



The Impact of Behavioral and Economic Drivers on Gig Economy Workers

INFORMS 2019



Park Sinchaisri (Wharton)
Gad Allon (Wharton), Maxime Cohen (McGill)

Gig Economy



Gig Economy



freelancing **Upwork**



Gig Economy



freelancing **Upwork**



local tasks **TaskRabbit** **handy**



Gig Economy

freelancing



local tasks



ride-hailing



delivery



Gig Economy



freelancing **Upwork**



local tasks **TaskRabbit** **handy**

retail **(allwork)** **snag.work**
on demand

ride-hailing **lyft** **Uber** **Grab**

delivery **instacart** **DOORDASH**



Gig Economy

2017

57.3 Million
= 36% of US workforce



Gig Economy

2017

57.3 Million
= 36% of US workforce

2027

60% of work
+ \$2.7 trillion global GDP

Who are Gig Workers?

70% by choice

44% primary income

~50% millennials/gen z

Who are Gig Workers?

70% by choice



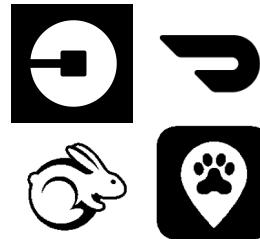
when to work?

44% primary income



how long?

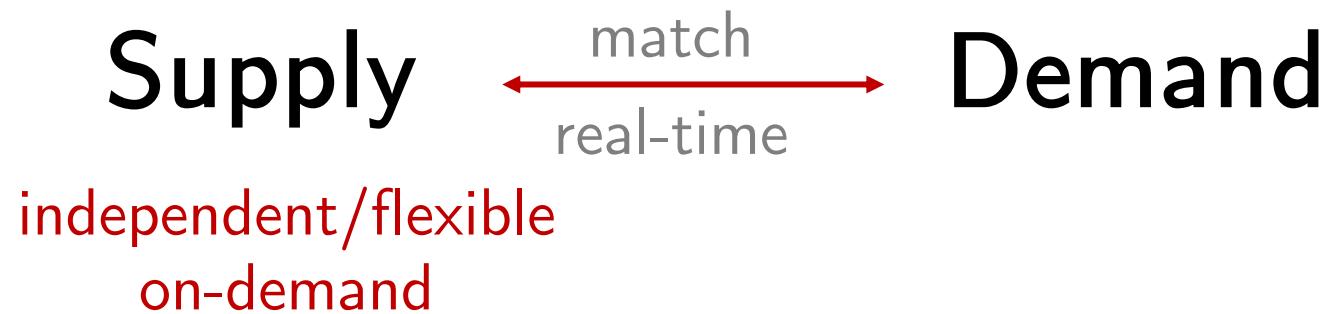
~50% millennials/gen z



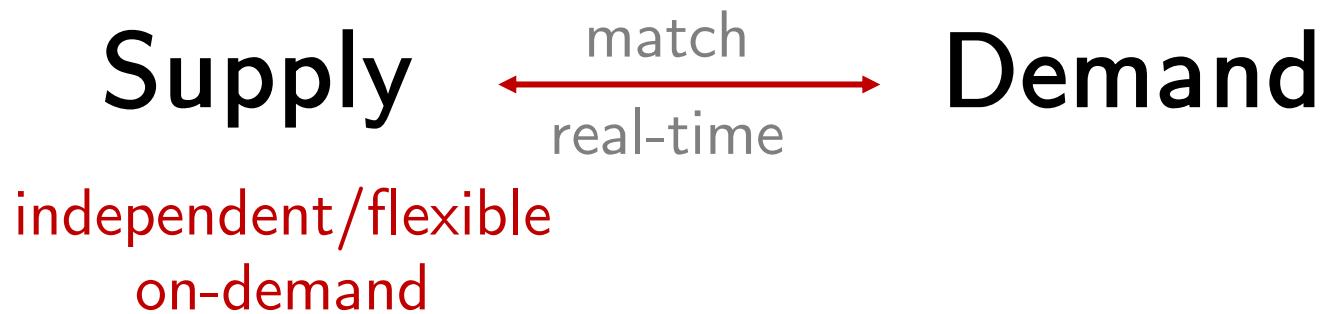
which platforms?

Workers decide work schedules

Gig Company



Gig Company



Workforce planning is challenging

Research Questions

How do gig economy workers
make labor decisions?

How can the platform influence
their decisions?

Research Questions

Econometrics

How do gig economy workers
make labor decisions?

How can the platform influence
their decisions?

Simulation

Structural

Outline

- **What has been done**

- Practice / labor economics / OM / hypotheses

- **Data and empirical strategy**

- Dealing with endogeneity and selection bias

- **Results**

- Impact of incentive and behavioral elements on labor decisions

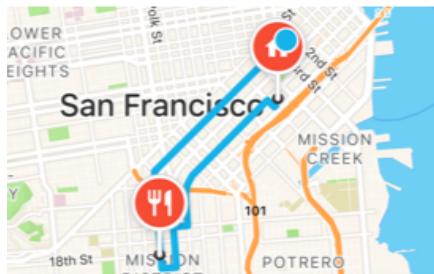
- **Implications**

- Simulation of optimal incentive re-allocation

In Practice

Real-time “surge pricing”

Deliver by 6:15pm Decline



Mission Chinese Food
\$22.78 subtotal (2 items)

BUSY PAY: +\$1.50

4.1 miles total

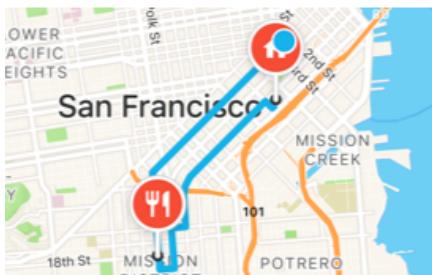
Accept Order



In Practice

Real-time “surge pricing”

Deliver by 6:15pm Decline



4.1 miles total

Accept Order

 DOORDASH

Pre-announced bonus

5:00 PM–6:00 PM

 +10% (5:00pm - 5:30pm)
+30% (5:30pm - 6:00pm)

6:00 PM–7:00 PM

 +30% (6:00pm - 6:30pm)
+40% (6:30pm - 7:00pm)

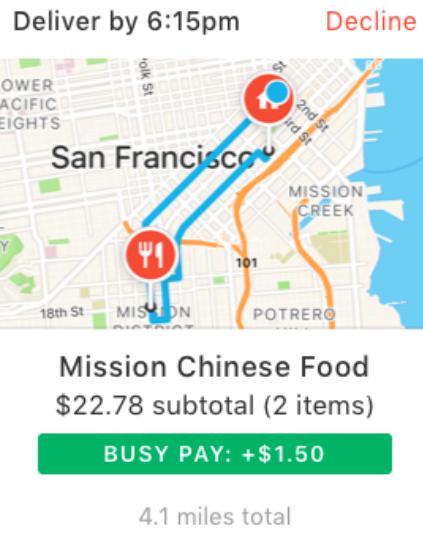
caviar

<https://dasherhelp.doordash.com/busy-pay>

https://courierhelp.trycaviar.com/customer/en/portal/articles/2821000-peak-hour-pay?b_id=9619/

In Practice

Real-time “surge pricing”



DOORDASH

Pre-announced bonus



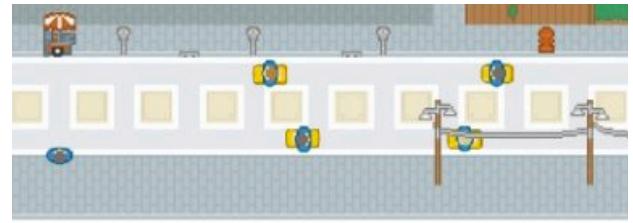
5:00 PM–6:00 PM
+10% (5:00pm - 5:30pm)
+30% (5:30pm - 6:00pm)



6:00 PM–7:00 PM
+30% (6:00pm - 6:30pm)
+40% (6:30pm - 7:00pm)

caviar

“You’re so close to your precious target”



How Uber Uses
Psychological Tricks to
Push Its Drivers’ Buttons

Theories of Labor Supply



Theories of Labor Supply

Neoclassical

- Maximize lifetime utility
- **Positive** income elasticities

Wage ↑
Work more

Theories of Labor Supply

Neoclassical

- Maximize lifetime utility
- **Positive** income elasticities

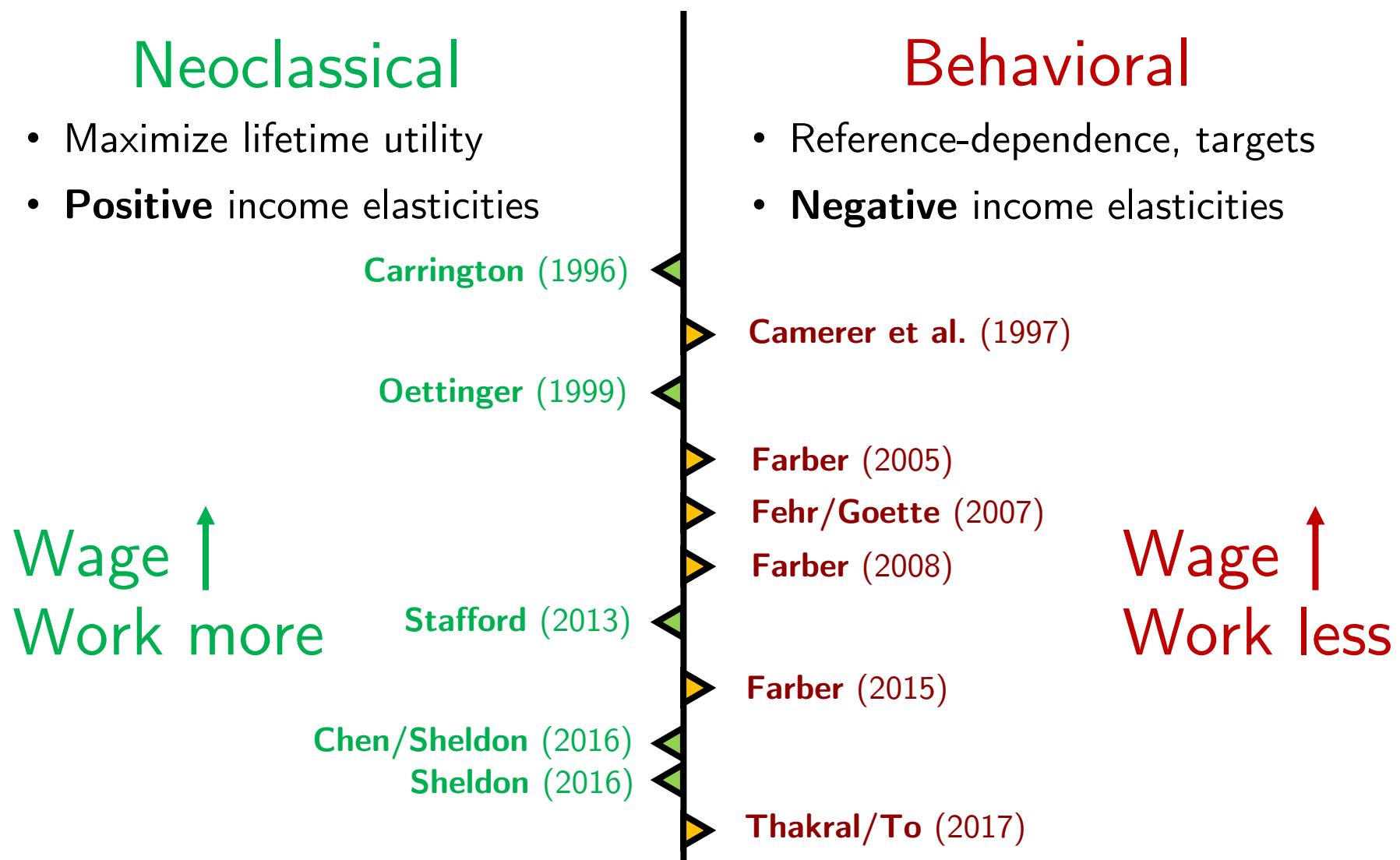
Wage ↑
Work more

Behavioral

- Reference-dependence, targets
- **Negative** income elasticities

Wage ↑
Work less

Theories of Labor Supply



Recent Work in OM

Theoretical

Dong & Ibrahim (2018)
Taylor (2018)
Cachon, Daniels & Lobel (2017)
Hu & Zhou (2017)
Ibrahim (2017)
Bimpikis, Candogan & Saban (2017)
Gurvich, Lariviere & Moreno (2016)
Tang et al. (2016)
Banerjee, Riquelme & Johari (2016)
Benjaafar et al. (2015)
...

Empirical

Sun, Wang & Wan (2019)
Kabra, Belavina & Girotra (2018)
Karacaoglu, Moreno & Ozkan (2018)
Chen, Chevalier, Rossi & Oehlsen (2017)
Cui, Li & Zhang (2017)
Li, Moreno & Zhang (2016)
...

Our Paper

- Behavioral drivers of decisions
- Rich data with complete description of the supply side

Drivers of Work Decisions

We are interested in three effects

Hourly Wage

Income Target

Time Target

on two work decisions:

Work or not?

If so, how long?

Drivers of Work Decisions

We are interested in three effects

“ISF”

Hourly Wage

“HSF”

Hours So Far
/Time Target

on two work decisions:

Work or not?

If so, how long?

Drivers of Work Decisions

We are interested in three effects

“ISF”

Hourly Wage

“HSF”

Hours So Far
/Time Target

H1: Positive

Carrington (1996),
Oettinger (1999), Stafford (2015)

H2: Negative

Farber (2008),
Thakral & To (2017)

H3: Negative

Crawford & Meng (2011), Farber (2015),
Agarwal et al (2015), Brachet et al (2012),
Collewet & Sauermann (2017)

on two work decisions:

Work or not?

If so, how long?

Data

US ride-hailing firm

Drivers are guaranteed an hourly

Base Rate

+

Promotions

“Offer”

Data

US ride-hailing firm

Drivers are guaranteed an hourly Base Rate + Promotions



Shift-level financial incentives and driving activity *for all*

Data

US ride-hailing firm

Drivers are guaranteed an hourly Base Rate + Promotions



Shift-level financial incentives and driving activity *for all*

5.5M

Observations

358

Days

Oct 2016 – Sep 2017

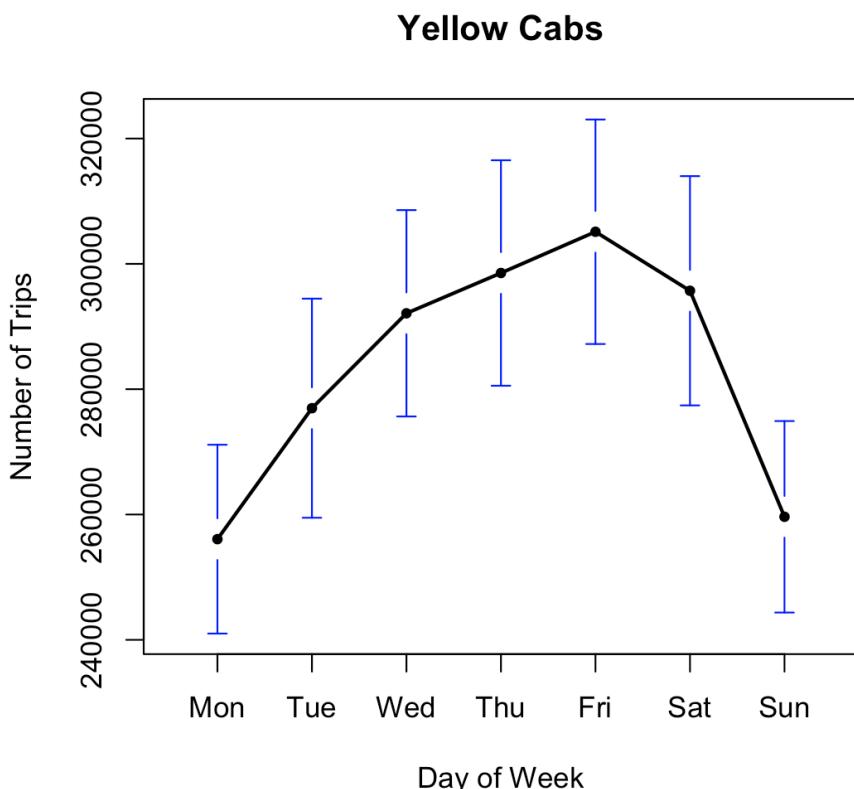
7,826

Unique drivers

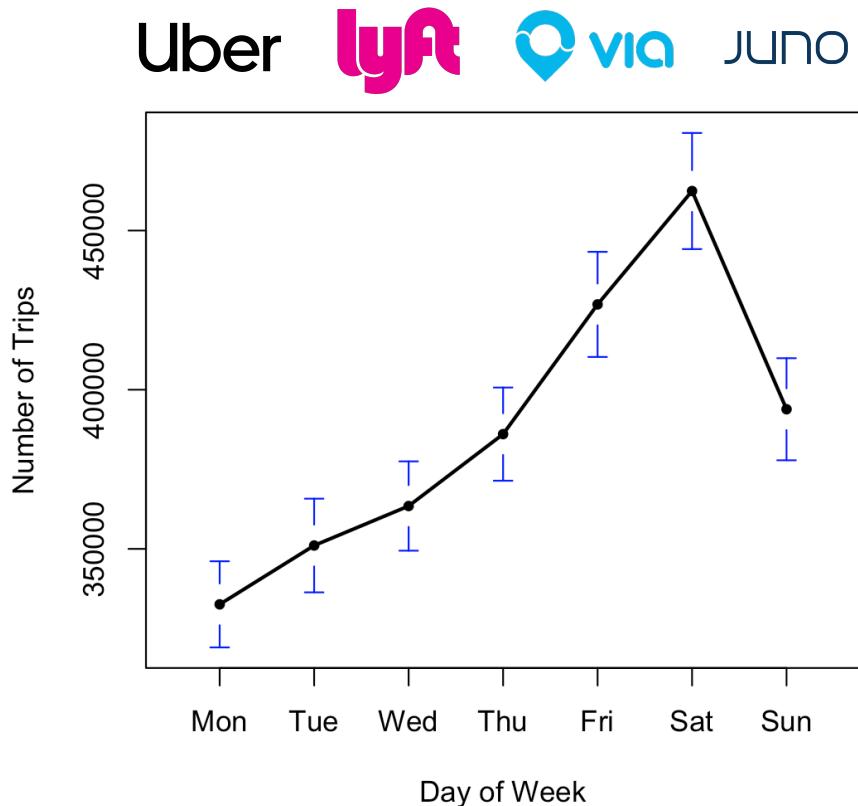
SUV/Sedan/Van

+ TLC Data

Trip records collected by Taxi & Limousine Commission



101M yellow cab trips

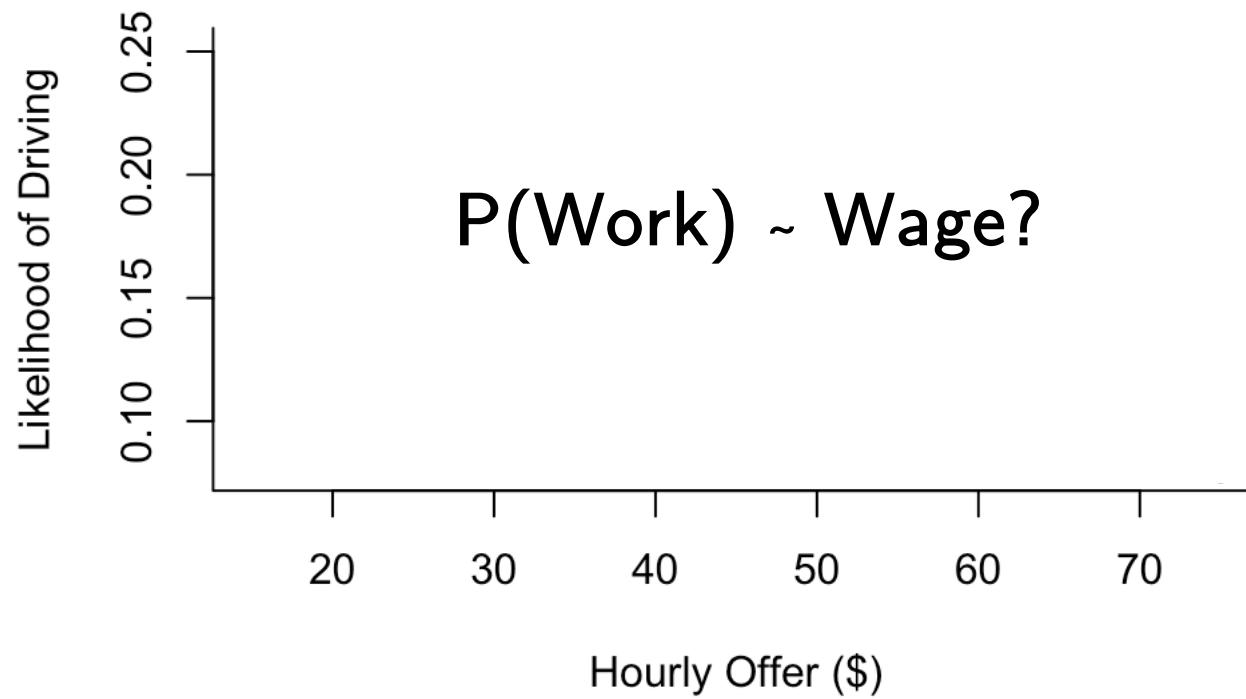


139M ride-hailing trips

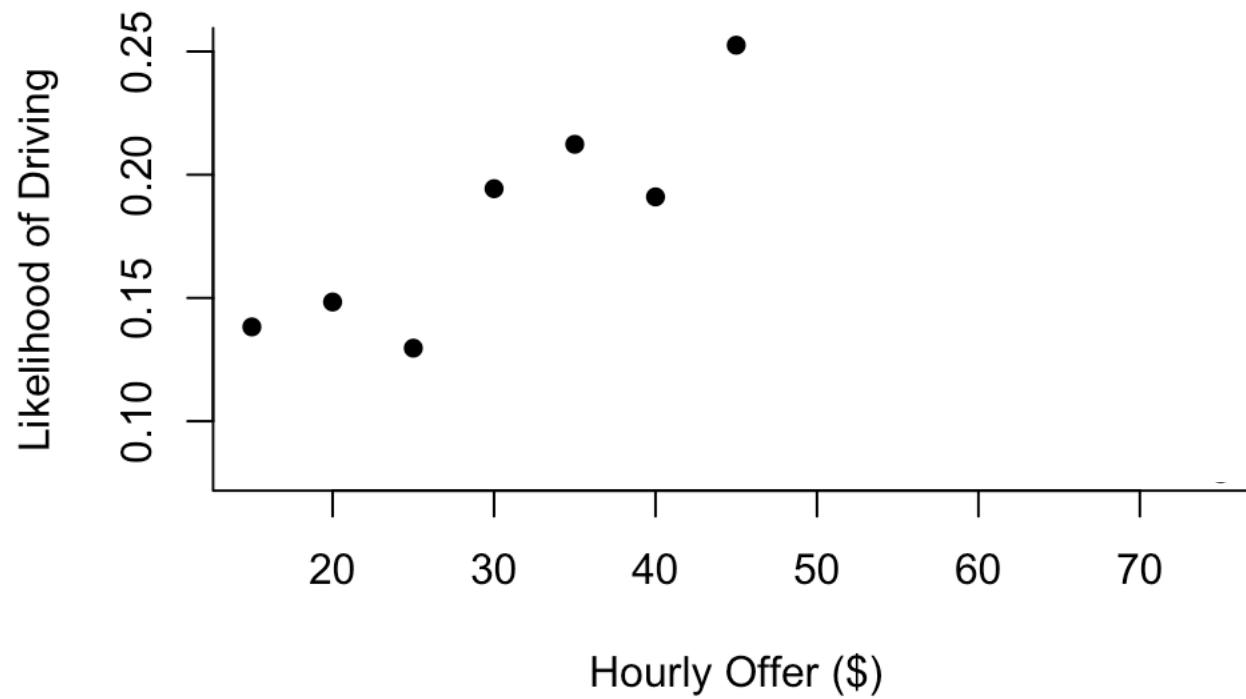


Empirical Strategy Challenges

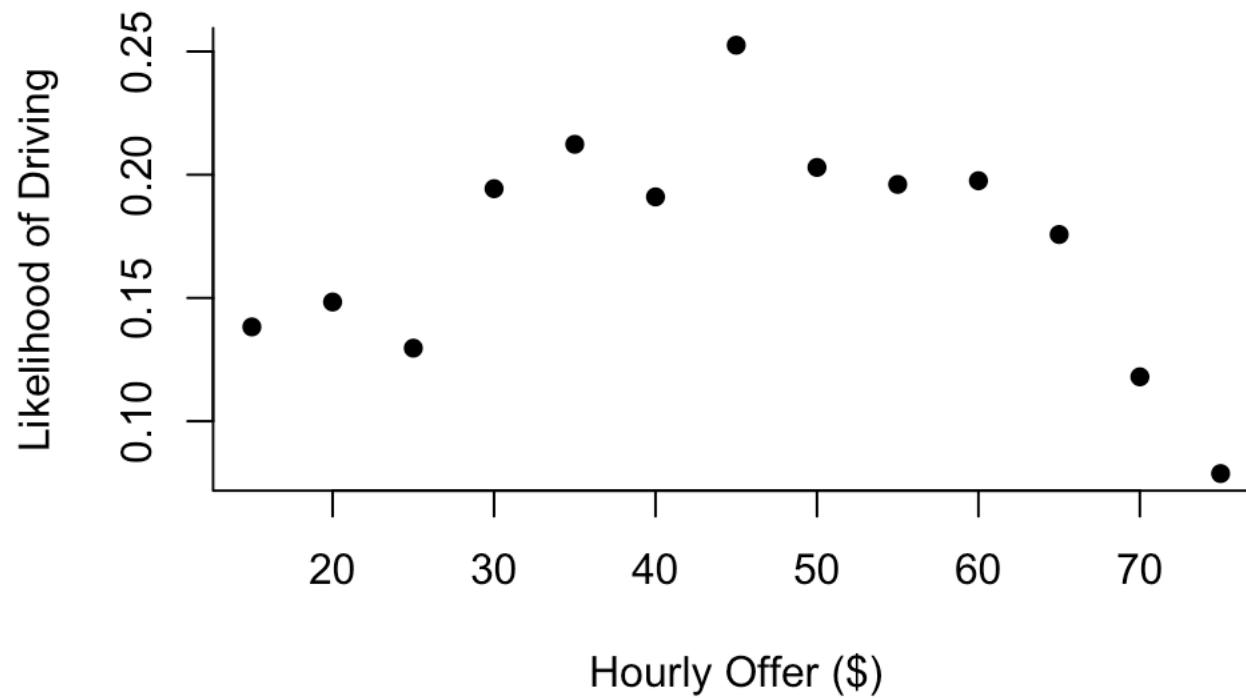
Empirical Strategy Challenges



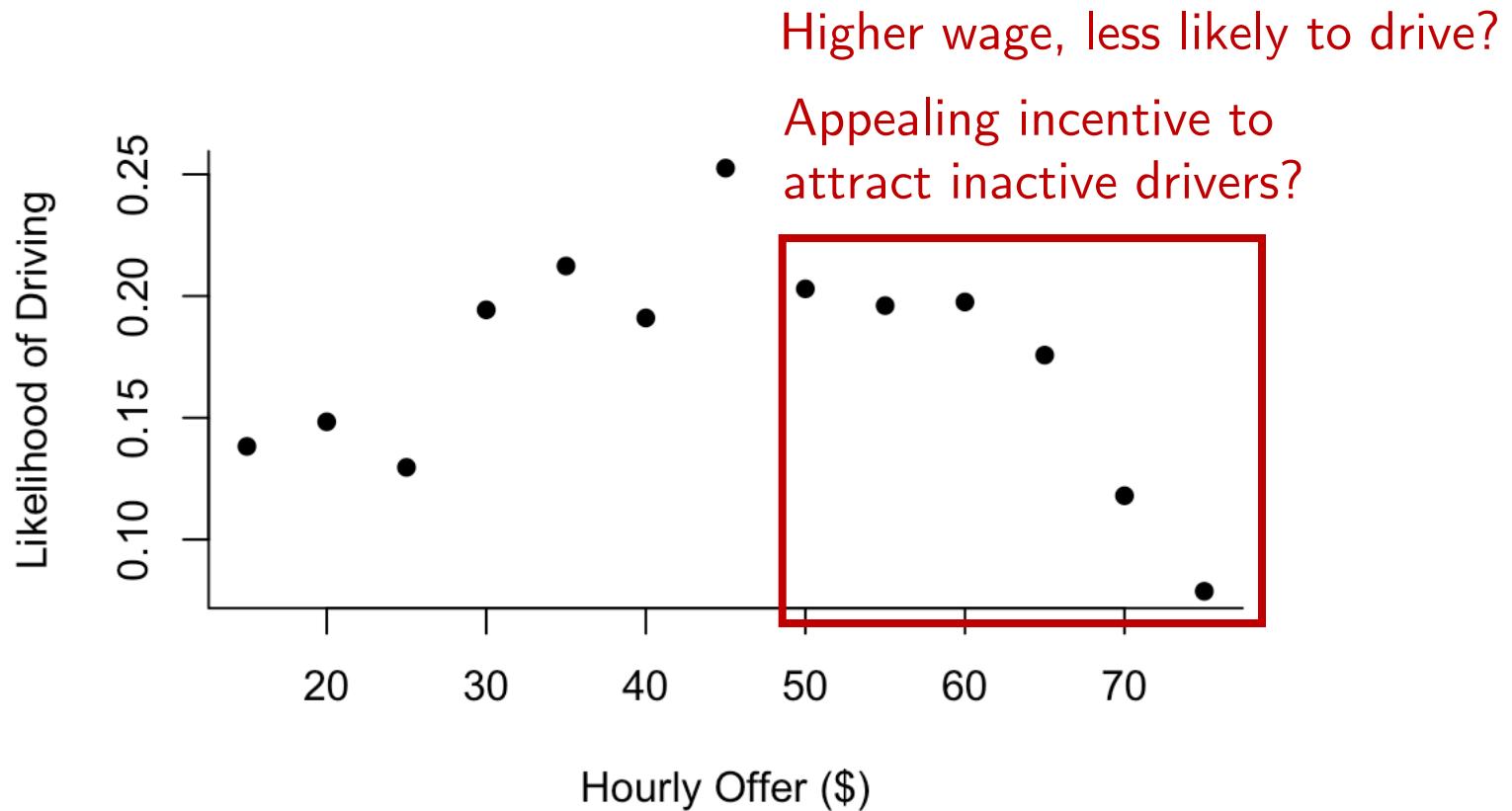
Empirical Strategy Challenges



Empirical Strategy Challenges

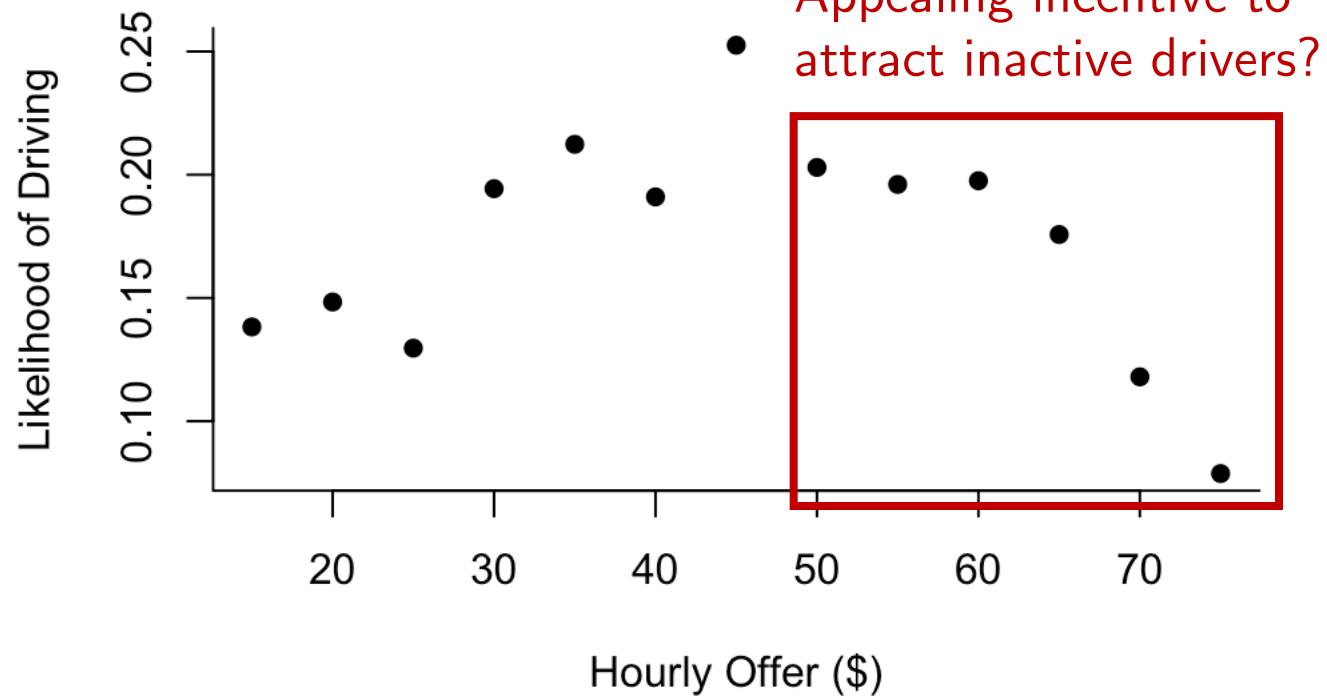


Empirical Strategy Challenges



Empirical Strategy Challenges

Simultaneity



Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable

Hourly offer

Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable

Instrument

Hourly offer

Average offers of “co-workers”

Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable

Instrument

Hourly offer

Average offers of “co-workers”

= currently available
+ made similar decisions
+ different vehicle type

Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable

Instrument

Hourly offer

Average offers of “co-workers”



sedan



Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable

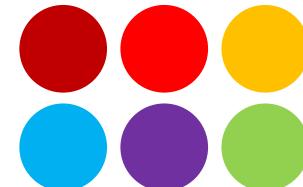
Instrument

Hourly offer

Average offers of “co-workers”



sedan



non-sedan

Empirical Strategy Challenges

Simultaneity

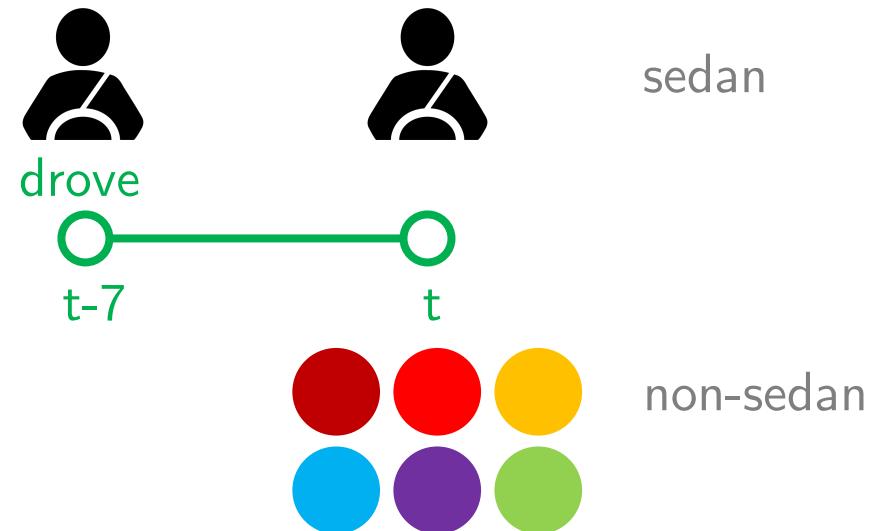
Solution: Instrumental Variables

Endogenous Variable

Hourly offer

Instrument

Average offers of “co-workers”



Empirical Strategy Challenges

Simultaneity

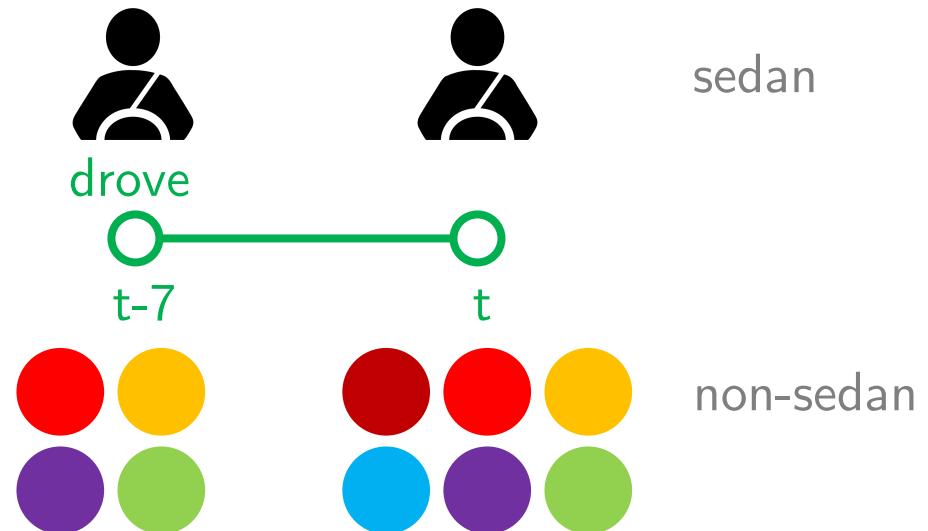
Solution: Instrumental Variables

Endogenous Variable

Instrument

Hourly offer

Average offers of “co-workers”



Empirical Strategy Challenges

Simultaneity

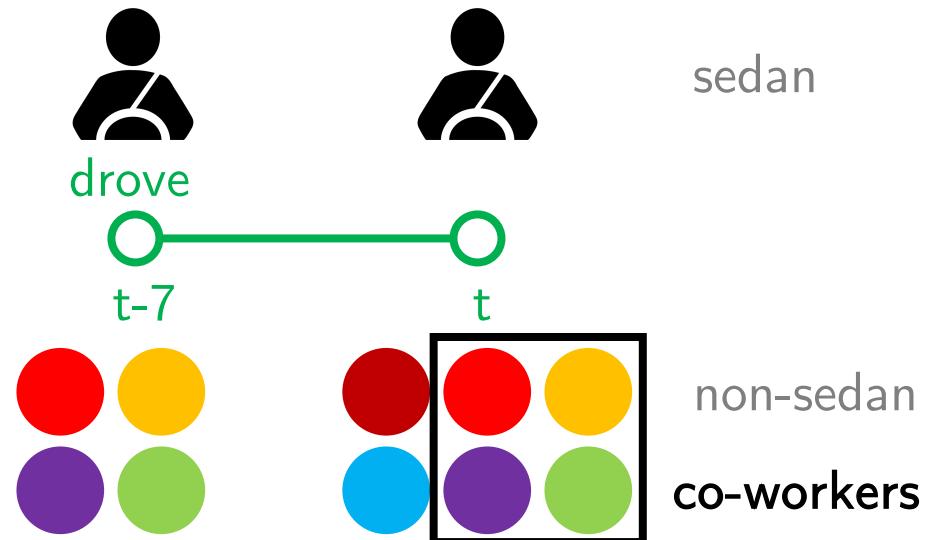
Solution: Instrumental Variables

Endogenous Variable

Instrument

Hourly offer

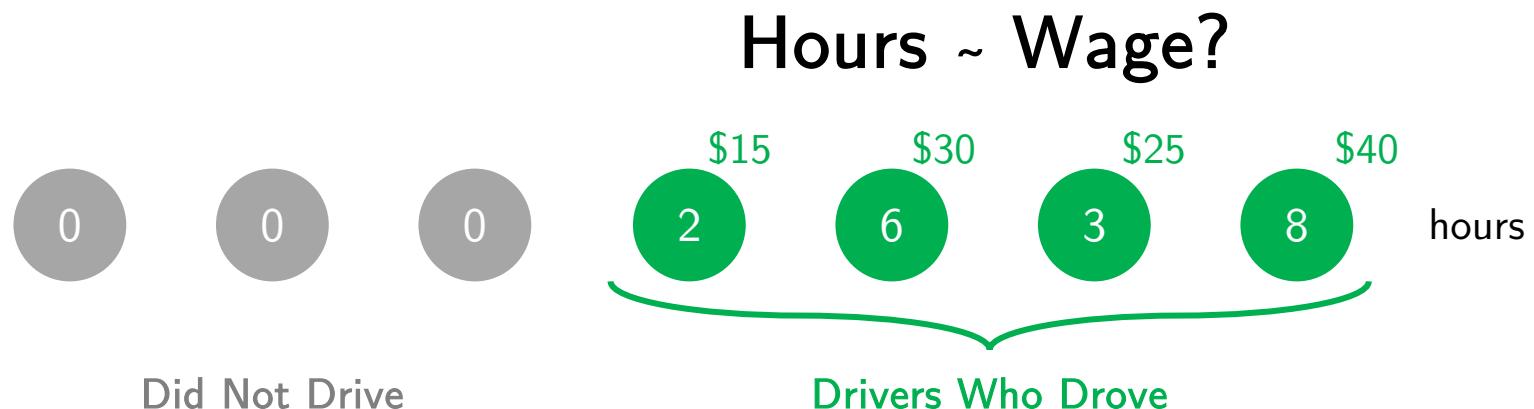
Average offers of “co-workers”



Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

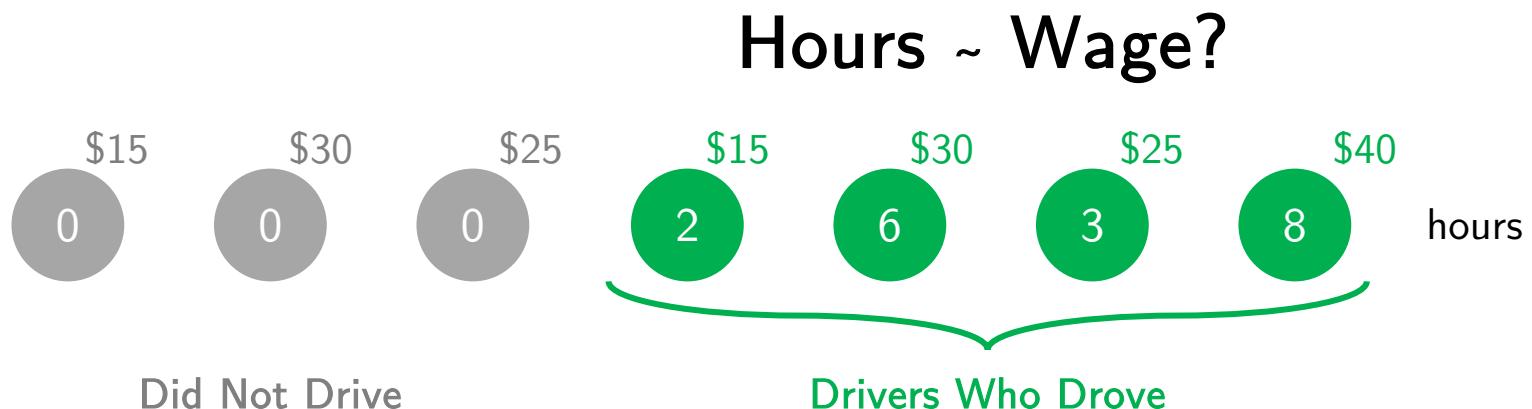


Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Decision to work is **not random**



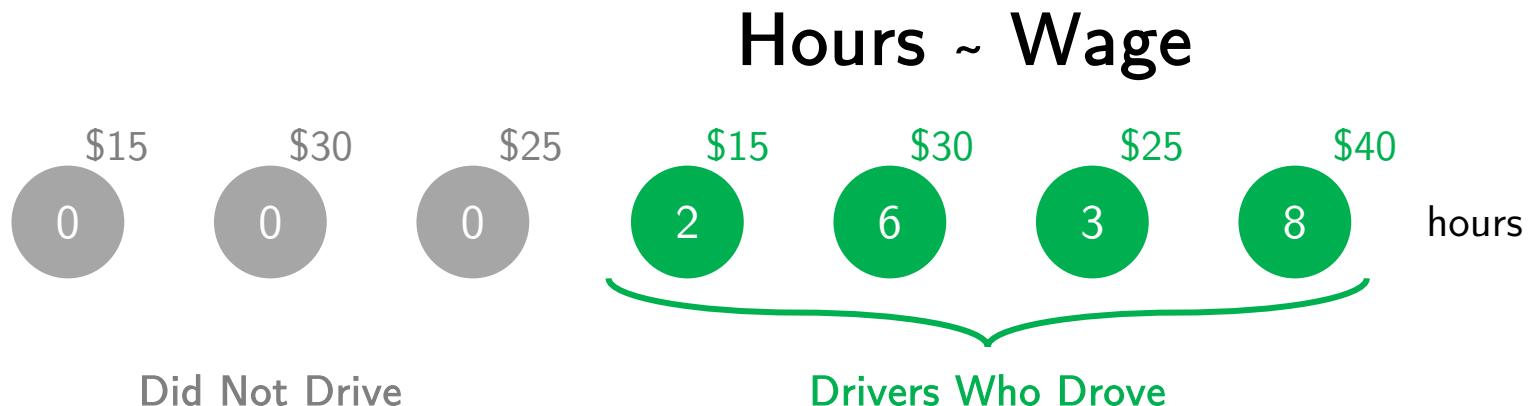
Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Selection Bias

Solution: Modified Heckman Two-Stage Method
("Heckit" - Heckman 1979)



Empirical Strategy

Heckman + IV

1 Work or not?

Control Function Probit:

$P(\text{drive})$ on *Offer*

Empirical Strategy

Heckman + IV

1 Work or not?

Control Function Probit:

$P(\text{drive}) \text{ on } \textit{Offer}$

+ Controls_1

Demand {

Short-term Habits {

Long-term Habits {

Empirical Strategy

Heckman + IV

1) Work or not?

Control Function Probit:

$P(\text{drive}) \text{ on } \textit{Offer}$

+ Controls_1



*Bias corrected with panel jackknife (Hahn & Newey 2004)

Empirical Strategy

Heckman + IV

1 Work or not?

Control Function Probit:

$P(\text{drive}) \text{ on } \textit{Offer} + \text{ISF}$ + Controls_1

Income So Far

= cumulative income since beginning of day

Empirical Strategy

Heckman + IV

1) Work or not?

Control Function Probit:

$P(\text{drive}) \text{ on } \textit{Offer} + \text{ISF} + \text{HSF} + \text{Controls}_1$

Income So Far

Hours So Far

= cumulative active hours
since beginning of day

Empirical Strategy

Heckman + IV

1 Work or not?

Control Function Probit:

$P(\text{drive}) \text{ on } \textit{Offer} + \text{ISF} + \text{HSF} + \text{Controls}_1$

Income So Far Hours So Far

Conditional
on working

2 How long to work?

2SLS with Fixed Effects

Hours on $\textit{Earning}$ + ISF + HSF + + Controls₂

Empirical Strategy

Heckman + IV

1 Work or not?

Control Function Probit:

$P(\text{drive}) \text{ on } \textit{Offer} + \text{ISF} + \text{HSF} + \text{Controls}_1$

Income So Far Hours So Far

Conditional
on working

2 How long to work?

2SLS with Fixed Effects

Hours on $\textit{Earning} + \text{ISF} + \text{HSF} + \text{IMR} + \text{Controls}_2$

Inverse Mills Ratio
= correct for selection bias

Empirical Strategy

Heckman + IV

1 Work or not?

CF: Regress hourly offer on IVs. Keep residuals

Probit: Estimate $P(\text{drive})$

$$P(\text{Drive}_{i,t} = 1 | \mathbf{X}_{i,t}) = \Phi(\alpha_{0,t} + \alpha_w w_{i,t} + \boldsymbol{\alpha} \mathbf{X}_{i,t} + \alpha_e \hat{e}_{i,t})$$

Bias corrected with
panel jackknife
(Hahn & Newey 2004)

Inverse Mills Ratio (IMR)

$$\lambda(c_z) = \frac{\phi(c_z)}{1 - \Phi(c_z)}$$

Conditional on
working

2 How long to work?

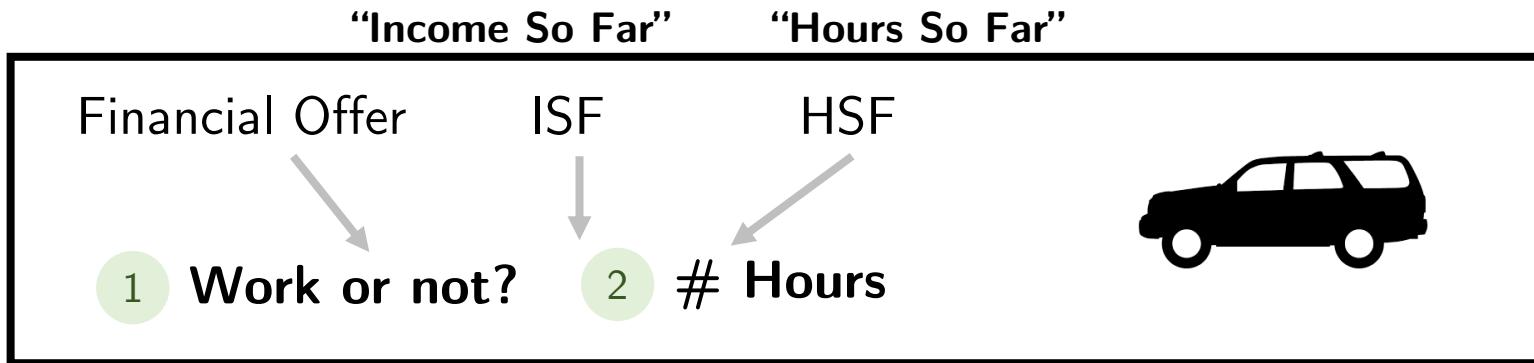
IV: Estimate hourly earning from IVs

OLS: Estimate hours

$$f(\text{Hour}_{i,t}) = \beta_{0,i} + \beta_w w_{i,t} + \boldsymbol{\beta} \mathbf{Z}_{i,t} + \theta \lambda_{i,t} + u_{i,t}$$

Adjust standard errors to account for the fact that IMR is an estimate
(and hence random) covariate in the above model.

Results



Within-Day

Midday



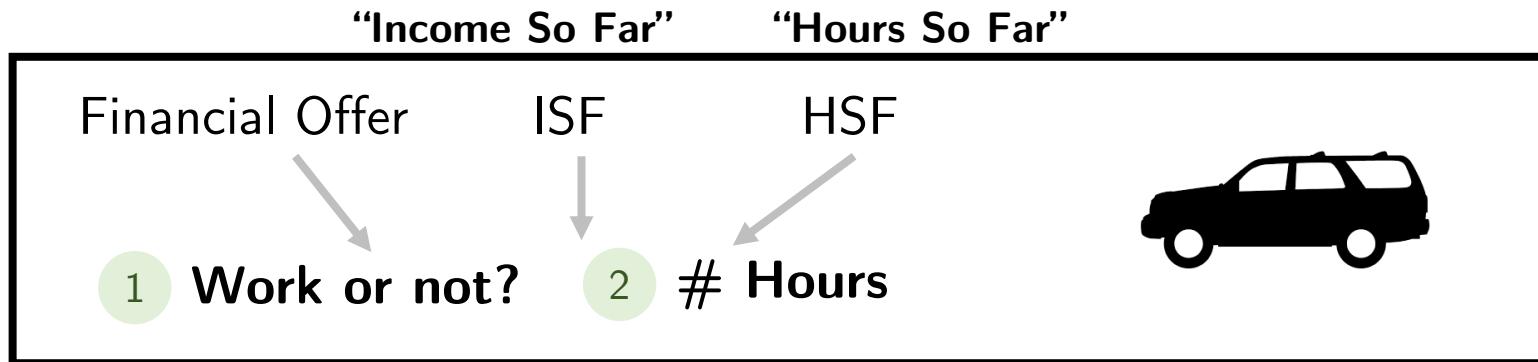
Late Night

Across-Days

Tuesday

Sunday

Results



Within-Day

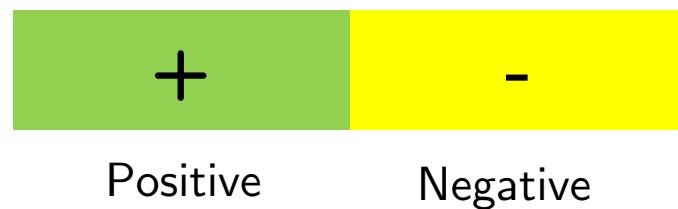
Midday

Late Night

Across-Days

Tuesday

Sunday



Results Within Day

1

Work or not?

SUV	Mean	IV-F	N
Midday	0.343	372.9	124,769
PM-Peak	0.277	345.1	131,910
PM-OPeak	0.182	320.6	130,651
Late Night	0.117	379.0	125,382

Results Within Day

1

Work or not?

SUV	Mean	IV-F	Offer	N
Midday	0.343	372.9	+	124,769
PM-Peak	0.277	345.1	-	131,910
PM-OPeak	0.182	320.6	+	130,651
Late Night	0.117	379.0	+	125,382

Financial incentives have a generally positive impact.

Results Within Day

1

Work or not?

SUV	Mean	IV-F	Offer	ISF	N
Midday	0.343	372.9	+	-	124,769
PM-Peak	0.277	345.1	-	-	131,910
PM-OPeak	0.182	320.6	+	-	130,651
Late Night	0.117	379.0	+	-	125,382

Income Targeting:

The more you earned,
the less likely you'll work
a new shift.

Income targeting has a
consistently negative
impact.

Results Within Day

1

Work or not?

SUV	Mean	IV-F	Offer	ISF	HSF	N
Midday	0.343	372.9	+	-	+	124,769
PM-Peak	0.277	345.1	-	-	+	131,910
PM-OPeak	0.182	320.6	+	-	+	130,651
Late Night	0.117	379.0	+	-	+	125,382

Inertia:

The longer you've been active, the more likely you'll work another shift.

Inertia has a **consistently positive impact**.

Results Within Day

1

2

Work or not?

Hours

SUV	Mean	IV-F	Offer	ISF	HSF	N	Mean	IV-F	N
Midday	0.343	372.9	+	-	+	124,769	4.987	180.2	45,329
PM-Peak	0.277	345.1	-	-	+	131,910	2.421	58.5	39,592
PM-OPeak	0.182	320.6	+	-	+	130,651	0.731	50.5	26,699
Late Night	0.117	379.0	+	-	+	125,382	1.996	39.91	17,137

Results Within Day

1

2

Work or not?

Hours

SUV	Mean	IV-F	Offer	ISF	HSF	N	Mean	IV-F	Earn	ISF	HSF	R ²	N
Midday	0.343	372.9	+	-	+	124,769	4.987	180.2	+	-	+	0.552	45,329
PM-Peak	0.277	345.1	-	-	+	131,910	2.421	58.5	+	-	+	0.244	39,592
PM-OPeak	0.182	320.6	+	-	+	130,651	0.731	50.5	+	-	+	0.281	26,699
Late Night	0.117	379.0	+	-	+	125,382	1.996	39.91	+	-	+	0.296	17,137

The three effects are consistent in both stages

Results Within Day

1

2

Work or not?

Hours

SUV	Mean	IV-F	Offer	ISF	HSF	N	Mean	IV-F	Earn	ISF	HSF	R ²	N
Midday	0.343	372.9	+	-	+	124,769	4.987	180.2	+	-	+	0.552	45,329
PM-Peak	0.277	345.1	-	-	+	131,910	2.421	58.5	+	-	+	0.244	39,592
PM-OPeak	0.182	320.6	+	-	+	130,651	0.731	50.5	+	-	+	0.281	26,699
Late Night	0.117	379.0	+	-	+	125,382	1.996	39.91	+	-	+	0.296	17,137

% Change in P(Work)

Change in Minutes Worked

SUV	+\$10 Offer	+\$10 ISF	+1h HSF	+\$10 Earn	+\$10 ISF	+1h HSF
Midday	0.82	-5.73	57.21	0.51	-4.87	109.53

Results Across Days

Results Across Days

1 Work or not?	SUV	Mean	IV-F	N
Tuesday	0.409	43.6		28,883
Wednesday	0.418	55.9		21,965
Thursday	0.426	73.4		29,233
Friday	0.412	74.0		20,294
Saturday	0.203	98.1		15,788
Sunday	0.162	82.2		13,025

Results Across Days

1 Work or not?

	SUV	Mean	IV-F	Offer	ISF	HSF	N
Tuesday	0.409	43.6	+	+	+	+	28,883
Wednesday	0.418	55.9	+	+	+	+	21,965
Thursday	0.426	73.4	+	+	+	+	29,233
Friday	0.412	74.0	+	+	+	+	20,294
Saturday	0.203	98.1	-	-	-	+	15,788
Sunday	0.162	82.2	-	-	-	+	13,025

Inertia has consistent positive effect.

Financial incentive and cumulative income switch from positive to negative later on.

Results Across Days

1 Work or not?

	SUV	Mean	IV-F	Offer	ISF	HSF	N
Tuesday	0.409	43.6	+	+	+	+	28,883
Wednesday	0.418	55.9	+	+	+	+	21,965
Thursday	0.426	73.4	+	+	+	+	29,233
Friday	0.412	74.0	+	+	+	+	20,294
Saturday	0.203	98.1	-	-	-	+	15,788
Sunday	0.162	82.2	-	-	-	+	13,025

2 # Hours

	IV-F	Earn	ISF	HSF	R ²	N
Tuesday	18.3	-	-	+	0.422	9,482
Wednesday	26.2	-	-	+	0.422	10,120
Thursday	34.6	-	-	+	0.412	9,894
Friday	33.7	+	-	+	0.436	9,283
Saturday	19.1	-	+	-	0.398	4,372
Sunday	15.1	+	+	-	0.390	3,240

Hours not decided
at the day level

Results Summary

Neoclassical
Financial Incentive

As day/week proceeds...



encourages working

Results Summary

As day/week proceeds...

Neoclassical
Financial Incentive

encourages working

Behavioral
Income Target

discourages working later on

Results Summary

As day/week proceeds...

Neoclassical
Financial Incentive

encourages working

Behavioral
Income Target

discourages working later on

New
Inertia

encourages working

Robustness Tests

- Isolating ISF and HSF effect
 - Positive HSF (inertia) effect dominates ISF (targeting) effect.
- Nonlinearity in ISF and HSF
- Instrumenting for ISF and HSF
- Controlling on types of promotions
- Other approaches to sample selection
 - Two-part models: insights stay the same in both parts
 - Dahl's correction: using B-splines instead of IMR
- Looking at granular log-on/log-off/break data
 - Smaller sample of sedan and SUV drivers
 - The longer the previous active session is,
the longer the current active session is.

Outline

- What has been done
 - Practice / labor economics / OM / hypotheses
- Data and empirical strategy
 - Dealing with endogeneity and selection bias
- Results
 - Impact of incentive and behavioral elements on labor decisions
- Implications
 - Simulation of optimal incentive re-allocation

Optimal Targeted Incentive



Optimal Targeted Incentive



Optimal Targeted Incentive

Ranking each driver by her
minimum work-inducing incentive

= how much to trigger working decision



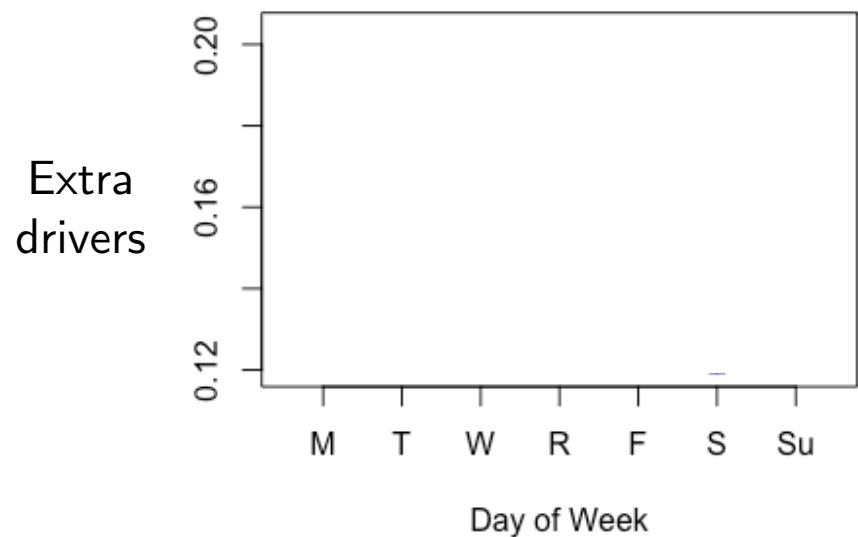
Reallocating Incentives

Compared to current practice from Jan to Sep 2017 out-of-sample
(Using data from Oct 2016 to right before the focal date as training)

Reallocating Incentives

Compared to current practice from Jan to Sep 2017 out-of-sample

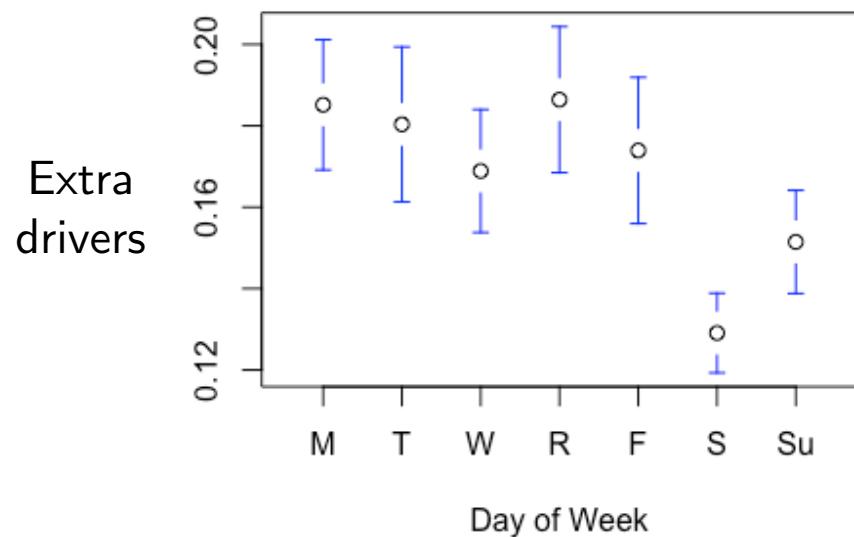
Given the same budget



Reallocating Incentives

Compared to current practice from Jan to Sep 2017 out-of-sample

Given the same budget



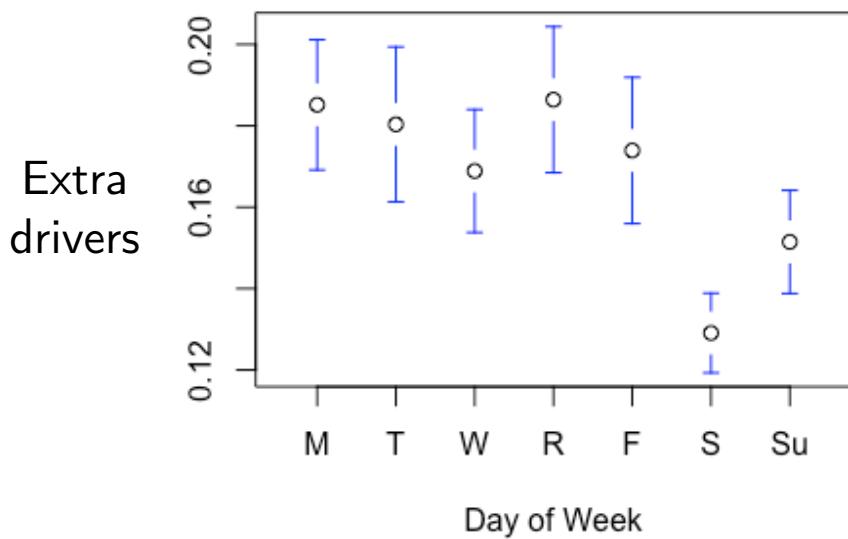
Can recruit **17% more drivers**

Average promo: 1.61x

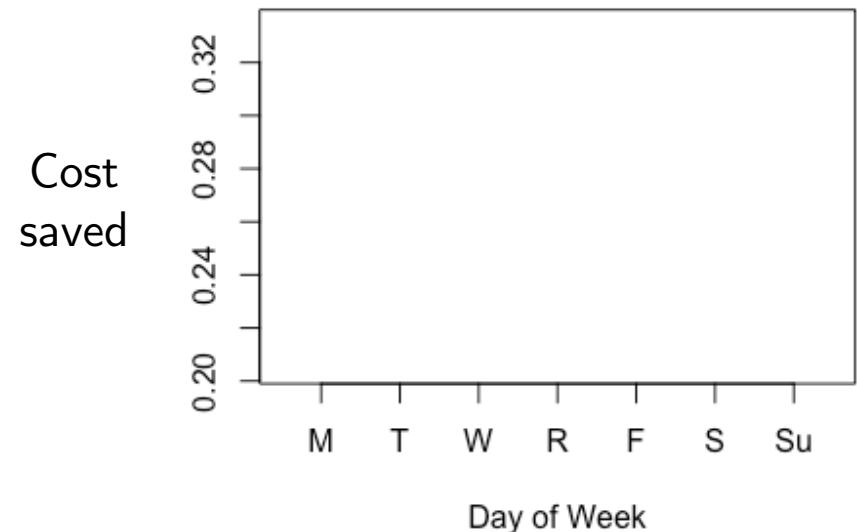
Reallocating Incentives

Compared to current practice from Jan to Sep 2017 out-of-sample

Given the same budget



Given the same capacity



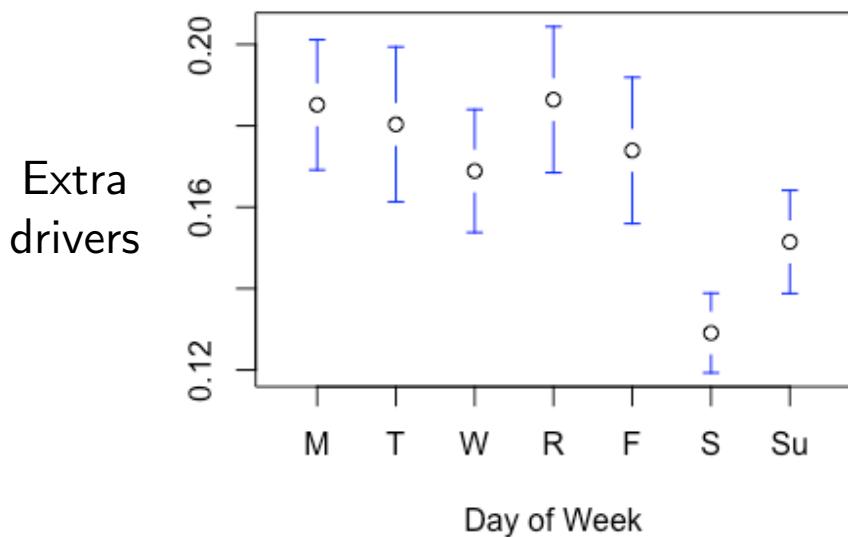
Can recruit **17% more drivers**

Average promo: 1.61x

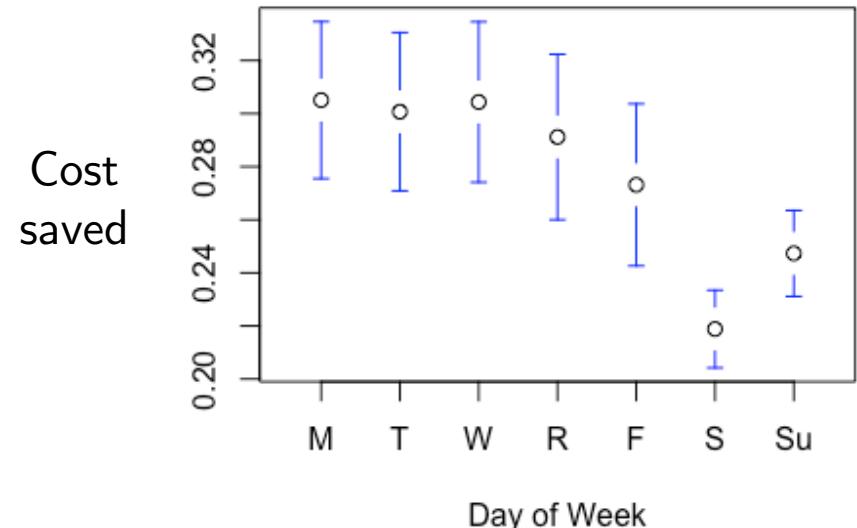
Reallocating Incentives

Compared to current practice from Jan to Sep 2017 out-of-sample

Given the same budget



Given the same capacity



Can recruit **17% more drivers**
Average promo: 1.61x

Costs 28% less to maintain capacity

Summary

How do gig economy workers make labor decisions?

Approach

- Shift-level data from ride-hailing company
- Modified Heckman estimation w/ IVs and fixed effects

Findings

As day/week proceeds...

Neoclassical

Financial Incentive

encourages working

Behavioral

Income Target

discourages working later on

New phenomenon

Inertia

encourages working

Implications

- Compared to current practice, our approach can improve service capacity without incurring extra cost or maintain the same capacity at a lower cost

Appendix

Results Within Day/Effect Size

1			2		
% Change in P(Work)			Change in Minutes Worked		
+\$10 Offer	+\$10 ISF	+1h HSF			
			+\$10 Earn	+\$10 ISF	+1h HSF

Results Within Day/Effect Size

1

SUV	+\$10 Offer	+\$10 ISF	+1h HSF
Midday	0.82	-5.73	57.21
PM-Peak	-4.39	-0.57	15.27
PM-OPeak	0.27	-0.36	6.43
Late Night	0.34	-0.22	3.32

2

Change in Minutes Worked	+\$10 Earn	+\$10 ISF	+1h HSF
0.51	-4.87	109.53	
13.64	-0.24	18.96	
1.72	-0.08	1.18	
14.87	-0.11	1.35	

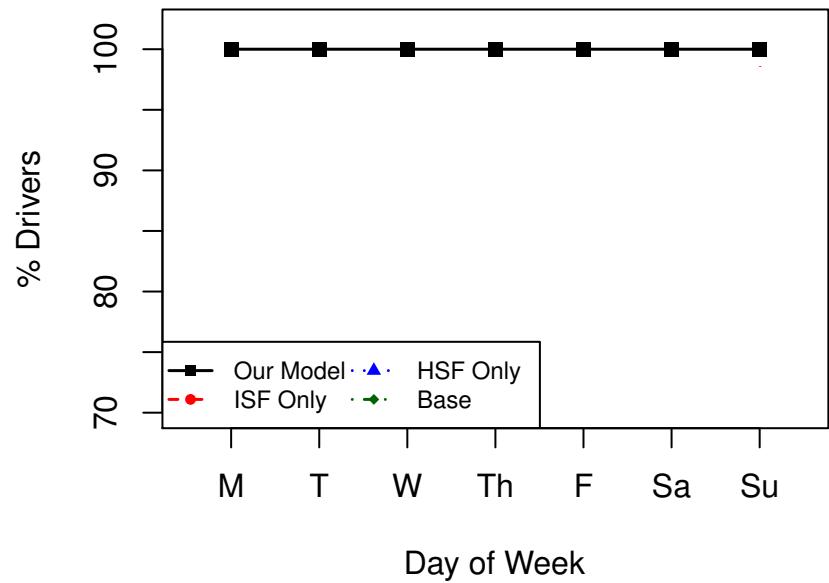
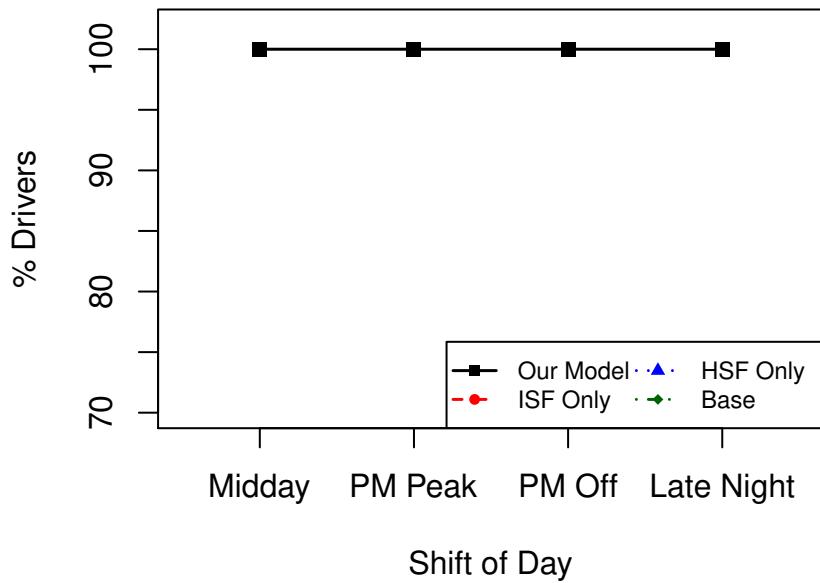
Results Within Day/Effect Size

	1			2		
	% Change in P(Work)			Change in Minutes Worked		
SUV	+\$10 Offer	+\$10 ISF	+1h HSF		+\$10 Earn	+\$10 ISF
Midday	0.82	-5.73	57.21		0.51	-4.87
PM-Peak	-4.39	-0.57	15.27		13.64	-0.24
PM-OPeak	0.27	-0.36	6.43		1.72	-0.08
Late Night	0.34	-0.22	3.32		14.87	-0.11
						109.53
						18.96
						1.18
						1.35

For an average SUV driver during Midday,
 +\$10 hourly wage → P(work) increases by 1% / extends duration by 30s
 +\$10 cumulative income → P(work) drops 6% / shortens duration by 5m
 +1 cumulative hours → P(work) increases by 57% / extends duration by 2h

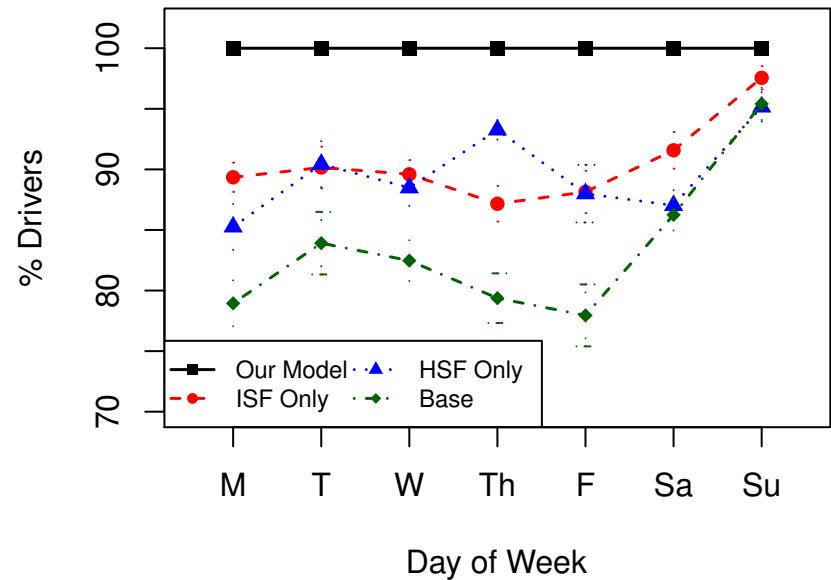
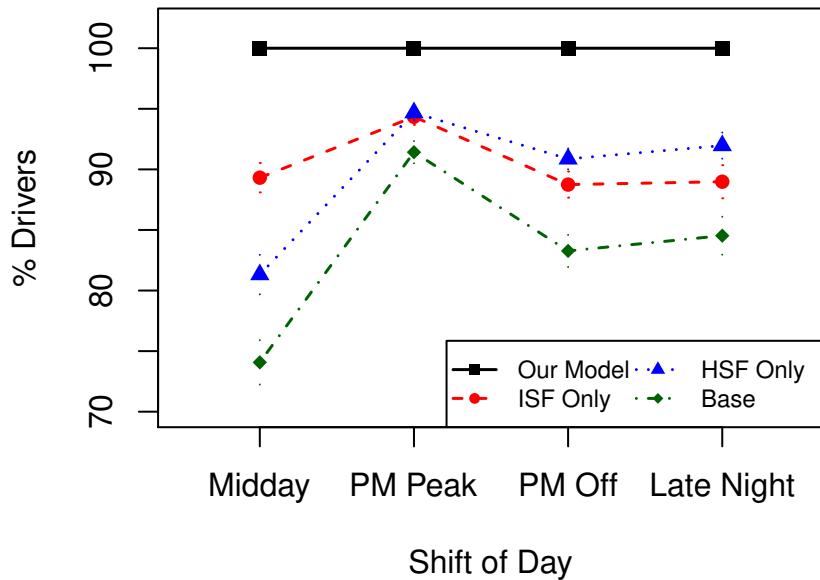
Ignoring Behavioral Factors

Assuming our model is correct, how many workers the firm would fail to attract if it did not incorporate income targeting and inertia?



Ignoring Behavioral Factors

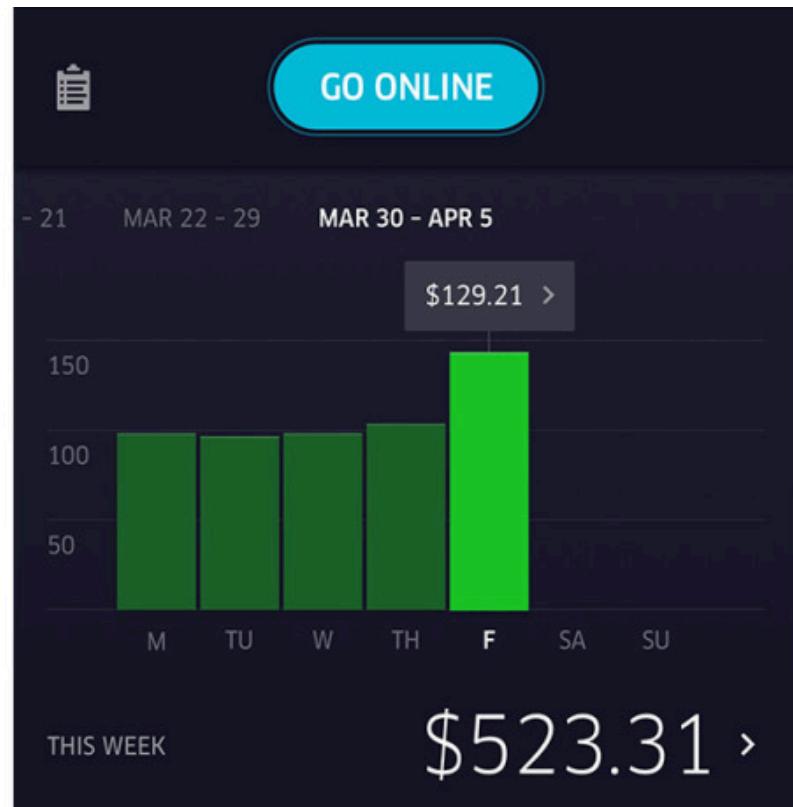
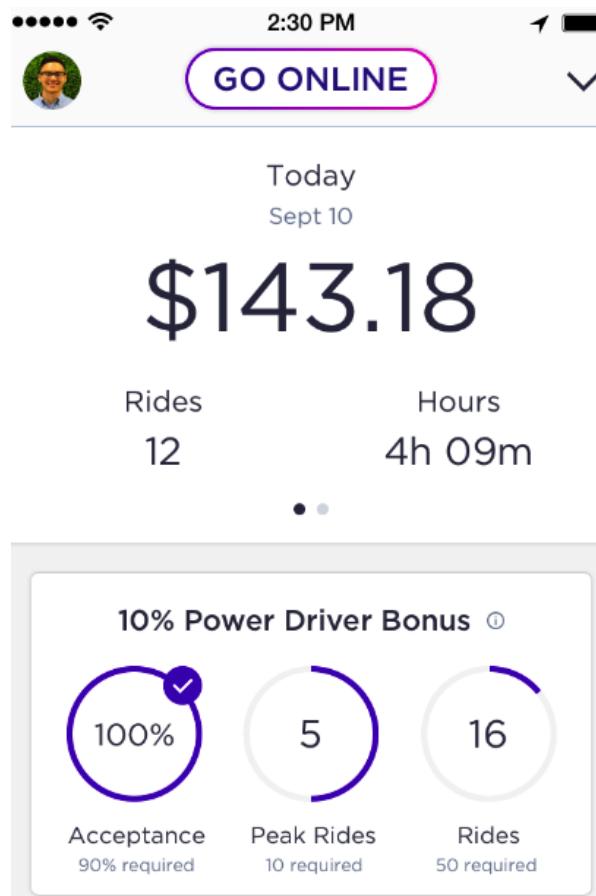
Assuming our model is correct, how many workers the firm would fail to attract if it did not incorporate income targeting and inertia?



Average loss in capacity:

Ignoring: Income targeting 10.32%, inertia 9.63%, or both 16.70%

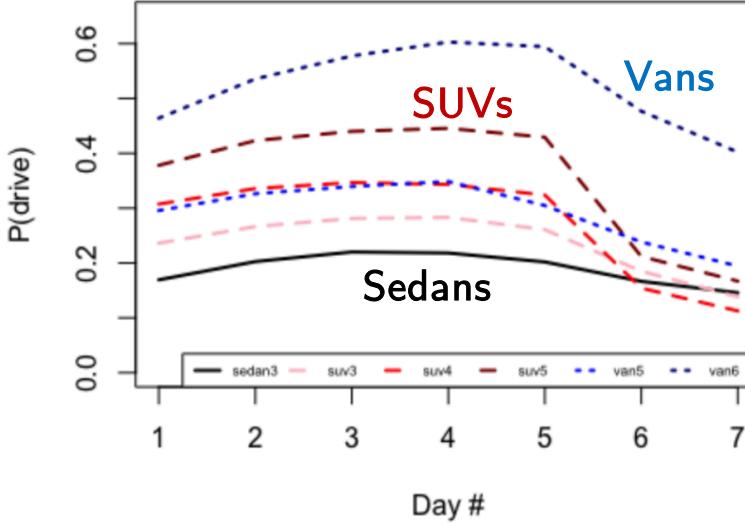
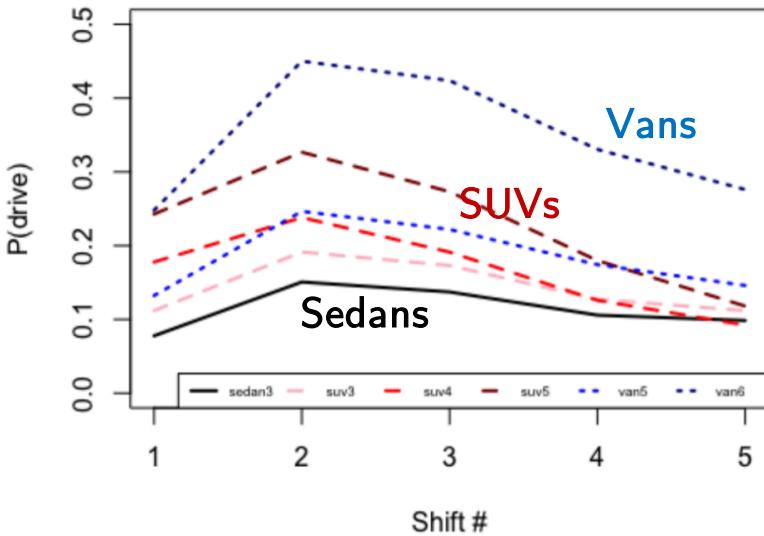
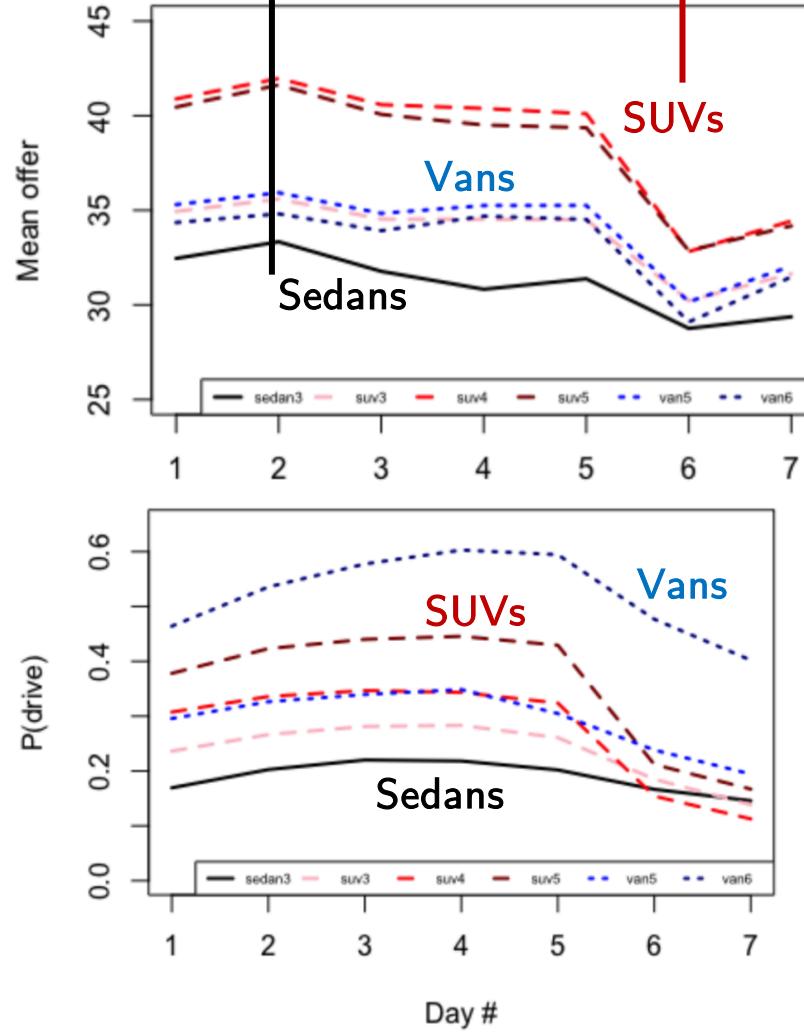
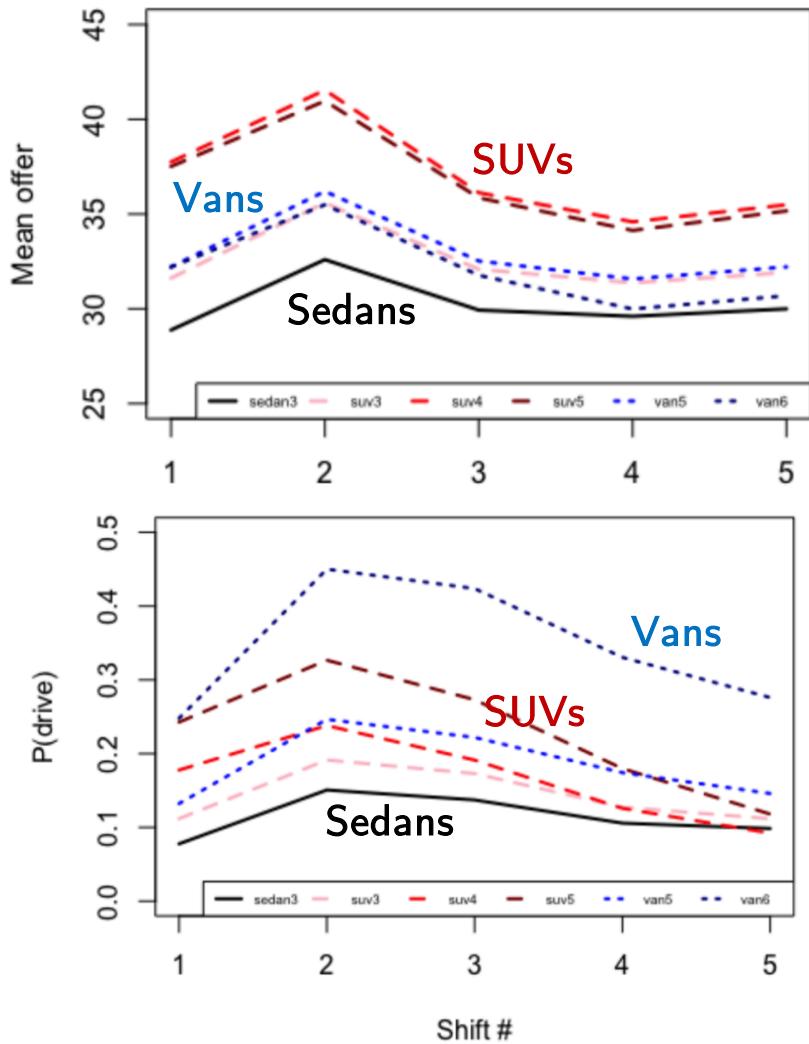
Driver's View



Drivers

5.33 hrs/day, 12.65 hrs/wk

4.86 hrs/day, 5.86 hrs/wk



Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable

Instrument

Hourly offer

Average offers of “co-skippers”

Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable

Instrument

Hourly offer

Average offers of “co-skippers”

Level of inactivity for drivers eligible to work today

last worked
6+ days ago

last worked
3-5 days ago

last worked
2 days ago

last worked
yesterday

Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable

Instrument

Hourly offer

Average offers of “co-skippers”

Level of inactivity for drivers eligible to work today

last worked
6+ days ago

last worked
3-5 days ago

last worked
2 days ago

last worked
yesterday

Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable	Instrument
Hourly offer	Average offers of “co-skippers”

Level of inactivity for drivers eligible to work today



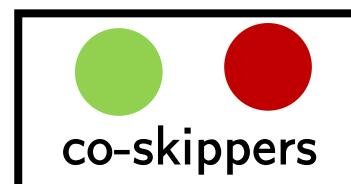
Empirical Strategy Challenges

Simultaneity

Solution: Instrumental Variables

Endogenous Variable	Instrument
Hourly offer	Average offers of “co-skippers”

Level of inactivity for drivers eligible to work today



non-sedan

Gig Economy x Retail



- **Retail candidates value flexible shift patterns** and shorter work weeks over compensation and benefits.
- Benefits of adopting flexible workforce: **Seasonality, resolving high turnover, matching consumer trend, high quality/fresh perspective**
- Many have already adopted/worked closely with gig companies
 - Delivery business: Walmart x Uber/Lyft, GM x Lyft, Apple x Didi
 - Flexible staffing: IKEA x TaskRabbit, Samsung x Upwork



Heckman Sample Selection

Suppose that the pattern of missingness (I'll refer to this as censored hereafter) is related to the latent (unobserved) process

$$\mathbf{z}^* = \mathbf{w}\gamma + \mathbf{u}$$

From this process, the researcher can observe

$$\begin{aligned} z_i &= 1 \text{ if } z_i^* > 0 \\ &= 0 \text{ if } z_i^* \leq 0 \end{aligned}$$

or $z_i = 1$ (y_i not censored) when

$$u_i \geq -\mathbf{w}_i\gamma$$

The probability of y_i not censored is

$$\begin{aligned} Pr(u_i \geq -\mathbf{w}_i\gamma) &= 1 - \Phi(-\mathbf{w}_i\gamma) \\ &= \Phi(\mathbf{w}_i\gamma) \end{aligned}$$

if we are willing to assume that $\mathbf{u} \sim N(\mathbf{0}, \mathbf{I})$. Note for identification purposes in the Heckman Model we restrict $Var(u_i) = 1$. Also note that $1 - \Phi(-\mathbf{w}_i\gamma) = \Phi(\mathbf{w}_i\gamma)$ by symmetry of the standard normal distribution.

Heckman Sample Selection

Having constructed a model for censoring, we can construct "amounts" equation as follows. Denoting \mathbf{y} as the not censored (observed) dependent variable, the censoring model defines what is in the estimation sample as

$$y_i = y_i^* = \mathbf{x}_i\beta + \epsilon_i \text{ observed, if } z_i = 1$$

Finally, the joint distribution of the errors in the selection (u_i) and amounts equation (ϵ) is distributed iid as

$$\begin{bmatrix} u_i \\ \epsilon_i \end{bmatrix} \sim Normal \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & \sigma_\epsilon^2 \end{bmatrix} \right)$$

To see how the selection and amounts model are related, consider

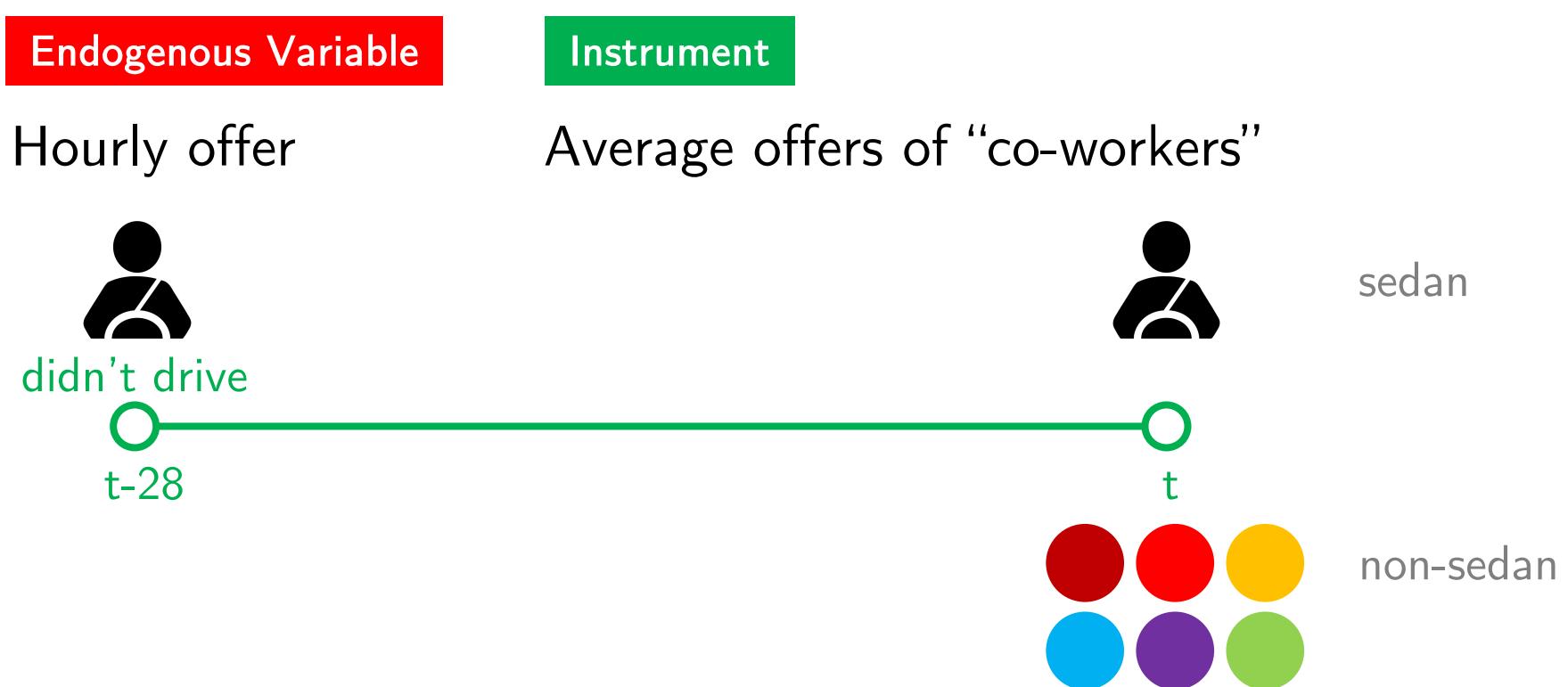
$$\begin{aligned} E(y_i | y_i \text{ observed}) &= E(y_i | z^* > 0) \\ &= E(y_i | u_i > -\mathbf{w}_i\gamma) \\ &= \mathbf{x}_i\beta + E(\epsilon_i | u_i > -\mathbf{w}_i\gamma) \\ &= \mathbf{x}_i\beta + \rho\sigma_\epsilon \frac{\phi(\mathbf{w}_i\gamma)}{\Phi(\mathbf{w}_i\gamma)} \end{aligned}$$

What is immediately apparent is that the conditional mean ($E(y_i | y_i \text{ observed})$) differs from the unconditional mean ($\mathbf{x}_i\beta$) only if $\rho \neq 0$ since all the other elements in the far right hand term (i.e., the variance of the error in the amounts equation, σ_ϵ , and the Inverse Mills Ratio, $\frac{\phi(\mathbf{w}_i\gamma)}{\Phi(\mathbf{w}_i\gamma)}$) in the preceding equation are strictly positive. So if the errors in the amounts and selection equations are uncorrelated ($\rho = 0$) we can safely apply ordinary least squares to uncover unbiased estimates for β and can ignore endogenous selection effects and the selection equation portion of the model.

Empirical Strategy Challenges

Simultaneity

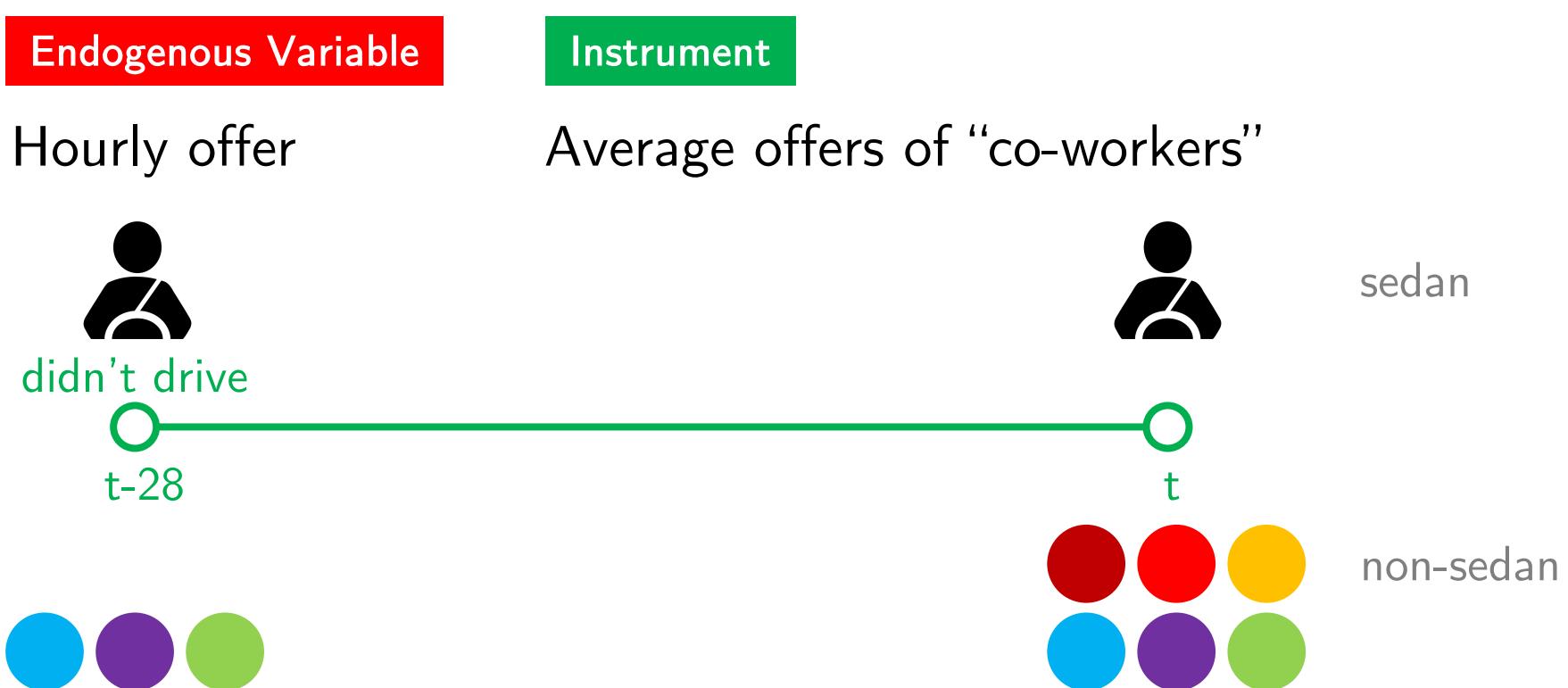
Solution: Instrumental Variables



Empirical Strategy Challenges

Simultaneity

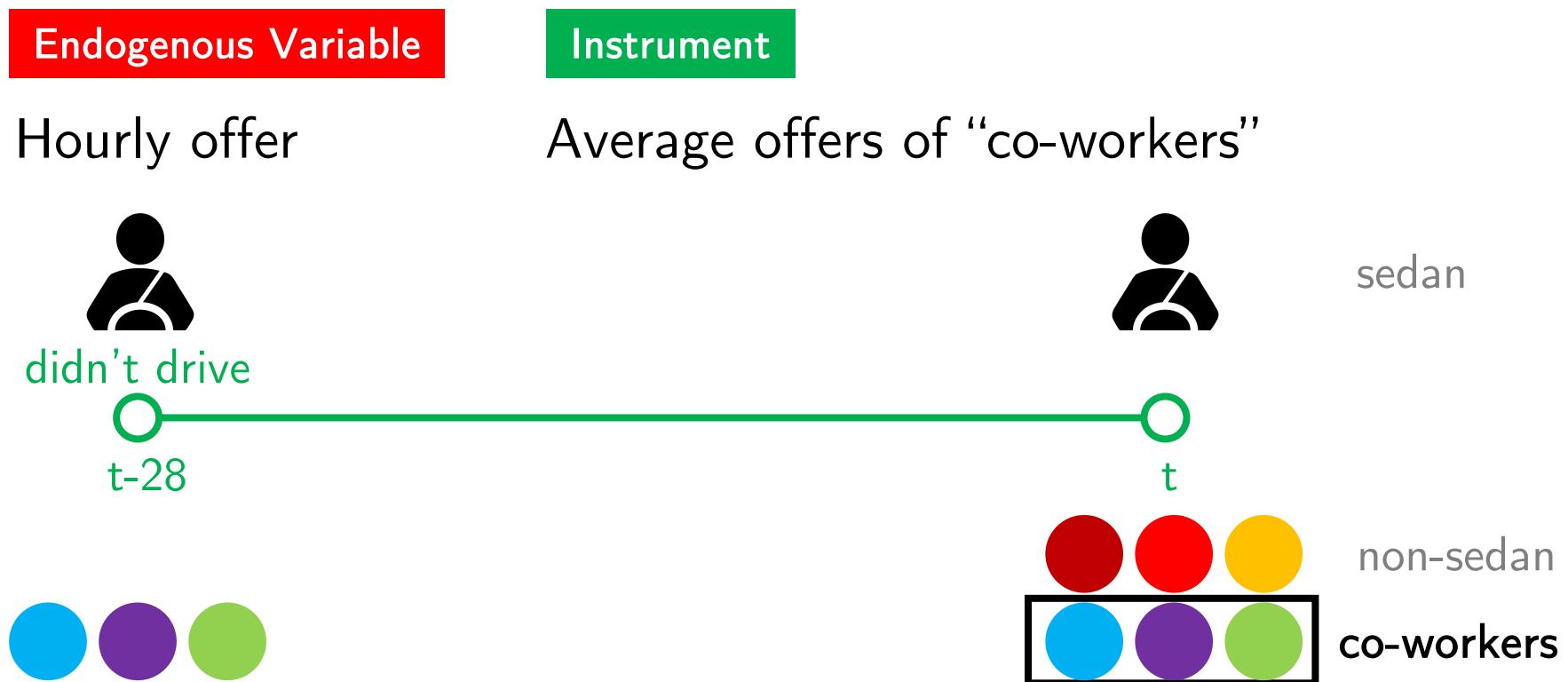
Solution: Instrumental Variables



Empirical Strategy Challenges

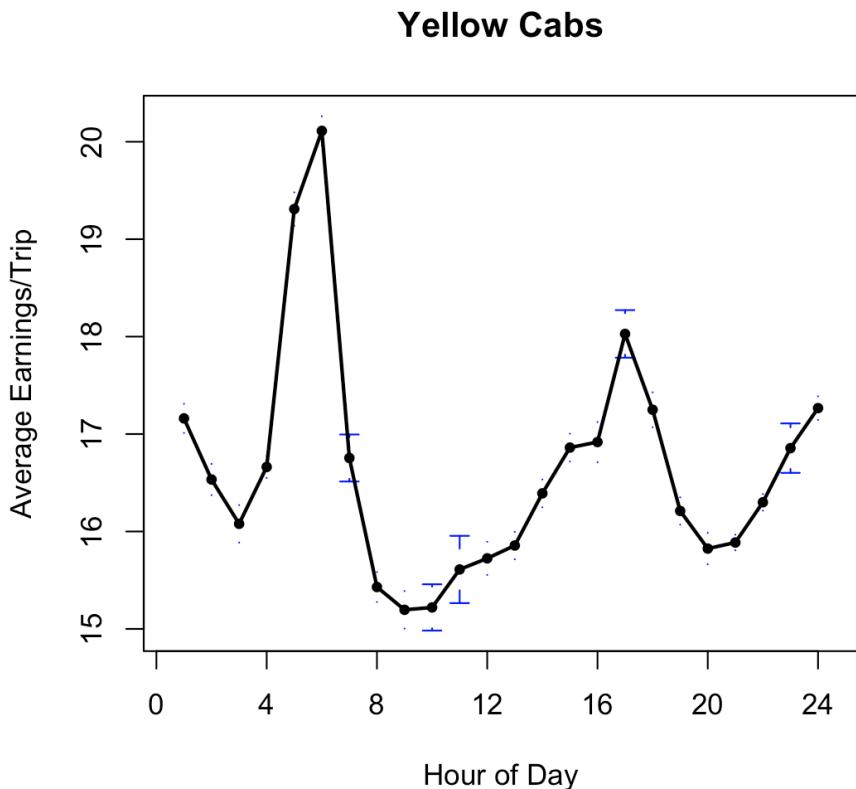
Simultaneity

Solution: Instrumental Variables



TLC Data

Fares/earnings for all yellow cab trips



101M yellow cab trips

