Gender Differences in Labor Supply: Experimental Evidence from the Gig Economy*

Sydnee Caldwell — UC Berkeley

Emily Oehlsen — University of Oxford, Open Philanthropy

July 2022

Abstract

We use field experiments to study market-level and firm-level labor supply. Market-level elasticities govern how labor supply responds to temporary productivity shocks. Firm substitution elasticities determine wage markdowns in frictional markets. We find that women are twice as elastic to the market as men. This is true even among high hours individuals. We find no evidence that women are less likely to switch between firms when relative wages change. Our results suggest that in environments without differences in firm location or amenities, firms with market power have little incentive to pay equally productive women less.

JEL: J22, J31, J42

Keywords: labor supply, monopsony, gig economy, gender wage gap

---

*Caldwell: scaldwell@berkeley.edu; Oehlsen: emily.oehlsen@economics.ox.ac.uk.

Caldwell thanks Daron Acemoglu, Joshua Angrist, and Patrick Kline for guidance and support. Special thanks go to the staff of Uber's Houston city team for hosting and assisting with our study and to Christopher Ackerman, Phoebe Cai, Javier Feinmann, Anran Li, and Tadej Svetina for excellent research assistance. We thank David Autor, David Card, Arindrajit Dube, Jonathan Hall, Jonathan Kolstad, Dan Knoepfle, Alexandre Mas, Jonathan Meer, Elizabeth Mishkin, Suresh Naidu, Ricardo Perez-Truglia, Raffaele Saggio, Heather Sarsons, Todd Sorensen, and Steve Tadelis for helpful comments. We also thank seminar audiences at Arizona State University, Berkeley-Haas, the Bank of Italy, the Federal Reserve Bank of Dallas, MIT, Oxford, U Mass Amherst, the University of Melbourne, U Nevada-Reno, Toronto-Rotman, SOLE, the U Chicago Advances in Field Experiments conference, the CesIfo Gender in the Developed & Developing World conference, the National Tax Association Meeting, the AEA/ASSA annual meetings, the Columbia Junior Micro-Macro Labor Conference, and the EIEF Junior Applied Micro workshop. This study is registered in the AEA RCT Registry as trial no. AEARCTR-0001656. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374 (Caldwell) and by the National Science Foundation Dissertation Improvement Grant No. 1729822 (Caldwell). Some of the results were previously circulated as "Monopsony and the Gender Wage Gap: Experimental Evidence from the Gig Economy." The views expressed are those of the authors and do not necessarily reflect those of Uber Technologies, Inc. Caldwell's work on this project was carried out under a data use agreement executed between MIT and Uber. Oehlsen is a former employee of Uber Technologies, Inc. This manuscript was not subject to prior review by any party, except to verify that it does not contain any confidential data or information. All mistakes are our own.

> "Perfect discrimination is probably rare in buying labor but imperfect discrimination may often be found. For instance there may be two types of workers (for example, men and women or men and boys) whose efficiencies are equal but whose conditions of [labor] supply are different. It may be necessary to pay the same wage within each group, but the wages of the two groups (say of men and of women) may differ."
> —Joan Robinson (1933)

1 Introduction

Labor supply elasticities are important in many economic models. Market-level Frisch elasticities are key parameters in macro business cycle models because they govern how labor supply, and therefore output, responds to temporary productivity shocks. Firm-level elasticities are key to wage-setting models; when firms have market power, wages depend both on worker productivity and on how quickly workers leave or join a firm when relative wages change. Many have informally argued that, even if women are more elastic to the market, women may be less elastic at the firm-level: women may be less likely than men to leave a firm for high wages elsewhere.[1] This could lead firms to pay women less than men.

Ride-share provides a unique setting in which to estimate both firm- and market-level labor supply elasticities. By introducing temporary, exogenous variation in wages in a setting where Uber is the only available firm, we can estimate market-level responses. Variation in access to Lyft—driven by the temporary withdrawal of Lyft from the Houston market and by differences in eligibility requirements—allows us to examine how labor supply elasticities vary when individuals have access to multiple firms. Exogenous, individual-level variation in wages allows us to rule-out competing explanations driving firm substitution, such as changes in firm-provided

[1] This could happen if women are more loyal to their employers (i.e. have higher switching costs), have less information about their outside opportunities, have different valuations for employer-provided amenities, or face smaller effective labor markets due to differences in commuting costs (Manning, 2003).

amenities.

We conducted a series of experiments in which we offered random samples of Uber drivers the opportunity to drive for one week with 10-50% higher earnings per trip. Both the week and generosity of the offer varied from driver to driver. While some drivers had access to a competing ride-share company, others did not. Though most Uber drivers are male, we structured our experiment to include roughly equal numbers of male and female drivers. We also included both part- and full- time drivers. The fact that our experiments were small, relative to the size of the market, allows us to abstract from general equilibrium concerns.

We start by presenting a simple model of individual labor supply that illustrates the relationship between market-level and firm-level labor supply elasticities. We show how labor supply elasticities estimated using data from a single firm may combine changes in workers' total hours worked and changes in workers' allocation of these hours across firms. Most macro-economic models require an estimate of how overall labor supply changes when wages change. We then show how to adapt standard monopsony mark-down equations for a flexible hours environment if workers can work for multiple employers ("moonlight") and if they cannot (Manning, 2003). When hours are flexible and workers work for one firm at a time, the amount of monopsony power in the market depends on both the traditional net recruitment elasticity and on the market-level hours elasticity.[2] Most prior work on monopsony has focused on the substitution elasticity, ignoring the elasticity of workers' hours to the market; most prior work on labor supply has ignored the role of firm substitution.

We begin our empirical analysis by estimating market-level elasticities for male

[2]The optimal wage in a standard dynamic monopsony model is the marginal product of labor, marked down by $1/\left(\eta^R - \eta^S\right)$, where the denominator is the difference between the recruiting and separation elasticities (Manning, 2003).

and female drivers. We do this using data from an experiment conducted while Lyft was out of the Houston market. Because our wage changes are temporary, we interpret these elasticities as Frisch elasticities. We find that Frisch elasticities are significantly larger for women than for men. In response to a ten percent increase in wages, female drivers work eight percent more hours ($\varepsilon = .8$), while male drivers work only four percent more hours ($\varepsilon = .4$). Extensive margin elasticities are modest, even among marginally attached drivers.

We consider several explanations for why female drivers might be more elastic, including differences in age and experience, in usual hours worked, and in outside options. However, even among the set of drivers we characterize as "full-time"—those who we observed driving (pre-treatment) more than 40 hours a week—, female drivers are significantly more elastic.

Our experimental results likely under-state differences in the labor supply elasticities of the average male and female worker. This is because we structured our experiment to minimize differences in usual (pre-experiment) hours worked between men and women in our sample. Elasticities are larger among low hours drivers, likely due to differences in the value of non-work time (Mas and Pallais, 2018). A simple re-weighting exercise suggests the gap in elasticities between active male and female drivers is larger than what we estimate (Blinder, 1973; Oaxaca, 1973).

In the final part of the paper we use data from this experiment and from from a second experiment, conducted after Lyft returned to the Houston market, to examine firm substitution. Two pieces of evidence indicate workers shift hours between firms. First, utilization rates—the fraction of the time a driver is active that he/she is on a trip—are higher among drivers with access to Lyft. This is consistent with high frequency multi-apping, a strategy in which drivers attempt to increase their hourly earnings by decreasing the amount of time they are idle. Second, drivers with access to Lyft are more elastic. This likely reflect the fact that, when the

Uber wage goes up, these drivers both increase their total number of hours worked and increase the share of these hours worked on Uber. Our results highlight the importance of considering market structure and workers' outside options when interpreting data—even experimental data—from a single firm or platform.

We use our estimates to calculate firm-optimal mark-downs and wage gaps (Robinson, 1933; Card et al., 2016). Our estimates of the elasticity of hours to the firm suggest optimal mark-downs on the order of 40%, in line with other recent estimates (Sokolova and Sorensen, 2021). However, unlike prior non-experimental work, we do not see any evidence that women are significantly less elastic than men; if anything, the women in our sample are more elastic (Hirsch, Schank, and Schnabel, 2010; Ransom and Oaxaca, 2010; Webber, 2016). This implies gig economy firms do not have any incentive to pay women less than men.

We also use the extensive margin estimates to back out firm substitution elasticities that can directly be compared to previous estimates from fixed-hours environments. These elasticities measure the extent to which drivers quit the outside firm in response to a wage increase. Our mean estimate of -2.3 is in line with other recent estimates of firm-specific elasticities outside of ride-share (Dube et al., 2020). The fact that these estimates are larger for women than they are for men—and that market hours elasticities are higher for women— suggests that flexible hours firms in environments without multiple job holding also would not have an incentive to pay women less than men.

The results in this paper contribute to a large literature on individual labor supply (see, e.g., Eissa and Liebman, 1996; Meyer and Rosenbaum, 2001). While several papers have examined gender differences in labor supply, there is little quasi-experimental evidence, and no experimental evidence, that the *Frisch* elasticities used in most business cycle models differ by gender (Killingsworth and Heckman, 1986; King and Rebelo, 1999; McClelland and Mok, 2012). This likely reflects the fact that it is difficult to find the type of wage variation necessary to identify Frisch elasticities: variation that is both temporary and exogenous.[3] While a few studies have exploited temporary wage variation in flexible hours settings, the populations in these studies were predominantly male (Oettinger, 1999; Farber, 2005; Fehr and Goette, 2007; Farber, 2015; Stafford, 2015). Our finding that women are significantly more elastic than men suggests prior microeconomic estimates of the Frisch elasticity—which come from predominantly male environments—may have been biased downward (Rogerson and Wallenius, 2009; Chetty et al., 2013).

The results also contribute to a small but growing literature on imperfect competition and the gender wage gap (see, e.g., Card, Cardoso, and Kline, 2016; Morchio and Moser, 2021).[4] Several papers have estimated firm-specific labor supply elasticities by examining how quits respond to changes in wages at other firms in the same industry and geographic market (see, e.g., Blau and Kahn, 1981; Barth and Dale-Olsen, 2009; Hirsch, Schank, and Schnabel, 2010). One challenge with this approach is that, without exogenous variation in wages, it is difficult to ensure that there are no coincident changes in firm-provided amenities, such as changes in hours flexibility or corporate culture. To the best of our knowledge, we are the first to use experimental wage variation to separately identify firm- and market- elasticities in the same setting. We are also the first to use experimental variation to compute firm substitution elasticities for both men and women.

Our work is also related to growing literature on ride-share (see, e.g., Athey, Castillo, and Chandar, 2019; Hall, Horton, and Knoepfle, 2021), and on the gender wage gap within ride-share (Cook et al., 2021). Our experiments are modeled after

[3]Most tax changes do not satisfy the second requirement. The tax holiday studied in Martinez, Saez, and Siegenthaler (2021) is a notable exception.

[4]The idea that imperfect competition could lead to a gender wage gap dates back to Joan Robinson's 1933 book, which introduced the term monopsony. She credits classics scholar BL Hallward with coining the term.

those used by Angrist, Caldwell, and Hall (2021) to analyze the relative value of the ride-share and taxi compensation models. Our paper complements contemporaneous work by Chen et al. (2020) and Chen et al. (2019), who use higher-frequency wage-variation—hourly, rather than weekly—to examine reservation wages and the value of job flexibility. Our work is distinct both in its focus on gender differences in labor supply and in its focus on firm substitution.

The rest of the paper proceeds as follows: the next section lays out a conceptual framework that illustrates the difference between firm-level and market-level labor supply elasticities and how differences in labor supply may contribute to wage gaps. Section 3 describes the empirical setting and experiments. Section 4 presents market-level labor supply elasticities. Section 5 presents estimates of platform substitution. Section 6 concludes.

2 Conceptual Framework

We start by illustrating how access to an outside firm may influence measurement of labor supply elasticities. We then show how differences in labor supply can contribute to a gender wage gap in a flexible hours environment.

2.1 Setup

Consider a simple inter-temporal labor supply model where, in period t, individual i consumes $c_{i,t}$ and works $H_{i,t} \geq 0$ hours. At time t, she chooses consumption and hours to maximize her present discounted value of utility,

$$u(c_t, H_t) + \sum_{k=1}^{\infty} \beta^k u(c_{t+k}, H_{t+k}).$$

Dropping i subscripts for simplicity, we can write the individual's inter-temporal budget constraint as $A_{t+1} = (1 + r_t)(A_t + y_t + w_t H_t - c_t)$. A_t represents assets in period t and r_t is the real interest rate from period t to $t + 1$. There is a constant (within-period) wage w_t and the individual earns an exogenous stream of non-labor income y_t. Utility is increasing in consumption and decreasing in hours worked.

We can write the individual's problem recursively as

$$V_t(A_t) = \max_{c_t, h_t} u(c_t, h_t) + \beta E_t \left[V_{t+1} \left(\underbrace{(1 + r_t)(A_t + y_t + w_t H_t - c_t)}_{A_{t+1}} \right) \right],$$

subject to the constraint that $H_t \geq 0$ (with corresponding Lagrange multiplier μ^{h_t}). The first order conditions can be simplified using the envelope theorem condition, $V_t'(A_t) = \beta(1 + r_t)V_{t+1}'(A_{t+1})$. This implies $\lambda_t := V_t'(A_t) = V_{t+1}'(A_{t+1})\beta(1 + r_t) =: \lambda_{t+1}$. The first order conditions are $\mu^{H_t} \geq 0, \mu^{h_t} H_t = 0$ and:

$$\begin{aligned} u_{c_t} &= \lambda_t, \\ -u_{H_t} &= w_t \times \lambda_t + \mu^{H_t} \end{aligned}$$

where u_{c_t} and u_{H_t} are the partial derivatives of the utility function with respect to c_t and H_t. These expressions define Frisch consumption and hours functions $c(w_t, \lambda_t, A_t)$ and $H(w_t, \lambda_t, A_t)$.[5]

2.2 Market-Level Labor Supply

Individuals' hours choices equate the marginal value of work (the marginal rate of substitution) with the wage: $-u_{H_t}/u_{c_t} = w_t + \mu^{H_t}/u_{ct}$. They do not work if the marginal rate of substitution, evaluated at 0 hours, is less than the offered wage w;

[5] We can accommodate preference heterogeneity by allowing $u(\cdot)$ to be a function of worker characteristics X_i.

when this occurs the restriction $H > 0$ is not binding and $\mu^{H_t} \neq 0$. Using the fact that $u_{c_t} = \lambda_t$, this implicitly defines the reservation wage as a function of the individual's marginal value of wealth: $w^r := -u_{H_t}(0)/\lambda_t$ and

$$h_t > 0 \iff w_t > w^r = -u_{H_t}(0)/\lambda_t.$$

If preferences are homogeneous the aggregate participation elasticity depends on the distribution of the marginal value of wealth in the population. Individuals may have lower values of λ_t—and therefore higher reservation wages—if they have greater assets or non-labor income. We define the extensive margin elasticity as the percent change in participation associated with a given change in log wages:

$$\eta := \frac{\int_{-u_{H_t}(0)/w^{new}}^{-u_{H_t}(0)/w^{old}} dF\lambda_t}{d\log w} \times \frac{1}{\int_{-u_{H_t}(0)/w^{old}}^{\infty} dF\lambda_t}.$$

The Frisch elasticity measures the change in hours worked in response to a change in wages $\partial \log H(w_t, \lambda_t, A_t)/\partial w_t$. Standard functional forms that feature separability in hours and consumption can motivate the standard log-log specifications often used to recover this elasticity. For instance, under the standard assumption that utility is quasi-linear and iso-elastic—$u(c_t, H_t) = c_t - \phi\frac{1}{1+1/\varepsilon}H^{1+1/\varepsilon}$—, there is a simple relationship between log hours and log wages:

$$\log H = \varepsilon \log w + \varepsilon \log \lambda_t - \varepsilon \log \phi.$$

Exogenous variation in wages that does not change λ_t—such temporary changes in the wage—can be used to identify ε.[6]

[6] If there are extensive margin effects and those induced to work complete fewer hours on average, estimates will be downward-biased.

2.3 Labor Supply with Multiple Firms

Many labor supply researchers use data from flexible hours environments because, outside of these environments, workers' choices are often constrained by firms (Dickens and Lundberg, 1993). However, in flexible hours markets individuals often hold multiple jobs. Suppose that the individual can work for multiple firms. One firm pays her $w_t h_t$ for h_t hours of work; the other pays her $S(r_t)$ for r_t hours of work where $S(\cdot)$ is increasing and concave.[7] This leads to a modified inter-temporal budget constraint: $A_{t+1} = (1 + r_t)(A_t + y_t + w_t h_t + S(r_t) - c_t)$. This modified setup also provides additional first-order conditions: $\mu^{r_t} \geq 0$, $\mu^{r_t} r_t = 0$, and:

$$-u_{r_t} = S'(r_t) \times \lambda_t + \mu^{r_t}.$$

Given the functional form above the is still a simple relationship between log wages and log total hours:

$$\begin{aligned}
\log(H_t) = \log(h_t + r_t) &= \varepsilon \log\left(w_t \lambda_t + \mu^{h_t}\right), \\
\log(H_t) = \log(h_t + r_t) &= \varepsilon \log\left(S'(r_t)\lambda_t + \mu^{r_t}\right).
\end{aligned}$$

Regressions of log total hours worked $(h_t + r_t)$ on $\log w$ identify ϵ if (1) individuals' counterfactual hours choices on the first job are positive (i.e. $\mu^{h_t} = 0$ and there are no first-job participation effects) and (2) the wage change does not influence λ_t.[8]

However, in many flexible hours environments, researchers do not observe total hours worked $(h_t + r_t)$ because they have data from a single firm. Regressions of $\log h_t$ on $\log w_t$, yield hours upward-biased estimates of the market-level elasticity

[7]Concavity could arise if the individual has access to multiple sources of outside employment, which differ in their productivity.

[8]For simplicity, we have not added a cost of multiple job-holding. However this can be accommodated by assuming workers pay a utility cost κ if they have positive hours in both jobs.

ϵ. This occurs because the elasticity of total hours worked (ε) is a weighted average of the elasticity of hours at primary (τ) and secondary jobs (s), weighted by the (pre-treatment) fraction of time spent at each job ($\phi = h/H$):[9]

$$
\begin{aligned}
\frac{\partial H}{\partial w}\frac{w}{H} &= \frac{\partial h}{\partial w}\frac{w}{H} + \frac{\partial r}{\partial w}\frac{w}{H} \\
&= \frac{\partial h}{\partial w}\frac{w}{(H\phi)(1/\phi)} + \frac{\partial r}{\partial w}\frac{w}{(H(1-\phi))/(1-\phi)} \\
\varepsilon &= \tau\phi + (1-\phi)\underbrace{s}_{<0}. \qquad (1)
\end{aligned}
$$

The estimated parameter (τ) is closer to ϵ whenever ϕ is close to 1 (i.e. when the researcher uses data from a near-dominant firm).[10] Appendix B shows there is an analogous relationship between the overall participation elasticity and the elasticities of employment at each firm.

2.4 Firm-Level Labor Supply

For some applications the relevant elasticity is one which explicitly includes firm substitution. In particular, in monopsony models firms face upward sloping labor supply curves; a firm's optimal wage depends on the elasticity of labor supply that it faces (Robinson, 1933; Manning, 2003). The optimal wage equates the marginal benefit from raising the wage (i.e. the additional hours supplied to the firm) with the marginal cost. Suppose a firm can hire $h(w)$ hours of labor at wage w. These $h(w)$ hours produce $Y(h(w))$ in output. The firm picks wages w to maximize profits:

$$\max_w Y(h(w)) - wh(w).$$

[9]The combined first order condition $S'(r_t) - w_t = 0$ says that—conditional on positive hours in both jobs—the driver works until the marginal returns at each job are equal. Conditional on working both jobs, hours in the second job are decreasing in w_t. To see this, apply the implicit function theorem to $S'(r_t) - w_t = 0$ and recall that $S'' < 0$.

[10]When outside hours are fixed, the relationship is simply $\epsilon = \tau\phi$.

Table 1: Market and Firm-Level Elasticities

	Market-Level Elasticity	Firm-Level Elasticity: With Moonlighting	Firm-Level Elasticity: Without Moonlighting
Elasticity	$\epsilon := \dfrac{\partial \log H}{\partial \log w}$	$\tau := \dfrac{\partial \log h}{\partial \log w}$	
Worker's Trade-Off	Labor-Leisure	Labor-Leisure and Inside vs. Outside Firm (Allocations of Hours)	Labor-Leisure and Inside vs. Outside Firm (Allocation of Employment)
Relationship with ϵ	—	$\tau = \frac{1}{\phi}\epsilon - \frac{1-\phi}{\phi}s$ where $\phi := h/H$ $s := \partial \log r / \partial \log w$	$\tau = \epsilon + \eta^{firm}$ where $N(w_t) :=$ firm employment$\eta^{firm} :=$ $N'(w_t)w_t/N(w_t)$
Measurement	Pre-Lyft Experiment	Post-Lyft Experiment	Pre-Lyft Experiment + Pre/Post Comparison

The profit maximizing wage is the marginal product of labor, marked down by the inverse hours elasticity,

$$w^* = \frac{Y'(h(w))}{1 + 1/\tau}.$$

The source of hours does not matter; additional hours worked may come at the expense of workers' leisure or employment at other firms.

Firms have incentives to offer different wages to groups of workers with different elasticities, even if those groups are equally productive. For instance, if the firm can offer different wages to (equally productive) male and female workers, the optimal wage gap is

$$\frac{w_m^*}{w_w^*} = \frac{1 + 1/\tau_w}{1 + 1/\tau_m}. \tag{2}$$

where τ_w and τ_m denote the elasticity of hours to the firm within each group.

When workers can hold multiple jobs, the elasticity of hours to the firm may reflect both changes in total hours worked and changes in the allocation of hours across firms. In some contexts, however, workers may be prohibited from working for multiple firms at the same time. This may be due to non-compete agreements or employer constraints on "moonlighting". In this case, the firm-specific elasticity will capture changes in market hours worked and changes in the allocation of *participation* across firms. Appendix B shows that, under standard assumptions, this is the sum of the net recruitment elasticity (which measures how employment at the firm changes) and on the market elasticity ϵ. This generalizes the result in Manning (2003): when hours are not flexible, τ is simply the net-recruitment elasticity. Appendix Table 1 summarizes the results of this section.

3 Empirical Setting and Data

3.1 Background on Ride-Share

Uber is a global Transportation Network Company (TNC) whose software connects drivers and riders. Uber launched its peer-to-peer operations in mid-2012 and as of 2019 had over 900,000 active drivers in the United States. In most cities in the United States there are few barriers to becoming a ride-share driver.[11]

Uber drivers can work whenever and wherever they chose (within Uber's service region). Throughout the course of our experiments, these drivers were paid per mile and per minute for each trip they completed. These rates increased at certain times of day and in certain locations due to Surge pricing. For many years Uber drivers paid a fixed fraction of their trip receipts to Uber in the form of the "Uber fee".[12] This fee varied across drivers based on their city and when they joined the platform.

[11] The most prominent counterexample is New York City, where ride-share drivers must obtain licenses from the Taxi & Limousine Commission. We do not use any data from New York City.

[12] In 2017 Uber loosened the link between driver and rider pay. Now riders and drivers face distinct per-minute and per-mile rates.

Appendix Figure A1 shows an example trip statement.

Many ride-share drivers drive for multiple platforms. This is known as "multi-apping". In 2021, The Rideshare Guy, a popular blog aimed at TNC drivers, estimated that two thirds of drivers drove for both Uber and Lyft (Campbell, 2022). At the time of our experiments, these two companies covered the vast majority of the on-demand gig market (Campbell, 2017). Drivers that have signed up for multiple platforms may shift between platforms on a shift-to-shift basis, driving for whichever app offers them the highest earnings when they start a shift. They may also engage in high frequency multi-apping: keeping both apps on during down time and accepting the first dispatch to come in. This is profitable because drivers only make money when a passenger is in the car; reducing the time between trips can lead to higher earnings.

While it is fairly straightforward to multi-app at low frequency, conversations in online forums, such as the one depicted in Appendix Figure A2, suggest that high frequency multi-apping requires a non-trivial amount of effort. Drivers must toggle between the two applications, turning off one as soon as they accept a trip on the other. Drivers exchange information on how to multi-app through word-of-mouth and through online forums; they also obtain information on how to multi-app through guides on ride-share blogs (Appendix Figures A3 provides one example). Several companies (e.g. Mystro, Upshift, and QuickSwitch) have developed third party apps to help drivers navigate between the two interfaces and multi-app at high frequency. The fact that these apps—recommended by popular ride-share blogs—are able to charge \$5/day, \$12/week, or \$100/year suggests both that multi-apping is profitable and that drivers find it non-trivial.[13]

[13] An advertisement from one of these companies (Appendix Figure A4) claims that they can help drivers increase their earnings by thirty-three percent. In our data, a driver would increase her earnings by this amount if she cut her waiting time in half.

In some cases, ride-share drivers are not eligible to drive for both Uber and Lyft. This can happen if only one platform operates in the driver's market. For instance, between November 2016 and May 2017, Lyft did not operate in Houston. However, even in cities where both platforms operate, some drivers are ineligible to work for one platform based on the age of their cars. For instance, for many years, Lyft required Boston drivers to use cars with a vehicle model year 2004 or newer, while Uber allowed vehicles as old as 2001 (Angrist, Caldwell, and Hall, 2021).[14]

3.2 Earnings Accelerator Experiments

We generated variation in wages and in access to Lyft through a set of randomized experiments. Ride-share companies routinely run promotions in which they change driver pay, without affecting the prices for riders. Our experiments were modeled after these promotions and were designed in collaboration with local city teams so that they looked like typical Uber promotions. They were advertised to drivers as the "Earnings Accelerator".

We conducted two "Earnings Accelerator" experiments in Houston. One was conducted in spring 2017, when Lyft was temporarily out of the market; the other was conducted in fall 2017, after Lyft had returned to the market. While Lyft was out of the market, Uber was essentially the only ride-share company operating in Houston.[15] We use the first experiment—sometimes referred to as the "main experiment"—to estimate market-level elasticities. We use the second experiment to estimate "firm-level" elasticities. The contrast between the first and second experiments identifies firm substitution. Because the second experiment was cut short, in some specifications we also use data from a third "Earnings Accelerator" exper-

[14]Both Uber and Lyft had additional requirements to drive for their "premium" services.

[15]While ride-share drivers today have many gig alternatives at Amazon, at food delivery companies, and at smaller ride-share companies, this was not the case in Houston when we conducted the experiment in mid-2017, as ride-share surveys at the time confirm (Campbell, 2017).

iment conducted in Boston (Angrist, Caldwell, and Hall, 2021).[16] Appendix Table A1 compares the three Earnings Accelerator experiments. Appendix Figure A5 depicts the design of the first Houston experiment.

The two Houston Earnings Accelerator experiment unfolded similarly. We first identified a set of drivers that satisfied two criteria ("eligible drivers"): (1) they were active on the Uber platform (had completed at least four trips in the past month), and (2) they averaged 5 and 40 hours per week in the four weeks prior to the experiment. So that we would have a mix of full-time and part-time drivers, we grouped drivers into bins based on their usual hours per week, and randomly selected subsets of drivers from each bin. The low group consisted of drivers that averaged between 5 and 15 hours per week, the high group consisted of drivers that (conditional on driving) averaged between 15 and 25 hours per week, and the very high group consisted of drivers that averaged more than 25 hours per week. Drivers in the very-high group worked more than part-time on the platform. Within each bin, we randomly selected drivers for inclusion in the experiment. We over-sampled women in each bin so that we would have roughly equal numbers of male and female drivers.[17] We selected 2020 drivers for inclusion in the main Houston experiment and 2100 drivers for inclusion in the second Houston experiment.

We offered drivers selected for the experiment the opportunity to earn X% more on every trip for a week. Half of the drivers in each hours-commission bin were offered the opportunity in week one; the other half were provided the same offer the following week. We told the drivers that this would result in an "X% higher

[16]We delayed the second Houston experiment due to Hurricane Harvey. Because Uber then made unexpected changes to the app, which removed our ability to change drivers' pay in a way that would be visible to drivers on a trip-by-trip basis, we only have one week of comparable data for this experiment. Angrist, Caldwell, and Hall used the Boston data to analyze the value of the proportional compensation contract embedded in ride-share, relative to the lease-based model used in the taxi industry (2021). Appendix C.2 provides more information.

[17]The Boston Earnings Accelerator experiment was structured similarly. However, it did not have a "very high" hours bins and did not over-sample female drivers.

payout". In the main experiment we implemented these increases as reductions in the Uber fee; as a result, X was either 39% or 33%.[18] While the treatment was active, drivers who accepted the offer saw increased earnings reflected in-app.

The treatment offers indicated that the promotion would apply to all trips that week, including those with Surge pricing. Drivers received these offers via e-mail and text message and through the Uber app itself. Figure 1 shows a sample e-mail and text message from the first Houston experiment. These messages (and the in-app notification) included links to Google Forms like those typically used in Uber promotions. The forms were pre-filled with a driver's unique Uber identifier and included detailed information on the promotion, as well as consent language. We sent the offers one week before the promotion went live; drivers had one week to accept the offer by clicking "yes" on the Google Form. Around 60% of the drivers in each experiment accepted our offer. While the offer should have been attractive to all drivers, Uber drivers receive many messages from Uber each week and some choose to ignore a portion of this messaging.[19]

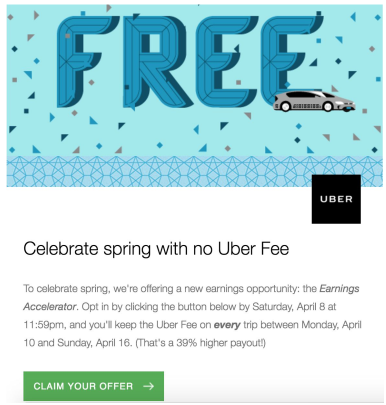
Each experiment was designed to be small—both in terms of drivers and number of trips by treated drivers—relative to the size of the market so that there would not be an impact on equilibrium wages. Given both the messaging we used and Uber's general approach to driver promotions, drivers would not have had reason to anticipate higher future wages as a result of their participation.[20]

[18]The fee was highly salient to drivers as it was included on each trip receipt and weekly pay statement (Angrist, Caldwell, and Hall, 2021). In the second Houston experiment, drivers were offered increases of between 10 and 50%, with no mention of the Uber fee. This is because the structure of pay had changed. More details are provided in Appendix C. In both cases, the increases were quoted in proportional terms to drivers.

[19]Appendix Figure A6 shows that drivers who drove the week the offers were sent out were more likely to accept the offer. This is not surprising because one way offers were delivered to drivers was through the Uber app, which drivers may not open if they do not drive. Among those who drove there was no relationship between hours worked and acceptance.

[20]A related concern in our second experiment (when Lyft was present) is that, when wage increases are small, there may be network effects. Intuitively, small changes in the offered wage could

Figure 1: Earnings Accelerator Messaging



Note: This is an example of the message that was sent to drivers selected for the Earnings Accelerator in the main Houston experiment. The link directed the driver to a more detailed opt-in page with information on how the incentive worked and with consent language.

In some specifications in Section 4 we also use data from a series of "taxi" experiments we conducted among drivers who accepted the initial offer in our main experiment. In the taxi experiments treated drivers were offered additional weeks of higher earnings in exchange for an up-front payment, much like the lease a taxi driver would pay to a medallion holder. While these offers were only attractive to drivers that intended to drive enough to pay off the lease, these treatments generated additional wage variation at much lower cost. The specifications that include these data are labeled. We refer to the initial treatment weeks—in which the treatment was offered at no up-front cost—as the "fee-free" driving weeks.

Appendix C provides detailed information on each Earnings Accelerator experiment, including sample counts, treatment offers, and messaging. Column 4 of Table A2 shows that the first and second week offers were balanced in the main experiment; Columns 2 and 4 of Table A3 show analogous results for the second Houston experiment and for the Boston experiment.

lead multi-apping drivers to drive more for *both* Uber and Lyft if it makes them more likely to start driving. By offering drivers reasonably large wage increases we are able to avoid this concern.

Table 2: Descriptive Statistics

	Active		Eligible		RCT	
	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Age	43.98	44.27	43.98	44.27	43.98	44.27
(standard deviation)	(12.53)	(11.90)	(12.53)	(11.90)	(12.53)	(11.90)
Driving Information						
Enrolled in Vehicle Solutions	0.07	0.10	0.09	0.13	0.09	0.15
Months Since Uber Signup	12.67	9.81	10.79	9.21	20.09	12.22
(standard deviation)	(10.03)	(8.58)	(8.83)	(8.03)	(9.47)	(9.29)
Vehicle Year	2012.97	2013.39	2012.90	2013.49	2013.42	2013.65
(standard deviation)	(3.12)	(2.60)	(3.24)	(2.60)	(2.52)	(2.47)
Usual Hours Worked						
Mean	18.92	10.32	18.46	13.44	20.67	18.62
Between 0 and 5	0.30	0.44	0.16	0.21	0.11	0.11
Between 5 and 15	0.28	0.33	0.37	0.46	0.31	0.35
Between 15 and 25	0.15	0.13	0.20	0.18	0.28	0.26
Between 25 and 35	0.10	0.06	0.12	0.09	0.12	0.17
Between 35 and 45	0.07	0.03	0.08	0.04	0.10	0.07
Over 45	0.10	0.02	0.07	0.02	0.08	0.04
Observations	12667	3149	8529	2112	972	1048

Note: This table describes active Houston drivers (columns 1 and 2), the subset of these drivers that were eligible for inclusion in our main experiment (columns 3 and 4), and the drivers included in this experiment (columns 5 and 6). Section 3.2 defines the samples.

3.3 Male and Female Drivers

Most ride-share drivers—like most taxi drivers—are male. Columns 1 and 2 of show that twenty percent of Table 2 active Uber drivers in Houston were female at the time of our experiment. Cook et al. (2021) document similar gender gaps in other large cities at other time periods.

Male and female Uber drivers are similar in age, but differ in experience. Among active Houston drivers, the average female driver has less than ten months of experience, relative to nearly thirteen months for men. Relative to male drivers, female drivers have somewhat newer cars.

The most significant difference between male and female drivers is that female drivers work many fewer hours. The mean among active female drivers is ten hours per week, relative to nearly nineteen hours per week among active male drivers. And while ten (seventeen) percent of male drivers work more than forty-five (thirty-five) hours per week, only two (five) percent of female drivers do. Appendix D.3 shows similar patterns among Houston taxi drivers.[21]

Columns 3-6 show that our inclusion criteria and stratified sampling narrow gender differences in hours worked but widen gaps in experience. In the analysis below we examine whether our estimates of labor supply elasticities would generalize to the full pool of active drivers, or whether the average estimates are biased because high-hours and high-experience drivers are overrepresented in our sample.

4 Market-Level Labor Supply

We use data from the first Houston Earnings Accelerator experiment, conducted when Lyft was out of the market, to estimate market-level labor supply elasticities.

4.1 First Stage Impact on Hourly Earnings

Our experimental offers increased hourly earnings for treated compliers by reducing the proportional commission they faced. This is analogous to a reduction in drivers' implicit tax rate. Columns 1 and 2 of Table 3 show that the average treatment offer corresponded to a .28–.31 log point (~35%) increase in the log net-of-commission rate ($\log(1-\tau_{it})$). The treatment was somewhat more generous for female drivers, who faced higher commissions on average because they had joined the platform more recently. Around sixty percent of drivers accepted the offer (Columns 3-4).[22]

[21] In this Appendix we also investigate whether there are differences in outside employment and find that, if anything, female drivers work a greater fraction of their total hours in their main job.

[22] There is no difference in opt-in rates (within gender) between drivers who faced different commissions (the p-values from within-gender tests are .93 for male drivers and .22 for female drivers).

Table 3: Earnings Accelerator Offers and Opt-In Decisions

	Offered Log(1-T)		1{Accepted Offer}		Log(1-T)		Log(Wage)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male	0.28***	0.28***	0.61***	0.57***	0.17***	0.16***	0.12***	0.14***
	(0.002)	(0.002)	(0.016)	(0.012)	(0.005)	(0.004)	(0.012)	(0.011)
Female	0.31***	0.31***	0.73***	0.60***	0.22***	0.19***	0.19***	0.20***
	(0.001)	(0.001)	(0.014)	(0.011)	(0.004)	(0.003)	(0.012)	(0.012)
Observations	4040	6746	4040	6746	4040	6746	2998	5031
Strata Dummies	✓	✓	✓	✓	✓	✓	✓	✓
Fee-Free Driving Data	✓	✓	✓	✓	✓	✓	✓	✓
Taxi Data		✓		✓		✓		✓

Note: This table presents estimates of the impact of the treatment on opt-in decisions (columns 1–2), the offered log net-of-commission rate (columns 3–4), the realized log net-of-commission rate (columns 5–6) and realized log hourly earnings (columns 7–8). We use a stacked model; each regression controls for the strata used for random assignment and for a female dummy. Odd columns use data from the two weeks of "fee-free" driving; even columns add data from the two weeks of taxi experiments. Standard errors, clustered by driver, are presented in parentheses. Levels of significance: *10%, ** 5%, and *** 1%.

Columns 5–8 show that there is a strong impact on $\log(1-\tau_{it})$ and $\log w_{it}$ for both male and female drivers. Because result, the impact on treated drivers' net-of-commission rates was roughly .2 log points (~22%). The impact on drivers' realized log hourly earnings (Columns 7–8) is smaller (though the difference is insignificant), suggesting marginal hours may have been somewhat less productive. Appendix D provides more detail on the first stage.

4.2 Baseline Estimates

We estimate labor supply elasticities by regressing indicators for positive hours worked (extensive margin) or the log of hours worked on the log net-of-commission rate $(\log(1-\tau_{it}))$ or on realized log hourly earnings $(\log w_{it})$. Because the realizations of $\log(1-\tau_{it})$ and $\log w_{it}$ are endogenous, we use the randomly assigned treatment offers Z_{it} as instruments.[23] The baseline over-identified model includes

[23] Appendix Table A5 shows that OLS estimates of labor supply elasticities are biased downwards. Appendix Table A4 shows that the treatment offers did not influence drivers' labor supply before or

separate offers for each week and treatment offer. In this model the X_{it} include the strata used for random assignment, as well as a number of baseline covariates.[24] We use a stacked model where we allow the coefficient on each covariate to vary by gender. Formally, we estimate the following model by two-stage least squares:

$$\begin{aligned} y_{it} &= \epsilon \log(1-\tau_{it}) + \beta X_{it} + \eta_{it}, \\ \log(1-\tau_{it}) &= \gamma Z_{it} + \lambda X_{it} + \upsilon_{it}. \end{aligned} \tag{3}$$

Column 1 of Table 4 shows that, in the full sample, the extensive margin elasticity for female drivers is double that of male drivers. In response to a ten percent increase in the offered wage, female drivers are two to three percent more likely to drive, relative to an (insignificant) one percent increase for men. Note that when y_{it} is an indicator for positive hours, the coefficient on $\log(1-\tau_{it})$ reveals the semi-elasticity; we convert this to an elasticity by dividing by the control proportion of drivers with positive hours.

We estimate intensive margin elasticities using the model in 3, with log hours replacing participation as the dependent variable. For comparison with the prior literature, we use both log hourly earnings (Columns 2 and 3) and $\log(1-\tau_{it})$ (Columns 4 and 5) as endogenous variables. The central estimates in Columns 3 and 5 show that, in response to a ten percent increase in hourly earnings, male drivers drive four to five percent more, while female drivers drive eight to nine percent more. The difference in elasticities is statistically significant.

However, because our treatment offers influence female drivers' decision to work (and therefore who is included in the regressions in Columns 3-6) our inten-

after treatment, consistent with there being no wealth effects. See Appendix D for more details.

[24] The baseline covariates include the number of months a driver has been on the Uber platform, an indicator for whether a driver uses Uber's "vehicle solutions" leasing program, the log of previous hours worked, a constant, and the interaction of these characteristics with the female dummy.

sive margin estimates do not reveal the Frisch elasticity for female drivers. Assuming drivers induced to drive work fewer hours, on average, than those who would have driven in the absence of treatment, the intensive margin estimates are a lower bound on the structural elasticity for these drivers.

In order to produce estimates that account for selection into participation among female drivers, we replace the dependent variable with the inverse hyperbolic sine of hours worked (Columns 6 and 7).[25] We also estimate Poisson regression models, which account for the bunching of hours at zero (Columns 8 and 9). All specifications suggest women are significantly more elastic than their male counterparts, with estimates roughly double those for men.

Because the Earnings Accelerator wage increases are temporary, we interpret the estimates as Frisch elasticities. Our estimated elasticity of 0.4 for male drivers is similar to previous estimates for men in flexible hours environments (Farber, 2005; Farber, 2015). The difference between our estimates for male and female drivers is both statistically significant and economically meaningful. However, both estimates are significantly below the elasticities typically used to calibrate macro-economic models.

The appendix documents robustness to using a just-identified model (Table A6), to controlling only for the strata used for random assignment (Table A6), and to using fares collected as the dependent variable (Table A7). It also shows the treatment shifts the entire distribution of hours worked (Appendix D.4).

4.3 Heterogeneity

We next explore heterogeneity across drivers. In doing so we also investigate several explanations for why female drivers might be more elastic. Table 5 and

[25] This can be interpreted like the log of hours worked as $\operatorname{asinh}(x) \to \ln(2x) = \ln(2) + \ln(x)$.

Table 4: Extensive and Intensive Margin Elasticities

	1{Drive}	Log(Hours)				Asinh(Hours)		Poisson Model	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Male	0.10	0.31	0.27*	0.51***	0.39***	0.50**	0.61***	0.29**	0.35***
	(0.09)	(0.22)	(0.15)	(0.17)	(0.14)	(0.22)	(0.20)	(0.14)	(0.12)
Female	0.26***	0.70***	0.51***	0.91***	0.80***	1.02***	1.13***	0.61***	0.70***
	(0.07)	(0.16)	(0.11)	(0.13)	(0.11)	(0.16)	(0.16)	(0.11)	(0.10)
p-value	0.194	0.145	0.194	0.055	0.018	0.059	0.037	0.076	0.018
First Stage F: Male	212	22	208	11	73	212	74	---	---
First Stage F: Female	452	34	477	14	155	452	141	---	---
Observations	3108	2766	2766	4650	4650	3108	5154	3108	5154
Endogenous Variable	Log(1-T)	Log(W)	Log(W)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)
Fee-Free Driving Data	✓	✓	✓	✓	✓	✓	✓	✓	✓
Taxi Data			✓		✓		✓		✓
Strata Dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: This table presents estimates of model 3. The dependent variable in columns 1-2 is an indicator for non-zero hours; the dependent variable in columns 2-5 is the log of hours worked; and the dependent variable in columns 6-7 is the inverse hyperbolic sine of hours worked. The endogenous variable is either log(earnings/hours) (Columns 2-3) or $\log(1-\tau_{it})$ (remaining columns). We scale the coefficients in Column 1 by baseline participation rates (0.737 for male drivers, 0.715 for female drivers) so that we recover elasticities, rather than semi-elasticities. The baseline covariates include the number of months a driver has been on the platform, whether a driver uses the "vehicle solutions" program and one lag of log earnings. We include each characteristic as well as its interaction with a female dummy. Within each treatment type there is a binary instrument for each combination of: commission group (2) and treatment week (2). Standard errors are clustered by driver. Levels of significance: *10%, ** 5%, and *** 1%. Appendix Table A6 presents additional specifications.

presents the main results, which come from a stacked model where—in order to boost power—we include data from the two additional "taxi" experiments. Each column presents results for the subgroup indicated in the column header.

Age, Experience, and Wealth We first examine heterogeneity across drivers with different demographic characteristics. We find that there is relatively little heterogeneity in elasticities by age or experience on Uber (Columns 1-4 of Table 5). As a result, heterogeneity in these characteristics does not explain the gender gaps; within each sample female drivers appear more elastic than their male counterparts.

We have limited information on drivers' outside income or wealth. However, the fact that female drivers have newer cars on average (Columns 5 and 6 of Table 2)—even among drivers in the highest hours stratum—weighs against the idea that differences in wealth or household income explain the gender differences in elasticities. If anything, we would expect drivers with greater wealth or household income to be less responsive to changes in the Uber wage.[26]

Driving Behavior Second, we explore heterogeneity by when drivers typically drive. As previous papers have documented, male and female drivers vary in when they work: female drivers are significantly less likely to work late at night (Cook et al., 2021). Appendix D shows this is also true among drivers in our sample.[27]

However, differences in when male and female drivers typically work do not drive differences in elasticities. We use pre-experimental data to construct samples of drivers who typically worked at certain times of day. Column 5 focuses on drivers who spend most (>60%) of their time driving late at night (between 10 PM

[26]In results not reported we have examined heterogeneity in wealth, as proxied by income in a driver's zip code. We found little heterogeneity.

[27]This may reflect differences in hourly reservation wages (Chen et al., 2020). Our treatment provides equal proportional incentives to work at all times of day. Appendix D describes usual driving patterns by sex and investigates when drivers chose to drive additional hours.

and 4 AM); Column 6 focuses on drivers who work primarily on the weekend; and Column 7 focuses on drivers who routinely work during school pick-up hours (between 2 and 5 PM) and who therefore seem less likely to have primary childcare responsibilities. Across all samples, female drivers are more elastic.

Hours Worked On and Off Uber Finally, we explore heterogeneity in elasticities across drivers with different usual hours worked. The remaining columns of of Table 5 show that elasticities are generally smaller for high hours drivers, and that this is true for both male and female drivers. These columns split drivers into the three hours strata used for random assignment. These strata consist of drivers who, before sample selection, averaged 5–15 hours/week (low), 15–25 hours/week (high), or more than 25 hours/week (very high). Panel A shows that the extensive margin effects are driven by drivers in the lower two strata; there is no impact on whether drivers in the "very high" hours stratum start driving. Within each stratum, female drivers are more elastic than their male counterparts. The fact that drivers in the lower-hours strata are more elastic is consistent with recent evidence in Mas and Pallais (2018) on the value of non-work time.

Gender differences in outside options are unlikely to drive the heterogeneity in elasticities that we observe. The final column of Table 5 focuses on the set of drivers who are unlikely to have outside employment: those who were observed driving at least 40 hours a week in one of the four weeks used for sample selection. This column shows that these drivers are similarly elastic to those in the overall "very high hours stratum". In order to explain gaps, female Uber drivers would need to be significantly *more* likely to have outside market employment.[28]

[28] If men are more likely than women to combine gig work with other market-based work, this would mean that we over-state elasticities for male drivers and therefore under-state the gap in elasticities. The descriptive statistics on Houston drivers presented in Appendix D.3 suggest that, if anything, female drivers have less outside work.

Table 5: Heterogeneity in Market-Level Elasticities

	By Months on Platform		By Age		Time of Day			By Hours Strata			"Full Time" Drivers
	Top 1/2	Bottom 1/2	<=35	>35	Late Night	Weekend	School Pick-Up	Low	High	Very High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
						A. 1{Drive}					
Male	0.09	0.17	0.06	0.11	0.02	-0.15	0.13	0.10	0.25	0.02	0.01
	(0.10)	(0.17)	(0.17)	(0.10)	(0.20)	(0.25)	(0.10)	(0.22)	(0.17)	(0.08)	(0.10)
Female	0.24*	0.27***	0.35*	0.23***	0.53***	0.92***	0.24***	0.49**	0.41***	0.04	-0.01
	(0.13)	(0.08)	(0.20)	(0.07)	(0.18)	(0.30)	(0.08)	(0.23)	(0.12)	(0.07)	(0.07)
Observations	1620	1488	834	2270	484	426	1698	818	1072	1218	828
						B. Asinh(Hours)					
Male	0.44*	0.96***	0.55	0.59**	0.48	0.77	0.75***	0.75**	0.96**	0.13	0.18
	(0.25)	(0.32)	(0.34)	(0.25)	(0.48)	(0.57)	(0.27)	(0.37)	(0.38)	(0.28)	(0.35)
Female	1.30***	1.08***	1.04***	1.14***	2.14***	2.44***	0.85***	1.36***	1.44***	0.66***	0.53**
	(0.28)	(0.19)	(0.40)	(0.17)	(0.42)	(0.51)	(0.20)	(0.35)	(0.27)	(0.21)	(0.26)
Observations	2653	2501	1344	3800	811	707	2812	1341	1765	2048	1409
						C. Poisson					
Male	0.55***	0.49**	0.22	0.38***	0.61*	0.83*	0.36**	0.52*	0.63***	0.16	0.12
	(0.18)	(0.21)	(0.23)	(0.14)	(0.35)	(0.47)	(0.14)	(0.31)	(0.24)	(0.15)	(0.17)
Female	1.01***	0.61***	0.69***	0.70***	1.06***	1.23***	0.51***	1.16***	0.90***	0.46***	0.32**
	(0.21)	(0.11)	(0.22)	(0.11)	(0.25)	(0.34)	(0.13)	(0.25)	(0.17)	(0.13)	(0.15)
Observations	2653	2501	1344	3800	811	707	2812	1341	1765	2048	1409

Note: This table presents estimates from an over-identified model where treatment indicators (interacted with sex) instrument for $\log(1-\tau_{it})$. All models control for the strata used for random assignment, date fixed effects, and for baseline covariates, interacted with a female dummy. We scale the coefficients in Panel A by the control participation rates in each gender-specific subgroup so that we recover elasticities, rather than semi-elasticities. "Late night" drivers complete >40% of their weekly hours between the hours of 10 PM and 4 AM. "Weekend" drivers complete >50% of their weekly hours on the weekend. "School pick-up" drivers drive for at least 20 minutes between 2 PM and 5 PM on >50% of the week-days they drive on. Samples are constructed using data from non-treatment weeks. Standard errors are clustered by driver. Panels B and C include data from the two "taxi" weeks. Levels of significance: *10%, ** 5%, and *** 1%.

4.4 Labor Supply Elasticities of the Average Driver

The previous section documented that male and female drivers with similar characteristics have different elasticities. However, male and female drivers differ in average characteristics. Because we designed our experiment to minimize differences in male and female hours worked, it is natural to ask whether our estimates for male and female drivers generalize to the full population of Uber drivers.

To shed light on this question, we run specifications where we allow the male and female elasticities ε^g to be a function of observable driver characteristics C_i. We then calculate

$$\varepsilon^p := \int \varepsilon^g(C_i) f^p(C_i) dC_i$$

for different populations p. To allow for both intensive and extensive margin adjustments we use a labor supply model which uses the inverse hyperbolic sine as the outcome variable. We estimate ϵ as a function of discrete characteristics C_i using fully interacted models.

We focus on two key characteristics—hours worked and experience—because these characteristics are the main differences between male and female drivers in our sample and between drivers in our sample and the full pool of "active" drivers (Table 2). We use the hours and experience distribution among the RCT sample to group drivers into hours and experience quantiles.[29]

We first re-weight ε^f and ε^m to accounts for the fact that the distribution of hours worked varies by gender—and the distribution of hours worked in our sample does not match the distribution of the overall population. The top panel of Figure 2 shows that re-weighting elasticities to account for heterogeneity in hours worked widens the observed gaps between male and female drivers. This is not surprising

[29]Columns 2 and 4 of Appendix Table A8 show that the compliers are selected on hours worked (work more hours), but not age or experience. This is especially true for female drivers. Appendix D provides more information.

given that female drivers drive fewer hours on average, and elasticities are larger among low hours drivers (Table 5). The second panel shows that, because more experienced drivers are more elastic, re-weighting elasticities by driver experience narrows, but does not close, the gap between male drivers (who are more experienced on average) and female drivers. However, when we weight both the male and female elasticities to match the distribution of experience among (sex-specific) active drivers, we see that there are still differences; the re-weighted female elasticities produce an estimate of 1.16, relative to .94 for the male elasticities (See Column 3 of Table A9 for point estimates and standard errors).[30]

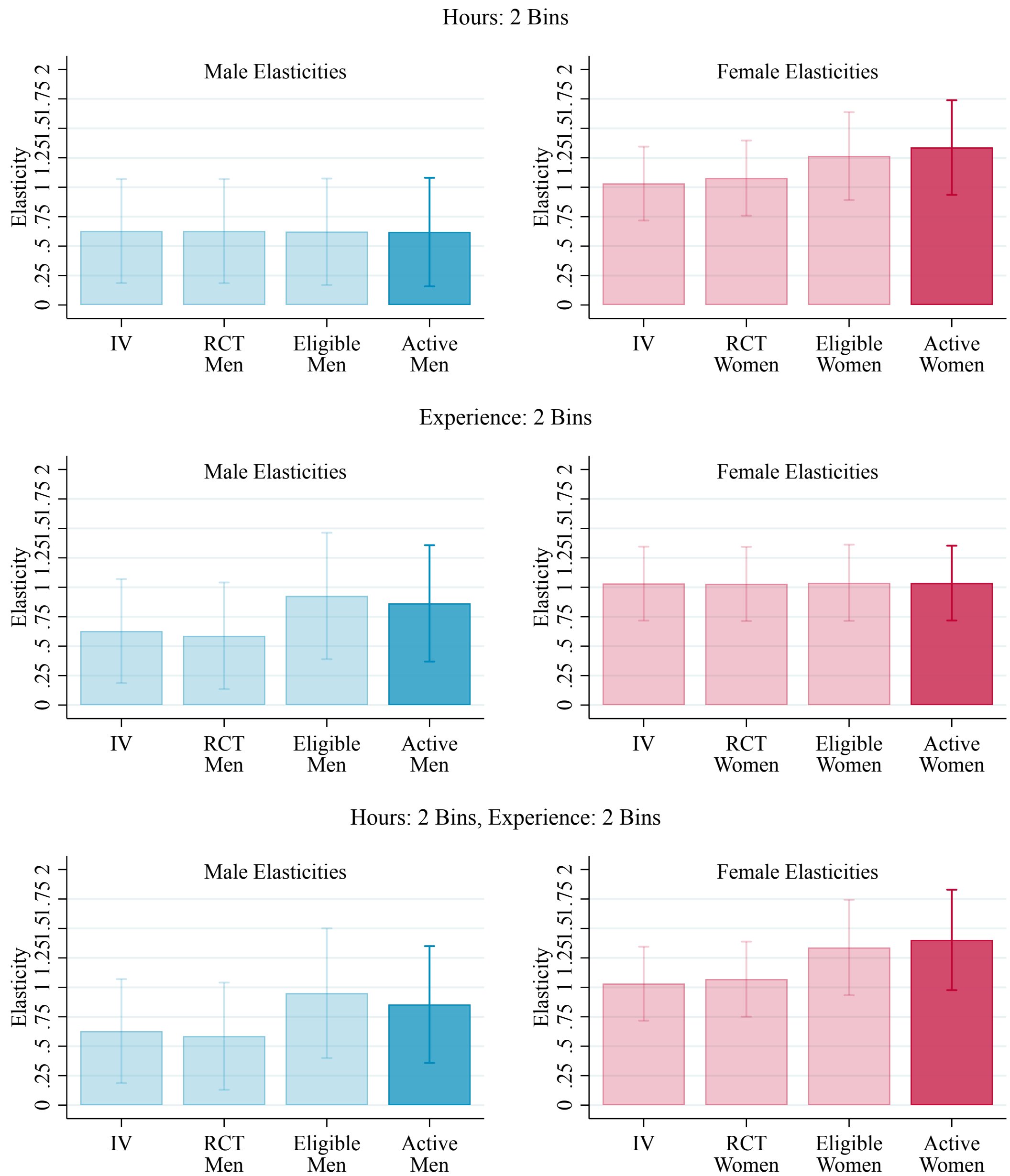
The final panel plots estimates that account for differences in both experience and hours. The final bars in each grouping (shown in darker colors) show that, even after once we allow for heterogeneity in both hours and experience, and re-weight the estimates to match that of the overall active driver population of that sex, the estimates for female drivers are larger than those for male drivers. Table A9 presents additional specifications.

5 Firm Substitution and Firm-Level Elasticities

We next examine firm substitution. We find that that access to Lyft is associated with higher measured elasticities, consistent with idea that workers shift hours between firms as relative wages change. Sections 5.2 and 5.3 examine the implications for firms' optimal mark-downs and wage gaps.

[30] We found similar results based on specifications that allow the elasticity to be a more continuous function of these variables.

Figure 2: Elasticities for the Average Driver



Note: We group drivers according to the categories listed in each of the three panels. We then regress the inverse hyperbolic sine of hours on $\log(1-\tau_{it})$, interacted with group dummies. We use treatment indicators for each group as instruments and control for strata, interacted with group dummies. We perform this regression separately by driver gender to get estimates of $\varepsilon^f(C_i)$ and $\varepsilon^m(C_i)$. We then calculated the weighted average of these covariate-specific elasticities for the groups indicated. The estimates in the left figure of each panel are re-weighted estimates of ε^m; the estimates in the right figure of each panel are re-weighted estimates of ε^f. We define each sample in Section 3.2.Whiskers denote 95% confidence intervals. Table A9 presents additional specifications.

Table 6: Utilization Rates by Gender and Access to Lyft

	All		Male		Female	
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Shifter	0.586	0.591	0.583	0.592	0.590	0.590
	[0.147]	[0.153]	[0.137]	[0.149]	[0.156]	[0.157]
	1776	1956	851	1006	925	950
Shifter	0.615	0.633	0.617	0.639	0.611	0.618
	[0.190]	[0.182]	[0.186]	[0.177]	[0.196]	[0.194]
	1850	3131	1164	2263	686	868
p-value for equality	<.001	<.001	<.001	<.001	0.018	<.001
Houston EA	✓	✓	✓	✓	✓	✓
Boston EA		✓		✓		✓

Note: This table shows the average daily utilization rates for shifters and non-shifters. Shifters are drivers in the 2nd Houston experiment and (in even columns) Boston drivers eligible to drive for Lyft. For each driver we calculate utilization by dividing the time a driver is on a dispatch or has a passenger in the car by the amount of time he/she is active on the app. We average daily utilization rates by driver before taking the average in each sample. We exclude data from weeks in which a driver had a treatment offer (regardless of acceptance). Standard deviations are in brackets.

5.1 Behavior of Drivers with Access to an Outside Firm

Two pieces of evidence suggest drivers shift hours between firms. First, while we cannot—given the nature of our data—directly observe drivers working for both platforms, we can use the fact that many of those who drive for both Uber and Lyft likely use the high frequency multi-apping strategy described in Section 3.1 to provide direct evidence on firm substitution.[31] We call drivers who could work for both Uber and Lyft "shifters" and those limited to Uber "non-shifters".

Because high frequency multi-apping raises hourly earnings by increasing the fraction of time a driver spends with a passenger in the car (the "utilization rate"), drivers with access to Lyft should have higher utilization rates. For each driver we divide the amount of time a driver spends with a passenger in the car by the amount

[31] This data limitation does not bias our estimates: the ideal division is between drivers with and without access to the outside firm, regardless of whether they take advantage of the option.

of time the app is on (signaling the driver is available to accept or is actively on a trip). Table 6 shows that utilization rates are indeed significantly higher among drivers with shifters than non-shifters.

Second, consistent with the model in Section 2, drivers are more elastic in the second Houston experiment, conducted after Lyft returned to the Houston market. Table 7 presents labor supply elasticities for drivers with and without access to Lyft. Across all specifications we see that shifters appear more elastic than the non-shifters. Our preferred estimates in Column 7, which only use data from Houston but account for intensive and extensive margin adjustments, show an average elasticity of .8 for non-shifters and 1.5 for shifters.[32]

These results suggest that labor supply elasticities estimated in flexible hours environments in which workers are able to work for multiple firms may reflect both changes in both hours worked and change in the allocation of hours across firms. Appendix Table A10 shows that the contrast between shifters and non-shifters is robust to alternative specifications that use a single instrument, and that control only for the strata used in random assignment. In Appendix E we also corroborate this finding using data from an Uber-run promotion that incentivized drivers to exceed specified trip thresholds.[33]

5.2 Implied Markdowns

Section 2 shows that firms with market power should mark-down wages by the inverse elasticity of labor supply to the firm. We use the estimates presented in the previous section to estimate optimal mark-downs under two scenarios.

[32] Re-weighting exercises similar to those in Section 4.4 suggest that the differences in elasticities between shifters and non-shifters are not explained by differences in usual hours worked or experience. This is not surprising given how the experiments were structured.

[33] These data allow us to compare the responsiveness of drivers concurrently operating in the same market conditions, some of whom have access to a competing platform, and some of whom do not.

Table 7: Labor Supply with Multiple Firms

	1{Drive}		Log(Hours)		Asinh(Hours)		Poisson	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Non-Shifter	0.19***	0.19***	0.42***	0.42***	0.82***	0.82***	0.47***	0.48***
	(0.06)	(0.06)	(0.09)	(0.09)	(0.13)	(0.13)	(0.09)	(0.09)
Shifter	0.37***	0.26***	0.81***	1.10***	1.48***	1.53***	0.87***	1.01***
	(0.13)	(0.06)	(0.21)	(0.11)	(0.33)	(0.17)	(0.18)	(0.12)
p-value	0.202	0.408	0.091	<.001	0.066	<.001	0.051	<.001
Observations	4700	7172	4155	6369	4700	7172	4700	7172
Endogenous Variable	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)
Houston EA	✓	✓	✓	✓	✓	✓	✓	✓
Boston EA		✓		✓		✓		✓

Note: The dependent variable in columns 1–2 is an indicator for non-zero hours; the dependent variable in columns 3–4 is the log of hours worked; and the dependent variable in columns 5-6 is the inverse hyperbolic sine of hours worked. Columns 7-8 present a Poisson regression model. In columns 1–2 we scale the coefficients by baseline participation rates so that they can be interpreted as elasticities. All models control for the strata used for random assignment and for baseline covariates, interacted with whether a driver is a "shifter". The model uses dummies for each treatment offer. Standard errors are clustered by driver. Levels of significance: *10%, ** 5%, and *** 1%.

First, we use the hours elasticity for shifters to calculate mark-downs in a setting where workers can choose their hours, and where workers can work for multiple firms. The estimates in Column 7 of Table 7 suggest that a firm would optimally set wages $1 - 1/(1 + 1/1.5) \approx 40\%$ below workers' marginal product. It is worth noting here that this mark-down—which comes from adding a single competitor—is roughly thirty percent smaller than the mark-down a pure monopsonist would set. Using the estimates for non-shifters presented in Column 7 of Table 7, we calculate that a firm with no competitors in a flexible hours environment would optimally set wages $1 - 1/(1 + 1/.8) \approx 56\%$ below workers' marginal product.[34]

Second, we calculate the optimal mark-down for a setting where workers cannot choose their hours, and can only work at a single firm at a time. This is the typical

[34]Labor market institutions and norms may constrain firms' wage-setting in practice (Sokolova and Sorensen, 2021). Previous research has indicated ride-share markets can re-equilibrate through utilization rates (Hall, Horton, and Knoepfle, 2017).

situation outside of the gig economy. Because our experiment focuses on existing Uber drivers, we cannot estimate this mark-down simply using the extensive margin responses of shifters and non-shifters. It is likely that we understate the extensive margin adjustments that would occur if higher wages were offered to all ride-share drivers. To address this concern, we calculate the extensive margin *substitution* elasticity. This measures how changes in participation at the outside firm respond to a change in wages at Uber.[35] This is the standard elasticity used in dynamic models *without* flexible hours (Sokolova and Sorensen, 2021).

We calculate the firm substitution elasticity using the participation analogue of Equation 1 (derived in Appendix B). This equation relates the overall participation elasticity to the participation elasticities at each firm:

$$\underbrace{\frac{\partial P}{\partial w}\frac{w}{P}}_{\text{overall}} = \underbrace{\frac{\partial p^{prim}}{\partial w}\frac{w}{p^{prim}}}_{\text{Uber}} \times \phi + (1-\phi)\underbrace{\frac{\partial p^{sec}}{\partial w}\frac{w}{p^{sec}}}_{\text{outside firm}}$$

where ϕ is the initial market share. We calculate the firm substitution elasticity using the market share $\phi = .93$, as reported in Koustas (2018). Using the shifter and non-shifter estimates in Column 1 of Table 7 and the estimate $\phi = .93$, we obtain a firm substitution elasticity of -2.3. This says that, in response to a 10% increase in the Uber wage, overall employment at the outside firm decreases by 23%. This would suggest mark-downs on the order of 30%.

5.3 Implications for Wage Inequality

An open question is whether firms with market power have an incentive to wage discriminate against women. Firms outside the gig economy may create such gaps

[35] The generosity of our treatment offers makes it unlikely drivers would find it profitable to multi-app when treated.

by choosing to negotiate with workers or by responding to outside offers; previous research has documented that women are less willing to negotiate and often ask for less (see, e.g., Babcock and Laschever, 2009; Biasi and Sarsons, 2022).

The structure of financial incentives may also lead to wage gaps within flexible hours or gig firms. This is because many flexible hours firms offer wages that vary over the course of the day or with total hours worked; if women are less willing to work at certain times of day or less able to work long hours, gender-blind pay could lead to a gender wage gap.[36] Such a firm could potentially increase (decrease) the gender pay gap by increasing (decreasing) the premium for late-night hours or for over-time. Within driving occupations such as taxi and ride-share, drivers typically earn fares that depend on both per-mile and per-minute rates.

However, our results suggest that flexible hours firms do not have an incentive to wage discriminate against women. Table 8 shows that, within-gender, elasticities are larger among shifters than non-shifters. If anything, female shifters are *more* elastic than their male counterparts. This would weigh against the idea that, in markets with flexible hours and multiple job holding, firms would have incentives to set lower wages for women.

This is also true for flexible hours firms in environments where workers cannot hold multiple jobs. As discussed in Section 2, in these environments, the relevant elasticity combines the market-level hours responses and the extensive margin employment choices. Section 4 showed that *market* level hours and participation elasticities are higher among female drivers. Our results also indicate that female drivers are more elastic Columns 1 and 2 show that, if anything, female drivers are more likely to shift where they work. The estimates in Columns 1 and 2 of Table 8 imply estimates of -0.19 and -3.43 for male and female drivers respectively. This

[36]For instance, Bolotnyy and Emanuel (2022) document that gender-blind pay leads to a gender wage gap among Massachusetts Bay Municipal Transit Authority drivers.

Table 8: Elasticities with Multiple Firms: Results by Driver Gender

	1{Drive}		Log(Hours)		Asinh(Hours)		Poisson	
	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Non-Shifter	0.11	0.26***	0.28*	0.53***	0.52**	1.04***	0.30**	0.62***
	(0.09)	(0.07)	(0.15)	(0.11)	(0.22)	(0.16)	(0.14)	(0.11)
Shifter	0.18***	0.52***	1.19***	0.80***	1.47***	1.68***	1.00**	1.10***
	(0.07)	(0.15)	(0.13)	(0.23)	(0.19)	(0.33)	(0.44)	(0.27)
p-value	0.46	0.15	0.00	0.28	0.00	0.09	0.12	0.10
Observations	4639	2533	4167	2202	4639	2533	4639	2533
Model	Over-ID	Over-ID	Over-ID	Over-ID	Over-ID	Over-ID	Over-ID	Over-ID
Endogenous Variable	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)
Strata Controls	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓	✓	✓

Note: Shifters are drivers in the 2nd Houston experiment and Boston drivers eligible to drive for Lyft. The dependent variable is indicated at the top of each panel. In columns 1-2 we scale the coefficients by baseline participation rates so that we recover elasticities (rather than semi-elasticities). All models control for the strata used for random assignment and baseline covariates, interacted with whether a driver is a "shifter". The over-identified model uses dummies for each treatment offer. Standard errors are clustered by driver. Levels of significance: *10%, ** 5%, and *** 1%. See Section 5.1 for more information.

suggests, if anything, women are more likely to switch between firms in response to relative wage changes.

6 Discussion and Conclusion

We used a series of field experiments we conducted at Uber to provide new evidence on the elasticity of labor supply, both to the market and to the firm.

We found that, at the level of the market, women are more elastic than men. Using data from the experiment we conducted when a ride-share alternative was unavailable, we document that, in response to a ten percent increase in wages, women are nearly two percent more likely to drive, compared to one percent for men. Conditional on driving, women drive eight percent more hours, compared to four percent more for men.

While elasticities are larger for low-hours drivers, this difference does not explain gender differences in our sample. Rather, gender gaps emerge even between "full-time" male and female drivers. A simple exercise that re-weights our estimates to match the distributions of hours and experience in the overall population of Houston drivers indicates the differences we find underestimate the differences between the average male and female driver (Oaxaca, 1973).

We also found that drivers shift hours between platforms (firms) when given the opportunity to do so. While many researchers have turned to flexible hours environments to avoid concerns regarding hours constraints (see, e.g., Kahn and Lang, 1991; Dickens and Lundberg, 1993), our results suggest that estimates in these settings may not recover market-level elasticities if workers are able to hold multiple flexible jobs. More generally, our results highlight the importance of considering workers' outside options and market structure when interpreting data—including experimental data—provided by a single firm or platform.

We found no evidence that women are less likely to engage in platform shifting than men. In a monopsonistic labor market, workers' wages depend both on their productivity and the elasticity of their labor supply to the firm; workers who are less elastic earn lower wages. Our experiment reveals firm-level hours elasticities that are larger for women than for men. It also suggests firm net-separation elasticities on the order of -1 (male drivers) and -3 (female drivers). Our results therefore indicate that, at least in the gig economy, firms do not have a strong incentive to pay women lower wages. To the extent that women may be drawn to the gig economy due to a desire for flexible work arrangements, this is particularly encouraging.

One limitation of our design is that our estimates of firm substitution come from a predominantly male labor market in which most firm-amenities do not vary across firms. Outside of the gig economy firms may have more (or less) wage-setting power over female workers than male workers if female workers have stronger preferences over non-wage amenities (Caldwell and Danieli, 2022), are more constrained in geographic location (Le Barbanchon, Rathelot, and Roulet, 2021), or have worse information about their outside options. While bargaining is not common among low-wage workers (see, e.g. Hall and Krueger, 2012; Lachowska et al., 2022), gender differences in bargaining may explain wage gaps among higher wage workers (see, e.g., Biasi and Sarsons, 2022; Roussille, 2020). Our results show that, in a setting without differences in firm-provided amenities, women are no less strategic in responding to changes in relative wages.

References

Angrist, Joshua D., Sydnee Caldwell, and Jonathan V. Hall (2021). "Uber vs. Taxi: A Driver's Eye View". In: *American Economic Journal: Applied Economics* 13.3, pp. 272–308.

Athey, Susan, Juan Camilo Castillo, and Bharat Chandar (2019). "Service Quality in the Gig Economy: Empirical Evidence About Driving Quality at Uber". In: *Available at SSRN 3499781*.

Babcock, Linda and Sara Laschever (2009). *Women Don't Ask: Negotiation and the Gender Divide*. Princeton University Press.

Barth, Erling and Harald Dale-Olsen (2009). "Monopsonistic Discrimination, Worker Turnover, and the Gender Wage Gap". In: *Labour Economics* 16.5, pp. 589–597.

Biasi, Barbara and Heather Sarsons (2022). "Flexible wages, bargaining, and the gender gap". In: *Quarterly Journal of Economics* 137.1, pp. 215–266.

Blau, Francine D and Lawrence M Kahn (1981). "Race and Sex Differences in Quits by Young Workers". In: *ILR Review* 34.4, pp. 563–577.

Blinder, Alan S (1973). "Wage Discrimination: Reduced Form and Structural Estimates". In: *Journal of Human Resources*, pp. 436–455.

Bolotnyy, Valentin and Natalia Emanuel (2022). "Why do women earn less than men? Evidence from bus and train operators". In: *Journal of Labor Economics* 40.2, pp. 283–323.

Caldwell, Sydnee and Oren Danieli (2022). *Outside Options in the Labor Market*. Tech. rep. Working Paper.

Campbell, Harry (2017). *2017 Rider Survey*. Tech. rep. URL: https://docs.google.com/document/d/1QSUFSqasfjM9b9UsqBwZlpa8EgqNj6EBfWybFBSHj3o/edit.

Campbell, Harry (2022). *How Many Uber Drivers Are There?* Tech. rep. URL: https://theridesshareguy.com/how-many-uber-drivers-are-there/.

Card, David, Ana Rute Cardoso, and Patrick Kline (2016). "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women". In: *Quarterly Journal of Economics* 131.2, pp. 633–686.

Card, David et al. (2016). "Firms and Labor Market Inequality: Evidence and Some Theory". In: *Journal of Labor Economics* 53.1, S13–S70.

Chen, Kuan-Ming et al. (2020). *Reservation Wages and Workers' Valuation of Job Flexibility: Evidence from a Natural Field Experiment*. Working Paper 27807. NBER.

Chen, M Keith et al. (2019). "The Value of Flexible Work: Evidence from Uber Drivers". In: *Journal of Political Economy* 127.6, pp. 2735–2794.

Chetty, Raj et al. (2013). "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities". In: *NBER Macroeconomics Annual* 27.1, pp. 1–56.

Cook, Cody et al. (2021). "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Uber Drivers". In: *The Review of Economic Studies* 88.5, pp. 2210–2238.

Dickens, William T, Shelly J Lundberg, et al. (1993). "Hours Restrictions and Labor Supply". In: *International Economic Review* 34.1, pp. 169–192.

Dube, Arindrajit et al. (2020). "Monopsony in Online Labor Markets". In: *American Economic Review: Insights* 2.1, pp. 33–46.

Eissa, Nada and Jeffrey B Liebman (1996). "Labor Supply Response to the Earned Income Tax Credit". In: *Quarterly Journal of Economics* 111.2, pp. 605–637.

Farber, Henry S. (2005). "Is Tomorrow Another Day? the Labor Supply of New York City Cabdrivers". In: *Journal of Political Economy* 113.1, pp. 46–82.

— (2015). "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers". In: *Quarterly Journal of Economics* 130.4, pp. 1975–2026.

Fehr, Ernst and Lorenz Goette (2007). "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment". In: *The American Economic Review* 97.1, pp. 298–317.

Flood, Sarah et al. (2021). *Integrated Public Use Microdata Series, Current Population Survey: Version 9.0*. Tech. rep. https://doi.org/10.18128/D030.V9.0. Minneapolis, MN.

Hall, Jonathan V, John J Horton, and Daniel T Knoepfle (2017). *Labor Market Equilibration: Evidence from Uber*. Tech. rep. Working Paper, 1–42.

— (2021). "Pricing in Designed Markets: The Case of Ride-Sharing". In:

Hall, Robert E and Alan B Krueger (2012). "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-The-Job Search". In: *American Economic Journal: Macroeconomics* 4.4, pp. 56–67.

Hirsch, Boris, Thorsten Schank, and Claus Schnabel (2010). "Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked Employer-Employee Data from Germany". In: *Journal of Labor Economics* 28.2, pp. 291–330.

Kahn, Shulamit and Kevin Lang (1991). "The Effect of Hours Constraints on Labor Supply Estimates". In: *The Review of Economics and Statistics*, pp. 605–611.

Killingsworth, Mark R and James J Heckman (1986). "Female Labor Supply: A Survey". In: *Handbook of Labor Economics* 1, pp. 103–204.

King, Robert G and Sergio T Rebelo (1999). "Resuscitating Real Business Cycles". In: *Handbook of Macroeconomics* 1, pp. 927–1007.

Koustas, Dmitri (2018). *Consumption Insurance and Multiple Jobs: Evidence from Rideshare Drivers*. Working Paper. University of Chicago.

Lachowska, Marta et al. (2022). "Wage Posting or Wage Bargaining? A Test Using Dual Jobholders". In: *Journal of Labor Economics* 40.S1, S469–S493.

Le Barbanchon, Thomas, Roland Rathelot, and Alexandra Roulet (2021). "Gender Differences in Job Search: Trading off Commute Against Wage". In: *Quarterly Journal of Economics* 136 (1), pp. 381–426.

Manning, Alan (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

Martinez, Isabel Z, Emmanuel Saez, and Michael Siegenthaler (2021). "Intertemporal Labor Supply Substitution? Evidence from the Swiss Income Tax Holidays". In: *American Economic Review* 111.2, pp. 506–46.

Mas, Alexandre and Amanda Pallais (2018). "Labor Supply and the Value of Non-Work Time: Experimental Estimates from the Field". In: *American Economic Review: Insights*.

McClelland, Robert and Shannon Mok (2012). "A Review of Recent Research on Labor Supply Elasticities". In: *Congressional Budget Office*.

Meyer, Bruce D and Dan T Rosenbaum (2001). "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers". In: *Quarterly Journal of Economics* 116.3, pp. 1063–1114.

Morchio, Iacopo and Christian Moser (2021). "The gender pay gap: Micro sources and macro consequences". In: *Available at SSRN 3176868*.

Oaxaca, Ronald (1973). "Male-Female Wage Differentials in Urban Labor Markets". In: *International Economic Review*, pp. 693–709.

Oettinger, Gerald S (1999). "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors". In: *Journal of Political Economy* 107.2, pp. 360–392.

Ransom, Michael R and Ronald L Oaxaca (2010). “New Market Power Models and Sex Differences in Pay”. In: *Journal of Labor Economics* 28.2, pp. 267–289.
Robinson, Joan (1933). *The Economics of Imperfect Competition*.
Rogerson, Richard and Johanna Wallenius (2009). “Micro and Macro Elasticities in a Life Cycle Model with Taxes”. In: *Journal of Economic Theory* 144.6, pp. 2277–2292.
Roussille, Nina (2020). *The central role of the ask gap in gender pay inequality*. Tech. rep. University of California, Berkeley Working Paper.
Sokolova, Anna and Todd Sorensen (2021). “Monopsony in labor markets: A meta-analysis”. In: *ILR Review* 74.1, pp. 27–55.
Stafford, Tess M (2015). “What Do Fishermen Tell Us That Taxi Drivers Do Not? An Empirical Investigation of Labor Supply”. In: *Journal of Labor Economics* 33.3, pp. 683–710.
Webber, Douglas A (2016). “Firm-Level Monopsony and the Gender Pay Gap”. In: *Industrial Relations: A Journal of Economy and Society* 55.2, pp. 323–345.

Gender Differences in Labor Supply:

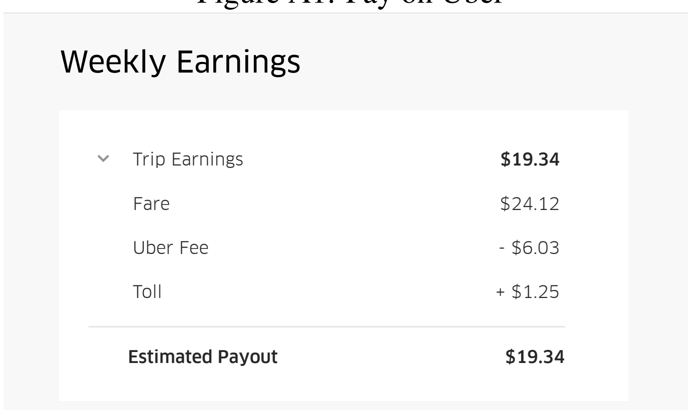
Experimental Evidence from the Gig Economy

Online Appendix

Sydnee Caldwell and Emily Oehlsen

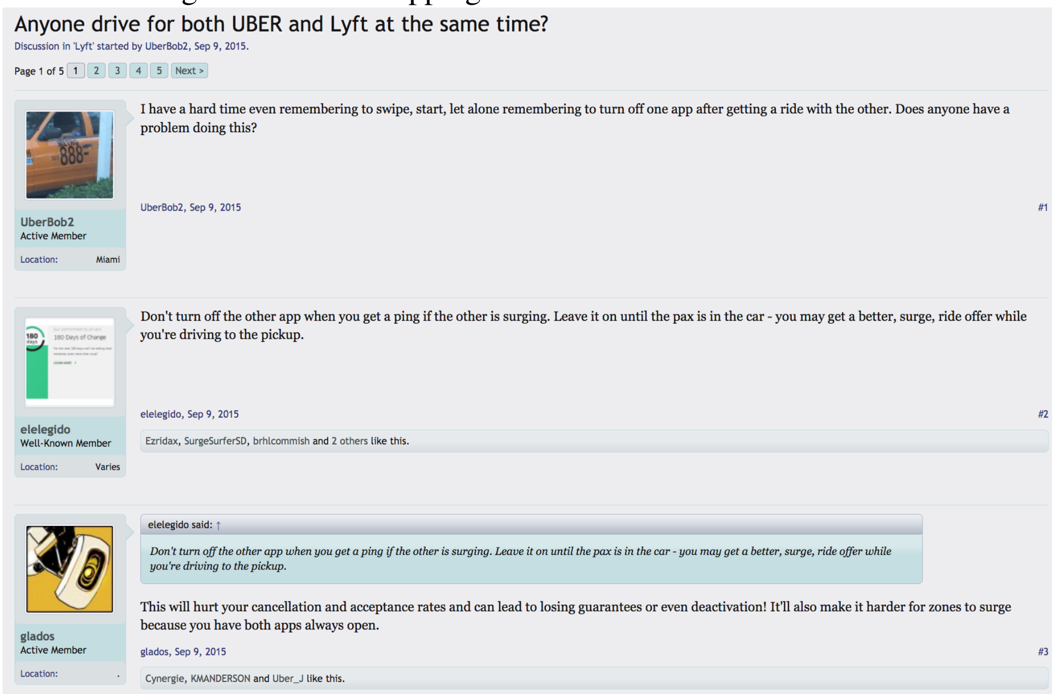
A Appendix Tables and Figures

Figure A1: Pay on Uber



Note: This picture shows that, when we ran the experiment, drivers' weekly pay statements listed (1) how much they collected in trip receipts, (2) how much of this went to Uber in the form of the Uber fee, (3) what, if anything, they earned in Uber promotions, and (4) what, if any, reimbursements they received for tolls. Their estimated payout was the sum of these four items. The structure of drivers' weekly earning statements has since changed.

Figure A2: Multi-Apping Information from Driver Forums



Anyone drive for both UBER and Lyft at the same time?

Discussion in 'Lyft' started by UberBob2, Sep 9, 2015.

Page 1 of 5 1 2 3 4 5 Next >

UberBob2
Active Member
Location: Miami

I have a hard time even remembering to swipe, start, let alone remembering to turn off one app after getting a ride with the other. Does anyone have a problem doing this?

UberBob2, Sep 9, 2015 #1

elelegido
Well-Known Member
Location: Varies

Don't turn off the other app when you get a ping if the other is surging. Leave it on until the pax is in the car - you may get a better, surge, ride offer while you're driving to the pickup.

elelegido, Sep 9, 2015 #2

Ezridax, SurgeSurferSD, brhlcommish and 2 others like this.

glados
Active Member
Location: -

elelegido said: ↑

Don't turn off the other app when you get a ping if the other is surging. Leave it on until the pax is in the car - you may get a better, surge, ride offer while you're driving to the pickup.

This will hurt your cancellation and acceptance rates and can lead to losing guarantees or even deactivation! It'll also make it harder for zones to surge because you have both apps always open.

glados, Sep 9, 2015 #3

Cynergie, KMANDERSON and Uber_J like this.

Note: This is a screenshot from “Uber People”, an online forum and discussion board where drivers discuss ride-share related topics. The forum is not affiliated with Uber Technologies, Inc. or any other ride-share company. The conversation highlights that drivers are interested in multi-apping but find it requires a non-trivial amount of effort.

Figure A3: Multi-Apping Guide

How to run the apps at the same time

Once you have everything set up you will need to actually open and run both apps at the same time. This is pretty easy for the most part, but there are a few tricks that will help you out.

First, close out any apps you have open. Driving for Lyft and Uber at the same time requires you to keep both apps open until you get a request, which puts quite a bit of strain on your phone, sucks the battery dry, and uses a ton of data in the process. So bring a charger, stay on task, and be prepared to pull down a lot of data. And I mean, a TON.

Second, try to minimize surfing the web or being active on social media. Keep in mind you have two apps that are constantly talking to their respective platforms, so the chance of your phone crashing or glitching out becomes much higher than usual. The last thing you want is to get a ride request and have your phone freeze up or crash, costing you both time and money.

It is also worth noting that when you have both of the apps open, you will need to have Uber open on the main screen, with Lyft running in the background behind it. Uber will automatically close out after a minute or two if it is running in the background, but Lyft will stay open.

Accepting ride requests

When you get a ride request, make sure to accept one and immediately log right out of the other. It won't take long to get a ride request, and on a busy night you may get two at the same time. If this happens, pick the once closest to you and decline the other. As every driver knows, the best way to make money is in volume, and the less dead miles you have, the better.

Note: This is a screenshot from rideshareapps.com's guide on how to drive for Lyft and Uber at the same time. This is intended to illustrate that multiple non-Uber/Lyft affiliated forums provide information to drivers on how to maximize earnings via multi-apping.

Figure A4: Example of a Third Party Multi-Apping Application

News Video Events Crunchbase

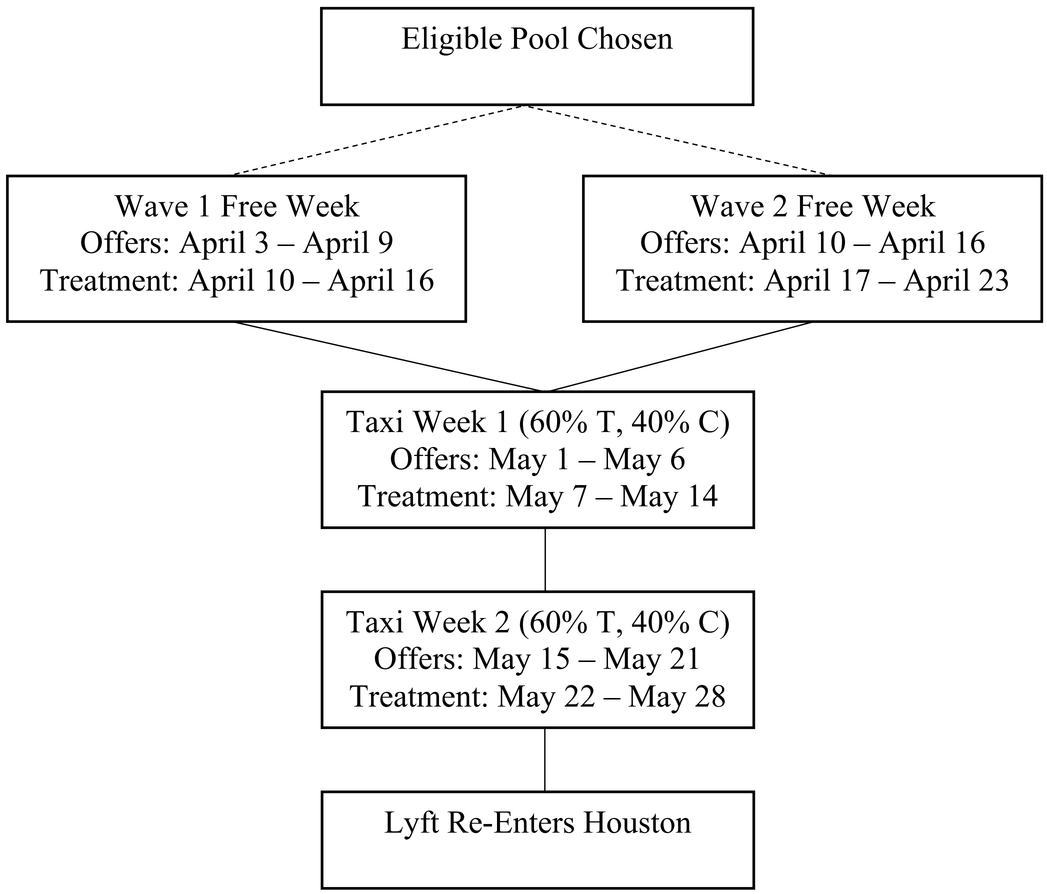
Mystro is an app aiming to bring in more bacon for Uber and Lyft drivers

Posted Aug 2, 2017 by Sarah Buhr (@sarahbuhr)

Mystro hopes to put 33 percent more money per year in the pockets of Uber and Lyft drivers through an app.

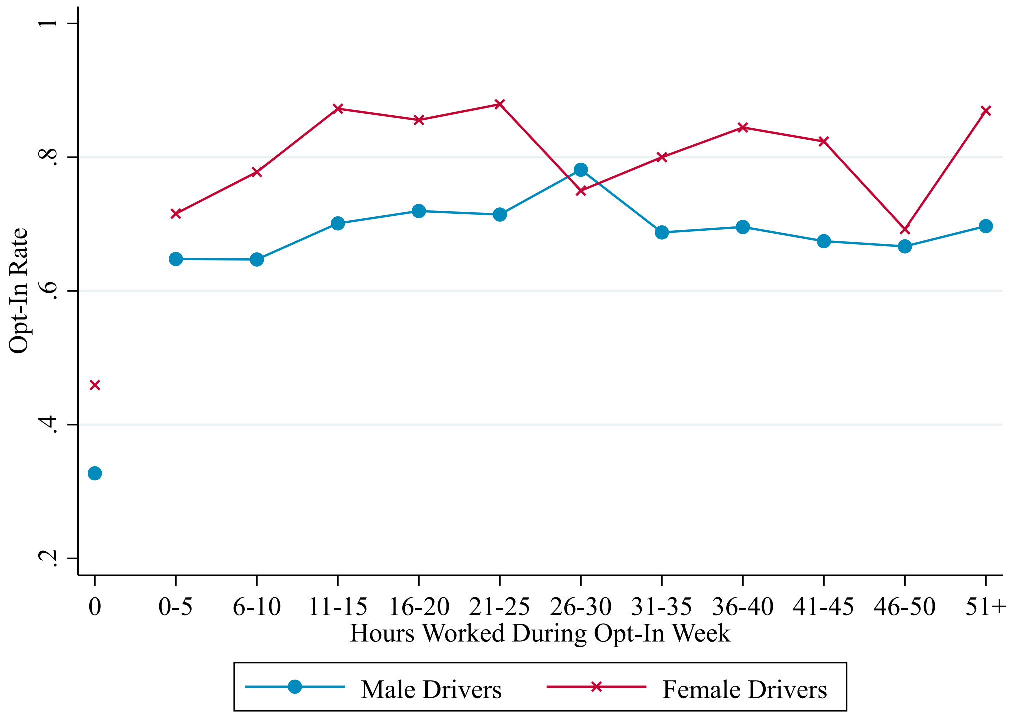
Note: This is a screenshot of a TechCrunch article discussing a third party app, Mystro, which helps drivers quickly switch between competing ride-share platforms. This is one of many third-party apps that help drivers maximize their earnings through multi-apping. As of August 10, 2019 (before the COVID-19 pandemic), Mystro charged $5/week or $12/month or $100/year for this service.

Figure A5: Design



Note: This diagram shows the design of the first Houston Earnings Accelerator experiment. All dates are from 2017.

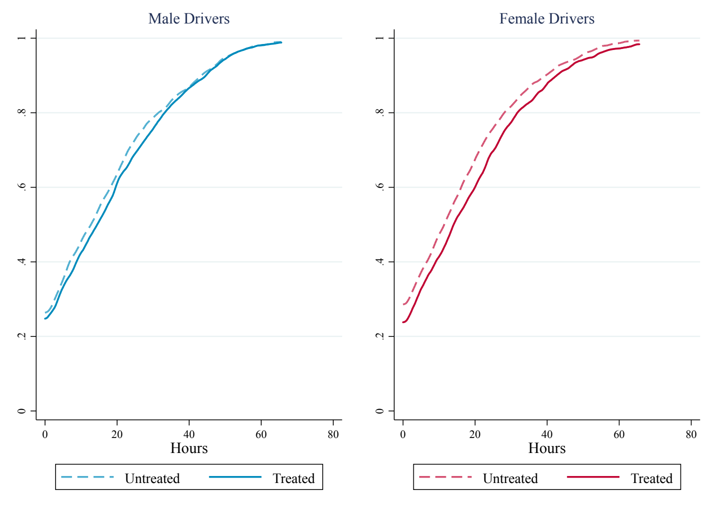
Figure A6: Opt-In Rates by Opt-In Week Hours



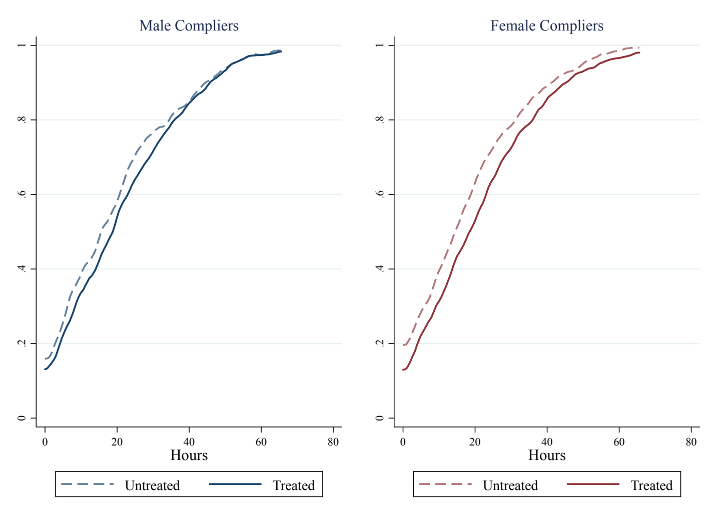
Note: This figure plots the proportion of drivers who accepted our initial (non-taxi) offer during the first Houston experiment, separately by gender and by hours worked during opt-in week. The far left dots plot the proportions for drivers who drove 0 hours during opt-in week. The remaining dots plots proportions for drivers whose hours fell within the intervals $(a, b]$ as indicated.

Figure A7: Effects on the Distribution of Hours Worked

A. All Drivers



B. Compliers



Note: This figure shows the impact of the wage offers on the distribution of hours worked for drivers in the main Houston experiment. Panel A presents the distribution of hours worked for untreated and treated drivers of each gender. Panel B presents the distribution of hours worked for untreated and treated compliers of each gender. The complier distributions are computed using estimates of $\alpha_0(v)$ and $\alpha_1(v)$ for $v \in [0, 80]$ from sex-specific regressions of the form:
$1[h_{it} < v](1 - D_{it}) = X_i'\beta_0(v) + \alpha_0(v)(1 - D_{it}) + u_{0iv}$ and
$1[h_{it} < v]D_i = X_i'\beta_1(v) + \alpha_1(v)D_{it} + u_{1iv}$ (Abadie 2003). D_{it} is an indicator for whether driver i accepted an offer in week t. We instrument for $(1 - D_{it})$ and D_{it} using the randomly assigned treatment offers Z_{it}. Each regression controls for the strata used for random assignment.

Table A1: Earnings Accelerator Experiments

	Boston	**Houston 1 ("Main Experiment")**	**Houston 2**
Timeline	Aug-Oct 2016	Apr-May 2017	Sept-Oct 2017
Sample Size	1600	2020	2100
Male	1368	972	1283
Female	232	1048	817
Shifters			
Definition	Car Year <=2003	No Drivers	All Drivers
# Shifters	1405	0	2100
# Non-Shifters	195	2020	0
Commission Rate	20% or 25%	25% or 28%	25% or 28%
Hours Groups			
Low	5-15 Hrs/Wk	5-15 Hrs/Wk	5-15 Hrs/Wk
High	15-25 Hrs/Wk	15-25 Hrs/Wk	15-25 Hrs/Wk
Very High	✕	25-40 Hrs/Wk	25-40 Hrs/Wk
Analysis			
Frisch Elasticity	✕	✓	✕
Firm Substitution	✓	✓	✓

Note: This table compares the three Earnings Accelerator experiments. The first two panels show the dates of the experiments and the sample sizes, broken down by gender. The third panel shows who is defined as a "shifter" in each experiment and the fourth describes the commission rates. The fifth panel defines the hours strata. The final panel describes what analysis uses each experiment. More information on these experiments is provided in Appendix Section C.2.

Table A2: Earnings Accelerator Balance: Main Experiment

	Sample Selection			Randomization Balance	
	Eligible Mean	RCT Mean	Experimental Difference	Wave1-Wave2	Taxi Treated-Control
	(1)	(2)	(3)	(4)	(5)
Female	0.20	0.52	---	0.02	0.01
			---		(0.017)
Months on Platform	44.13	44.13	0.00	-0.27	0.01
			(0.000)		(0.475)
Vehicle Solutions	0.10	0.12	0.00	0.01	-0.01
			(0.010)		(0.013)
Vehicle Year	10.48	16.01	0.24	0.30	0.06
			(0.162)		(0.185)
Hours Week Prior to Offer	16.84	18.45	-0.59	-0.37	-0.52
			(0.388)		(0.620)
Earnings Week Prior to Offer	262.74	296.14	-4.11	-6.88	-9.68
			(7.324)		(8.934)
F-statistic			1.49	0.73	0.35
p-value from F-test			0.20	0.62	0.91
Drivers	10641	2020	10641	2020	1355

Note: Columns 1 and 2 present the mean value of the indicated characteristic for drivers eligible for inclusion in the experiment (Column 1) and drivers included in the experiment (Column 2). Drivers were eligible if they completed at least four trips in the prior month and had average hours per week (conditional on driving) that fell between 5 and 40 hours. Column 3 shows the strata-adjusted difference between those selected and not selected for the experiment. Columns 4 and 5 show strata-adjusted differences between drivers in waves 1 and 2 of the experiment (Column 4) and between drivers offered and not offered a taxi contract (Column 5). Column 5 includes 2 observations for each of the 1355 drivers who accepted initial offer of fee-free driving and were therefore eligible for inclusion in the taxi phase of the experiment: one for each week of taxi offers. Levels of significance: *10%, ** 5%, and *** 1%.

Table A3: Earnings Accelerator Balance: Additional Experiments

	Boston Experiment		Second Houston Experiment	
	Sample Mean	Week 1-2 Balance	Sample Mean	Treatment-Control Balance
	(1)	(2)	(3)	(4)
Female	0.14	-0.02	0.39	0.01
		(0.019)		(0.020)
Months on Platform	11.14	-0.01	8.75	-0.19
		(0.250)		(0.250)
Vehicle Solutions	0.08	0.00	0.08	0.01
		(0.016)		(0.012)
Vehicle Year	2009.27	0.13	2013.63	0.07
		(0.192)		(0.119)
Hours Week Prior to Offer	13.87	-0.01	14.82	0.71
		(0.690)		(0.619)
Earnings Week Prior to Offer	242.90	-2.92	205.76	3.75
		(12.325)		(8.444)
F-statistic		0.48		0.82
p-value from F-test		0.82		0.55
Drivers		1600		2100

Note: This is a combined balance table for the two additional Earnings Accelerator experiments used in this paper. Columns 1-2 present statistics for the Boston experiment; columns 3-4 present statistics for the second Houston experiment, conducted after Lyft had returned to the market. Columns 1 and 3 present sample means for drivers in each experiment. Columns 2 and 4 present the strata-adjusted difference between drivers offered treatment in the first week and the drivers offered treatment driving in the second week. Levels of significance: *10%, ** 5%, and *** 1%.

Table A4: Treatment Effects During Non-Treatment Weeks

	Before Wave 1		Treatment Weeks		After Wave 2	
	(1)	(2)	(3)	(4)	(5)	(6)
			A. 1{Drive}			
Treated	-0.01	-0.01	0.03***	0.03***	0.01	0.01
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Observations	2020	2020	4040	4040	1355	1355
			B. Log Hours			
Treated	-0.04	-0.04	0.09***	0.09***	0.03	0.03
	(0.04)	(0.04)	(0.02)	(0.02)	(0.04)	(0.04)
Observations	1554	1554	2998	2998	1066	1066
Strata Dummies	✓	✓	✓	✓	✓	✓
Baseline Covariates		✓		✓		✓

Note: This table examines the impact of the treatment on driving behavior before and after treatment. Each estimate comes from a regression of the indicated variable on a dummy for whether the driver was treated, controlling for the strata used for random assignment (all columns) and for baseline covariates (even columns). The data come from the first Houston experiment, conducted before Lyft re-entered the market. Standard errors are clustered by driver. Levels of significance: *10%, ** 5%, and *** 1%.

Table A5: Non-Experimental Estimates of Labor Supply Elasticities

	Non-Experimental Weeks				Experimental Weeks			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. OLS							
Male	0.18**	0.12	0.23***	0.22***	0.26**	0.25**	0.16**	0.23***
	(0.09)	(0.09)	(0.07)	(0.08)	(0.11)	(0.11)	(0.08)	(0.09)
Female	0.10	0.11	0.10*	0.24***	0.17**	0.17**	0.27***	0.36***
	(0.08)	(0.08)	(0.06)	(0.06)	(0.08)	(0.08)	(0.06)	(0.07)
Observations	9870	9870	7651	9744	5031	5031	4648	4842
	B. 2SLS (Leave-One-Out Log Wages)							
Male	-0.27	-0.35	-0.09	-0.17	0.63**	0.59**	-0.04	0.16
	(0.23)	(0.23)	(0.14)	(0.24)	(0.26)	(0.26)	(0.19)	(0.23)
Female	-0.63**	-0.63**	-0.29	-0.76**	0.23	0.22	0.46***	0.42
	(0.28)	(0.28)	(0.19)	(0.31)	(0.19)	(0.19)	(0.14)	(0.26)
Observations	9854	9854	7637	9728	5023	5023	4640	4833
Week Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Months on Uber		✓	✓			✓	✓	
Lag of Log(Hours)			✓				✓	
Driver Fixed Effects				✓				✓

Note: This table presents estimates of labor supply elasticities that come from regressions of log hours on log wages (earnings/hours worked). Panel B presents specifications where we instrument log wages with the leave-out mean of the wages of other drivers of the same gender that week. Standard errors are clustered at the driver level. Columns 1-4 use data from 8 non-experimental weeks surrounding our main (pre-Lyft) Houston experiment. Columns 5-8 use data from the 2 weeks of fee-free driving and 2 weeks of taxi experiments. Because the dependent variable is log hours, each regression excludes drivers with non-positive hours. Levels of significance: *10%, ** 5%, and *** 1%.

Table A6: Robustness of Frisch Labor Supply Elasticities

	1{Drive}	Log(Hours)				Asinh(Hours)		Poisson	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	A. Just-Identified Model								
Male	0.11	0.45*	0.28*	0.56***	0.43***	0.52**	0.63***	0.29**	0.42***
	(0.09)	(0.23)	(0.15)	(0.17)	(0.14)	(0.22)	(0.21)	(0.14)	(0.12)
Female	0.25***	0.74***	0.53***	0.98***	0.88***	1.03***	1.17***	0.62***	0.81***
	(0.07)	(0.16)	(0.11)	(0.13)	(0.11)	(0.17)	(0.17)	(0.11)	(0.10)
p-value from test of equality	0.194	0.296	0.171	0.054	0.013	0.061	0.046	0.061	0.012
Observations	3108	2766	2766	4650	4650	3108	5154	3108	5154
	B. Strata Only								
Male	0.12	0.40*	0.28*	0.55***	0.41***	0.57**	0.66***	0.37*	0.39**
	(0.09)	(0.23)	(0.16)	(0.19)	(0.15)	(0.23)	(0.25)	(0.21)	(0.19)
Female	0.31***	0.57***	0.45***	0.93***	0.80***	1.06***	1.27***	0.65***	0.75***
	(0.07)	(0.16)	(0.12)	(0.14)	(0.12)	(0.16)	(0.18)	(0.15)	(0.15)
p-value from test of equality	0.115	0.544	0.413	0.091	0.043	0.087	0.048	0.285	0.118
Observations	4040	2998	2998	5031	5031	4040	6750	4040	6750
Endogenous Variable	Log(1-T)	Log(W)	Log(W)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)
Fee-Free Driving Data	✓	✓	✓	✓	✓	✓	✓	✓	✓
Taxi Data			✓		✓		✓		✓

Note: This table presents results analogous to those in Table 4. Panel A presents results from a just-identified model that uses the same baseline covariates as Table 4. Panel B presents results from the baseline model, controlling only for the strata used for random assignment. Standard errors are clustered by driver. Levels of significance: *10%, ** 5%, and *** 1%.

Table A7: Earnings Elasticities

	1{Positive Fares}	Log(Fares)				Asinh(Fares)		Poisson	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	A. Just-Identified Model								
Male	0.11	0.40*	0.58***	0.25	0.45***	0.71*	0.77**	0.26*	0.43***
	(0.09)	(0.24)	(0.17)	(0.16)	(0.15)	(0.39)	(0.36)	(0.15)	(0.12)
Female	0.24***	0.84***	1.13***	0.61***	1.01***	1.50***	1.56***	0.68***	0.91***
	(0.07)	(0.16)	(0.13)	(0.13)	(0.12)	(0.29)	(0.29)	(0.11)	(0.10)
p-value	0.290	0.125	0.010	0.080	0.005	0.102	0.089	0.022	0.002
Observations	3108	2715	4581	2715	4581	3108	5154	3108	5154
	B. Over-Identified Model								
Male	0.11	0.36	0.62***	0.25	0.41***	0.70*	0.82**	0.25*	0.35***
	(0.09)	(0.23)	(0.17)	(0.16)	(0.15)	(0.38)	(0.34)	(0.14)	(0.12)
Female	0.24***	0.81***	1.09***	0.60***	0.92***	1.49***	1.55***	0.66***	0.79***
	(0.07)	(0.16)	(0.12)	(0.13)	(0.12)	(0.29)	(0.27)	(0.11)	(0.10)
p-value	0.276	0.098	0.018	0.089	0.006	0.099	0.095	0.024	0.004
First Stage F: Male	212	22	11	205	72	212	74	---	---
First Stage F: Female	452	35	14	481	157	452	141	---	---
Observations	3108	2715	4581	2715	4581	3108	5154	3108	5154
Endogenous Variable	Log(1-T)	Log(W)	Log(W)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)	Log(1-T)
Fee-Free Driving Data	✓	✓	✓	✓	✓	✓	✓	✓	✓
Taxi Data			✓		✓		✓		✓
Strata Dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
Baseline Covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: This table presents results from specifications analogous to those in Table 4 where the outcome variable is either an indicator for positive earnings that week (Column 1), the log of fares collected (Columns 2-5), or the inverse hyperbolic sine of earnings (Columns 6-7). Columns 8 and 9 present results from Poisson regression models. Panel A presents results from a just-identified model that uses the same baseline covariates as Table 4. Panel B presents results from the baseline model, controlling only for the strata used for random assignment. Standard errors are clustered by driver. Levels of significance: *10%, ** 5%, and *** 1%.

Table A8: Complier Characteristics

	Male Drivers		Female Drivers		p-value for equality of (1) and (3)
	$E[X]$	$\frac{E[X\|D_{i1} > D_{i0}]}{E[X]}$	$E[X]$	$\frac{E[X\|D_{i1} > D_{i0}]}{E[X]}$	
	(1)	(2)	(3)	(4)	(5)
Age	43.84	1.00	44.27	1.02	0.44
Months Since Signup	20.09	0.99	12.22	1.02	<.01
Vehicle Solutions	0.09	1.06	0.15	1.12	<.01
Commission	24.17	1.00	26.53	1.00	<.01
Usual Hours Worked (Pre-Selection)					
Mean	20.67	1.03	18.62	1.05	<.01
Between 0 and 5	0.11	0.70	0.11	0.76	0.75
Between 5 and 15	0.31	1.05	0.35	0.96	0.05
Between 15 and 25	0.28	1.02	0.26	1.04	0.53
Between 25 and 35	0.12	1.11	0.17	1.10	<.01
Between 35 and 45	0.10	0.99	0.07	1.11	<.01
Over 45	0.08	1.03	0.04	1.08	<.01

Note: This table compares the characteristics of fee-free driving compliers to the pool of experimental drivers. Columns 1 and 3 report sample means for male and female drivers. Columns 2 and 4 show how the compliers compare to their same-gender counterparts. Column 5 shows the p-value from a test of equality of the sample means in columns 1 and 3. Complier characteristics are computed following (Abadie 2003).

Table A9: Elasticities for the Average Driver

	Hours		Experience		Hours & Experience
	(1)	(2)	(3)	(4)	(5)
Female Elasticities					
RCT Women	1.08***	1.47***	1.03***	1.05***	1.07***
	(0.16)	(0.21)	(0.16)	(0.18)	(0.16)
Eligible Women	1.26***	1.55***	1.04***	1.08***	1.34***
	(0.19)	(0.23)	(0.16)	(0.18)	(0.21)
Active Drivers	1.34***	1.59***	1.04***	1.08***	1.40***
	(0.21)	(0.25)	(0.16)	(0.18)	(0.22)
Male Elasticities					
RCT Men	0.63***	0.94***	0.59**	0.62**	0.59**
	(0.23)	(0.30)	(0.23)	(0.28)	(0.23)
Eligible Men	0.62***	0.74**	0.93***	0.86***	0.95***
	(0.23)	(0.31)	(0.27)	(0.31)	(0.28)
Active Drivers	0.62***	0.60*	0.86***	0.84***	0.85***
	(0.24)	(0.33)	(0.25)	(0.30)	(0.25)
Hours Bins	2	3			2
Experience Bins			2	3	2

Note: This table presents estimates of the average driver's labor supply elasticity. We group drivers according to the categories listed at the bottom of each column. We then regress the inverse hyperbolic sine of hours on $\log(1-\tau_{it})$, interacted with group dummies. We use treatment indicators for each group as instruments and control for strata, interacted with group dummies. We perform this regression separately by driver gender to get estimates of $\varepsilon^f(C_i)$ and $\varepsilon^m(C_i)$. We then calculated the weighted average of these covariate-specific elasticities for the groups in the indicated row. The estimates in the first panel are re-weighted estimates of ε^f; the estimates in the second panel are re-weighted estimates of ε^m. We define each sample in Section 3.2. Standard errors are calculated using the delta method. Levels of significance: *10%, ** 5%, and *** 1%. Columns 1, 3, and 5 are plotted in Figure 2.

Table A10: Robustness of Shifter and Non-Shifter Elasticities

	Male			Female		
	1{Drive}	Log(Hours)	Asinh(Hours)	1{Drive}	Log(Hours)	Asinh(Hours)
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Just-ID Model					
Non-Shifter	0.18**	0.30***	0.27*	0.46***	0.67***	1.04***
	(0.09)	(0.07)	(0.15)	(0.12)	(0.22)	(0.16)
Shifter	0.00	0.51*	1.16***	0.68***	1.25***	1.51***
	(0.23)	(0.28)	(0.14)	(0.26)	(0.22)	(0.42)
Observations	3227	2913	4598	2428	5963	3377
	B. Strata Only					
Non-Shifter	0.14	0.30***	0.30*	0.47***	0.63***	1.03***
	(0.09)	(0.07)	(0.16)	(0.12)	(0.23)	(0.16)
Shifter	0.01	0.51*	1.14***	0.69***	1.26***	1.53***
	(0.23)	(0.28)	(0.14)	(0.25)	(0.21)	(0.42)
Observations	3227	2913	4598	2428	5963	3377

Note: Shifters are drivers in the 2nd Houston experiment and (when noted) Boston drivers eligible to drive for Lyft. The dependent variable in Panel A is the log of hours worked; the dependent variable in Panel B is the the inverse hyperbolic sine of hours worked. All models control for the strata used for random assignment. The over-identified model uses dummies for each treatment offer. Standard errors are clustered by driver. Levels of significance: *10%, ** 5%, and *** 1%. Baseline results are presented in Table 7.

B Theoretical Appendix

B.1 Firm and Market Employment Elasticities

In Section 2 we showed how hours estimates from flexible hours environments may be biased (relative to the market-level elasticity) if a researcher only has access to data from a single firm.

In this appendix we derive the analogous result for participation elasticities. We assume that individuals can only work for one firm at a time. We use p^{prim} and p^{sec} to denote the probability of participating at the primary and secondary job. We use P to denote the probability of participating in either job (i.e. $P = p^{prim} + p^{sec}$). We define $\phi := p^{prim}/P$.

The overall participation elasticity can be written as follows:

$$\begin{aligned} \frac{\partial P}{\partial w}\frac{w}{P} &= \frac{\partial p^{prim}}{\partial w}\frac{w}{P} + \frac{\partial p^{sec}}{\partial w}\frac{w}{P} \\ &= \frac{\partial p^{prim}}{\partial w}\frac{w}{(P\phi)\,(1/\phi)} + \frac{\partial p^{sec}}{\partial w}\frac{w}{(P(1-\phi))\,/(1-\phi)} \\ &= \frac{\partial p^{prim}}{\partial w}\frac{w}{p^{prim}} \times \phi + (1-\phi)\frac{\partial p^{sec}}{\partial w}\frac{w}{p^{sec}} \end{aligned} \tag{4}$$

The last line shows the main result: the overall participation elasticity is a weighted average of the participation elasticities at each job, weighted by the probabilities the individual was working at that job initially. If the market-level elasticity is positive, the two expressions on the right do not cancel. Rather, participation at the primary firm increases both due to firm substitution and due to shifts in who is working.

B.2 Monopsonistic Wage Gaps With an Outside Firm

Suppose firms pick wages w to maximize their earnings, subject to an aggregate labor supply function for hours $L(w)$:

$$\max_{w} Y(L(w)) - wL(w).$$

The firm's profit maximizing wage is the marginal product of labor, marked down by the total hours elasticity,

$$w^* = \frac{Y'(L(w))}{1 + 1/\tau}.$$

where $\tau = \frac{d \log L(w)}{d \log w}$.[37]

In the perfectly competitive benchmark $\epsilon = \infty$ and individuals' respond completely and instantaneously to changes in relative wages. As individuals become less elastic (whether due to firm heterogeneity, concentration, or search costs), τ decreases and workers earn lower wages (Manning, 2003; Card et al., 2016). If the firm can price discriminate between two groups–say men and women–a wage gap in favor of the more elastic group will emerge.

Additional hours may come either from workers who are new to the firm or from existing employees. Suppose that, for a given wage w_t, $N(w_t)$ individuals work for the firm, providing

$$L = \int_0^{N(w_t)} h(i, w_t) di$$

[37]This is analogous to expressions used in monopoly pricing models in industrial organization, where the profit-maximizing markup depends on the inverse elasticity of demand (the "Lerner index").

hours of labor. Hours respond to wages according to

$$\frac{dL}{dw_t} = \frac{d}{dw_t}\int_0^{N(w_t)} h(i,w_t)di = \underbrace{h(N(w_t),w_t)N'(w_t)}_{\text{change in employees}} + \underbrace{\int_0^{N(w_t)} \frac{\partial}{\partial w_t}h(i,w_t)di}_{\text{change in hours}}$$

by Leibniz's rule. The first term is the change in hours that occurs because some workers join (or leave) the firm in response to the change in wages. The second term is the change in hours for workers whose firm location is unaffected by the change in wages. In elasticity terms this is

$$\frac{d\log L}{d\log w_t} = \frac{h(N(w_t),w_t)N'(w_t)}{L}w + \frac{\int_0^{N(w_t)} \frac{\partial}{\partial w_t}h(i,w_t)di}{L}w$$

This expression simplifies if we assume that, conditional on working for this firm, workers have identical preferences, i.e. $h(i,w_t) = h(w_t)$ for all i. Under this assumption, $L = N(w_t)h(w_t)$ and we can write

$$\begin{aligned} \tau = \frac{d\log L}{d\log w_t} &= \left(\frac{d\log N(w_t)}{d\log w_t} + \frac{d\log h(w_t)}{d\log w_t}\right) \\ &= \frac{N'(w_t)}{N(w_t)}w + \frac{h'(w_t)}{h(w_t)}w \\ &= \eta^{firm} + \epsilon \end{aligned}$$

In this case the relevant elasticity depends on both ϵ and the net-recruiting elasticity η^{firm}. In a setting where hours are not flexible, η^{firm} is the relevant parameter (Manning, 2003).

We can summarize the results of this section in two propositions.

Proposition 1. *If workers can flexibly choose their hours, a monopsonist would* *choose the wage gap:*

$$\frac{w_m^*}{w_w^*} = \frac{1 + 1/\tau_w}{1 + 1/\tau_m} \tag{5}$$

where τ includes intensive (hours) and extensive (firm choice) margin adjustments.

Proposition 2. *If workers cannot flexibly choose their hours, a monopsonist would choose the wage gap:*

$$\frac{w_m^*}{w_w^*} = \frac{1 + 1/\eta_w^{firm}}{1 + 1/\eta_m^{firm}} \tag{6}$$

where η reflects the change in the number of workers at the firm. Note that this could reflect both differences in retention (workers remaining at this firm instead of switching to another firm) and differences in recruitment (differences in the extent to which workers are pulled from other firms).

C Empirical Appendix

C.1 Variable Definitions

Hours A driver is considered to be working whenever their Uber app is on and they have indicated that they are available for a dispatch.

Utilization Drivers' hours can be split into three distinct periods: time waiting for a trip, time traveling to a pickup, and time on a trip. The utilization rate is the fraction of hours spent in the second two periods.

Earnings Uber distinguishes between gross earnings– which include promotional incentives—and net earnings—which subtract the amount the driver paid in Uber fees. Both of these measures are not net of costs the driver may incur, including gas or depreciation to the driver's vehicle. We focus on gross earnings. Assuming that driver costs are proportional to hours driven, this is without loss of generality.

C.2 The Earnings Accelerator

This section provides more detail on the implementation of the three Earnings Accelerator experiments.

C.2.1 Houston: Spring 2017 (Main Experiment)

In spring 2017 we conducted an Earnings Accelerator experiment in Houston, Texas. Uber launched operations in Houston in July 2013 and by the spring of 2017 had over 15,000 active drivers (drivers who had completed at least four trips in the previous month). Lyft entered the market in February 2014 but suspended operations in August 2017 after the Houston City Council passed new TNC regulations which mandated a stricter background check for drivers. They had fully withdrawn from the Houston market by November 2017. Uber remained operational in Houston despite the new regulation.

There were three phases: (1) the selection of eligible drivers, (2) "fee-free" offers, and (3) taxi offers. Drivers were eligible for inclusion in the experiment if they completed at least 4 trips in the prior month (were "active" drivers) and if their average hours per week, conditional on driving, were between 5 and 40 hours per week in the month before the experiment. Within this sample of eligible drivers, we randomly selected 2020 drivers for inclusion in the experiment within six strata defined by the interaction of hours bandwidth and gender. Houston drivers faced either a 20% or 28% commission. "Low-hours" drivers drove an average of 5-15 hours/week; "high-hours" drivers drove an average of 15-25 hours/week; "very high hours" drivers drove an average of 25-40 hours/week. We over-sampled female drivers so that we would have roughly equal proportions of male and female drivers within each hours stratum.

We offered half of the drivers in the experiment the opportunity to drive fee-free in one week, and the other half of the drivers the same opportunity the next week. Column 3 of Table A2 shows that drivers offered fee-free driving in wave 1 were statistically indistinguishable from those offered fee-free driving in wave 2. Drivers were notified about the Earnings Accelerator offer via e-mails, text message, and in-app notification. The messaging was crafted so as to mimic that used for standard Uber promotions. The in-app notification stayed at the top of each driver's Uber app for the entire opt-in period, and drivers received reminder e-mails and text messages throughout the week. Each message contained a link to a Google Form, which provided more information on the incentive. The form indicated the exact time the incentive would be active (Monday 4 A.M. for one week, following the standard Uber week) and informed the drivers that if they opted in to the Earnings

Accelerator, their data would be used by academic researchers. One thousand, three hundred and fifty-five drivers accepted our offer and were included in the third phase of the experiment.

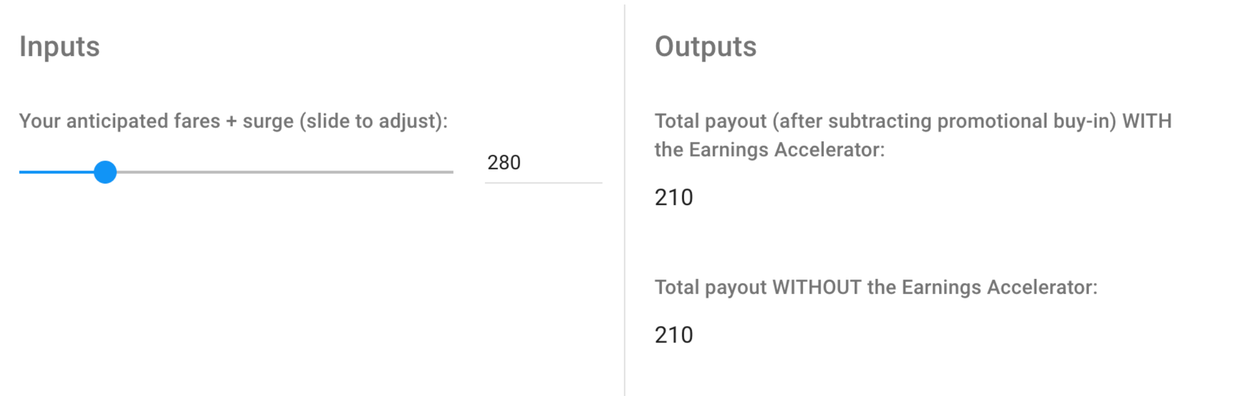
At the time we ran the experiment, drivers' trip receipts typically showed three things: the amount collected from the rider, the amount collected by Uber (due to the proportional fee), and the amount they were paid (the difference between the two). Drivers who accepted the offer of fee-free driving were able to see in-app that their fees were zero (see Figure A1 for a sample trip receipt). They also received e-mail, text message, and in-app reminders throughout the week that the "Earnings Accelerator [was] on" and that they were earning more on every trip.

The third phase of the experiment consisted of two weeks of taxi offers. The figure below (Figure C1) shows the sliders that were provided to drivers in the taxi treatments, to assist them with gauging trade-offs. Table B1 shows the taxi contracts offered in each of the two taxi weeks, along with the probability of selection and the percentage of drivers who accepted our offers. Each taxi contract provided the driver with fee-free driving in exchange for a one-time payment (listed in Columns 1 and 5). Columns 4 and 5 of Table A2 shows that the taxi treatment and control groups were balanced.

C.2.2 Houston: Fall 2017

We conducted a second Earnings Accelerator experiment in Houston, several months after Lyft re-entered the market. In May 17, 2017, the Texas State Legislature passed bill, H.B. 100, with a super-majority in the Senate (21-9), and on May 29, the Governor signed it, immediately removing mandatory fingerprinting. Lyft announced its intention to resume operations and re-entered Houston at 2 p.m. C.T. on May 31, 2017.

Figure C1: Earnings Accelerator Taxi Offer



Note: Each driver who was offered a taxi contract was sent a slider that allowed them to compare the earnings they would receive if they accepted the offer (net of the lease) to the earnings they would normally receive. The slider was set to load at the breakeven (the place where treated and untreated earnings would be identical).

Table B1: Taxi Treatments and Opt-In

	20% Fee Class			28% Fee Class		
Bandwidth (1)	Lease (2)	Treatment Fraction (3)	Opt-In Rate (4)	Lease (5)	Treatment Fraction (6)	Opt-In Rate (7)
			Week 1			
Very High	$100	60%	36%	$120	60%	36%
High	$40	60%	47%	$50	60%	46%
Low	$15	60%	47%	$15	60%	52%
			Week 2			
Very High	$65	60%	47%	$90	60%	39%
High	$35	60%	45%	$35	60%	46%
Low	$10	60%	57%	$10	60%	52%

Note: This table presents the taxi treatments offered to drivers included in the taxi phase of the main (pre-Lyft) Earnings Accelerator experiment. Only drivers who accepted fee-free driving were included. The treatment fraction refers to the fraction of consented drivers in each hours bandwidth and commission who were offered a given taxi offer. The opt-in fraction is the fraction of drivers offered the treatment who accepted it.

Drivers were eligible for inclusion in the third iteration of the Earnings Accelerator if – as before – they had completed at least four trips in the prior month (were "active" drivers), if their average hours per week, conditional on driving, were between 5 and 40 hours per week, and if they had completed a trip in Houston after Uber re-started operations following Hurricane Harvey.[38] The messaging, notifications, and timeline were similar to the first Houston experiment. The one key change was reference to Uber's fee: Uber changed its policy in June 2017 to loosen the link between rider fares and driver earnings—this was called "up front pricing"—and removed the concept of the Uber "fee". What drivers earned per trips did not change; it remained a function of a base fare plus a per-mile rate and a per-minute rate. As a result, we did not mention the "Uber fee" in the second Houston experiment. Instead we focused our messaging on the proportional increase in earnings. Column 3 of Table A3 shows that drivers were balanced across waves one and two.

We included 2100 drivers in the second Houston experiment, for which we have one week of opt-in data and one week of treatment data. During opt-in week (the week of September 18) treatment group drivers were offered one of four multipliers on total earnings at no cost (the equivalent of fee-free driving): 1.2x, 1.3x, 1.4x, and 1.5x. Those who opted-in saw the treatment in-app (as a proportional increase in the base fare, per-mile rate, and per-minute rate) the following week. The delivery of the treatment was analogous to that in the other two Earnings Accelerator experiments, and was visible on drivers' apps in real time.

After the first experiment was completed, Uber changed (without the authors' prior knowledge) the driver app, making it impossible for us to deliver treatments in

[38] We delayed the start of the experiment until the total volume of trips had rebounded to pre-hurricane levels. However, we did not want to include drivers who had stopped driving because they had been personally impacted by the hurricane.

a way that allowed drivers to see or track their higher earnings in real time. Instead, drivers would only see their non-treated earnings until the treatment was over. The change only affected our ability to show *Earnings Accelerator treatment payments* in real-time, and did not impact driver's ability to see, e.g., Surge in real-time. These weeks also occurred coincident with a change in the Uber app–which very likely affected drivers' behavior absent treatment. Because we have no experiment that compares the impact of these ex-post payments pre- and post-Lyft (and the treatments we were able to deliver after the app change were not comparable to those in the Boston or first Houston experiments), we do not include data from weeks after the app change. We view this as a conservative choice. A previous draft of this paper documented similar results when these weeks were included.

C.2.3 Boston experiment

The experiments we conducted in Houston were based on an Earnings Accelerator experiment conducted by Angrist, Caldwell, and Hall (2021) August-October 2016. As in both of the Houston experiments, there were three phases: (1) the selection of eligible drivers, (2) "fee-free" offers, and (3) taxi offers. Because the taxi offers differed from those used in Houston we only include data from the 2 weeks of "fee-free driving" offers.[39] These offers were comparable to those included in both Houston experiments.

Drivers were eligible for inclusion in the Boston experiment if they had completed at least 4 trips in July 2016 (were "active" drivers) and if their average hours per week, conditional on driving, were between 5 and 25 hours per week. Drivers were grouped into two bandwidths based on their average hours per week. "Low-

[39] One concern is that the taxi offers are only attractive to high hours drivers. As a result, the set of compliers depends on the precise parameters of the taxi offer. We find elasticities differ by usual hours worked.

hours” drivers drove an average of 5-15 hours/week and “high-hours” drivers drove an average of 15-25 hours/week. Note that unlike in the Houston experiments there was no “very high” hours group.

1600 eligible were selected for inclusion in the experiment, within strata defined by (1) average hours driven in July, (2) driver fee class (commission rate), and (3) vehicle model year. All 1600 of these drivers were offered one week of fee-free driving, half in one week (wave 1), and half in the next (wave 2). Column 3 of Table A3 shows that drivers offered fee-free driving in wave 1 were statistically indistinguishable from those offered fee-free driving in wave 2. The lower proportion of female drivers in Boston reflects the fact that the authors of Angrist, Caldwell, and Hall (2021) did not over-sample female drivers.

D Additional Results and Discussion

D.1 Driver Opt-In Decisions

The offer of fee-free driving was generous, both in absolute terms, and relative to standard Uber promotions. However, not everyone accepted our offer. Opt-in rates were extremely high, relative to normal Uber promotions. Conversations with local city teams suggest that drivers sometimes ignore Uber's promotional messaging.

Figure A6 plots the opt-in rate by the number of hours a driver worked during the week in which they received the treatment offers ("opt-in week"). This figure shows that, for both male and female drivers, drivers who drove during opt-in week were about thirty percentage points more likely to accept our offer. This is not surprising because one of the ways we delivered the treatment offer was via in-app message. Drivers who did not open the driver app during opt-in week would not have seen this message.

The fact that there is no relationship between hours worked and opt-in decisions—conditional on positive hours worked—is consistent with the idea that drivers did not decide whether to accept the offer based on the total (rather than proportional) payoff they expected. Rather, drivers who drove during opt-in week were more likely to accept the offer because they saw the in-app promotional messaging.

We do not have data on which link (the e-mail text message or Uber app) a driver clicked on in order to access the offer acceptance form.

D.2 First Stage Impact on Hourly Earnings and the Net-of-Commission Rate

Drivers' hourly earnings depend both on what they collect in trip receipts and on the commission they face. Our treatment increased drivers' hourly earnings by giving them the opportunity to face a lower commission on each trip or by giving them a proportional "bonus" on each trip. This is analogous to lowering the tax rate on drivers' income. Drivers' realized hourly earnings, W, depends both on their hourly trip receipts w and on the commission they face, τ: $W = (1 - \tau) \times w$. Drivers' hourly earnings can change due to changes in either changes in the commission rate or changes in hourly trip receipts:

$$\Delta \log W \quad = \quad \Delta \log(1 - \tau) + \Delta \log w$$

While our treatment targets per-trip earnings, average hourly earnings may change if marginal hours are less profitable than average hours.

The impact on wages depends on both the impact of the treatment on treated drivers and on experimental take-up. Column 1 of Table 3 shows that roughly 60% of male drivers accepted the offer, relative to roughly 70% of female drivers. These opt-in rates are high relative to those for typical Uber promotions.

Female drivers in our sample faced higher commissions on average because they had joined the platform more recently. This led female drivers to receive more generous offers, on average. Column 3 of Table 3 shows that the average female driver was offered a .31 log point increase in hourly earnings, compared with .28 for male drivers. Table D1 shows that differences in treatment generosity do not explain gender differences in opt-in rates; there is no within-gender difference in opt-in rates between drivers who faced different commission rates.

Table D1: Earnings Accelerator Opt-In Rates by Gender and Commission

	Full Sample	By Commission Rate 20%	By Commission Rate 28%
	(1)	(2)	(3)
Male	0.61***	0.61***	0.61***
	(0.016)	(0.023)	(0.022)
	1944	930	1014
Female	0.73***	0.76***	0.72***
	(0.014)	(0.031)	(0.015)
	2096	386	1710

Note: This table presents opt-in rates for the main (pre-Lyft) Houston experiment, separately by driver gender and commission rate. Within each sample we regress an indicator for whether the driver accepted the offer on an indicator for whether the driver was treated (i.e. given an offer) and on the strata used for random assignment. Each cell presents the coefficient on treatment, the standard errors (clustered by driver) in parentheses, and the sample size. The p-value for equality between the opt-in rates for male (female) drivers of different commission rates is .93 (.22). Levels of significance: *10%, ** 5%, and *** 1%.

The impact of the treatment on the log net-of-commission rate is simply the average offer, multiplied by the experimental opt-in rate. Column 5 of Table 3 shows that the average impact on the log net-of-commission rate was .22 for female drivers, relative to .17 for male drivers. The impact on log hourly earnings (column 7) is somewhat smaller, though we cannot detect a significant difference between the two specifications. The smaller coefficient likely reflects differences in the productivity of marginal hours. However the fact that the difference between columns 5 and 7 is small suggests that marginal hours were not significantly less productive than average hours.

D.3 External Validity and Drivers' Outside Options

We use data from both the American Community Survey and the Current Population Survey to shed light on whether male and female drivers differ in outside work arrangements.

First, in Table D2 we use data from the 2017-2019 American Community Surveys to compare the patterns of hours worked among male and female drivers and male and female taxi drivers in Houston. Columns 1 and 2 use data on the pool of active Houston Uber drivers. Columns 3 and 4 focus on Houston drivers (occupation codes 9141 and 9142) and Columns 5 and 6 focus on Houston taxi driver (occupation code 9142). This table shows that, among Houston drivers and Houston taxi drivers, there is also a gender gap in hours worked. This is true when both looking at means and when examining the proportion who work more than 35 or more than 45 hours per week.[40]

We next use data from the 2010-2018 Current Population surveys to examine the characteristics of male and female taxi drivers in the Houston metro area (Flood et al., 2021). Following Cook et al. (2021), we identify taxi drivers using occupation code 9140. We find that approximately twenty percent of the drivers are female, similar to the fraction among Houston Uber drivers. The average Houston driver works full time: 35 hours a week. However, female drivers work fewer hours per week (first row). They are much less likely to work more than twenty-five hours per week (second row). Women also work a larger share of their hours at their main job (third row).

D.4 Impact on the Distribution of Hours Worked

Our treatment increased workers' hourly earnings by reducing the implicit tax rate they face. Panel A of Figure A7 shows that drivers who received the treatment offers drove more than drivers who did not receive the offer. The distribution of hours worked for treated lines (solid line) lies to the right of that for untreated

[40]Uber does not collect data on driver education. Previous research has documented that the average ride-share driver is more educated than the average worker: roughly a third have college degrees, compared with a quarter among workers in the same markets (Hall and Krueger, 2018).

Table D2: Comparing the RCT Sample to Houston Taxi Drivers

	Houston Uber Drivers		Houston Drivers (ACS)		Houston Taxi Drivers (ACS)	
	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Age	43.98	44.27	44.81	42.47	43.62	42.52
High School or More	---	---	82%	88%	92%	90%
Some College or More	---	---	41%	48%	63%	68%
Married	---	---	56%	41%	48%	37%
Income	---	---	$46,961	$21,116	$36,660	$18,314
Household Income	---	---	$91,523	$79,832	$78,861	$93,083
Usual Hours Worked						
Mean	3.66	3.22	46.02	34.37	42.14	34.80
Between 0 and 5	12.67	9.81	0.00	0.01	0.00	0.01
Between 5 and 15	10.03	8.58	0.03	0.08	0.07	0.13
Between 15 and 25	2012.97	2013.39	0.04	0.13	0.06	0.20
Between 25 and 35	3.12	2.60	0.07	0.18	0.11	0.07
Between 35 and 45	18.92	10.32	0.41	0.42	0.41	0.28
Over 45	22.67	11.26	0.45	0.18	0.36	0.30

Note: This table compares the mean characteristics of Houston Uber drivers to drivers in the Houston metropolitan area. We use pooled data from the 2017-2019 American Community Survey. Columns 3 and 4 present characteristics for individuals in driving occupations (occupation codes 9141 and 9142). Columns 5 and 6 present characteristics for taxi drivers (occupation code 9142). We compute usual hours worked for Uber drivers using pre-Earnings Accelerator data.

Table D3: Main Job Hours and Total Hours Among Houston Drivers

	(1)	(2)	(3)	(4)
Total Hours Worked	-1.92	-3.06	-2.86	---
	(1.92)	(1.93)	(1.92)	---
Work >= 25 Hours	-0.08	-0.11 *	-0.11 *	-0.06
	(0.06)	(0.06)	(0.06)	(0.04)
Main Job/Total	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Education Dummies		✓	✓	✓
Marital Status			✓	✓
Total Hours Worked				✓

Note: This table uses data from the 2010-2018 Current Population Surveys to examine hours worked among Houston metro area (IPUMS metro area code 3362) taxi drivers (IPUMS occupation code 9140). Each row shows the coefficient on a female dummy from a regression of the characteristic indicated in the row on a female dummy and on the characteristics indicated at the bottom of the table. Regressions are weighted using CPS sampling weights. Levels of significance: *10%, ** 5%, and *** 1%.

drivers (dashed line). There is a small change in the probability a worker drives (gap between the solid and dashed lines at zero hours). The shift in the hours distribution is larger for female drivers than for male drivers.

Panel B of Figure A7 shows a similar pattern emerges among the treated and untreated compliers. We estimate complier hours distributions using regressions of the form

$$
\begin{aligned}
1[h_{it} < v](1 - D_{it}) &= X_i'\beta_0(v) + \alpha_0(v)(1 - D_{it}) + u_{0iv} \\
1[h_{it} < v]D_i &= X_i'\beta_1(v) + \alpha_1(v)D_{it} + u_{1iv},
\end{aligned}
$$

where the outcome variable is an indicator for whether the driver drove fewer than v hours that week ($1[h_{it} < v]$), multiplied by the driver's opt-in decision (D_{it}) or $(1 - D_{it})$ (Abadie 2003). Because opt-in is endogenous, we instrument for $(1 - D_{it})$

and D_{it} using the randomly assigned treatment offers Z_{it}. Panel B plots estimates of $\alpha_0(v)$ and $\alpha_1(v)$. Panel B shows that the treatment also shifts the distribution of compliers' hours worked outward. The shift is larger for female drivers than for male drivers.

D.5 Robustness to Potential Measurement Error in Hours

Our hours measure captures the time a driver has their app turned on. This captures both the time a driver spends waiting for dispatch, the time a driver spends en route to a trip, and the time a driver spends with a passenger in the car. Drivers are not directly compensated for each of these windows: they receive no money while they are waiting for a trip. To address potential mis-measurement of hours, we repeat all of our results with drivers' fareboxes. Because this outcome is used to pay drivers, it likely has minimal measurement error. Table A7 shows that we obtain similar results when using log farebox instead of log hours at the outcome variable.

D.6 Impact on When Drivers Work

D.6.1 Hours Worked Before and After Treatment

A natural question is whether treated drivers reduce their hours worked in the week before or after treatment in order to shift hours into higher-earnings weeks. Appendix Table A4 shows that the treatment offers did not influence drivers' labor supply before or after treatment, consistent with there being no wealth effects. Angrist, Caldwell, and Hall (2021) report similar results based on the Boston Earnings Accelerator experiment. They also document that the treatment did not influence passengers' star ratings, a measure of trip quality.

In results not reported we used alternate control groups to verify the robustness of the results in Table A4. In particular, we obtain virtually identical results when we use drivers' log farebox instead of drivers' log wages as the dependent variable.

D.6.2 Time of Day and Day of Week

Because Uber drivers can drive at any time of day (and any day of the week), it is interesting to examine where the marginal hours of work come from, separately for male and female drivers. Prior work by Cook et al. (2021) documented that there are significant differences in the normal hours worked by gender. Figure D1 below shows the impact of the treatment on whether a driver works at each hour of the day (Panel A), and how much he/she works at that hour (Panel B). We estimate these treatment effects using the following two-stage least squares setup:

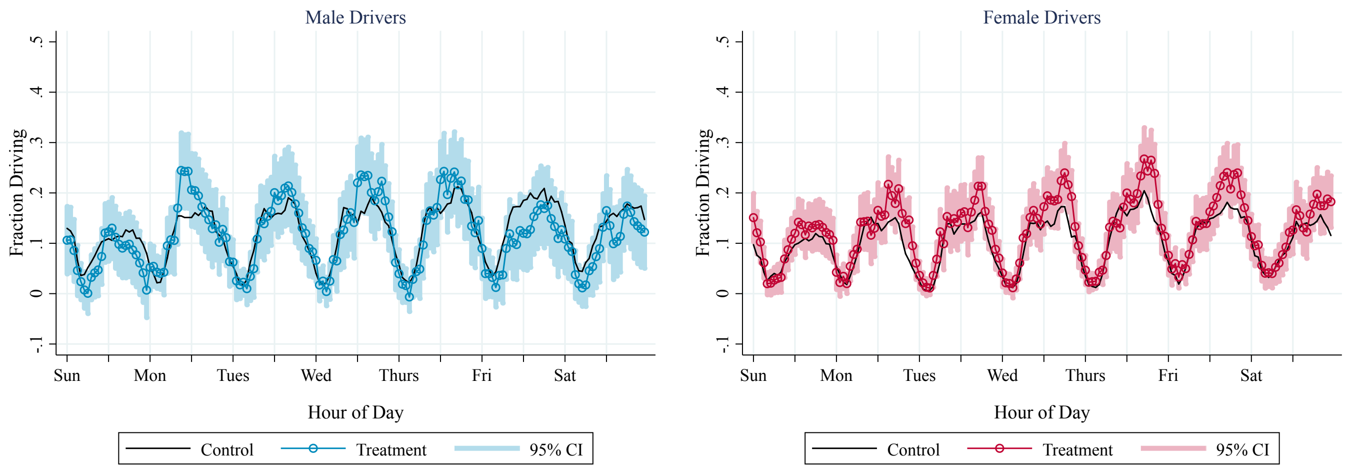
$$\begin{aligned} y_{it} &= D_{it} + \beta X_{it} + \eta_{it}, \qquad (7) \\ D_{it} &= \gamma Z_{it} + \lambda X_{it} + \upsilon_{it}, \end{aligned}$$

where y_{it} is either an indicator for whether the driver is active or the total number of minutes worked that hour, D_{it} is an indicator for whether the driver accepted the treatment offer, and Z_{it} is an indicator for whether the driver was offered the treatment. The X_{it} include the strata used for random assignment.

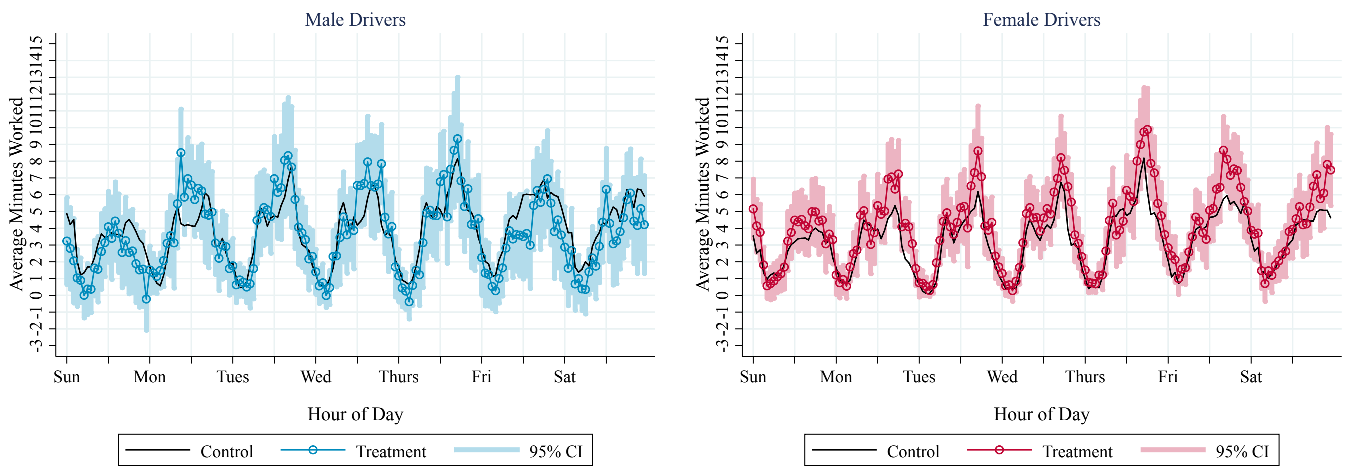
The control means for male and female drivers show a regular pattern: more drivers are active during the day than at night. There is a regular peak at the time most workers end their full-time jobs. Women are less likely to drive late at night. The round circles plot the control mean plus the treatment effect for each hour. The only times when the (shaded) 95% confidence interval does not intersect the control mean are week-day afternoons (especially for female drivers). Overall, the fact that the 95% confidence intervals intersect the control mean for most outcomes, indicates there is no systematic pattern in *when* drivers drive additional hours.

Figure D1: Treatment Effects by Time of Day

Panel A: Effect on Driving



Panel B: Effect on Minutes Active



Note: This figure shows the impact of the treatment on when drivers work. The outcomes in Panel A are indicators for whether a driver is driving at a given hour of the day; the outcomes in Panel B are average minutes worked by hour. The black lines in each figure denote the control means. The dark lines (red/blue) show the control mean + treatment effect. The treatment effects come from regressions of the outcome variable on an indicator for whether the driver is treated. All regressions control for the strata used for random assignment. 95% confidence intervals for each regression are shaded.