Driving the Drivers:

Algorithmic Wage-Setting 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.
- Gig workers become increasingly aware that their bosses are algorithms that prioritize some objectives
  - The opposite of flexibility. News
- Provide the first empirical study of algorithmic wage-setting and its impact on worker behavior and welfare.

Research Question

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

Preview of Findings

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

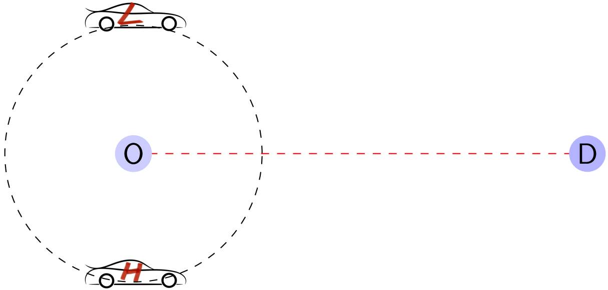
Literature

- How algorithms affect market outcomes
  - Assad, Clark, Ershov and Xu (2020), Kleinberg, Ludwig and Mullainathan (2020), Calvano, Calzolari, Denicolo and Pastorello (2020)
- Labor literature on compensation and incentives
  - 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)
- IO techniques
  - Rysman (2004, 2009), Hotz and Miller (1993), Arcidiacono and Miller (2011)

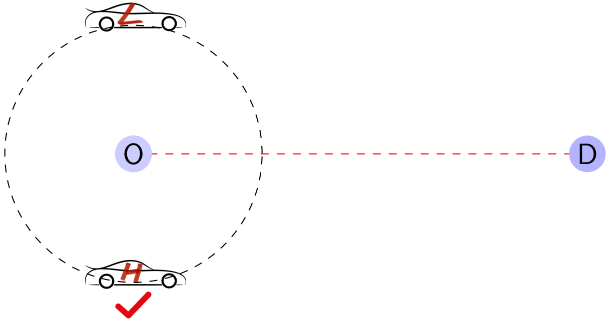
Outline

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

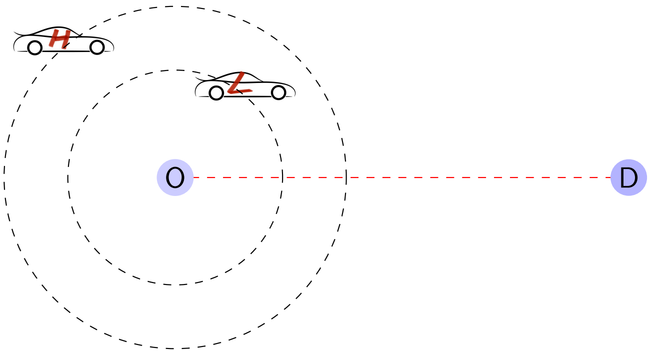
Preferential Algorithm



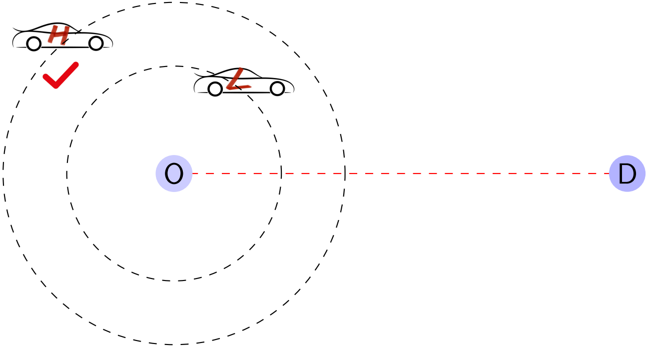
Preferential Algorithm



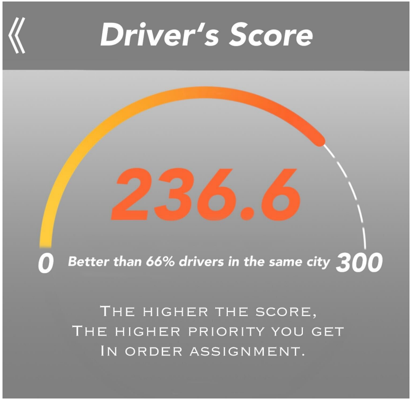
Preferential Algorithm



Preferential Algorithm



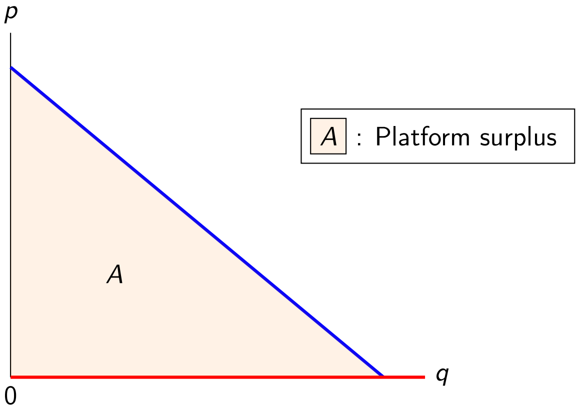
Preferential Algorithm



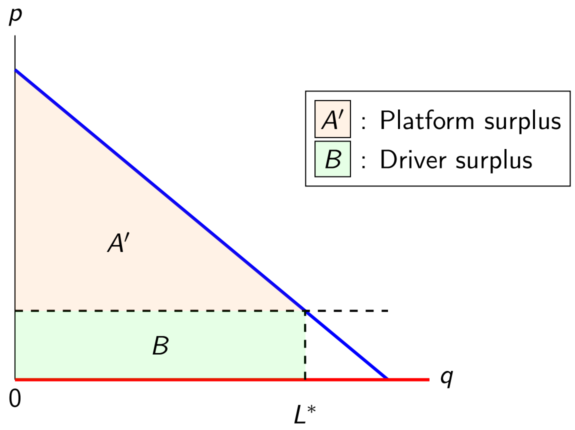
Outline

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

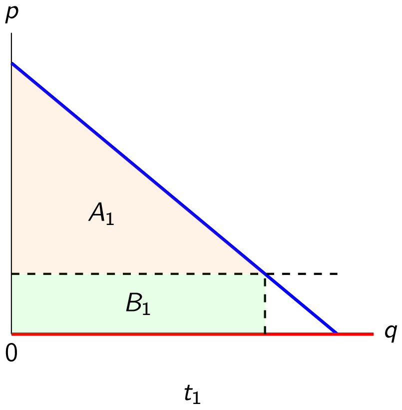
First-degree Price Discrimination

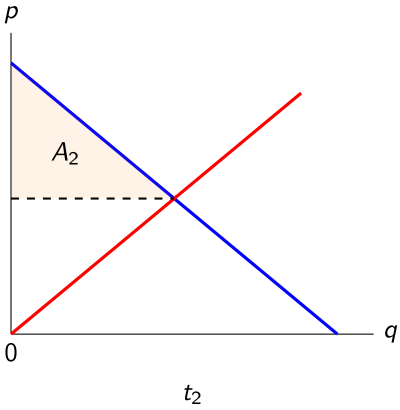


Imperfect Price Discrimination Over Drivers

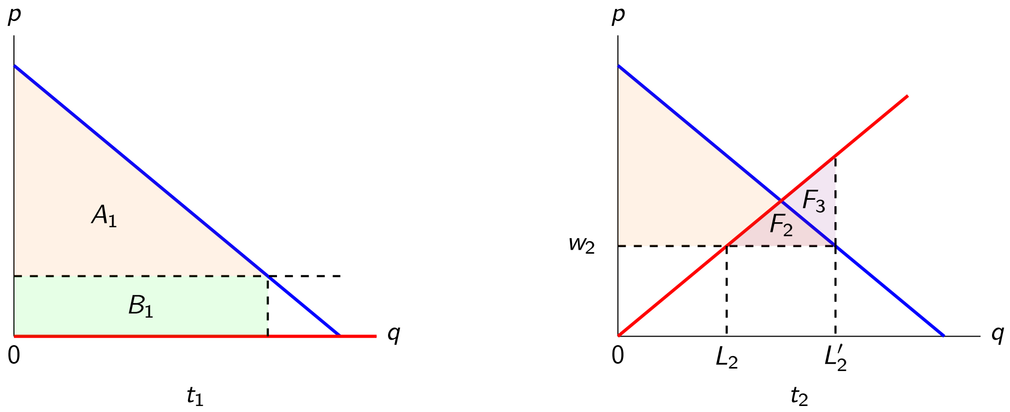


Cross-Time Elasticity Differentials





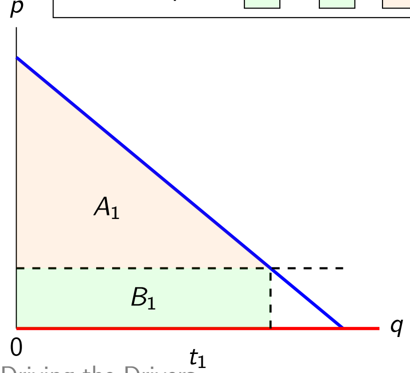
Incentive Wages ($F_2 + F_3$) versus Preferential Algorithm

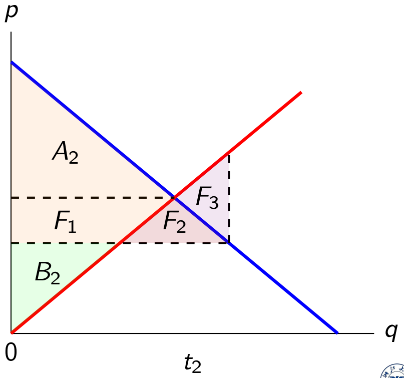


Welfare Implication of the Preferential Algorithm

Platform surplus : $\boxed{A_1} + \boxed{A_2} + \boxed{F_1} + \boxed{F_2}$

Driver surplus : $\boxed{B_1} + \boxed{B_2}$ - $\boxed{F_2}$ - $\boxed{F_3}$





Outline

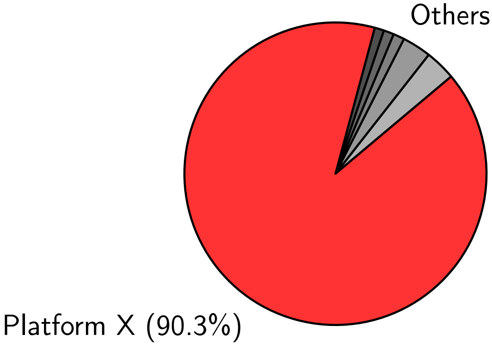
1. Institutional Background and Preferential Algorithm ✓
2. Theoretical Motivation ✓
3. Reduced-Form Evidence
4. A Model of Dynamic Labor Supply
5. Results

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)
% Off-peak (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-performing Drivers: drivers who commit to work for at least two consecutive hours during incentivized hours (midday or night) DriverType

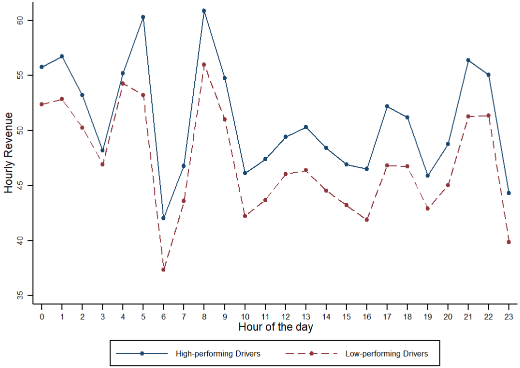
- S_1: 10am-12pm
- S_2: 11am-1pm
- ...
- S_5: 2pm-4pm
- S_6: 7pm-9pm
- ...
- S_{16}: 5am-7am

The rest: S_0: Low performing

High-Performing versus Low-Performing

	High-performing (1)	Low-performing (2)
Panel I: Driver/Vehicle Characteristics		
% femal	2.2%	3.5%
% non-local	69%	53%
Age	37.2	37.4
Panel II: Performance (in a month)		
Work Days	17	5
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-performing	3.886***	3.794***	3.851***
	(0.0397)	(0.0393)	(0.0391)
Constant	46.49***	46.57***	47.24***
	(0.0376)	(0.0372)	(0.0701)
Day-Hour FE	N	Y	Y
Origin FE	N	N	Y
Destination FE	N	N	Y
Observations	4,182,318	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-performing	0.125***	-0.0023***	0.748***	1.579***	-2.140***
	(0.0018)	(0.0004)	(0.0003)	(0.0187)	(0.0221)
Constant	1.468***	0.0894***	12.85***	32.35***	17.04***
	(0.00313)	(0.0005)	(0.0212)	(0.0334)	(0.0395)
Mean of Low-performing	1.76 (orders)	8.2%	13.4 (km)	29.3 (min)	20.4 (min)
High-performing compared to Low-performing	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-performing 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

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

Wage Differentials with Finer Grids

Dependent Variables	Hourly Wage (OLS)				IV
	(1)	(2)	(3)	(4)	(5)
High-Performing	2.742***	2.704***	2.705***	2.731***	8.99***
	(0.0391)	(0.0453)	(0.0448)	(0.0448)	(0.9614)
Constant	47.98***	21.38	23.90	47.56***	41.85***
	(0.0701)	(22.75)	(22.51)	(0.0427)	(0.8755)
Time Controls:					
Day-Hour	Y	Y			
15Minute			Y		
Location Controls:					
Origin/Destination	Y				
Grid		Y	Y		
Grid-15Minute				Y	Y
Observations	3,160,528	3,160,528	3,160,528	3,160,528	3,160,528
R-squared	0.053	0.075	0.094	0.097	(omitted)

Driver Cancellation and Driver Speed

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

Outline

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

A Model of Dynamic Labor Supply

The Driver's Problem

- Individual choices: finite-horizon dynamic

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

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

- a "warm-up" cost κ.
- $\tau \in \{H, L\}$
- $t \in \{1, 2, \cdots, 24\}$

The Platform

- $\overrightarrow{P}$: rider fare by the hour
- $\overrightarrow{s}$: the share of orders assigned to H-drivers
- choice of $(\overrightarrow{P}, \overrightarrow{s})$ to maximize profit:

$$\begin{aligned} \max_{(\overrightarrow{P}, \overrightarrow{s})} \quad & (1-\eta) \sum_t P_t D_t(P_t) \\ \text{s.t.} \quad & D_t(P_t) s_t \le \lambda_{Ht} N_{Ht} \\ & D_t(P_t)(1 - s_t) \le \lambda_{Lt} N_{Lt}. \end{aligned}$$

Passenger Demand

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

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

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

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 $j \in \{S_0^L, S_1^H, \ldots, S_{16}^H\}$ DriverDef

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

2. the exact schedule (DDC)

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

Outline

1. Institutional Background and Preferential 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

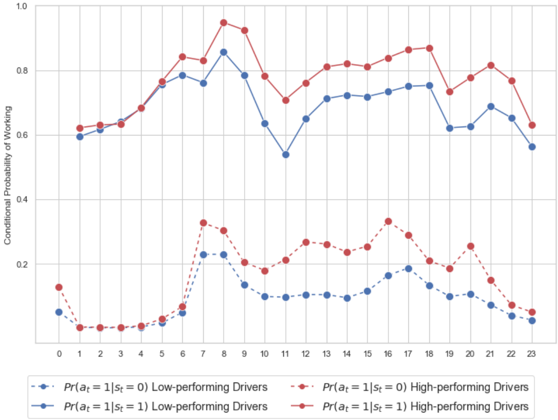
- Main parameters $\boldsymbol{\theta}$
  - hourly reservation value $\{O_t\}$, where $t = 1, \ldots, 24$
  - the warm-up cost κ
  - normalization term σ

- The MSM estimate $\widehat{\boldsymbol{\theta}}$

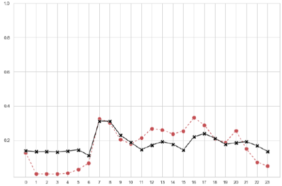
$$\min_{\boldsymbol{\theta}} \quad \left[\widehat{\mathsf{CCP}} - \mathsf{CCP}(\boldsymbol{\theta})\right]' \Omega \left[\widehat{\mathsf{CCP}} - \mathsf{CCP}(\boldsymbol{\theta})\right],$$

where Ω is a positive definite matrix.

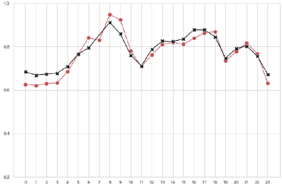
Conditional Probability of Working (from data)



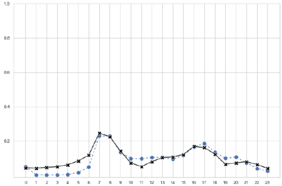
Model Fit



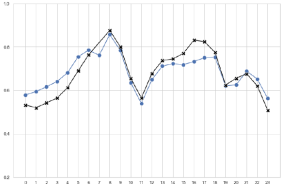
(a) High-Performing Drivers, State = 0



(b) High-Performing Drivers, State = 1

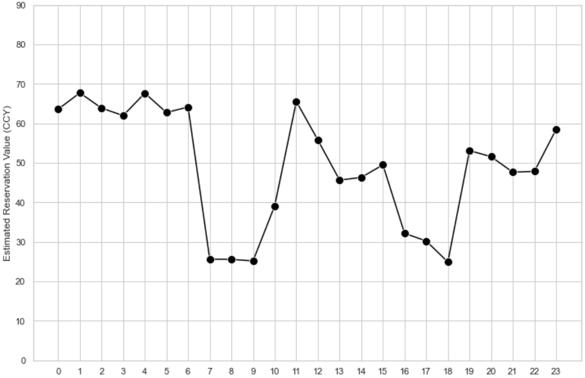


(c) Low-Performing Drivers, State = 0

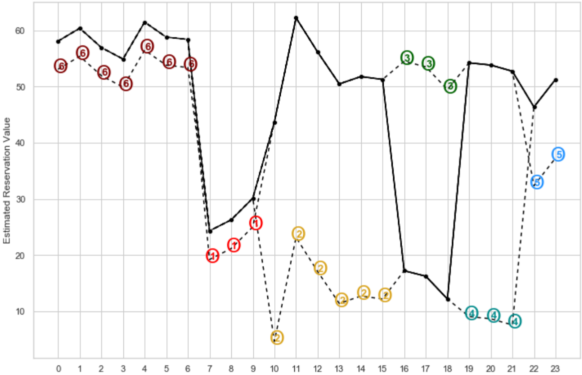


(d) Low-Performing Drivers, State = 1

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

Elimination of Preferential 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

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

$$U_{0t} = O_t + \sigma \cdot \epsilon_{0t},$$

Who Gains and Who Loses?

Changes in	Fixed Price	Reoptimized Price
Platform revenue	-12.16%	-1.42%
Consumer surplus	-12.16%	-1.42%
Driver surplus	0.14%	0.49%
Total surplus	-7.16%	-0.64%
Decomposition of Per-Driver Surplus		
High-performing driver (non-switcher)	-0.63%	-0.16%
Low-performing driver (non-switcher)	0.69%	0.99%
Switcher (from H-type to L-type)	3.51%	3.81%
Prob (HP schedule)	-11.48%	-9.98%

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.

News about Preferential Algorithm

THE GLOBE AND MAIL

Uber gets almost everything it wants in Ontario's Working For Workers Act

"In two hours of work, I used to get five or six orders. Now I'm getting one or two," he said. Then he started seeing threads on Reddit and Twitter in which other walkers complained they had experienced a decline in orders. Mr. Chowdary ended up taking on a job as a dishwasher in a restaurant, which paid him a consistent minimum wage for hours worked, but no tips.

Toy Model [Model]

Without preferential algorithm. In each period: [solution]

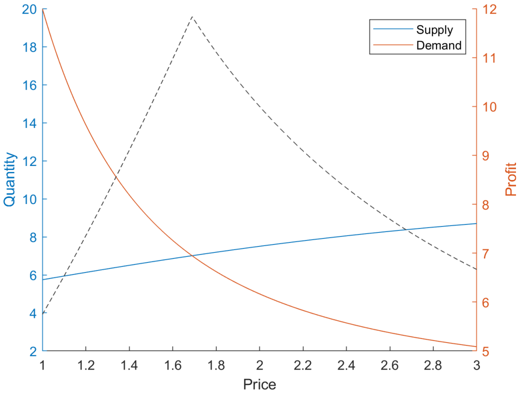
- Demand is $D_t(P_t) = \delta_t P_t^{-\epsilon_t}$.
- Supply is $S_t(P_t) = M_t \cdot \frac{\exp(\eta P_t)}{\exp(\eta P_t)+\exp(O_t)}$.
- Platform's decision is $\max_{P_t}(1-\eta)P_t \cdot \min\{D_t(P_t), S_t(P_t)\}$

With preferential algorithm: [solution]

- Demand is $D_1(P_1) = \delta_1 P_1^{-\epsilon_1}$, $D_2(P_2) = \delta_2 P_2^{-\epsilon_2}$
- Supply is $S_1 = S_2 = S(P_1, P_2) = M \cdot \frac{\exp(\boldsymbol{\eta}(\boldsymbol{P}_1+\boldsymbol{P}_2))}{\exp(\boldsymbol{\eta}(\boldsymbol{P}_1+\boldsymbol{P}_2))+\exp(\boldsymbol{O}_1+\boldsymbol{O}_2)}$.
- Platform's decision is

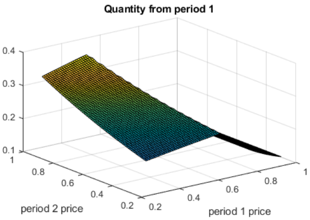
$$\max_{P_1,P_2}(1-\eta)P_1 \cdot \min\{D_1(P_1), S(P_1,P_2)\} + (1-\eta)P_2 \cdot \min\{D_2(P_2), S(P_1,P_2)\}$$

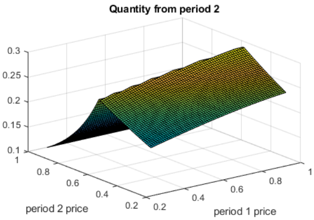
Without Preferential Algorithm

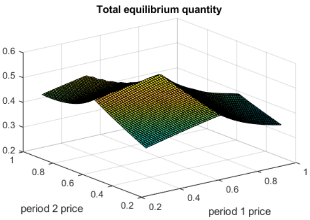


▸ Model

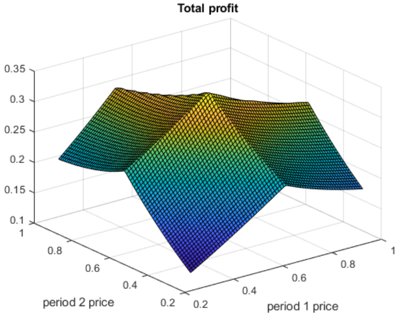
With Preferential Algorithm







With Preferential Algorithm



Model

Labor Cost

Figure: Difference of Labor Cost η With and Without Preferential Algorithm

		Reservation value at period 1, O_1					
		1.0	1.2	1.4	1.6	1.8	2.0
O_2	1.0	-0.12	-0.08	-0.04	0.00	0.04	0.07
	1.2	-0.08	-0.08	-0.04	0.00	0.03	0.07
	1.4	-0.04	-0.04	-0.04	0.00	0.03	0.07
	1.6	0.00	0.00	0.00	0.00	0.03	0.07
	1.8	0.04	0.03	0.03	0.03	0.03	0.06
	2.0	0.07	0.07	0.07	0.07	0.06	0.06

Profit

With Preferential Algorithm versus Without (when ϵ changes)

Figure: Profit of the Platform When Changing Elasticity of Demand

		Elasticity of Demand at Period 1, ε_1					
		1.5	1.9	2.3	2.7	3.1	3.5
ε_2	1.5	30.9	29.2	28.1	27.4	26.9	26.5
	1.9	29.2	27.4	26.3	25.6	25.1	24.7
	2.3	28.1	26.3	25.2	24.5	24.0	23.6
	2.7	27.4	25.6	24.5	23.7	23.2	22.8
	3.1	26.9	25.1	24.0	23.2	22.7	22.3
	3.5	26.5	24.7	23.6	22.8	22.3	21.9

(a) With Preferential Algorithm

		Elasticity of Demand at Period 1, ε_1					
		1.5	1.9	2.3	2.7	3.1	3.5
ε_2	1.5	1.5	1.7	1.9	2.1	2.1	2.2
	1.9	1.7	2.0	2.1	2.2	2.3	2.3
	2.3	1.9	2.1	2.2	2.3	2.4	2.4
	2.7	2.1	2.2	2.3	2.4	2.4	2.5
	3.1	2.1	2.3	2.4	2.4	2.5	2.6
	3.5	2.2	2.3	2.4	2.5	2.6	2.6

(b) Difference (With - Without)

Profit

Driver Surplus by Group under "Fair Pay"

Changes in Driver Surplus	Driver Group						
	Group 0	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Panel I: Short-term							
Total	0.08%	0.05%	-0.36%	0.20%	-0.22%	0.00%	0.07%
H-Schedule	-0.41%	-0.43%	-0.50%	-0.14%	-0.44%	-0.38%	-0.42%
L-Schedule	0.35%	0.38%	0.86%	0.12%	0.57%	0.37%	0.36%
Panel II: Long-term							
Total	0.29%	0.28%	-0.02%	0.22%	0.08%	0.23%	0.29%
H-Schedule	-0.14%	-0.15%	-0.17%	-0.04%	-0.13%	-0.12%	-0.14%
L-Schedule	0.54%	0.58%	1.19%	0.16%	0.86%	0.57%	0.56%

Comparative Statics

Demand Elasticity	Changes in (With - Without)			
	Platform Revenue/ Consumer Surplus	Driver Surplus	Driver Surplus (Low-performing)	Average Wage
Benchmark	1.44%	-0.49%	-0.98%	-7.26%
$\epsilon \times 1.1$	2.13%	-0.47%	-0.52%	-6.55%
$\epsilon \times 1.2$	2.60%	-0.40%	-0.29%	-5.78%
$\epsilon \times 1.3$	2.89%	-0.32%	-0.17%	-5.03%
$\kappa \times 1.1$	1.45%	-0.49%	-0.93%	9.64%
$\kappa \times 1.2$	1.46%	-0.49%	-0.86%	9.71%
$\kappa \times 1.3$	1.47%	-0.48%	-0.79%	9.76%