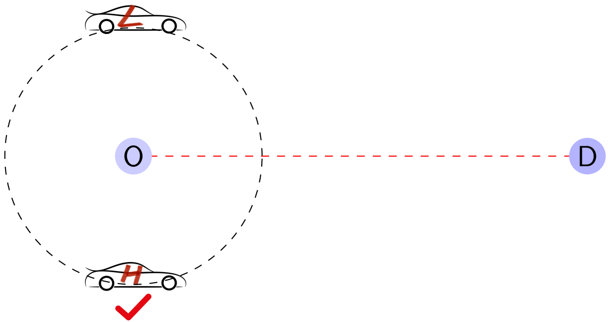
- How algorithms affect market outcomes
  - Assad, Clark, Ershov and Xu (2024), Rambachan, Kleinberg, Ludwig and Mullainathan (2020), Calvano, Calzolari, Denicolo and Pastorello (2020)
- Labor literature on compensation and work flexibility
  - Lazear (2018), Katz and Krueger (2019), Mas and Pallais (2017)
  - wage differential: Blau and Kahn (2017), Aaronson and French (2004)
- Literature on taxi and ride-hailing
  - Chen, Rossi, Chevalier and Oehlsen (2019), Liu, Wan and Yang (2019), Castillo (2020) Frechette, Lizzeri, and Salz (2019), Cook et al. (2021), Buchholz (2022)
- IO techniques
  - Rysman (2004, 2009), Hotz and Miller (1993), Arcidiacono and Miller (2011)

Outline

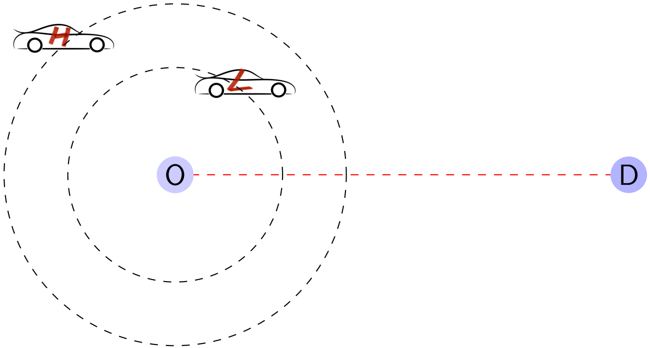
Preferential Assignment Algorithm



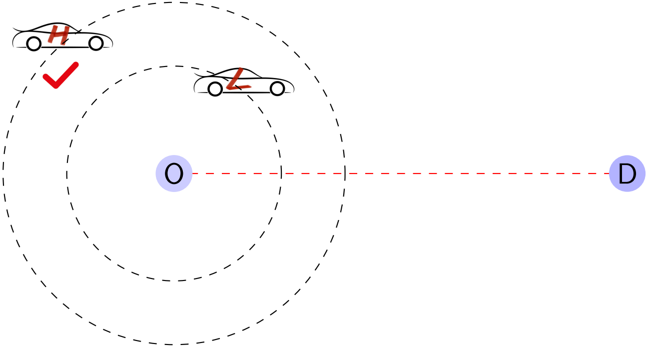
Preferential Assignment Algorithm



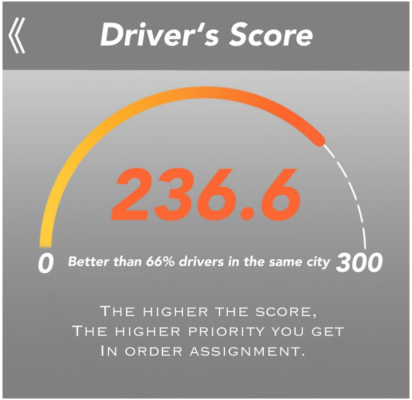
Preferential Assignment Algorithm



Preferential Assignment Algorithm



Preferential Assignment Algorithm



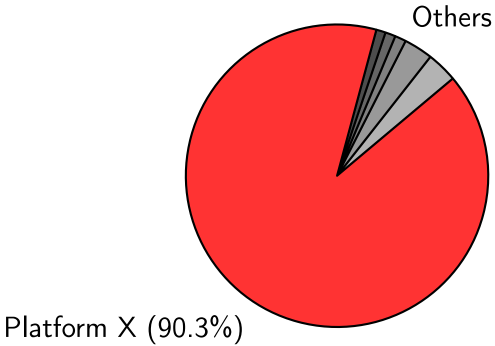
Outline

Data

- All the completed transactions in December 2018 of a major city in Asia
- Departure, destination, and distance of a trip, the time spent picking up and transporting passengers, and the price paid by the driver.
- Drivers' attributes such as age, gender, and birth location

Highly Concentrated

- Platform X in year 2020: 493 millions users, 15 million drivers



Summary Statistics (Driver-Hour)

	Mean	Std. Dev.	Min	25 Pctl	Median	75 Pctl	Max
Hourly Wage	49.98	24.52	0	32.83	47.42	62.74	286.86
Earning Time (minutes)	30.60	12.01	0	21	31	40	60
Pickup time (minutes)	10.62	6.67	0	6	10	15	60
Idle Time (minutes)	18.78	14.32	0	6	17	29	60
Number of Orders	1.89	1.11	0	1	2	3	9
Distance (km)	14.11	7.41	0	8.78	13.1	18.2	94.13
Number of Observations				4,182,318			

Who Earn Higher Hourly Wages?

Hourly Wage	(1)	(2)
# Work Hours in a Month	0.003***	0.003***
	(0.000)	(0.000)
% Incentivized Hours		18.724***
		(0.170)
Constant	54.918***	39.201***
	(0.126)	(0.190)
Observations	4,182,318	4,182,318
R-squared	0.040	0.043
Day-Hour FE	Y	Y
Origin/Destination FE	Y	Y

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

Type of Drivers

Incentivized hours: (1) midday 10am-4pm (2) evening 7pm-7am (next day).

High-Score Drivers: drivers who commit to working for at least two consecutive hours during incentivized times.

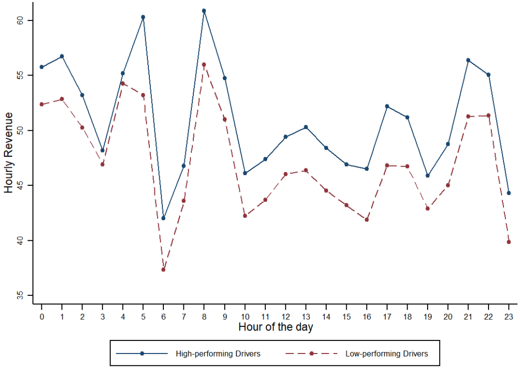
- S_1: 10am-12pm
- S_2: 11am-1pm
- ...
- S_5: 2pm-4pm
- S_6: 7pm-9pm
- ...
- S_{16}: 5am-7am

The rest: S_0: Low-Score Drivers

High-Score versus Low-Score Drivers

	High-Score (1)	Low-Score (2)
Panel I: Driver/Vehicle Characteristics		
% non-local	69%	53%
Age	37.2	37.4
Panel II: Performance (in a month)		
Work Hours	159	26
# orders	301	46
Monthly Revenue	7,985	1,202
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	23,712	16,392
Share Drivers	59.1%	40.9%

Hourly Wage Differentials



Dependent Variables	Hourly Wage		
	(1)	(2)	(3)
High-Score	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	4,182,318	4,182,318	4,182,318
R-squared	0.002	0.039	0.050

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

How? Driving Forces of Wage Differential

Dependent Variables	# Orders	Cancellation Rate (Rider)	Drive Dist	Earning Time	Idle Time
	(1)	(2)	(3)	(4)	(5)
High-Score	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)
Mean of Low-Score	1.76 (orders)	8.2%	13.4 (km)	29.3 (min)	20.4 (min)
High-Score compared to Low-Score Drivers	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 all columns except for column (2), we use completed transactions for the analysis. Completed transactions are available from Dec. 1st, 2018 to Dec. 31st, 2018. In column (2), we also include canceled orders to compute rider cancellation rates. Information on canceled order is available from Dec. 1st, 2018 to Dec. 10th, 2018. Standard errors are in parentheses. In all specifications, we control for day-hour fixed effect, origin district fixed effect and destination district fixed effect.

Summary

- High-Score drivers get assigned more rides
- Less idle time
- Assigned to riders with lower cancellation rates

Competing Hypotheses

- Strategically choose where to work
- Strategically cancel orders
- Drive faster or know the routes better

Service Areas

	Origin		Destination	
District	Low-Score	High-Score	Low-Score	High-Score
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%

Wage Differentials with Finer Grids

Dependent Variables	Hourly Wage			
	(1)	(2)	(3)	(4)
High-Score	3.850***	3.862***	3.693***	3.721***
	(0.0391)	(0.0385)	(0.0382)	(0.0375)
Constant	47.24***	46.51***	47.29***	46.64***
	(0.0701)	(0.0364)	(.0684)	(0.0355)
Controls:				
Day-Hour	Y	Y		
Day-Hour-15Minute			Y	Y
Origin/Destination	Y		Y	
1km $\times$ 1km Grid		Y		Y
Observations	4,182,318	4,182,318	4,182,318	4,182,318
R-squared	0.050	0.095	0.094	0.141

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

Driver Cancellation and Driver Speed

Dependent Variables	Probability of Cancellation (Driver)	Speed
	(1)	(2)
High-Score	-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-Score 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

Outline

1. Institutional Background and Preferential Assignment Algorithm ✓
2. Theoretical Motivation ✓
3. Reduced-Form Evidence ✓
4. A Model of Dynamic Labor Supply
5. Results

The Timeline: Step 1

The Platform announces prices

- $\overrightarrow{P}$: rider fare by the hour
- $\overrightarrow{W}^H, \overrightarrow{W}^L$: the wages by the hour and driver's schedule

$$\underbrace{W_t^H}_{\substack{\text{high-score} \\ \text{wage rate}}} = \underbrace{\eta}_{80\%} \underbrace{P_t D_t(P_t)}_{\text{total fares}} \underbrace{s_t}_{\substack{\text{high-score} \\ \text{share of trips}}} \frac{1}{N_t^H}$$

Riders choose ride-hailing or other options by the hour: $D_t(P_t)$

The Timeline: Step 2

Each driver chooses a work schedule in two steps

1. a work schedule type $j \in \{S_0^L, S_1^H, \ldots, S_{16}^H\}$

$$N_j = N \cdot \frac{\exp(EV_j)}{\sum_k \exp(EV_k)}$$

2. the exact schedule (DDC)

$$N_{jt} = N_j \times \Pr(\text{work in hour t}|\text{work schedule j})$$

A Two-Sided Market

1. Passenger demand (measured in earning time)

$$Q_t = \delta_t P_t^{-\epsilon}$$

2. Drivers' dynamic labor supply

$$W_t^H = \eta P_t D_t(P_t) s_t \frac{1}{N_t^H}, \quad W_t^L = \eta P_t D_t(P_t)(1 - s_t)\frac{1}{N_t^L}$$

3. Platform matches drivers and passengers

$$D_t(P_t)s_t \leq \lambda_t^H N_t^H, \quad D_t(P_t)(1 - s_t) \leq \lambda_t^L N_t^L$$

The Driver's Problem

- Individual choices: finite-horizon dynamic
  - obtain utility from either working or outside option

$$U_{it,1}^{\tau} = \underbrace{W_t^{\tau}}_{\substack{\text{schedule-type-specific} \\ \text{wage rate}}} + \sigma \cdot \epsilon_{it,1}$$

$$U_{it,0}^{\tau} = \underbrace{O_t + \eta_{s(i)t}}_{\text{outside option}} + \sigma \cdot \epsilon_{it,0}$$

    where $s(i)$ denotes driver type.
  - a "warm-up" cost κ: transit from "not working" to "working"

- Aggregate labor supply by the hour

$$N_t^H = \mathcal{N}_t^H(\vec{P}, \vec{s}; \boldsymbol{\theta}), \quad N_t^L = \mathcal{N}_t^L(\vec{P}, \vec{s}; \boldsymbol{\theta})$$

Outline

1. Institutional Background and Preferential Assignment Algorithm ✓
2. Theoretical Motivation ✓
3. Reduced-Form Evidence ✓
4. A Model of Dynamic Labor Supply ✓
5. Results

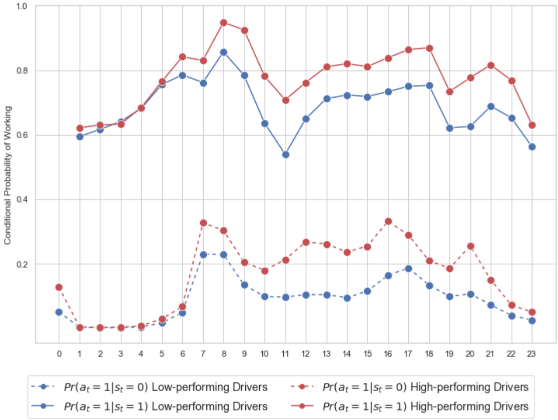
CCP-Based Estimation of $\overrightarrow{O}$ and κ

- Main parameters $\boldsymbol{\theta}$
  - hourly reservation value $\{O_t\}$, where $t = 1, \ldots, 24$
  - Unobserved heterogeneity $\eta_{s(i)_t}$
  - the warm-up cost κ
  - normalization term σ
- The MSM estimate $\widehat{\boldsymbol{\theta}}$

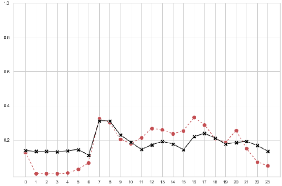
$$\min_{\boldsymbol{\theta}} \quad \left[\widehat{\mathsf{CCP}} - \mathsf{CCP}(\boldsymbol{\theta})\right]' \Omega \left[\widehat{\mathsf{CCP}} - \mathsf{CCP}(\boldsymbol{\theta})\right],$$

where Ω is a positive definite matrix.

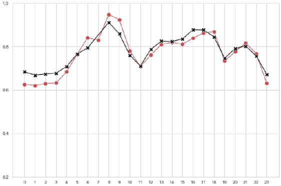
Conditional Probability of Working (from data)



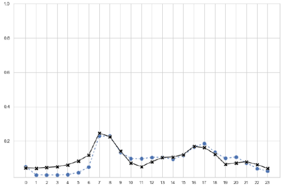
Model Fit



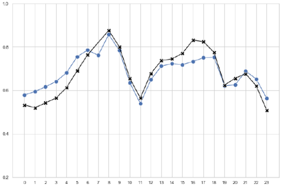
(a) High-Performing Drivers, State = 0



(b) High-Performing Drivers, State = 1

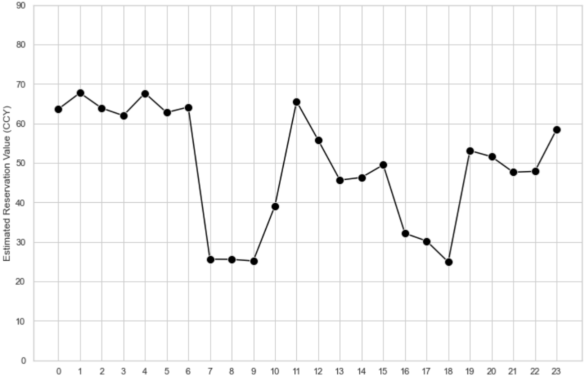


(c) Low-Performing Drivers, State = 0

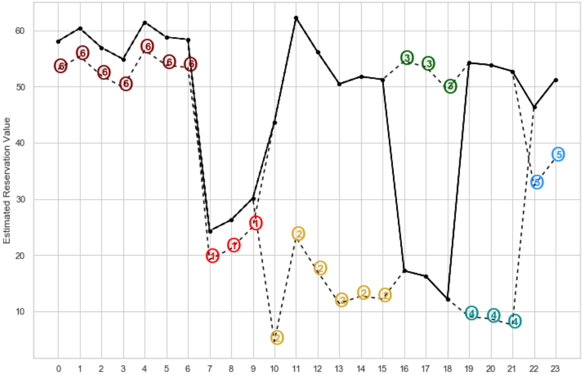


(d) Low-Performing Drivers, State = 1

Estimated Reservation Values



Estimated Reservation Values



Estimated Reservation Values

Table: Estimation Results of Unobserved Heterogeneity

	Group 0	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Population density of each group	0.07	0.06	0.18	0.42	0.18	0.04	0.05
Probability of *H*-Type	76.7%	78.7%	96.5%	49.6%	93.4%	82.8%	81.0%
Average Reservation Value	46.2	45.6	36.5	50.9	40.6	45.1	44.8

- Driver groups 2 and 4 are more likely to consist of older, non-local drivers.
- Driver group 3 is more likely to consist of younger, local, male drivers.

Elimination of Preferential Assignment Algorithm ("Fair" Pay)

- Non-preferential algorithms: "Fair pay"

$$\widetilde{W}_t = \eta P_t D_t(P_t) \underbrace{\frac{1}{N_t}}_{\text{rnd asgmt}}$$

- Given the new hourly wages, drivers solve a new DDC

$$U_{1t} = \underbrace{\widetilde{W}_t}_{\substack{\text{non-preferential} \\ \text{wage rate}}} + \sigma \cdot \epsilon_{1t},$$

$$U_{0t} = O_t + \eta_{s(i)t} + \sigma \cdot \epsilon_{0t},$$

Who Gains and Who Loses?

- An additional 10% of drivers would switch to flexible schedules, reducing their total work hours.
- Platform revenue decreases by 12% without a fare adjustment.
- Ride fares increase by 7.79%
- 3.81% surplus gain for drivers who switch to flexible schedules.

Conclusion

- Document preferential algorithm based on hourly work schedule. Drivers favored by the algorithm earn 8 percent more hourly than the other drivers.
- Construct and estimate a two-sided market model with dynamic labor supply. Platform revenues will decrease by 12 percent, and the total surplus will decrease by 7 percent if we eliminate the preferential algorithm but fixed the price.
- Without the preferential algorithm, an additional 10 percent of drivers will switch to flexible schedules. Young, male, and local drivers benefit more from the non-preferential algorithm.