

Reservation Wages and Workers' Valuation of Job Flexibility: Evidence from a Natural Field Experiment*

Kuan-Ming Chen[†]

Ning Ding[‡]

John List[§]

Magne Mogstad[¶]

September 2, 2020

Abstract: Recent changes in labor arrangements have increased interest in estimating and understanding the value of job flexibility. We leverage a large natural field experiment at Uber to create exogenous variation in expected market wages across individuals and over time. Combining this experiment with high frequency panel data on wages and individual work decisions, we document how labor supply responds to exogenous changes in expected market wages in a setting with virtually no restrictions on driver labor allocation. We find that there is i) systematic heterogeneity in labor supply responses both across drivers and within a driver over time, ii) significant fixed costs of beginning a shift, and iii) high rider demand when it is costly for drivers to work. These three findings motivate a model of labor supply with heterogeneous preferences over work schedules, adjustment costs, and statistical dependence between market wages and the costs of driving. We recover the labor supply elasticities and reservation wages of this dynamic labor supply model via a combination of experimental estimates and other data moments. We then perform counterfactual analyses that allow us to examine how preference heterogeneity and adjustment costs influence the responses of workers' to wage incentives as well as infer drivers' willingness to pay for the ability to customize and adjust their work schedule. We also show that a static approach to the driver's dynamic problem delivers materially different estimates of workers' labor supply elasticities and their value of job flexibility.

Keywords: Adjustment costs, dynamic model of labor supply, job flexibility, labor supply elasticities, reservation wages, value of time

*The opinions expressed in this paper are those of the authors alone and do not reflect the views of Uber. We appreciate the constructive comments and suggestions from Oystein Daljord, Jonathan Hall, Vishal Kamat, Libby Mishkin, Jesse Shapiro, Winnie van Dijk, and Kerry Smith.

[†]Department of Economics, University of Chicago. E-mail: kmchen@uchicago.edu

[‡]Department of Economics, University of Chicago. E-mail: nding@uchicago.edu

[§]Department of Economics, University of Chicago, NBER. E-mail: jlist@uchicago.edu

[¶]Department of Economics, University of Chicago; Statistics Norway; NBER; IFS. E-mail: magne.mogstad@gmail.com

1 Introduction

The last decade has witnessed a dramatic increase in the prevalence of flexible working, either via workers entering the Gig Economy or historically traditional jobs becoming more flexible, allowing the worker to choose specific hours or where to work. This phenomenon is not new, as people have been choosing time and place of work for centuries. For example, the Medieval years saw many of the working-class live where they worked. Operating businesses out of their residences was commonplace for the weaver, silversmith, tailor, seamstress, and shoemaker. Not until the Renaissance would administrative offices begin as the modal place of work, and the Industrial Revolution saw an even greater exchange of home for office for the typical worker.

The roots of the modern movement back to flexible working arrangements can be traced to the 1930s, when Kellogg’s deviated from the standard 8 hours per day, five days a week work week in the 1930s (Hunnicut, 1996). The movement was temporarily sidetracked due to Presidential decree during the war, but it would soon return: by 1972 Hewlett Packard offered flexible working arrangements at its Waltham MA plant. By 1978, West Germany had adopted the term “flex time,” effectively establishing policies to balance work life and family life (Santovec, 2014).

We are perhaps witnessing the next major chapter in the work/home relationship due to health concerns and growth in the Gig Economy, where both place and time of work is more flexible.¹ With this movement arise several first order questions of both policy and practical import. How do labor supply elasticities and reservation wages vary across days of the week and hours of the day? To what extent do labor supply elasticities and reservation wages differ between people such as men and women or old and young? While both economists and policymakers are keenly interested in these questions, credible answers have been hindered by a lack of high frequency panel data on wages and work decisions as well as by the difficulty of identifying how labor supply elasticities and reservation wages vary between people and over time.

The goal of our paper is to answer the above questions while addressing both the measurement and the identification challenges. The context of our study is the largest ride-sharing company, Uber. Our work draws on three strengths of this environment.² First, Uber allows a driver to work anytime she is willing to accept the wage she would be paid in the market. Second, we have access to high frequency panel data on the wage an individual is paid and her decision to work.³ Third, via a large natural field experiment we observe reactions to exogenous variation in expected market wages across individuals and over time.

Combining the panel data with the experiment, we first estimate individuals’ labor supply responses to exogenous changes in expected market wages. These experimental findings motivate and guide our modeling of the labor supply of the drivers. The primitives of the model are recovered from a combination of the experimental estimates and other data moments. We use the estimated model to compute how labor supply elasticities and reservation wages vary between people and over time and to perform counterfactual analyses. These analyses allow us to infer the drivers’ willingness to pay for flexible work arrangements, and to examine

¹According to the survey data of Katz and Krueger (2019), 8.4 percent of American workers participate in independent contractor work as their primary job, a 22% increase over the last decade. Oyer (2016) estimates that nearly one third of all U.S. workers participate in independent work as a primary or secondary activity.

²Hall and Krueger (2018) describe the labor market for Uber’s drivers. They find that drivers cite flexibility as a reason for working for Uber and that many drivers report that Uber is a part-time activity secondary to more traditional employment.

³As in Cook et al. (2019) and Chen et al. (2019), we calculate the “wage” in an hour as a driver’s total earnings in that hour divided by minutes worked (i.e. the number of minutes for which a driver has the app on and is available for accepting requests).

how preference heterogeneity and adjustment costs influence the effectiveness of driver incentives that Uber may offer.

In Section 2, we describe the labor market for Uber’s drivers and the natural field experiment. The analyses of the experiment yields three main findings, which we present in Section 3. The first main finding is that the labor supply responses vary systematically both across people and over time. In order to discover these heterogeneous effects, we apply the method of Chernozhukov et al. (2018) to estimate an instrumental variables model with a full set of interactions between the endogenous regressor of interest, wages, and the pre-determined covariates, gender, hours of the day, and days of the week. A clear pattern of heterogeneity emerges: Labor supply is most responsive in the evenings; men and older drivers have, on average, larger responses than other drivers. Taken together, these results suggest it is key to allow preferences of the drivers to vary by gender and age for each demographic group across hours of the day and days of the week.

The second main finding is that drivers do not only increase their labor supply during the periods with exogenously higher wages but also in the hours preceding and following these periods. This finding of anticipatory and persistent responses to increases in expected wages is consistent with forward looking drivers with fixed costs of starting to drive. In the presence of such adjustment costs, a static labor supply model is insufficient to analyze the behavior of the drivers. A dynamic model is needed to capture the connection between the decision to drive in the current period and future utility.

The third main finding is that unobserved determinants of wages, if ignored, lead to a significant downward bias in the estimated labor supply responses. In particular, OLS estimates show much weaker associations between labor supply and wages than the experimental estimates. This downward bias is consistent with demand being high when it is costly for the drivers to work. Including fixed effects for workers, days of the week, and hours of the day reduces the bias, but the labor supply elasticities remain too small. This finding suggests that idiosyncratic factors, such as weather conditions and entertainment events, may create high demand while, at the same time, make driving more costly or difficult.

The experimental estimates provide key data points for recovering reservation wages, labor supply elasticities and the value of job flexibility, but do not by themselves tell us these quantities. To do so, we develop, in Section 4, a dynamic model of labor supply. This model builds on the experimental findings and accommodates important features of the market, including uncertainty about wages and costs of driving in the future, the possibility of a job other than driving for Uber, fixed costs of starting to drive, and substitution possibilities to the key competitor, Lyft. When taking the model to the data, we allow for both observed and unobserved heterogeneity across drivers, and we allow market wages to be correlated with the cost of driving in a given period. Even with these considerations, it is possible to prove identification of the primitives of the model given the panel data and the experiment that creates exogenous variation in expected market wages.

We use the EM algorithm to find the maximum likelihood estimates of the model parameters. The parameter estimates suggest significant costs of starting to drive and considerable observed and unobserved heterogeneity in preferences. Conditional on age and gender, there appears to be three types of drivers: The ‘infrequent driver’ who only drives occasionally; the ‘full-time driver’ who drives regularly both in the evening and during the day; and ‘the evening driver’ who rarely drives during the day, possibly because she has a daytime job. To assess the importance of substitution between Lyft and Uber, we compare the results for all drivers to those we obtain from a subsample of drivers who are ineligible to drive for Lyft. It is reassuring

to find that both the experimental estimates and the estimated model parameters do not materially change when we restrict attention to this subsample.

The model delivers three key insights, presented in Section 5. The first insight from the model is that reservation wages, and thus the shadow prices of time, vary a lot both over time and across people.⁴ For an average driver at a typical day, the reservation wage is relatively low during the day, starts increasing in the late evening, peaks at around 4 a.m., and then declines gradually until 9 a.m.. By way of comparison, there is little variation in reservation wages across days of the week: On average, the reservation wage is only a few percent lower on weekdays than during the weekends. Holding day of the week and hour of the day fixed, there is also a great deal of variation in reservation wages across people. On average, reservation wages of women are 106 percent higher than male reservation wages. There is also a great deal of heterogeneity conditional on observables: The infrequent drivers have much higher reservation wages than the full-time drivers, while the evening drivers demand relatively high wages to drive during the day.

The second insight from the model is that drivers would demand much higher wages if they had to commit to pre-set work schedules. We quantify the importance of two distinct types of job flexibility. One is the ability to set a customized work schedule, so that each driver may plan to work only when her expected reservation wage is lower than the expected wages. We quantify the value of this type of flexibility by removing certain hours of the day or days of the week from the choice set of the workers. Our findings suggest that drivers are particularly averse to restrictions on what hours of the day to work. By way of comparison, constraining drivers to work only on the weekends or only on weekdays would require a modest increase in wages. The other type of flexibility we consider is the ability to adjust the schedule from day to day or even hour to hour in response to unexpected changes to offered wages or costs of driving. We measure the value of this flexibility by restricting drivers to stick to the work schedule they prefer before observing any shocks to wages and preferences. Our findings suggest that Uber drivers, especially those who are female, benefit significantly from the possibility to adapt work schedules to unexpected events. Taken together, these results suggest that job flexibility is a central component of the total compensation of ride-sharing companies like Uber.

The third insight from the model is that preference heterogeneity and adjustment costs are essential to accurately predict the effectiveness of work incentives to drivers. Ride-sharing companies are regularly using driver incentives to influence when and how much drivers work. Our findings suggest that the influence of such incentives are grossly overstated in a static labor supply model. The reason is that it is costly to start driving and, therefore, it is much cheaper to induce drivers to continue to drive for a few more hours than to induce drivers to start a shift. It is also cheaper to incentivize drivers to increase their labor supply during the day than at night and on Fridays and Saturdays than on other days of the week.

Our paper is primarily related to a large literature on labor supply. The models, data, and findings have been summarized and critiqued in multiple review articles including Pencavel (1986), Killingsworth and Heckman (1986), Blundell and MaCurdy (1999), Keane (2011) and Chetty et al. (2011). Most models of labor supply are concerned with the problem of choosing how much to work, not when to work. In many of these models, there are no hours restrictions, and each worker supplies labor until the wage she would face in the market equals the value she places on her time, the reservation wage. When taking such models

⁴It is important to recognize that a driver's reservation wage should be interpreted as a shadow price of time that reflects not only leisure possibilities but also alternative economic activities such as home production or other jobs.

to the data, labor supply elasticities and reservation wages are typically inferred from differences in work hours across people given their observed wages. There are, however, several concerns with this revealed preference argument. One of these concerns is that both theory and evidence suggest restrictions on hours choices stemming from the demand side of the market. This concern motivates a large body of work that incorporates hours restrictions in models of labor supply under the assumption that the analyst has full or partial knowledge about the probability distributions of either offered or desired hours of work.⁵

To avoid making questionable assumptions about hours restrictions, we take advantage of the fact that Uber is a platform on which drivers, once approved, are free to choose their work hours. There are no minimum-hours requirements and only modest constraints on maximum hours. As a result, our estimates of extensive margin labor supply elasticities and reservation wages are not confounded by hours restrictions from the demand side of the market. Instead, the estimated elasticities capture the sensitivity of the decision to supply labor in a given hour to anticipated and exogenous changes in hourly market wages. The wage changes we consider are modest and temporary, so that lifetime wealth is approximately unchanged. Thus, the elasticities we report can arguably be interpreted as approximating extensive margin Frisch elasticities.⁶ On average over time and across driver, we find an extensive margin elasticity of 0.46. This estimate is significantly larger than what is typically reported in micro studies that ignore or make assumptions about hours restrictions from the demand side of the market (Chetty et al., 2011).

The closest study to ours is the important work of Chen et al. (2019). Like us, they take advantage of the fact that Uber has virtually no hours restrictions.⁷ Thus, Chen et al. (2019) argue, one can recover how reservation wages vary across people and time by relating the probability an individual drives in a given time period to the mean prevailing market wage for that period. Using a multivariate probit model with time-varying thresholds for work decisions, they estimate driver-specific reservation wages, and then decompose these reservation wages into predictable and unpredictable components. Armed with the estimates from this static labor supply model, they calculate the surplus from driving for Uber and the surplus changes that would result from requiring the driver to instead work specific patterns of hours.

Our paper complements and extends the model and analyses of Chen et al. (2019) in several important ways. First, motivated by our natural field experimental results, we develop, identify, and estimate a dynamic model of labor supply with fixed costs of starting a shift. Second, we allow market wages be correlated with the unobserved cost of driving in a given period. Third, we use a natural field experiment to identify the primitives of the model. Fourth, we allow permanent heterogeneity both by the drivers' observable characteristics and according to their unobserved latent types. Empirically, we find that these modeling choices are important to match the data as well as for the estimated reservation wages and the counterfactual analyses. Our paper also offers a complementary perspective on the heterogeneity in reservation wages. Chen et al. (2019) model and estimate the heterogeneity in reservation wages as arising from idiosyncratic preferences. We show there is a systematic and predictable pattern in the reservation wages by not only the day of week or hour of the

⁵See Beffy et al. (2019) and the references therein for details.

⁶See Attanasio et al. (2018) for a discussion of how extensive margin Frisch labor supply elasticities can be defined and estimated.

⁷There are also other papers using data from Uber. Angrist et al. (2017) study how workers' view the commission-based compensation model of Uber as compared to traditional taxi compensation contract. Cohen et al. (2016) estimate consumers' demand and surplus from Uber rides. Cook et al. (ming) study the determinants of the gender earnings gap amongst Uber drivers.

day but also according to the gender, age, and type of driver. This pattern is useful to better understand who benefits from flexible work arrangements, and, as a result, it may also help improve the design of driver incentives and inform discussions over recent policy proposals about regulation and pay rules for ride-sharing companies.

Our paper also relates to a body of work on the labor supply of taxi drivers. The primary goal of this work is to estimate the wage elasticity of daily hours of work to test if labor supply behavior is consistent with reference dependence. The work is summarized and critiqued in Farber (2015). He also replicates and extends existing work. His findings suggest that reference dependence is not an important factor in the daily labor supply decisions of taxi drivers.

Some of our findings are similar to those reported in Farber (2015). For example, much of the variation in hourly wages is predictable based on the day of the week and the hour of the day, and drivers are more likely to work when market wages are high. Other findings differ. For instance, Farber (2015) finds that the probability of ending a shift depends strongly on hours worked. We do not find support for such fatigue being empirically important for the behavior of Uber drivers. However, the environment and decision problem of Uber drivers differ in important ways as compared to taxi drivers. In particular, accumulated hours worked in a given day tend to be a lot higher for taxi drivers, and, as a result, fatigue could be more salient for whether they continue driving or end a shift. By comparison, the labor supply of Uber drivers is best described by a combination of adjustment costs in terms of starting to drive and heterogeneous reservation wages, especially by hour of the day and type of driver.

Another literature to which we relate is the research on how individuals value workplace amenities such as job flexibility. Survey evidence shows that workers state that they are willing to take lower pay for more flexible jobs (e.g., Hamermesh (1999); Pierce (2001); Eriksson and Kristensen (2014); Maestas et al. (2018)). However, recovering the workers' actual valuation of job flexibility from naturally occurring data has proven difficult for several reasons. One challenge is that firms may pay differently simply because they employ workers of different quality. A second challenge is that observed wage variation across firms may reflect workplace amenities other than job flexibility. Most research to date tries to address these issues by controlling for worker and firm characteristics, hoping that any remaining wage variation across firms is due to job flexibility.⁸

Even if these controls were sufficient to address concerns about omitted variables bias, it is important to observe that additional assumptions or data are needed to draw inference about workers' valuation of job flexibility. Wage differentials across firms could reflect imperfect competition in the labor market, not workplace amenities. Additionally, in standard models of equalizing differences, such as Rosen (1986), the observed wage differentials are the market prices of amenities, providing only information on the valuation of marginal workers. Lamadon et al. (2019) develop, identify and estimate an equilibrium model of the U.S. labor market with two-sided heterogeneity where workers view firms as imperfect substitutes because of heterogeneous preferences over workplace amenities. The estimated model makes it possible to distinguish between and draw inference about imperfect competition, compensating differentials, and the distribution of worker preferences over amenities. The empirical findings suggest one needs to be cautious in extrapolating the valuation of amenities among marginal workers, as measured by the compensating differentials, to the

⁸Mas and Pallais (2017), Chen et al. (2019) and He et al. (2019) review this literature.

valuation of inframarginal workers, who extract a significant amount of surplus or rents from workplace amenities. The importance of worker heterogeneity in the value of amenities like job flexibility is consistent with both our findings and those in Chen et al. (2019).

Our findings on job flexibility complement recent evidence that uses a stated preference approach to infer workers’ preferences based on their choices between pairs of exogenously assigned hypothetical jobs with different combinations of amenity levels and pay (e.g., Mas and Pallais, 2017; Wiswall and Zafar, 2017). For example, Mas and Pallais (2017) use a discrete choice experiment in hiring for a U.S. call center to estimate the willingness to pay for alternative work arrangements relative to traditional office positions. A significant number of workers state that they are willing to give up a substantial share of their wages to avoid a schedule set by an employer on short notice. By comparison, the stated willingness to pay for choosing when to work is relatively low.

He et al. (2019) provide complementary evidence from a revealed preferences approach to estimating workers’ valuation of flexibility. They combine data from a natural field experiment conducted on a Chinese job board with survey and observational data. The experimental job ads differ randomly in offering jobs that are flexible regarding when and where one works. Both the survey evidence and the experimental estimates suggest that workers are willing to take lower pay for more flexible jobs. For instance, application rates are significantly higher for flexible jobs, conditional on the salary offered. He et al. (2019) argue that a natural field experiment offering real jobs to real job seekers has several advantages over alternative approaches.⁹ The participants in the natural field experiment are actually searching for jobs, properly incentivized to respond in ways most likely to get them the jobs they want, and unaware they were under scrutiny in a scientific study. The natural field experiment that we study have the same advantages. In addition, we can measure the participants’ valuation of flexibility in terms of their actual work decisions, unlike the job board experiment that does not capture the final outcomes of the search process (such as callbacks for interviews, job offers, and actual remuneration).

Finally, methodologically, we join a set of recent studies that combines field experiments with structural methods to uncover key counterfactuals (see, e.g., DellaVigna et al., 2012; 2016). In doing so, we highlight how the combination of theory and field experiments can be used to evaluate a wide range of economic issues (see also Laibson and List (2015)).

2 Background and Experiment

We now review the labor market for Uber’s drivers before describing the natural field experiment.

2.1 Uber Marketplace

Uber’s rideshare platform is the largest service provider in the ride-sharing market in the U.S. In 2016, for example, it had a market share of about 83% of the U.S. consumer ride-sharing market. Uber and Lyft combined owned nearly 97 percent of this market. Uber connects riders and drivers through its app. Once a ride request is made, the app contacts nearby drivers for the ride. Drivers would see the rider’s location.

⁹See Harrison and List (2004) for a broader discussion of the advantages of natural field experiments.

While drivers are incentivized to maintain a high acceptance rate, drivers can decide whether to accept this trip.

Drivers are effectively free to choose when and how much to work. There are no minimum-hours requirements and only modest constraints on maximum hours. Drivers are paid according to a fixed, non-negotiated formula. As described in detail later, workers earn a base fare per trip plus amounts for how long and how far they drive. On top of this standard fare, Uber offers fare multipliers when demand for rides is sufficiently high compared to the supply of drivers (surge hereafter) or if the drivers are participating in the randomized experiment. Both the variation in surge multipliers and the randomness in the arrival of rider requests lead to variability in effective market wages per hour.

In most cities in the U.S., there are relatively few barriers to becoming an Uber driver. While the exact requirements vary from city to city, drivers must typically fill out online paperwork, undergo a background check, and meet certain driver and vehicle requirements. In the years (2016-2018) and cities (Boston, Chicago, San Francisco) we consider, one of these requirements is that the vehicle’s model year is 2001 or newer. By comparison, Lyft required that the vehicle must have a year model of 2003 or newer. As a robustness check, we take advantage of this difference in eligibility requirements to assess the importance of substitution between Lyft and Uber.

2.2 Experiment Setup: Guaranteed Surge Level

Our natural field experiments arises from the so-called Guaranteed Surge Levels (GSL hereafter). The GSL is essentially a fare multiplier that Uber randomly offers to a subset of drivers to increase their expected market wages during certain hours. Drivers who were active in the past 28 days and have completed at least 40 trips are eligible to receive fare multipliers through GSL.

The experimental setup is as follows: Uber divides each week into 2 blocks: Block 1 starts from Monday 4:00 a.m. and ends on Friday 3:59 a.m. while Block 2 goes from Friday 4:00 a.m. to 3:59 a.m. the next Monday. Over the course of a given block, GSL is switched on for a subset of hours. In Figure 1, we show an example of Block 1. During the example block, the highlighted hours, such as Monday 5:00–6:00 a.m., are chosen as experiment hours where drivers receive hour-specific fare multipliers. We refer to consecutive experiment hours as an experiment window. We refer to the schedule of GSL experiment hours within a block as a GSL menu. These menus vary across blocks. Drivers learn about the GSL menu, via email and/or the Uber app, the night before a block starts. Figure 2 shows an example of a GSL menu with multiple experiment windows, making clear to drivers when and for how long GSL will be switched on in an upcoming block.

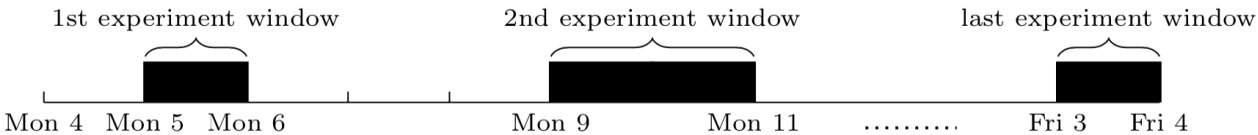


Figure 1: Example of a GSL Menu of Block 1 with Multiple Experiment Windows

Notes: Highlighted hours denote the hours when the GSL experiment is switched on.

For each block, eligible drivers are randomly assigned to treatment and control groups where the treatment

group receives 0.1 higher GSL fare multipliers than the control group for all the experiment hours within the block. Consider again the example in Figure 1. Monday 5:00–6:00 a.m. and Monday 9:00–11:00 a.m. are two experiment windows. Suppose the control group drivers receive $1.1\times$ fare multiplier in the first window and $1.3\times$ fare multiplier in the second window. Since the treatment group always receives 0.1 higher GSL fare multipliers than the control group, the treated drivers would then be receiving $1.2\times$ fare multiplier in the first window and $1.4\times$ fare multiplier in the second window. At the end of each block, drivers are re-randomized into treatment and control groups for the next block.

On average, an experiment window lasts for about 5 hours, and there are about 7 experiment windows per block. Across blocks, there is variation in the days and hours of the experiment windows. In total, around 40 percent of the hours in our sample are subject to the GSL experiment. Thus, the GSLs generate considerable variation in expected wages at different days of the week and at various hours of the day.

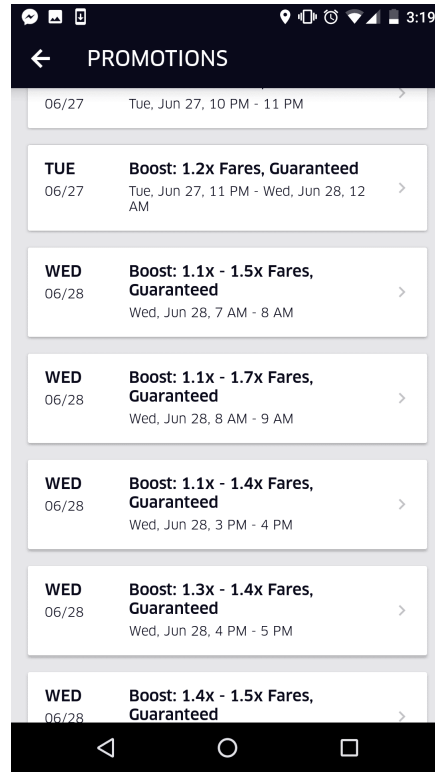


Figure 2: Example of a GSL Announcement to Drivers

Notes: The screenshot shows a GSL announcement to drivers on the Uber Driver App. The announcement clearly shows the hours when the GSL will be switched on and the corresponding GSL multiplier.

2.3 Trip Earnings, Wages and Work Decisions

The unit of observation in our analysis is an individual driver at a given hour. Thus, we measure labor supply and wages on an hourly basis. We define total minutes worked per hour as the number of minutes for which a driver has the app on and is available for accepting requests in that hour. In other words, a driver is said to be working if she is actively searching for rider requests.¹⁰

¹⁰This is the same definition of working as in Chen et al. (2019). A driver is active if the driver-side app is turned on and she is available to accept requests for rides. This is to be distinct from a “browsing” mode in which the app is on but the driver has not indicated a willingness to accept rides.

The observed hourly wage rate is measured as a driver’s total earnings divided by the number of minutes worked in an hour, multiplied by 60. The total earnings in an hour are measured as the sum of trip earnings a driver receives, where each trip earnings is determined by the following formula:

$$\text{Trip Earning} = \max(\text{GSL}, \text{Surge}) \times \text{Baseline Fare}.$$

In this formula, GSL is the experimental fare multiplier, Surge is the demand-driven fare multiplier, and BaselineFare is the baseline trip earnings following Uber’s fixed compensation rule:

$$\text{Baseline Fare} = \text{Fixed Price} + (\text{Price per Minute} \times \text{Minutes}) + (\text{Price per Mile} \times \text{Miles}).$$

As in Chen et al. (2019), our measure of wages does not net out the variable costs of operating a vehicle. Therefore, our reservation wages should be interpreted as a gross quantity. It is important to observe, however, that labor supply decisions are based on the differences between expected wages and reservation wages, which do not depend on assumptions regarding the incorporation of time-invariant operating costs.

Drivers base labor supply decisions on expected hourly wages rather than the realized wages that we observe. To construct measures of expected wages, we predict the hourly wage a driver is likely to face in each hour. As a first step, we calculate the wage multiplier, defined as $\max(\text{GSL}, \text{Surge})$, from our detailed data on GSL and Surge. Next, we calculate the pre-multiplier wages as the observed hourly wages divided by the calculated wage multipliers. We then fit the following regression model to the panel data on pre-multiplier wage:

$$\tilde{W}_{it} = \alpha_i + \kappa_{h(t)} + \epsilon_{it}$$

where t is an hour, $h(t)$ is the hour of the week at t , i denotes a driver, α_i and $\kappa_{h(t)}$ are driver and hour-of-week fixed effects, and \tilde{W}_{it} is the pre-multiplier hourly wage. For each hour t , we fit the model with the panel data up to $t - 1$, and use the estimated $\hat{\alpha}_i$ and $\hat{\kappa}_{h(t)}$ to compute a predicted value for \tilde{W}_{it} for every worker i in each hour t . The predicted hourly wage is then constructed as the product of the predicted pre-multiplier wage and the calculated wage multiplier.

To assess how well our prediction model performs, we compare it to alternative approaches using a cross-validation procedure. This procedure repeatedly divides samples into a training sample and a testing sample. For each approach, we use the training samples to estimate a prediction model, and then we use the estimated model to form a prediction and calculate the out-of-sample mean squared errors on the testing samples. In addition to our current prediction model, we consider several alternative approaches, including a matching procedure with a K-means clustering method. Our prediction model performs considerably better than the alternative approaches.

3 Data and Experimental Findings

In this section, we describe the data and present the findings from the experiment.

3.1 Description of the Data

3.1.1 Sample and Variables

Our analyses are based on panel data of UberX and Uber Pool drivers who are eligible for the GSL experiment in Boston, Chicago, and San Francisco. In Boston and Chicago, we observe all these drivers. In San Francisco, we have data for a random subsample of 35 percent of the eligible drivers. In Boston and San Francisco, the duration of our data spans from October 2016 to March 2018. In Chicago, we have only one year of data, covering October 2016 to March 2017. For each driver, we observe gender, age, type of vehicle, minutes worked per hour, trip earnings, and fare multipliers.

Our sample covers 333,172 drivers. A quarter of these are female, and the median age is about 38. The average observed wage is \$19.29 per hour. In Figure 3, we plot the predicted hourly wage over time according to gender and age. We define drivers as young if they are younger than 38 years old. Consistent with Chen et al. (2019), most of the heterogeneity in hourly wages is due to hours of the day. It is also evident that hourly wages tend to be higher in the weekends and that male drivers have only slightly higher hourly wage as compared to female drivers.

Figure 4 plots the probability of working over time for all drivers and by subgroup. The work probability varies considerably across hours of the week and days of the week. There are also distinct differences in the probability of working by age and gender. Conditional on age, male drivers are much more likely to work, especially during the daytime. Holding gender fixed, old drivers tend to work more than young drivers. In Figure 5, we show the average amount of hours worked per week by age and gender. In expectation, drivers work nearly five hours per week. However, males drivers work twice as many hours per week as female drivers, and young drivers work 36 percent less than older drivers.

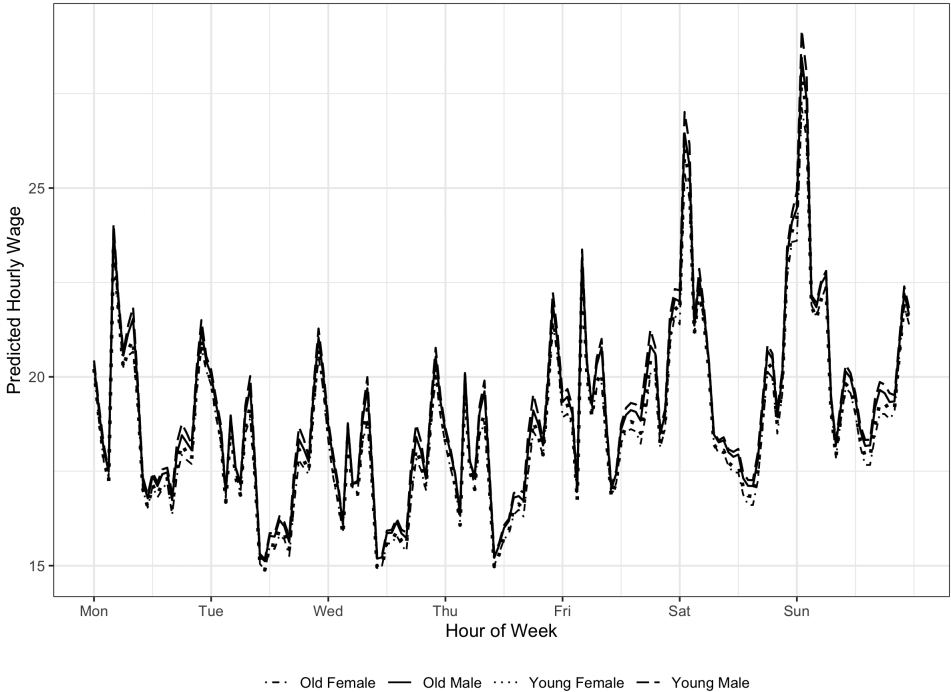


Figure 3: Predicted Hourly Wages Across Hours of the Week by Demographic Groups

Notes: In this figure, we compute the average predicted wage at every hour of the week and for each demographic group.

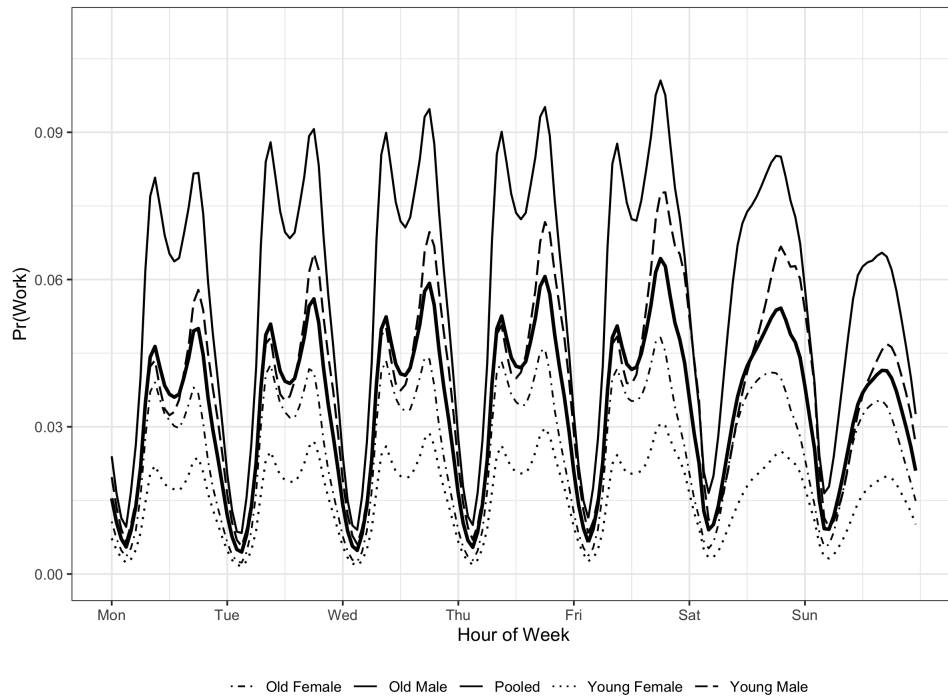


Figure 4: The Probability of Working Across Hours of the Week by Demographic Groups

Notes: In this figure, we compute the share of active drivers at every hour of the week and for each demographic group. We define active drivers as those who work any positive number of minutes in a given hour. "Pooled" refers to the combined sample across demographic groups.

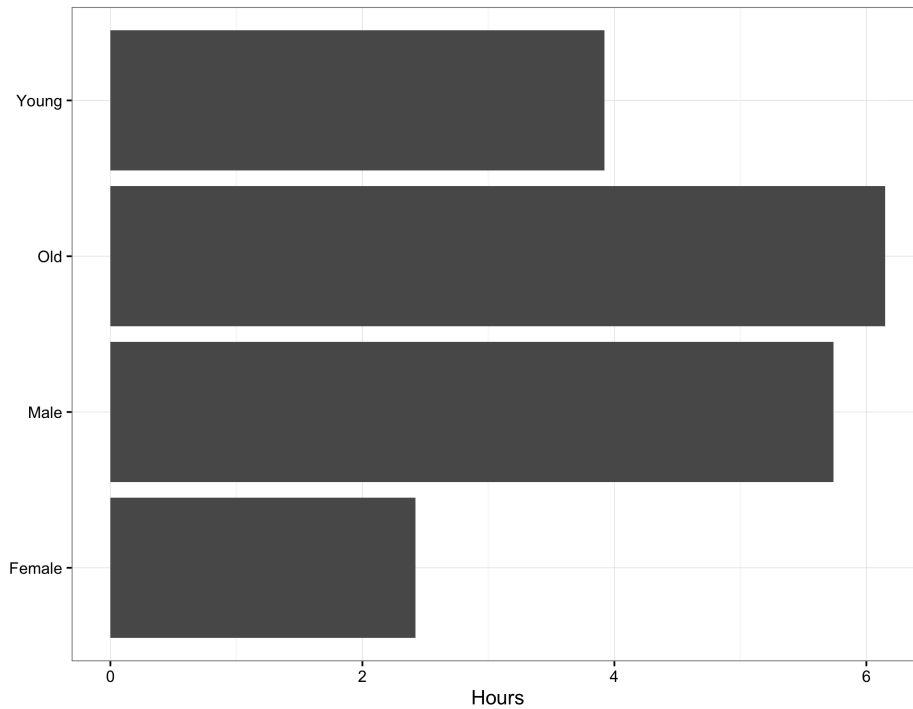


Figure 5: Average Number of Hours Worked per Week by Demographic Groups

Notes: In this figure, we compute the average number of hours worked for each demographic group. An hour worked is defined as an hour where a driver works for any positive number of minutes.

3.2 Checking Covariate Balance

In a properly implemented, randomized experiment with a sufficiently large sample size, we expect the treatment and control groups to be balanced in their distribution of pre-treatment variables. To assess this, we check the covariate balance by regressing treatment status in the GSL experiment on the pre-treatment characteristics of the drivers:

$$D_{it} = X'_{it}\beta + u_{it}$$

where D_{it} is an indicator variable that is equal to 1 if driver i is assigned to the treatment group in the GSL experiment at time t , and X_{it} is a vector of covariates that include gender, age, number of trips completed, and past wages. All these covariates are measured in the week before the randomization. Table 1 reports the estimates. In Column 1, we regress treatment status on each characteristic separately. There is no evidence of systematic differences between drivers in the treatment and control groups in the characteristics considered. In Column 2, we regress treatment status on all the characteristics in a multiple regression. Consistent with the randomization, we cannot reject the null hypothesis that all coefficients are zero.

Pre-determined Var.	Separate Regression Var. on Treatment	Joint Regression Treatment on Var.
Female	-0.0004 (0.0003)	0.0000 (0.0000)
Age	-0.0000 (0.0000)	-0.0007 (0.0004)
Trips Completed	-0.0000 (0.0000)	0.0001 (0.0000)
Wage Last Week	0.0001 (0.0000)	-0.0000 (0.0000)
N (Blocks)	22,318,255	22,318,255
R^2		0.0000
F Statistic		1.8450
p value - F		0.1171

Table 1: Balance Tests

Notes: In Column 1, we regress each driver characteristic on the treatment status separately, and each row represents a separate regression. In Column 2, we regress the treatment status on all the characteristics in one regression. We use the F-test to examine whether one or more of the coefficients of these four pre-determined variables are significantly different from zero.

3.3 First Stage Results: Effects of GSL on Expected Market Wages

Figure 6 presents the treatment effects of the GSL experiment on the predicted hourly wages during the experiment hours. These effects will be the first stage estimates in the instrumental variables (IV) estimation of the effects on labor supply of exogenous changes in predicted wages. The shaded area in the figure indicates the hours with the experiment switched on. The figure shows that assignment to the GSL experiment increases the average predicted hourly wage by around forty cents or, equivalently, around 1.8 percent. To examine if the wage effects are persistent across experiment hours, we divide each experiment window in half. The

estimates suggest little if any changes in the wage effects across hours within the experiment window.

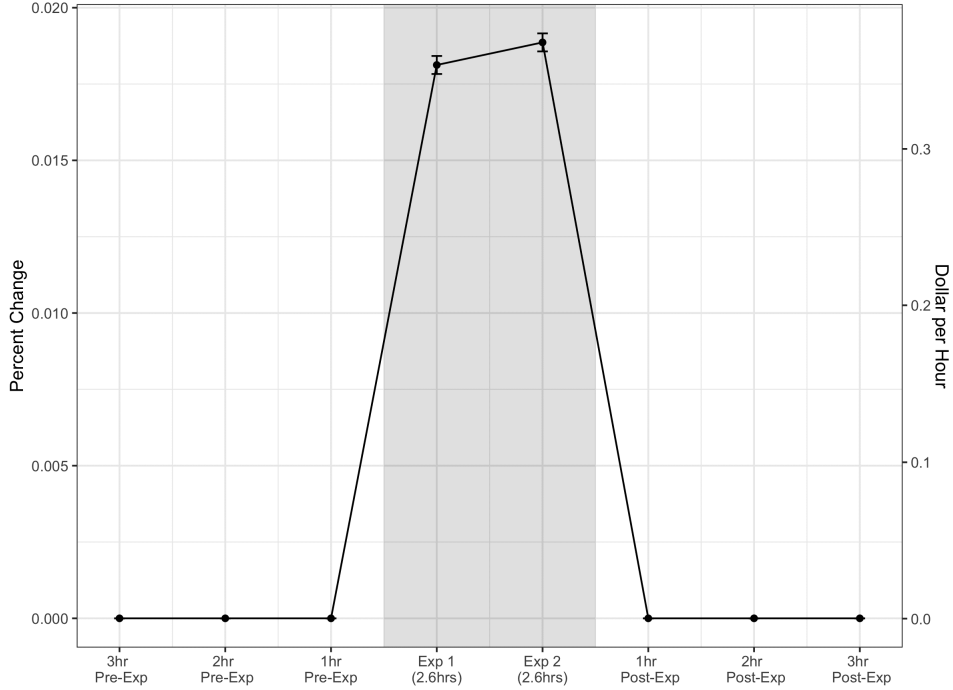


Figure 6: Treatment Effects of the GSL Experiment on the Predicted Hourly Wages

Notes: In this figure, we present the estimates of the GSL experiment on the predicted hourly wages during the experiment hours (in the hours before and after the experiment window, the effect is zero). The estimation is performed separately for the first half (Exp 1) and the second half (Exp 2) of the experiment windows. On average, the experiment windows last 5.2 hours. The shaded area in the figure indicates the experiment hours with the GSL switched on. The y-axis on the left shows percent differences between the treatment and the control groups. It is computed as the treatment effect on the predicted hourly wage divided by the predicted hourly wage of the control group. The y-axis on the right reports the estimated effects in dollars per hour. The bars indicate the 90% confidence intervals calculated from subsampling bootstrap.

3.4 Reduced Form Effects: Labor Supply Responses to GSL

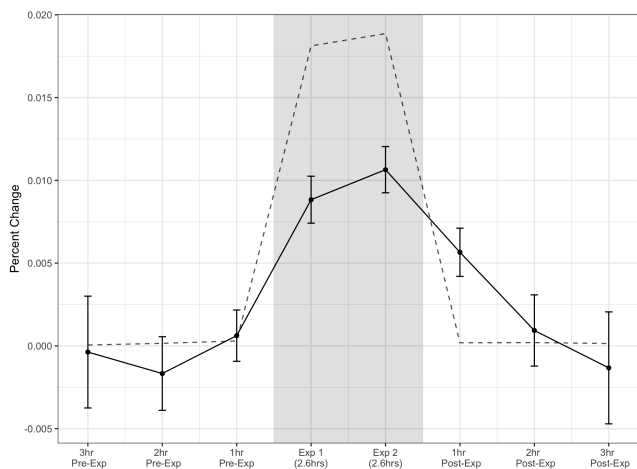
We now document that drivers respond to the GSL experiment both during the experiment windows and in the hours preceding and following these periods. These effects will be the reduced form estimates in the IV estimation of the effects on labor supply of exogenous changes in predicted wages. We consider labor supply responses along two margins: In each hour, we measure if the driver worked at all and the number of minutes she worked.

In Figure 7, we plot the effects of the GSL experiment on the labor supply responses of the drivers before, during, and after the experiment window. These effects are represented by the solid lines. The dotted lines represent the changes in wages due to the GSL experiment. We find significant changes in labor supply during the experiment window. During this window, the treated drivers increase the employment rate per hour and the hours of work by about one percent as compared to the control drivers. There is also some suggestive evidence of persistent effects outside the experiment window. For example, in the hour following the experiment window, the labor supply is half a percent higher for the treated group as compared to the control group.

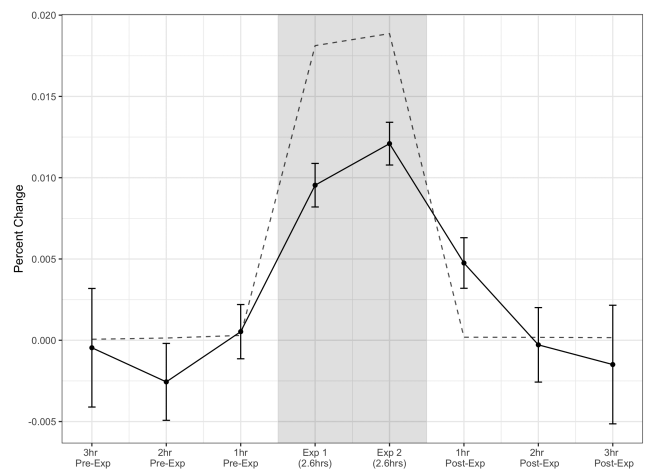
When interpreting the estimated effects in the hours preceding and following the experiment window, it

is important to recognize that most blocks have several GSL experiment windows. As a result, the hours preceding and following a given experiment window are likely to be confounded by other GSL experiment windows. For example, in 30 percent of the blocks in our sample, there exists at least two experiment windows that are no more than an hour apart. To address this concern, we take advantage of the re-randomization at the start of every block and restrict attention to the experiment windows preceding and following the re-randomization. By estimating the labor supply responses around these experiment windows, we avoid confounding anticipatory and persistent responses with other GSL experiment windows.

In Figures 8 and 9, we present the estimates from the experiment windows preceding and following the re-randomization. Even four hours prior to the experiment window, we see the treated drivers are more likely to be working as compared to drivers in the control group. The anticipatory response is most pronounced in the hour just before the experiment window. The same holds true for the persistent responses. In the first few hours following the experiment, the labor supply of the treated drivers remains significantly higher than the drivers in the control group. As time passes, these differences decline, and four hours after the experiment window the treated drivers work as much as the control drivers.



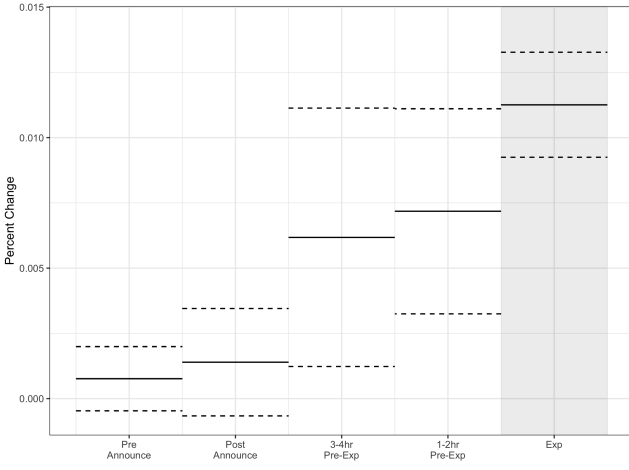
(a) Working



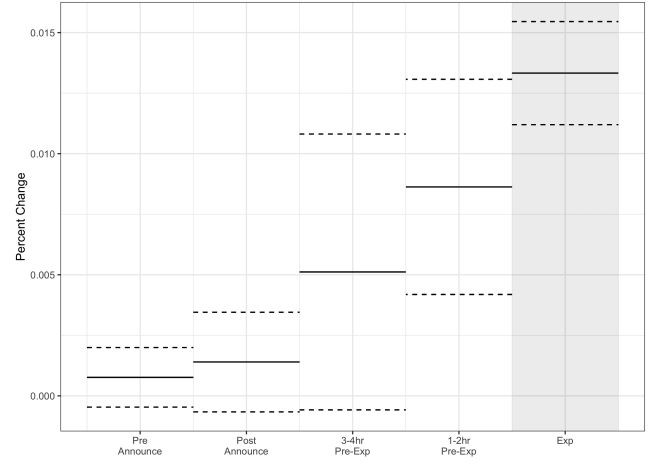
(b) Minutes Worked

Figure 7: Treatment Effects on Labor Supply Responses

Notes: In this figure, we illustrate the changes in the probability of working and minutes worked. We re-center all the experiment windows and plot the x-axis the same way as in Figure 6. The estimation is performed separately for the first half (Exp 1) and the second half (Exp 2) of the experiment windows. On average, the experiment windows last 5.2 hours. The shaded area in the figure indicates the experiment hours with the GSL switched on. The solid line represents the difference in labor supply responses, and the dashed line represents the difference in predicted hourly wages. The bars indicate the 90% confidence intervals.



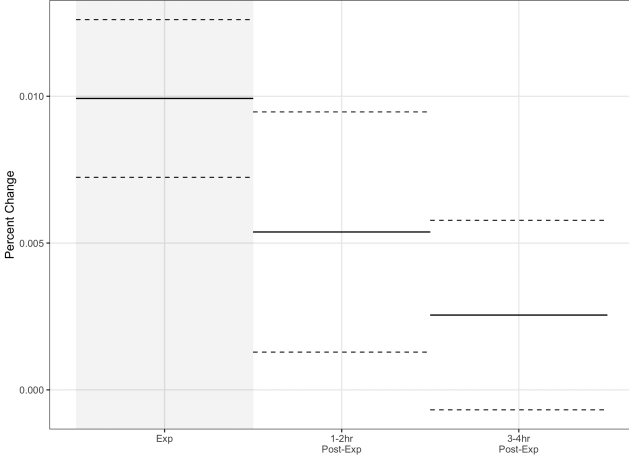
(a) Working



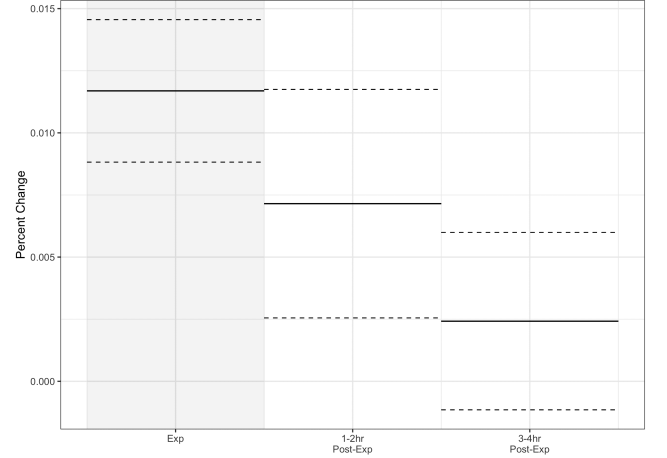
(b) Minutes Worked

Figure 8: Treatment Effects on Labor Supply Responses in the First Experiment Windows

Notes: We illustrate the labor supply responses immediately after re-randomization. We pool all the first experiment windows for estimation. The point estimates are constructed in the same way as Figure 7. The shaded area in the figure indicates the experiment hours in the first experiment windows. The “Pre Announce” period contains all hours before the announcement of the GSLs, while the “Post Announce” period contains the hours after announcement up to 5 hours before the first experiment windows. The dashed bars indicate the 90% confidence intervals.



(a) Working



(b) Minutes Worked

Figure 9: Treatment Effects on Labor Supply Responses in the Last Experiment Windows

Notes: We illustrate the labor supply responses immediately before re-randomization. We pool all the last experiment windows for estimation. The point estimates are constructed in the same way as Figure 7. The shaded area in the figure indicates the experiment hours in the last experiment windows. The dashed bars indicate the 90% confidence intervals.

3.5 IV Results: Labor Supply Responses to Exogenous Changes in Expected Market Wages

We now turn attention to the IV estimates of the labor supply responses to exogenous changes in expected market wages. These estimates are based on the reduced form and first stage estimates reported above. Table 2 presents the IV estimates. In the first three columns, we measure labor supply as any work in a given hour. The last three columns measure labor supply as minutes worked during an hour. In Columns 1 and 4, we use the data from all the experiment hours. Columns 2 and 5 restrict attention to the hours surrounding the

first experiment window in a block, whereas Columns 3 and 6 consider only the hours surrounding the last experiment window in a block.

The first stage estimates in Columns 1 and 4 are very precise and show that the GSL experiment raises the predicted hourly wage by nearly forty cents. More importantly, the IV estimates in these columns imply that a \$10 increase in hourly wages would raise the share of workers that drive in a given hour by 1.4 percentage points or, equivalently, by 27 percent. By comparison, this increase in hourly wages would raise the amount of minutes that an average driver works by about 30 percent, from 2.26 to nearly 3 minutes per hour. The columns other than 1 and 4 quantify the labor supply responses of the drivers in the hours preceding and following the exogenous changes in predicted wages. We find that these anticipatory and persistent responses are significant and economically relevant. For both measures of labor supply, the anticipatory and persistent responses are about a third of the size of the responses during the experiment hours.

To better understand what drives the increase in labor supply, it is useful to decompose the IV estimates into responses on the extensive (any work in a given hour) and the intensive margin (minuted worked conditional on working in a given hour). Concretely, we decompose the estimate in Column 4 of Table 2 as follows:

$$\underbrace{\frac{\partial}{\partial W_t} E(\text{MinutesWorked}_t)}_{0.680} = \underbrace{\frac{\partial}{\partial W_t} E(\text{MinutesWorked}_t | \text{Work}_t) \times Pr(\text{Work}_t)}_{0.064 \text{ (9.4\%)}} + \underbrace{E(\text{MinutesWorked}_t | \text{Work}_t) \times \frac{\partial}{\partial W_t} Pr(\text{Work}_t)}_{0.616 \text{ (90.6\%)}}$$

where Work_t is an indicator variable that is equal to one if the driver works in a given hour. Using the drivers in the control group, we calculate $E(\text{MinutesWorked}_t | \text{Work}_t)$ and $Pr(\text{Work}_t)$ while $\frac{\partial}{\partial W_t} E(\text{MinutesWorked}_t)$ and $\frac{\partial}{\partial W_t} Pr(\text{Work}_t)$ are taken from the estimates in Columns 4 and 1, respectively. The results suggest that responses at the extensive margin account for nearly all the increase in the amount of minutes worked during an hour. This finding suggests it is important to model the driver's decision to work or not in a given hour, not the amount of minutes she works within an hour.

3.6 Heterogeneity in Labor Supply Responses

So far, we have focused on the average labor supply responses across all drivers. However, these average impacts miss a lot: The labor supply responses vary systematically both across people and over time. In order to discover these heterogeneous effects, we apply the method of Chernozhukov et al. (2018) to estimate an IV model with a full set of interactions between the endogenous regressor of interest, wages, and the pre-determined covariates, gender, hour of the day, and day of the week. Since there are 168 hours per week, we have 168 bins for the time dimension. For age, we divide the sample into four equally sized groups: Younger than 30 years, 30 to 38 years, 38 to 48 years, and older than 48 years. Since gender is binary, we therefore get 1,344 mutually exclusive and collectively exhaustive groups.

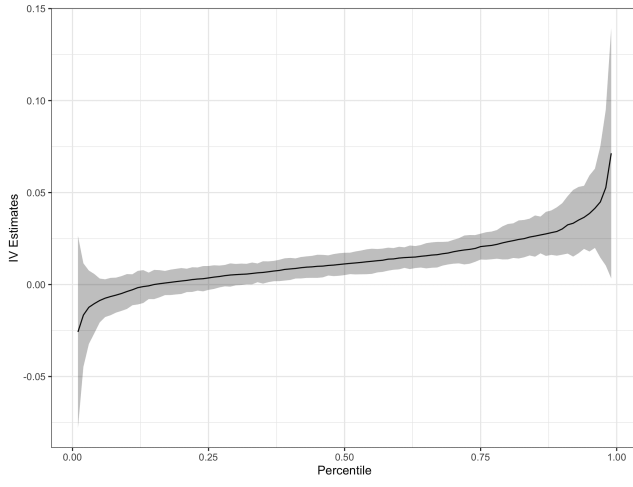
We sort these 1,344 IV estimates in an increasing order. In Figure 10 we illustrate these sorted IV estimates. The x-axis represents the percentile rank in the distribution of the estimated effects, and the y-

Dependent Var.	Working			Minutes Worked		
Panel A: IV						
	All	First	Last	All	First	Last
Experiment Hrs	0.0142 (0.0014)	0.0167 (0.0009)	0.0113 (0.0009)	0.6801 (0.0654)	0.8130 (0.0382)	0.5558 (0.0394)
Anticipation		0.0053 (0.0009)			0.2451 (0.0447)	
Persistence			0.0036 (0.0009)			0.1800 (0.0403)
Control Mean ¹¹ of Dep. Var.	0.0525 (0.0002)	0.0474 (0.0002)	0.0375 (0.0002)	2.2600 (0.0084)	1.9484 (0.0088)	1.5636 (0.0075)
Panel B: First Stage						
	All	First	Last	All	First	Last
First Stage	0.3608 (0.0037)	0.3193 (0.0057)	0.3288 (0.0062)	0.3608 (0.0035)	0.3193 (0.0054)	0.3288 (0.0061)
Control Mean of First Stage	19.5091 (0.0138)	20.3611 (0.0196)	19.7459 (0.0189)	19.5091 (0.0133)	20.3611 (0.0186)	19.7459 (0.0180)

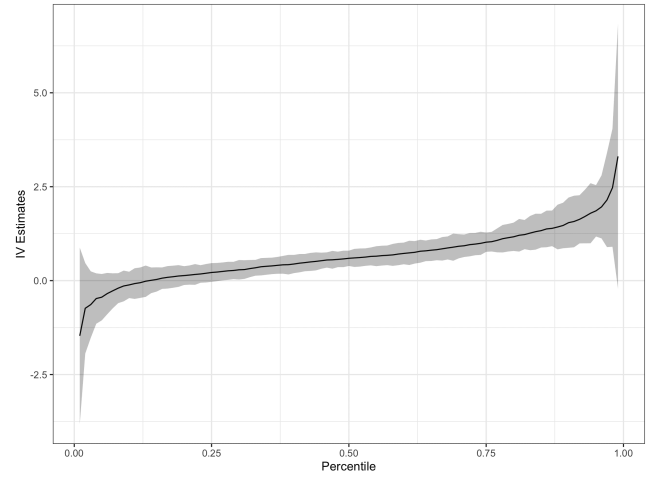
Table 2: Labor Supply Responses: IV and First Stage Estimates

Notes: “Control mean” is the expected outcome for the control group. IV is estimated as the increase in the probability of working per hour (or minutes worked per hour) per \$10 increase in the predicted hourly wage. The standard errors of the IV estimates and the first stage are estimated by bootstrap. Column 2 and Column 5 show the estimates for the first experiment windows after re-randomization. Column 3 and Column 6 show the estimates for the last experiment windows before re-randomization.

axis shows the estimated effect sizes. Panel (a) graphs the estimated effects for working and Panel (b) shows the estimated effects for minutes worked. It is evident that the IV estimates vary widely across groups. At the 25th percentile, the estimated effect for working is 0.0036, while the value at the 75th percentile is 0.021, almost 6 times as large. The estimated effect for minutes worked at the 75th percentile is almost 5 times as large as the value at the 25th percentile.



(a) Working



(b) Minutes Worked

Figure 10: Heterogeneity in IV Estimates

Notes: In this figure, we plot the distribution of the heterogeneous treatment effects. The solid line indicates the IV estimates measured as the increase in the probability of working (or minutes worked) w.r.t. a \$10 increase in the predicted hourly wage. The estimates and the 90% bootstrap uniform confidence bands are derived following Chernozhukov et al. (2018) based on the linear model with full saturation of observed heterogeneity in hours of the week, young versus old, and male versus female.

A natural question is what drives the heterogeneity in labor supply responses documented in Figure 10. To answer this question, we begin by examining the time dimension. In Figure 11, we plot the average responses across all drivers for different hours of the day and for weekend versus weekdays. There is substantial heterogeneity along the time dimension. Most of this variation comes from the differences across hours of the day rather than days of the week. When performing formal statistical tests, we can strongly reject the null hypotheses of equal average responses across hours of the day and across weekends versus weekdays.

As shown in Figure 12, there is also considerable heterogeneity by gender and age. On average, male and old drivers have larger responses than young and female drivers. The differences across gender are larger than those by age. Again, the null hypotheses of equal average responses by gender or age are strongly rejected in the data.

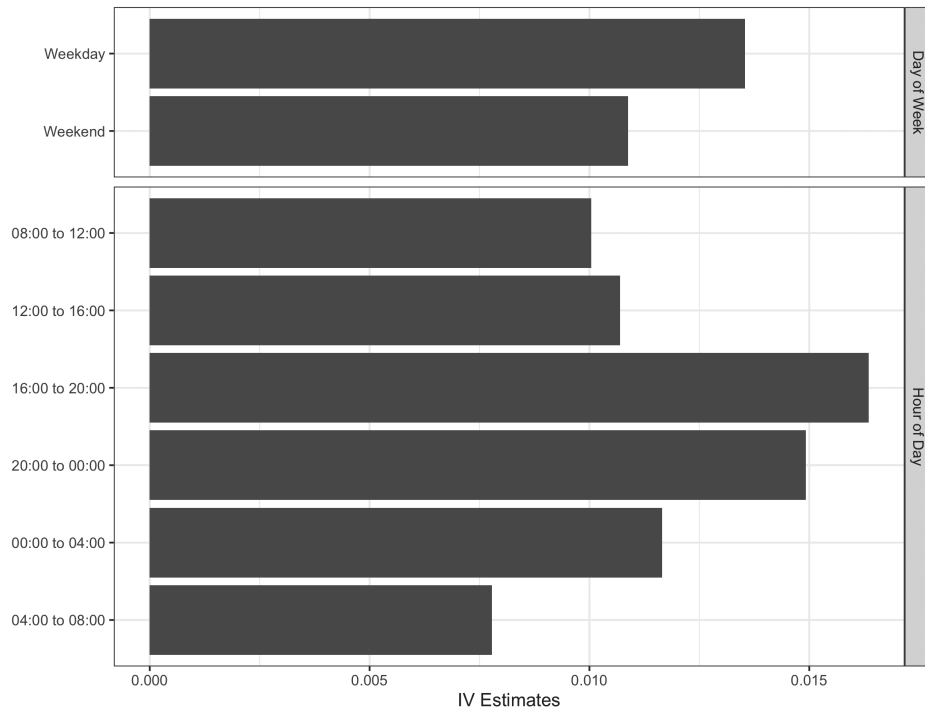


Figure 11: Heterogeneity in IV Estimates over Time

Notes: The figure compares the IV estimates across hours of the day, and weekdays versus weekend. The unit of the IV estimates is the increase in the probability of working w.r.t. a \$10 increase in the predicted hourly wage.

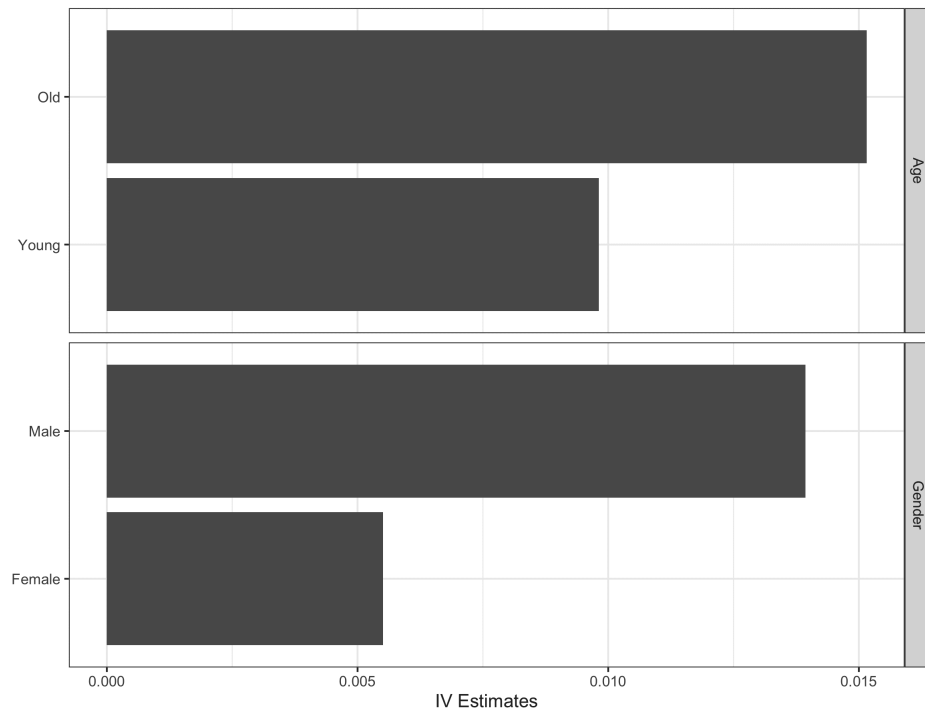


Figure 12: Heterogeneity in IV Estimates Across Demographic Groups

Notes: This figure compares the IV estimates across demographic groups. The unit of the IV estimates is the increase in the probability of working w.r.t. a \$10 increase in the predicted hourly wage.

We conclude the analysis of heterogeneity by examining how much of the heterogeneity in the estimated labor supply responses that can be explained or accounted for by various covariates. We begin by regressing

the 1,344 IV estimates reported in Figure 10 on indicator variables for time. We find that day of the week explains as little as 1.2 percent of the variation in the labor supply responses. By comparison, measuring time through indicators for hours of the day increases the R-squared to 12.4 percent. There are only small gains in explanatory power from including indicators for both hour of the day and day of the week in a separable fashion. By way of comparison, interactions between hours of the day and days of the week are empirically important to explain the pattern of heterogeneity in labor supply responses, increasing the R-squared from 13.7 percent to 34.9 percent. Also including indicator variables for gender and young (defined as younger than 38) further increases the R-squared by a few percentage points. By comparison, a flexible regression model with interactions between the day of the week, the hour of the day, gender and young explains nearly 70 percent of the variation in labor supply responses. Taken together, these results suggest it is key to let preferences of the drivers vary by gender and age, and for each demographic group, across hours of the day and days of the week.

Independent Variable	Explanatory Power (R^2)	
	Working	Minutes Worked
Day of Week	0.012	0.012
Hour of Day	0.124	0.124
Day of Week + Hour of Week	0.137	0.137
Day of Week \times Hour of Day	0.349	0.349
(Hour of Day \times Day of Week) + (Young \times Gender)	0.384	0.384
(Hour of Day \times Day of Week) \times (Young \times Gender)	0.691	0.718

Table 3: Explanatory Power in Regressions of Labor Supply Responses on Covariates for Time and Demographics

Notes: We regress the IV estimates of labor supply responses on the time, age, and gender dummies, and we report the R-squared in this table. Each number in this table corresponds to a separate regression. Each regression is weighted by the inverse of the variance of the IV estimates.

3.7 Comparison with OLS estimates

Table 4 compares OLS estimates of the labor supply responses to the IV estimates we obtain using the experiment. These results suggest that unobserved determinants of wages, if ignored, lead to a significant downward bias in the estimated labor supply responses. In particular, the OLS estimates in Column 1 show much weaker associations between labor supply and wages than the IV estimates. This downward bias is consistent with demand being high when it is costly for the drivers to work. Including fixed effects for workers, days of the week, and hours of the day reduces the bias, as shown in Column 2. However, the labor supply elasticities remain too small. This finding suggests that idiosyncratic factors, such as weather conditions and entertainment events, may create high demand while, at the same time, make driving more costly or difficult.

Dependent Variable: Working			
	OLS	OLS+FE	IV
Estimates	0.0099 (0.0003)	0.0125 (0.0003)	0.0142 (0.0014)
Dependent Variable: Minutes Worked			
	OLS	OLS+FE	IV
Estimates	0.5223 (0.0157)	0.6296 (0.0180)	0.6801 (0.0654)

Table 4: OLS and IV Estimates of Labor Supply Responses

Notes: In this table, we report the OLS, OLS with fixed effects, and IV estimates of the labor supply responses. OLS is estimated by regressing the outcome variables on predicted hourly wages. The standard errors for the OLS estimates are clustered at the drivers level, and the standard errors for the IV estimates are estimated by bootstrap. The fixed effects include the driver fixed effect and the hour of week fixed effect.

4 Dynamic Model of Labor Supply

The experimental estimates provide key data points for learning about labor supply elasticities, reservation wages and the value of job flexibility, but do not by themselves tell us these quantities. In order to recover the labor supply elasticities and reservation wages and to infer the value of job flexibility, we now develop, identify, and estimate a dynamic model of labor. In this section, we present this model and discuss the parameter estimates. In the Appendix, we compare these estimates to those produced by more restrictive models, including a static labor supply model. This comparison highlights how several of our modeling choices – including adjustment costs, permanent observed and unobserved heterogeneity, and the field experiment to address wage endogeneity – are key not only to match the data but also for the estimates of the reservation wages and for the results from the counterfactual analyses.

4.1 Model Setup

4.1.1 Driver’s Problem

We model the driver as living infinitely many periods where each period is an hour. In each period t , the driver decides whether to work $a_{it} = 1$ or rest $a_{it} = 0$, taking into account both the current period payoff $U_{it}(a_{it})$ and how her choice in t will affect the payoffs in the future $\tau \geq t$. At each period t , a driver chooses a_{it} in order to maximize the expected sum of discounted flow payoffs:

$$\max_{a_{it}} \mathbb{E} \left[\sum_{\tau=t}^{\infty} \rho^{\tau} U_{i\tau}(a_{i\tau}) | a_{it} \right]$$

where i indexes a driver, τ indexes hour, $U_{i\tau}(a_{i\tau})$ is the flow payoff associated with choice $a_{i\tau}$, and ρ is the discount rate. The expectation is taken over the future values of $U_{i\tau}(a_{i\tau})$ given the current choice a_{it} for $\tau \geq t + 1$.

4.1.2 Preferences

We sort drivers into subgroups based on their age and gender, $X = (\mathbf{1}\{\text{Female}\}, \mathbf{1}\{\text{Young}\})$. For a driver in an observed subgroup $X = x$ who works in a given city, the flow payoff associated with action a_{it} is given by

$$U_{it} = \begin{cases} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \xi_{it} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{cases}$$

where i is a driver, t is a calendar hour (e.g., 2018/10/10, 9 a.m.) and $h(t)$ is an hour of a week at time t (e.g., Monday 9 a.m.). We also include city fixed effect in the empirical specification of U_{it} to allow for systematic differences in the costs of driving across cities. For notational simplicity, we suppress these fixed effects as well as the conditioning on X . A driver's flow payoff from work depends on the wage she may earn, w_{it} , and the time-specific shifter of the cost of driving at a given hour of the week, $\beta_{h(t)}$. If the driver did not work at $t - 1$, she needs to pay an adjustment cost to start to work, μ . The parameter $\eta_{j(i),h(t)}$ captures the unobserved type j of driver i at time $h(t)$. For example, full-time drivers who are more likely to drive at all times have higher $\eta_{j(i),h(t)}$ at all $h(t)$ than infrequent drivers, while evening drivers have higher $\eta_{j(i),h(t)}$ only in the evenings. Driver type $\eta_{j(i),h(t)}$ is known to the drivers themselves but unobserved to the analyst. A driver's flow payoffs from the choices at t are also affected by a set of choice-specific preference shocks, ξ_{it} , ϵ_{1it} and ϵ_{0it} , which are revealed to the driver at the beginning of t . The component ξ_{it} captures the unobserved preference shocks that may correlate with wages. For example, an individual may dislike to drive in periods with heavy traffic or poor weather and these conditions may also covary with the demand for rides and thus offered wages. The components ϵ_{1it} and ϵ_{0it} capture the idiosyncratic preference shocks that are independent of wages. Wage Equation

The equation for the predicted hourly wage of worker i in hour t is specified as follows

$$\begin{aligned} w_{it} &= \delta_{h(t)} + \delta_0 m_t + \delta_1 m_t \times z_{it} + u_{it} \\ u_{it} &\sim G_u(u_{it} | u_{it-1}, a_{it-1}, z_{it-1}, h(t-1), m_{t-1}, c(i)) \end{aligned}$$

where m_t is an indicator of whether t is an experiment hour and z_{it} is an indicator for being assigned to the treatment group in an experiment hour, $\delta_{h(t)}$ is a fixed effect for hour of week, and u_{it} represents the unobservables determinants of wages. By including m_t in the wage equation, we allow the wage to be different when a GSL experiment is switched on. The parameter δ_1 captures the exogenous change in the wage for the treated drivers during experiment hours.

The wage in our model evolves as a first-order Markov process. We allow persistence in wages by letting u_{it} depend on u_{it-1} , together with lagged choice a_{it-1} , lagged treatment z_{it-1} , lagged experiment hour m_{t-1} and city $c(i)$. We allow a driver's unobserved type $\eta_{j(i),h(t)}$ to depend nonparametrically on her initial wage and action. The endogeneity of wages arises if $Cov(\xi_{it}, u_{it}) \neq 0$. Thus, an exogenous wage process is a special case of our model in which the costs of working do not covary with market wages, $\delta_0 = \delta_1 = 0$.

4.1.3 Timeline and Information Set

Recall that 4 a.m. every Monday and Friday, drivers are randomized into treatment and control groups. Drivers are then informed about when and for how long GSL will be switched on in an upcoming block. We

specify the timeline within and between blocks as follows:

Timeline Between Blocks.

At the beginning of each block, a driver i learns her treatment status z_{is} and the experiment hours m_s for all hours s in the block. At the same time, the driver forms expectations about her treatment status and the experiment hours in future blocks. The driver then sequentially makes labor supply decisions, taking into account the current flow payoff and the continuation values. At the beginning of the next block, the drivers are re-randomized. Figure 13 presents an example of the timeline between blocks.

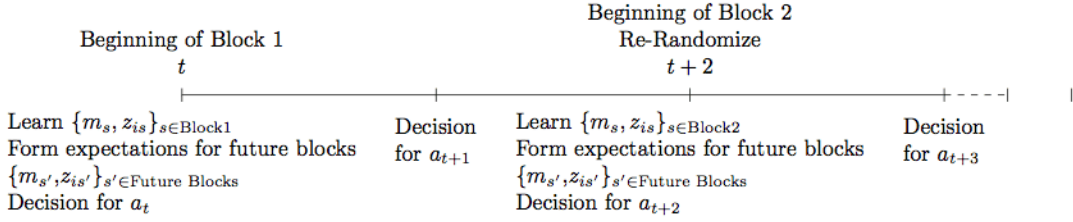


Figure 13: Example of Timeline with 2 Blocks and 2 Periods per Block

Timeline Within a Given Block.

The decision timeline within a block is given as follows: Let t be an arbitrary period within a given block. At the beginning of t , driver i learns the realizations of u_{it} (equivalently, her market wage w_{it}) and the realizations of the preference shocks $(\xi_{it}, \epsilon_{1it}, \epsilon_{0it})$. Based on these realizations, she forms expectations about future values of $u_{it'}$, $\xi_{it'}$, $\epsilon_{1it'}$, $\epsilon_{0it'}$. Next, she makes the work decision for period t , taking into account both the current flow payoffs and how her decision at t will affect her future flow payoffs.

4.2 Assumptions and Identification

Our identification argument combines a control function based on the experiment with fairly standard assumptions in the dynamic discrete choice literature. In this section, we briefly discuss the key assumptions and the outline of the identification argument. The details are in the Appendix.

The identification argument begins by making the following assumptions:

Assumption 1. Control Function Assumption

$$\begin{aligned}
 & \text{(Instrument Exogeneity)} \quad z_{it} \perp (\epsilon_{0it}, \epsilon_{1it}, \xi_{it}, u_{it}) \\
 & \text{(Joint Normality)} \quad \begin{pmatrix} u_{it} \\ \xi_{it} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho_{u\xi} \\ \rho_{u\xi} & 1 \end{pmatrix}\right)
 \end{aligned}$$

Under joint normality of u_{it} and ξ_{it} , we can rewrite ξ_{it} as

$$\xi_{it} = \frac{\rho_{u\xi}}{\sigma} u_{it} + \psi_{it} = \frac{\rho_{u\xi}}{\sigma} (w_{it} - \delta_{h(t)} - \delta_0 m_t - \delta_1 m_t \times z_{it}) + \psi_{it}$$

where $\psi_{it} \sim \mathcal{N}(0, 1 - \rho_{u\xi}^2)$ and $\psi_{it} \perp u_{it}$ by construction. We define a new state variable $\phi_{it} \equiv w_{it} - \delta_{h(t)} -$

$\delta_0 m_t - \delta_1 m_t \times z_{it}$. Thus, the flow payoffs of the problem become

$$U_{it} = \begin{cases} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \frac{\rho_{u\xi}}{\sigma} \phi_{it} + \nu_{it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{cases}$$

$$\begin{aligned} w_{it} &= \delta_{h(t)} + \delta_0 m_t + \delta_1 m_t \times z_{it} + u_{it} \\ u_{it} &\sim G_u(u_{it} | u_{it-1}, a_{it-1}, z_{it-1}, h(t-1), m_{t-1}, c(i)) \end{aligned}$$

where $\nu_{it} = \psi_{it} + \epsilon_{1it}$. In addition to Assumption 1, we make the following set of assumptions which are often invoked in the literature on dynamic discrete choice (see Rust (1987), Magnac and Thesmar (2002), Kasahara and Shimotsu (2009)):

Assumption 2. *Standard dynamic discrete choice assumptions*

- (IID) $\epsilon_{0it}, \epsilon_{1it}$ are iid across i, t
- (EXOG) $\nu_{it} \perp (w_{it}, \eta_{j(i),h(t)}, \phi_{it})$
- (CI-X) State transition probability F satisfies

$$F(w_{it+1}, \phi_{it+1} | a_{it}, a_{it-1}, w_{it}, \phi_{it}, h(t), j, \nu_{it}) = F(w_{it+1}, \phi_{it+1} | a_{it}, a_{it-1}, w_{it}, \phi_{it}, h(t))$$

- (DISTR) Distributional assumption on ϵ_{0it} and ν_{it} , and independence: $\epsilon_{0it} \perp \nu_{it}$.
- (DISCOUNT) ρ is known
- (REACH) All states are reachable at any given time

$$F(w_{t+1}, \phi_{t+1} | a_t, a_{t-1}, w_t, \phi_t, h(t), j) > 0 \quad \forall a_t, a_{t-1}, w_t, \phi_t, h(t), j$$

- (NTYPE) The number of types is small

Assumption 2 follows the standard assumptions in the dynamic discrete choice literature. The only difference is that, in our setup, the state-dependent unobserved preference shock, ξ_{it} , can be re-written as a combination of an observed state, ϕ_{it} , and an idiosyncratic component ψ_{it} , as a result of Assumption 1. Under the restrictions (IID), (EXOG), (CI-X), (DISTR), (DISCOUNT), we can identify the parameters in the flow payoff in the absence of unobserved heterogeneity (see Rust (1987), Magnac and Thesmar (2002)). The restrictions (REACH) and (NTYPE) are made to incorporate unobserved heterogeneity in the model. Under these restrictions, the structural parameters in the flow payoffs are identified (see Hotz and Miller (1993), Kasahara and Shimotsu (2009)). The restriction (REACH) imposes that the entire support of wages at $t+1$ has a positive probability conditional on the state at t . In our problem, the restriction requires that at each hour of the week and conditional on the lagged choice and the wage in the previous period, a driver may get any wage in the support with a strictly positive probability. The restriction (NTYPE) limits the number of unobserved types among drivers. We allow for three unobserved types of drivers. For the restriction (DISTR), we assume ϵ_{0it} and ν_{it} are both distributed as T1EV.

It is useful to observe that some of the restrictions in Assumption 2 are not that strong in our setting. For example, the restriction (CI-X) implies that, conditional on choices, the transition probability of the state variables is the same across unobserved types and independent of transitory shocks. To understand

this restriction, consider a driver who gets tired when working in period t . As a result, she may drive slower or take fewer trips and thus earn less if she chooses to continue driving in $t+1$. Restriction (CI-X) permits such a scenario. The restriction (DISCOUNT) requires the discount rate to be known to the analyst. In our setting, the data suggests that temporal decisions are primarily driven by heterogeneity in preferences over when to work $\beta_{h(t)}$, not discounting of future payoffs over a relatively short period of time. Thus, we think the discount rate plays a minor role for the behavior we observe. Assuming an annual interest rate as 5 percent, we set the hourly discount rate $\rho = 1/(1 + \frac{0.05}{365 \times 24})$.

4.2.1 Value Function and Structural Equation

We now describe the solution to the model and the structural equation. Under the standard conditions, the driver's problem can be characterized by the Bellman equation

$$V(s_{it}) = \max \{V(a_{it} = 1, s_{it}), V(a_{it} = 0, s_{it})\} \quad (1)$$

where s_{it} is a vector of observed state variables and the unobserved type, $V(s_{it})$ is the (ex-ante) value function for driver i who is in state s_{it} , and the choice-specific conditional value functions $V(a_{it} = 1, s_{it})$ and $V(a_{it} = 0, s_{it})$ are defined as follows

$$V(a_{it} = 1, s_{it}) = s'_{it}\theta + \xi_{it} + \epsilon_{1it} + \rho \mathbb{E}V(s_{it+1})$$

$$V(a_{it} = 0, s_{it}) = \epsilon_{0it} + \rho \mathbb{E}V(s_{it+1})$$

where $\theta = (\gamma, \beta_{h(t)}, \mu, \eta_{j(i),h(t)})$ is a vector of structural parameters in the flow payoff of work. The expectation is taken over future states and actions, $s_{i\tau}$ and $a_{i\tau} \forall \tau \geq t+1$, conditional on s_{it} and a_{it} , as well as future preference shocks $(\epsilon_{0i\tau}, \epsilon_{1i\tau}, \xi_{i\tau}) \forall \tau \geq t+1$, according to the evolution of states $F(s_{it+1}|s_{it}, a_{it})$ and the distribution of shocks. Thus, with equation (1), we can describe the driver's decision rule as follows: At the beginning of each period t , driver i learns state s_{it} , and chooses to work if and only if $V(a_{it} = 1, s_{it}) \geq V(a_{it} = 0, s_{it})$.

4.2.2 Estimation

The first step in our estimation procedure is to use OLS to estimate the parameters of the wage equation, $\delta_{h(t)}$, δ_0 , δ_1 , and then to obtain the residuals as measured by the empirical counterpart of u_{it} . We then non-parametrically estimate the transition probability of u_{it} given lagged state variables $(u_{it-1}, a_{it-1}, z_{it-1}, h(t-1), m_{t-1})$.

To explain how we estimate the other parameters of the model, it is useful to first consider the estimation procedure in the absence of unobserved types. Next, we describe how this procedure is modified to include the types $\eta_{j(i),h(t)}$.

Estimation of Dynamic Model without Unobserved Types. Estimation procedure follows Hotz and Miller (1993). Hotz and Miller (1993) shows that for each state vector $s \in S$ the log ratio of conditional choice probabilities can have the closed form solution as follows

$$\log \frac{P(a=1|s_{it}=s)}{P(a=0|s_{it}=s)} = [s + \rho(\tilde{s}(a=1, s) - \tilde{s}(a=0, s))]'\theta + \rho(\tilde{e}(a=1, s) - \tilde{e}(a=0, s)) \quad (2)$$

where $s = (w_{it}, \phi_{it}, a_{t-1}, h(t), j(i))$ denotes the state variables, and $\tilde{s}(a, s)$ and $\tilde{e}(a, s)$ are known functions of the state s , the conditional choice probabilities $P(a|s)$, and the state transition probabilities $F_s(s|a, s)$.

We first discretize ϕ_{it} distribution into decile grids, and use bin estimators to non-parametrically estimate the conditional choice probabilities, \hat{p} , and the wage transition probabilities, $\hat{F}_s(s'|a, s)$. We then construct estimators of \tilde{s} and \tilde{e} by applying the known functional forms to the estimators of the conditional choice probabilities and the state transition probabilities. Denote the estimators as $\hat{\tilde{s}}$ and $\hat{\tilde{e}}$, respectively. As described in detail in the Appendix, one may then form an OLS estimator of θ based on a closed-form estimation equation of (2).

Estimation of Dynamic Model with Unobserved Types. With unobserved driver types, we can no longer directly estimate θ using equation (2) since the unobserved type j enters both \tilde{s} and \tilde{e} . We instead follow the two-stage estimator of Arcidiacono and Miller (2011) to estimate the model with unobserved types by maximizing the log likelihood of the finite mixture model:

$$\{\hat{\theta}, \hat{\pi}\} = \arg \max_{\theta, \pi} \sum_{i=1}^N \ln \left[\sum_{j=1}^J \pi_j(s_{i1}) \prod_{t=1}^{T_i} l(a_{it}|s_{it}, j, \hat{p}, \theta) \right]$$

where $s_{it} = (w_{it}, \phi_{it}, a_{it-1}, h(t))$ is the vector of observed states, \hat{p} is a vector of empirical conditional choice probabilities, $\pi(s_{i1})$ is the population probability of type j conditional on initial state s_{i1} . The number of unobserved types, J , is assumed to be known. Let $l(a_{it}|s_{it}, j, \hat{p}, \theta)$ denote the likelihood contribution of driver i at time t . Based on equation (2), we can express the likelihood as follows

$$l(a_{it}|s_{it}, j, \hat{p}, \theta) = \frac{a_{it} e^{[s_{it} + \rho(\hat{\tilde{s}}(a=1, s_{it}) - \hat{\tilde{s}}(a=0, s_{it}))]'\theta + \rho(\hat{\tilde{e}}(a=1, s_{it}) - \hat{\tilde{e}}(a=0, s_{it}))} + (1 - a_{it})}{1 + e^{[s_{it} + \rho(\hat{\tilde{s}}(a=1, s_{it}) - \hat{\tilde{s}}(a=0, s_{it}))]'\theta + \rho(\hat{\tilde{e}}(a=1, s_{it}) - \hat{\tilde{e}}(a=0, s_{it}))}}.$$

We first initialize $\hat{\theta}, \hat{\pi}_j(s), \hat{q}_{ij}$ for all $\forall i, j$ by estimating the dynamic model without unobserved types. Next, we set $\hat{\pi}_j(s_1)$ equal to $\hat{\pi}_j(w_1, a_0)$. We then update $\hat{\pi}(s), \hat{q}_{ij}, \hat{p}(s, j)$ and $\hat{\theta} \forall i, j$ based on the EM algorithm. We refer to the Appendix for further details about the estimation procedure.

4.3 Model Fit and Estimation Results

Before we present the estimation results, we examine how well our estimated model fits the data. To examine the fit of the model, we focus on the probability of working conditional on the treatment status and other state variables.

Figure 14 plots the probability of working by the treatment status by hours of the week, days of the week, and hours of the day. We integrate the model predicted probabilities across all states, except for the treatment status, time, and observed heterogeneity. The model predicted probabilities fit the data counterparts very well. We next examine if the working pattern by treatment status across the four demographic groups is well captured by our model. Figure 15 shows that our model well predicts the probability of working by treatment status averaged across hours of week for each demographic group.

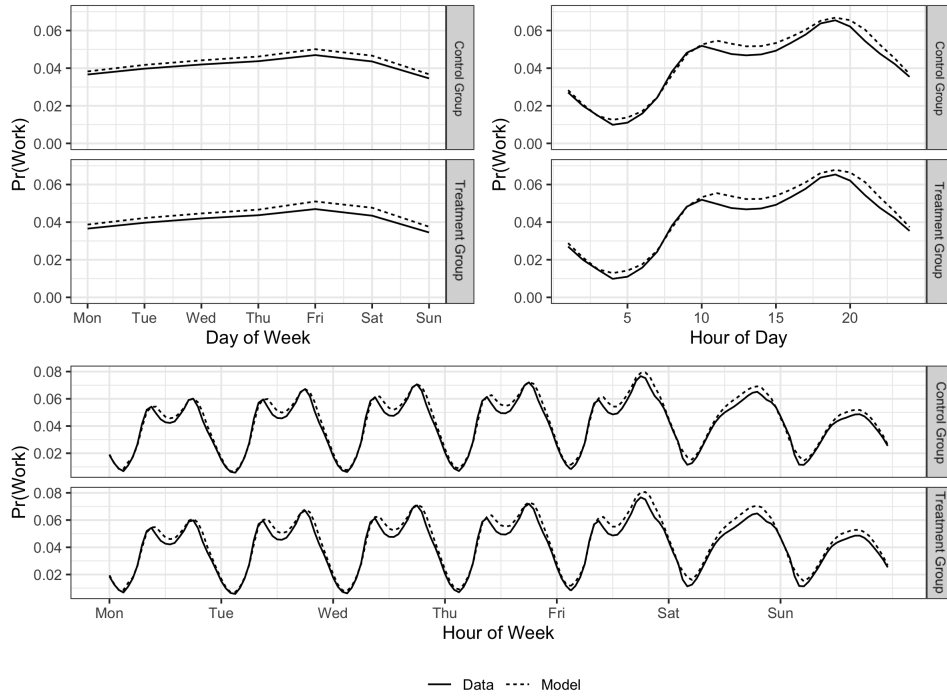


Figure 14: Model Fit of Probability of Working by the Treatment Status over Time

Notes: The solid line plots the data, and the dashed line plots the prediction from the model.

Figure 16 shows that our model captures reasonably well the three unobserved types of drivers after taking into account the observed heterogeneity. We use the observed data and the simulated data to plot the average work probabilities within a day of three groups of drivers: (1) infrequent drivers whose average hours worked during the day and at night both rank in the bottom 80 percent,¹² (2) evening drivers whose average hours worked in daytime rank in bottom 80 percent but the hours worked in the nighttime rank in the top 20 percent, and (3) the remaining drivers, who we refer to as frequent drivers because they tend to work a considerable amount both during the day and at night. The patterns of work vary distinctly over time across the three types of drivers. This is true both in the actual and the simulated data.

Our model also captures the important dynamic component, lagged choices, in a driver's labor supply decision. Figure 17 shows the probability of working conditional on the lagged work decision, integrating over observed heterogeneity and all other states. The probability of working differs significantly depending on whether a driver worked or not in the previous period. Through the lens of the model, this difference produces sizable costs of starting to drive.

One might worry that there are important sources of adjustment costs other than fixed costs of starting to drive. For example, Farber (2005) and Farber (2015) argue that fatigue is important to understand the behavior of taxi drivers. The reason is that the probability that a taxi driver ends a shift depends strongly on hours worked. In the Uber setting, there is little, if any, evidence of such dependence, as evident from Figure 18. This figure plots the probability of working in a given hour as a function of cumulative hours worked. Given the weak relationship between the probability of stopping to drive and cumulated hours worked, we decided against including fatigue in the model. Arguably, a small improvement in model fit does not justify to further enlarge the state space, which would significantly increase the computational costs of solving the

¹²We define the daytime to be 6 a.m.-9 p.m., and the nighttime to be 10 p.m.-5 a.m. next day.

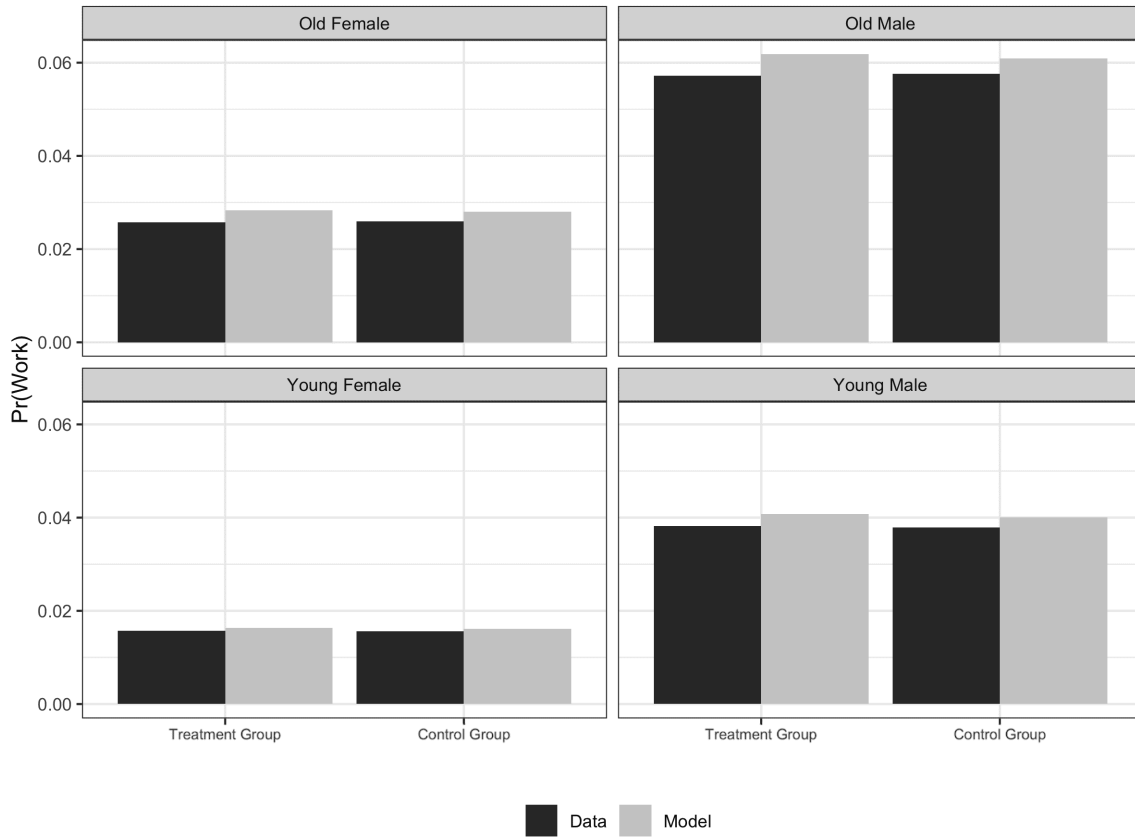


Figure 15: Model Fit of Probability of Working by the Treatment Status Across Demographic Groups
Notes: The black bar plots the data, and the grey bar plots the prediction from the model.

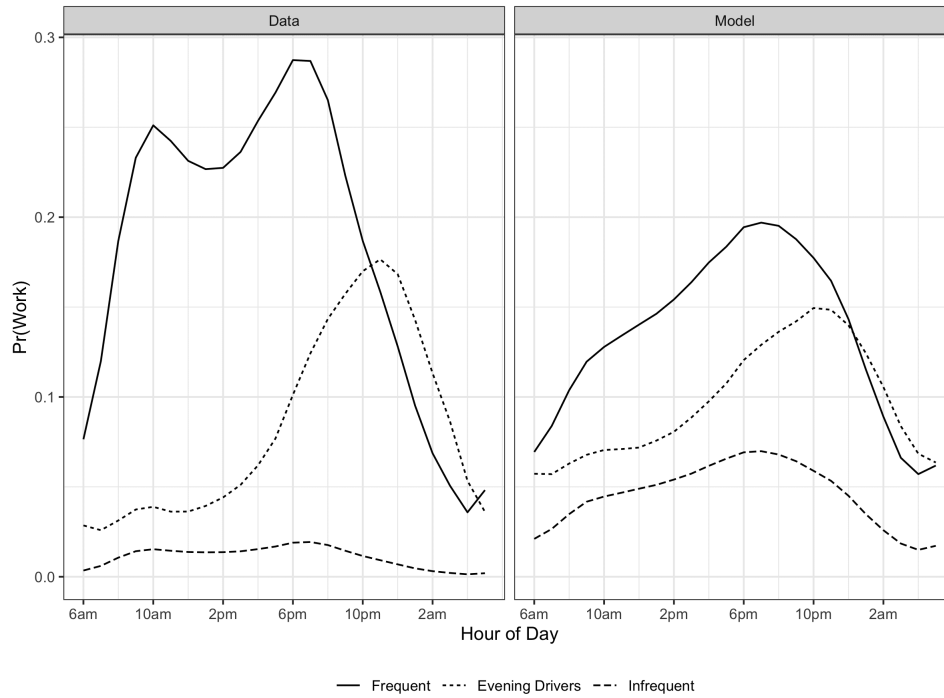


Figure 16: Model Fit of Average Probability of Working by Worker Types Across Hours of the Day
Notes: In this figure, we plot the average work probabilities within a day of three groups of drivers: (1) infrequent drivers whose average hours worked in the daytime and in the nighttime both rank in the bottom 80 percent, (2) evening drivers whose average hours worked in the daytime rank in the bottom 80 percent, but hours worked in the nighttime rank in the top 20 percent, and (3) the remaining frequent drivers. The ranks are calculated by pooling drivers from all four observed subgroups. The daytime is defined to be 6 a.m.-9 p.m., and the nighttime to be 10 p.m.-5 a.m. next day.

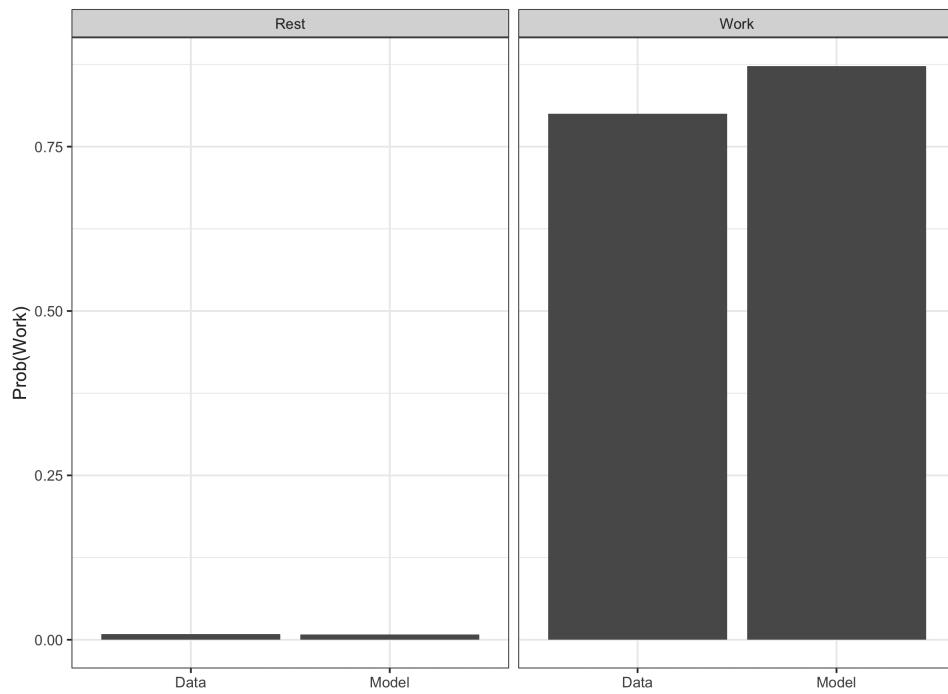


Figure 17: Model Fit of Probability of Working by Lagged Work Decisions

Notes: The “Rest” panel shows the probability of working conditional on drivers not working in the previous hour. The “Work” panel shows the probability of working conditional on drivers working in the previous hour.

model.

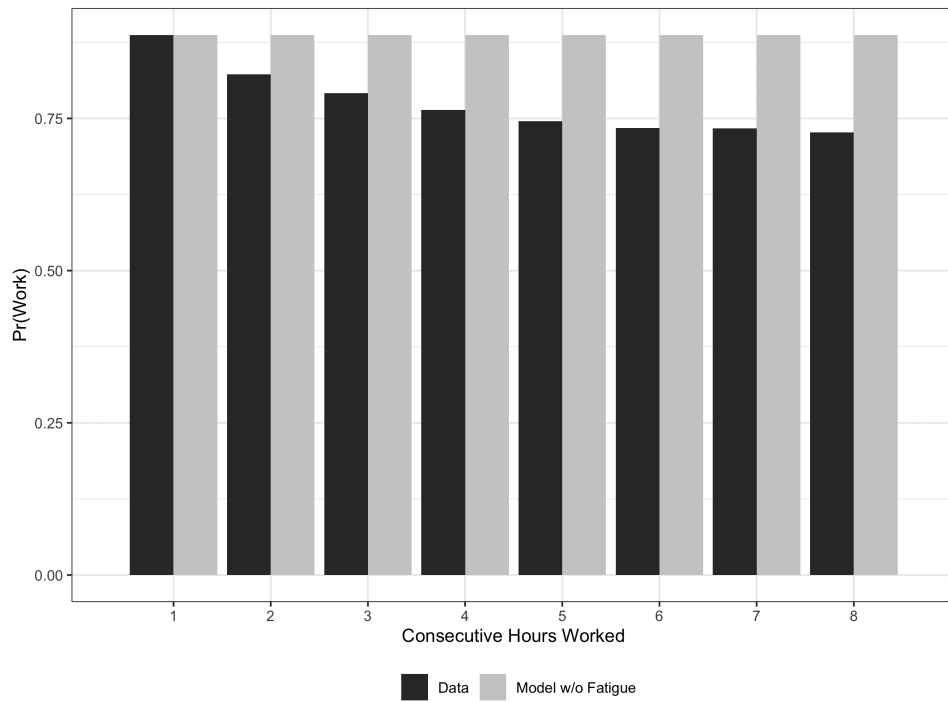


Figure 18: Model Fit of Probability of Working by Consecutive Hours Worked

Notes: We define consecutive hours worked as the total number of hours a driver works in a row. The consecutive hour resets to zero whenever a driver stops driving in an hour.

Estimation Results. In Table 5, we present the parameter estimates. The estimate of γ , which captures the drivers' sensitivity to wage changes, are positive across the four subgroups. However, the magnitudes vary across the groups. Male drivers are more responsive to exogenous wage changes than female drivers. Consider, for instance, young male drivers. All else equal, a 1% increase in market wages induces young male drivers to increase their probability of working by around 0.75%.

Our estimation shows considerable dispersion in the value of time $\beta_{h(t)}$ across hours of the week for each of the four subgroups. Recall that a large $\beta_{h(t)}$ in absolute value corresponds to a high cost of driving at $h(t)$, and consequently a low probability of working, all else equal. On average, the absolute value of the estimated $\beta_{h(t)}$ at the 90th percentile is twice as large as the absolute value of the estimated $\beta_{h(t)}$ at the 10th percentile.

We also find significant adjustment costs in our model as captured by the estimate of μ . For example: Among young male drivers at 8:00 a.m. on Monday, those who worked at 7:00 a.m. have a predicted probability of working around 0.86, while the probability is only 0.07 for those who did not work at 7:00 a.m. This state-dependency emphasizes the importance of incorporating the dynamic component in the driver's decision problem.

Our model allows for three unobserved types among drivers. The type parameter, η , shifts the cutoff in the work choice equation. We normalize η to 0 at all hours of the week for drivers of the baseline type. Thus, high η shifts up the value of work and thus increases the probability of working relative to the baseline type.

Our estimates of the η 's suggest three types of drivers. One type is likely to drive in the evening. Another type drives frequently both during the day and at night. The third type, the baseline type, drives infrequently. Evening drivers ($\eta_{1,Night}, \eta_{1,Day}$) have a much higher cutoff of work at night than frequent drivers ($\eta_{2,Night}, \eta_{2,Day}$). Across all the four demographic groups of drivers, the estimated $\eta_{1,Night}$ is almost twice as large as the estimate of $\eta_{2,Night}$. Consider, for example, young male drivers. Conditional on not working in the previous period, the model predicts evening drivers have a probability of working as high as 0.08 at midnight 12:00 a.m., whereas frequent drivers' work probability is only 0.01 and infrequent drivers' work probability is close to zero.

Table 5 also reveals that the estimates of the correction term $\frac{\rho_{u\xi}}{\sigma}$ are negative in all four demographic groups. Recall that $\rho_{u\xi}$ is the correlation coefficient between the preference shock ξ_{it} and the wage component u_{it} . Our estimation results indicate that the costs of working tend to co-move with the market wages. As shown later, it is important to take this endogeneity into account to obtain reliable estimates of the preference parameters.

			Weighted Average	Old Male	Young Male	Old Female	Young Female
Preference for Wage	γ		0.036	0.040	0.042	0.024	0.021
Time Preferences	β	$E[\beta_{h(t)}]$	-1.544	-1.612	-1.760	-1.257	-0.977
		$Sd(\beta_{h(t)})$	0.513	0.516	0.462	0.597	0.570
		$Median(\beta_{h(t)})$	-1.328	-1.375	-1.598	-0.972	-0.734
		$q_{10}(\beta_{h(t)})$	-2.295	-2.428	-2.351	-2.164	-1.861
		$q_{90}(\beta_{h(t)})$	-1.055	-1.151	-1.301	-0.653	-0.432
Adjustment Cost	μ		-6.377	-6.191	-6.269	-6.771	-6.864
Unobserved Types	η						
		$\eta(1, Night)$	1.857	1.847	1.901	2.039	1.571
		$\eta(1, Day)$	0.601	0.589	0.669	0.547	0.483
		$\eta(2, Night)$	1.178	1.120	1.237	1.288	1.067
		$\eta(2, Day)$	0.794	0.773	0.802	0.838	0.788
Selection Term	$\frac{\rho_{u\xi}}{\sigma}$		-0.039	-0.043	-0.044	-0.028	-0.024

Table 5: Estimates of Model Parameters

Notes: “Weighted average” is calculated by averaging the estimates of the four demographic groups weighted by the share of the drivers. “Young” is defined as those whose ages are less than or equal to the median age.

Implied Labor Supply Elasticities. To interpret the magnitude of the preference parameters, we use our estimated model to calculate the labor supply elasticity of each driver in every hour of the week. We define the labor supply elasticity in our model as the percent change in the probability of working in a given hour of the week for an anticipated and temporary one percent exogenous increase in the hourly wage. Formally, we follow Dagsvik (2020) and define the labor supply elasticity on the extensive margin as:

$$\frac{\partial Pr(a_{it} = 1|s_{it})}{\partial w_{it}} \frac{w_{it}}{Pr(a_{it} = 1|s_{it})} = f(V(a_{it} = 1, s_{it}) - V(a_{it} = 0, s_{it}) - (\nu_{it} - \epsilon_{0it})) \times \frac{\partial(V(a_{it} = 1, s_{it}) - V(a_{it} = 0, s_{it}) - (\nu_{it} - \epsilon_{0it}))}{\partial w_{it}} \frac{w_{it}}{Pr(a_{it} = 1|s_{it})} \quad (3)$$

where $V(\cdot)$ and s_{it} are the value function and the vector of state variables, ν_{it} and ϵ_{0it} are idiosyncratic shocks, and f is the probability density function of the choice which follows the logistic distribution (see Section 4). To calculate the elasticity, we first use the estimated model parameters to recover the value function V for each state vector s_{it} and choice a_{it} . With the value functions, we then compute the elasticities evaluated at the mean wage for every possible realization of the remaining state variables, using equation (3). We average the resulting elasticities weighted by the share of the states in the data and report them in Table 6.

Our labor supply elasticities range from 0.19 to 0.66, with a weighted average of 0.46. On average, male drivers have higher labor supply elasticities than female drivers. Even conditional on observables, there is substantial variation in the elasticity: Infrequent drivers have the highest elasticities, evening drivers the lowest. By way of comparison, our model implies much smaller differences in the labor supply elasticity over time than across drivers.

Panel A: Observed Heterogeneity					
	Weighted Average	Daytime	Evening	Weekday	Weekend
Old male drivers	0.48	0.47	0.50	0.45	0.56
Young male drivers	0.59	0.57	0.62	0.56	0.66
Old female drivers	0.23	0.23	0.24	0.21	0.28
Young female drivers	0.24	0.20	0.32	0.23	0.24
Panel B: Unobserved Types					
	Weighted Average	Daytime	Evening	Weekday	Weekend
Frequent drivers	0.31	0.35	0.23	0.27	0.44
Evening drivers	0.21	0.22	0.19	0.20	0.25
Infrequent drivers	0.52	0.48	0.60	0.51	0.54

Table 6: Extensive Margin Labor Supply Elasticities of an Anticipated and Temporary Percent Increase in Hourly Wages

Notes: “Weighted Averages” is calculated by averaging the estimates of the four demographic groups weighted by the share of the drivers. Young is defined as age less than or equal to the median age.

5 Insights from the Model

We now use the estimated model to compute the reservation wages and to perform counterfactual analyses. These analyses allow us to infer the drivers’ willingness to pay for the ability to customize and adjust their work schedule, and to examine how preference heterogeneity and adjustment costs influence the effectiveness of wage incentives that Uber can offer.¹³

5.1 Value of Time and Reservation Wages

A key objective of the model is to recover the reservation wages and study how they vary over time and across people. We start with presenting the value of time, $\beta_{h(t)}$, which is an important component of the reservation wages.

Figure 19a plots the weighted average of $\beta_{h(t)}/\beta_{\text{Saturday8a.m.}}$ across the four demographic subgroups. Our findings reveal that the value of time varies systematically during a typical week, with the value peaking at late nights around 4 a.m. To examine whether the variation mostly comes from hours of a day or days of a week, Figure 19b shows the average $\beta_{h(t)}$ per hour of the day (relative to 8 a.m.), while Figure 19c reports the average $\beta_{h(t)}$ per day of the week (relative to Saturday). The value of time varies a lot across hours within a day. By way of comparison, there is little variation in the value of time across weekdays. The value of time tends to be higher during weekends than weekdays, but the differences are rather small.

In Figure 20, we compare the value of time across the demographic groups. For each group, we normalize

¹³ Throughout the counterfactual analyses, we abstract from how the market wages may be affected by changes in the labor supply of the drivers. Taking into account such effects would require data on and a model of the demand side of the market.

the estimates of $\beta_{h(t)}$ by the value of time for older males on Saturdays at 8 a.m. Figure 20 shows that gender is the key dimension of observable heterogeneity when it comes to the value of time.

Knowledge of the value of time $\beta_{h(t)}$ is necessary but not sufficient to draw inference about reservation wages. We also need to take into account the unobserved preference component ξ_{it} , which may correlate with market wages. In Figure 21, we compute and plot the reservation wages against the expected market wages for an average driver during a typical week. The reservation wages stay low during the day, start to increase in the late evening, peak around 4 a.m., and then gradually decline until 9 a.m. The reservation wages are slightly higher during weekends compared to weekdays. While the market wage and the reservation wage tend to co-move across hours, the levels differ significantly. In particular, the average reservation wage is always considerably higher than the average market wage. This finding explains why no more than 4 percent of the drivers are choosing to work in an average hour during the week.

The reservation wages vary not only over time but also across people. Panel A of Table 7 presents the reservation wages across the four demographic groups and Panel B presents the reservation wages across the three types of drivers. We normalize the reservation wages by the daytime reservation wage of older male drivers. On average, the female drivers' reservation wages are twice as high as the male drivers' reservation wages. Furthermore, young drivers tend to have slightly higher reservation wages than older drivers. There is also a lot of heterogeneity across drivers conditional on age and gender. The frequent drivers have half as large reservation wages as compared to the infrequent drivers, and the evening drivers have a lower reservation wage in the evenings as compared to the daytime.

5.2 The Value of the Ability to Set Customized Work Schedules

Another key objective of the model is to infer the value of job flexibility. One important source of job flexibility is the driver's ability to set a customized work schedule, so she can plan when to work based on her expectations about reservation wages relative to market wages. We quantify the value of this job flexibility by restricting the driver's ability to set the work schedule. In particular, we remove certain hours of a day or days of a week from the choice set of a driver, and then solve the driver's problem given this restricted choice set.

For computational reasons, it is useful to consider a situation where each person drives a given number of consecutive hours per week. In our counterfactual analyses, we let each driver work 5 consecutive hours per week. This choice matches the average number of work hours among the drivers in our estimation sample. At the beginning of a week, each driver is required to choose one 5-hour block to work. We consider two scenarios: The benchmark case and the restricted case. In the benchmark case, drivers can choose among all possible 5-hour blocks. In the restricted case, we remove certain blocks from the choice set. The only difference between the benchmark case and the restricted case is the restriction on the drivers' choice set. Thus, by comparing a driver's utility in these two cases, we can calculate the wage multiplier that she would need to accept a restricted choice set.

In Figure 22, we plot the average wage multipliers that the drivers would demand to accept various restrictions on the choice set. In the left panel, we remove the drivers' preferred hours from their choice set. In the calculations behind the first bar, we remove the favorite 5-hour block. The resulting wage multiplier is 1.05. In other words, the average worker would require 5 percent higher wages to accept such a restriction on

the choice set. In the second bar, we remove the entire day containing the favorite 5 hours from the driver's choice set. The wage multiplier barely increases. In the third bar, we remove the preferred 5 hours from each day of the week. The wage multiplier now exceeds 1.1.

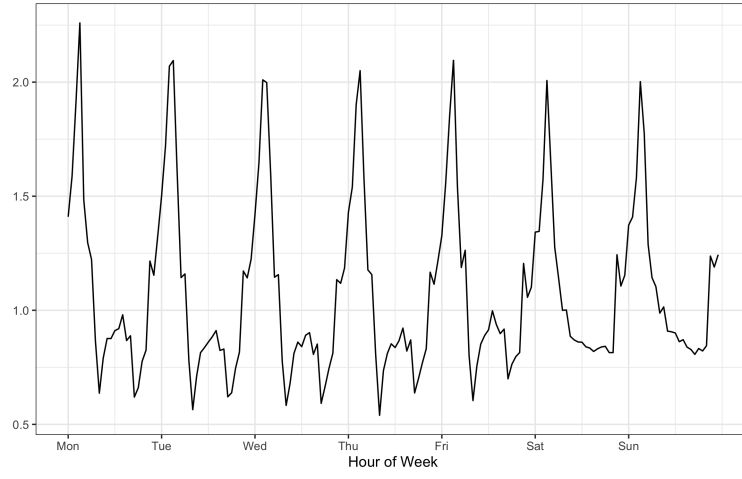
In the second and the third panel, we restrict the choice set of all drivers to certain days of a week or certain hours of a day. We find that removing the morning block (7 a.m. to 1 p.m.) results in a small wage multiplier. This is because relatively few drivers prefer to work during this time of the day. In contrast, the drivers would demand a large wage multiplier if the evening block (7 p.m. to midnight) would be removed from the choice set.

In Figure 23-25, we plot the wage multipliers by gender, age, and type of driver. On average, young drivers and female drivers require relatively high multipliers. This suggests that the ability to set customized work schedules is more important for these groups. We also find that frequent drivers and especially evening drivers would require large multipliers to accept restrictions on the possibility of driving in the evening and at night.

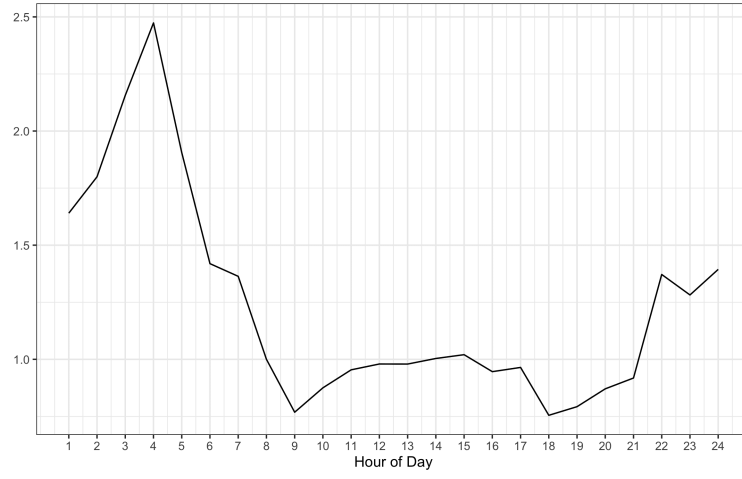
Panel A: Observed Heterogeneity				
	Daytime	Evening	Weekday	Weekend
Old male drivers	1.00	1.28	1.09	1.11
Young male drivers	1.08	1.28	1.14	1.16
Old female drivers	1.93	2.42	2.08	2.11
Young female drivers	2.35	2.88	2.51	2.57
Panel B: Unobserved Types				
	Daytime	Evening	Weekday	Weekend
Frequent drivers	1.00	1.54	1.17	1.20
Evening drivers	1.26	1.12	1.24	1.15
Infrequent drivers	2.33	2.73	2.45	2.49

Table 7: Reservation Wages by Driver Types and Time

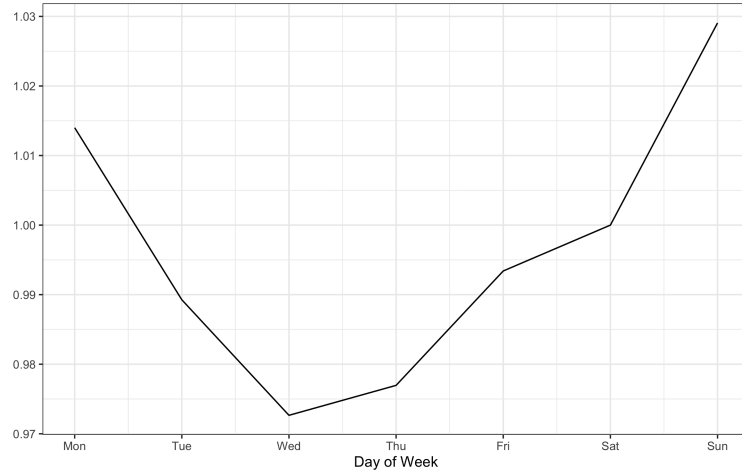
Notes: In Panel A, the reservation wage estimates are normalized by the mean reservation wage of old male drivers during the daytime. In Panel B, the reservation wage estimates are normalized by the mean reservation wage of frequent drivers during the daytime. The daytime is defined as 6 a.m.-9 p.m., and the evening is defined as 10 p.m.-5 a.m.



(a) $\beta_{h(t)}$ Relative to Saturday 8 a.m.



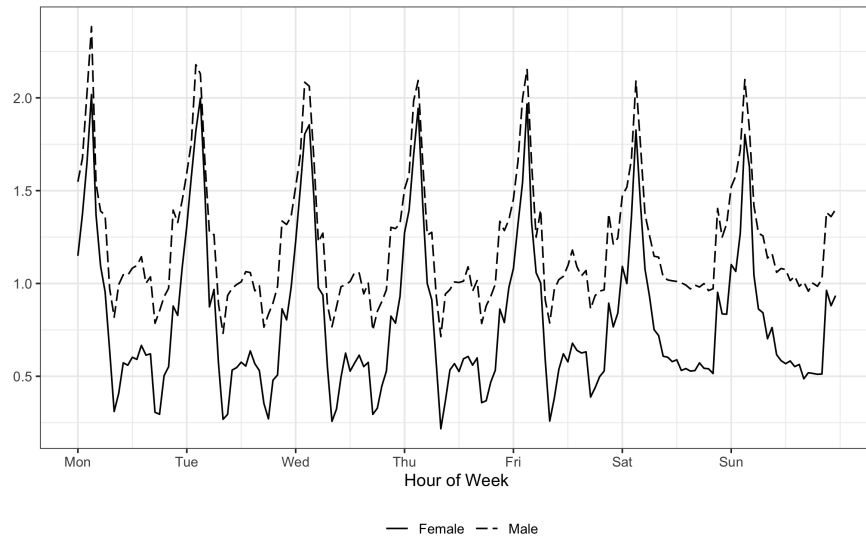
(b) $\beta_{h(t)}$ Relative to 8 a.m.



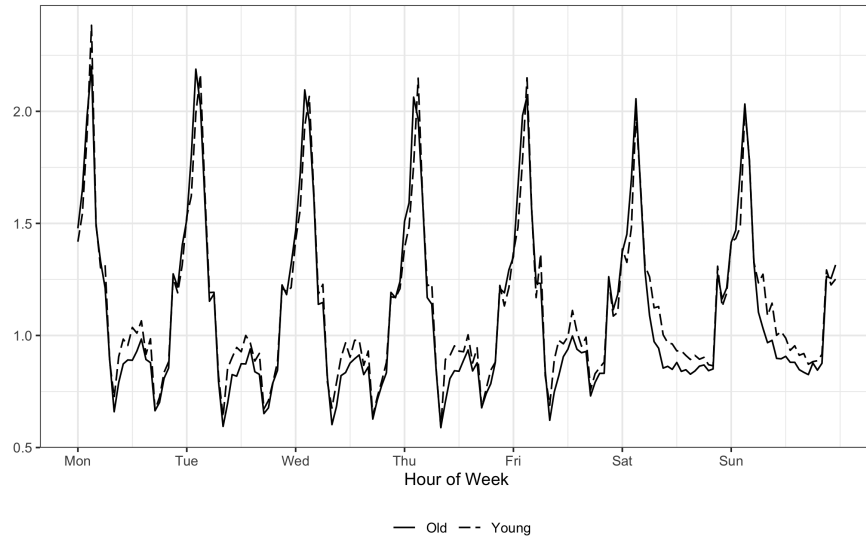
(c) $\beta_{h(t)}$ Relative to Saturday

Figure 19: Variation in the Value of Time, $\beta_{h(t)}$, over Time

Notes: In Figure (a), we compute for each hour of the week the weighted average of $\beta_{h(t)}/\beta_{Saturday8a.m.}$ using population shares of each demographic group. In Figure (b), we compute the average $\beta_{h(t)}$ per hour of the day (relative to 8 a.m.) weighted by the shares of each demographic group. In Figure (c), we compute the average $\beta_{h(t)}$ per day of the week (relative to Saturday) weighted by the population shares of each demographic group.



(a) Gender



(b) Age

Figure 20: Variation in the Value of Time, $\beta_{h(t)}$, Across Demographic Groups

Notes: We compute the weighted average of $\beta_{h(t)}$ relative to old males' value of time on Saturday 8 a.m. using population shares as weights.

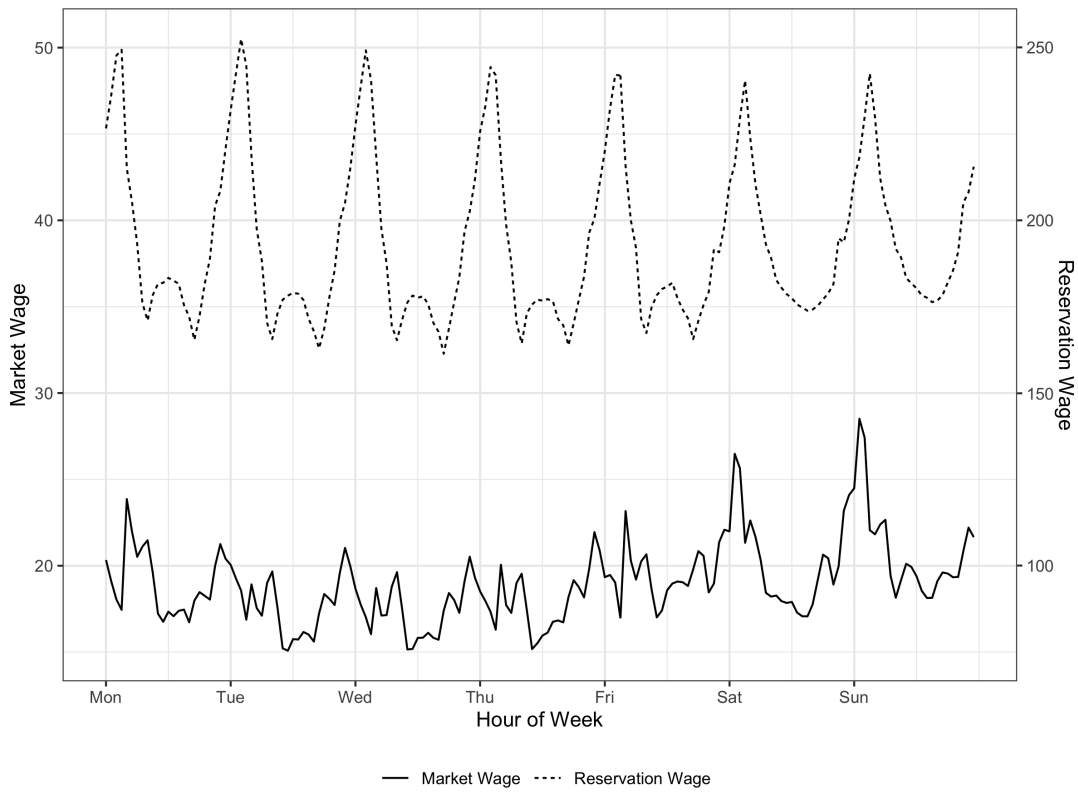


Figure 21: Comparison of Observed Market Wages and Reservation Wages

Notes: In this figure, we compute the reservation wage as the minimal wage needed to work given mean preference shocks and averaged across the state variables. The y-axis on the left represents the scale of the market wages, while the y-axis on the right represents the scale of the reservation wages.

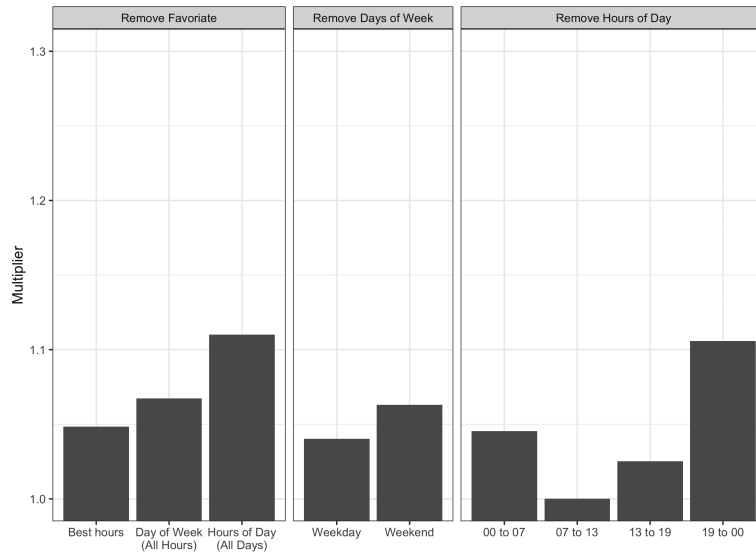


Figure 22: Wage Multipliers to Accept Restrictions on the Choice Set

Notes: The left panel removes each drivers' favorite 5-hour block from their choice sets. The "Best hours" indicates the best 5-hour block of the week for the drivers. The middle panel removes weekdays or weekends from drivers' choice sets. The right panel removes certain 5-hour blocks across the entire week from drivers' choice sets.

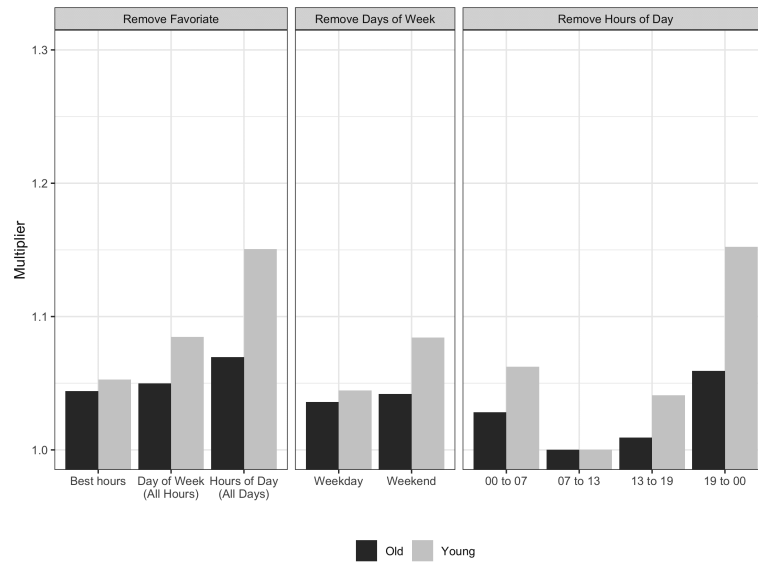


Figure 23: Wage Multipliers to Accept Restrictions on the Choice Set by Age

Notes: The left panel removes each drivers' favorite 5-hour block from their choice sets. The "Best hours" indicates the favorite 5-hour block of the week for the drivers. The middle panel removes weekdays or weekends from drivers' choice sets. The right panel removes certain 5-hour blocks across the entire week from drivers' choice sets.

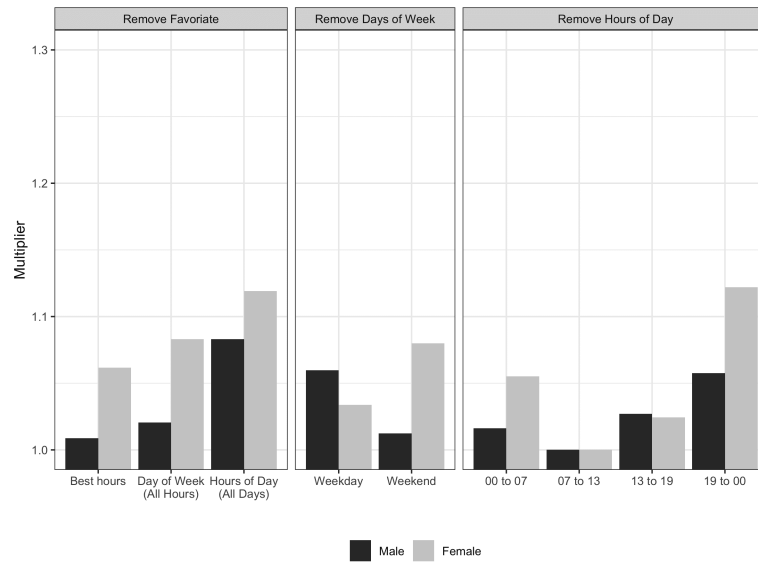


Figure 24: Wage Multipliers to Accept Restrictions on the Choice Set by Gender

Notes: The left panel removes each drivers' favorite 5-hour block from their choice sets. The "Best hours" indicates the favorite 5-hour block of the week for the drivers. The middle panel removes weekdays or weekends from drivers' choice sets. The right panel removes certain 5-hour blocks across the entire week from drivers' choice sets.

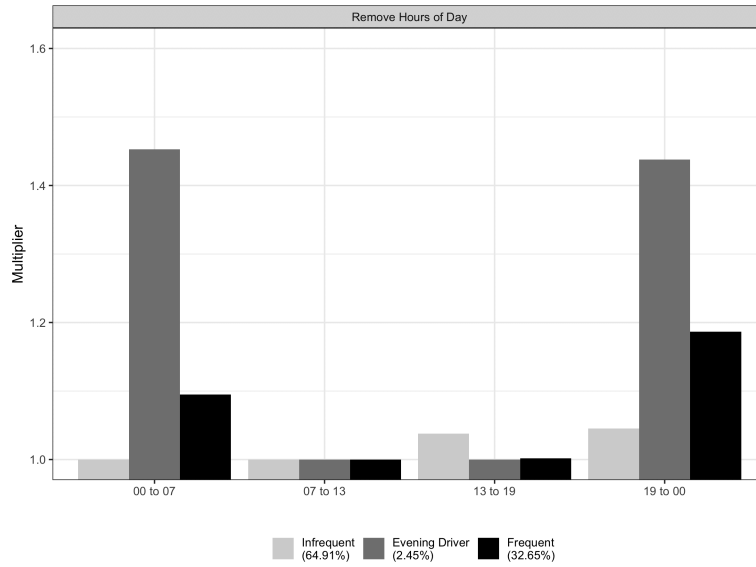


Figure 25: Wage Multipliers to Accept Restrictions on the Choice Set Across Unobserved Types

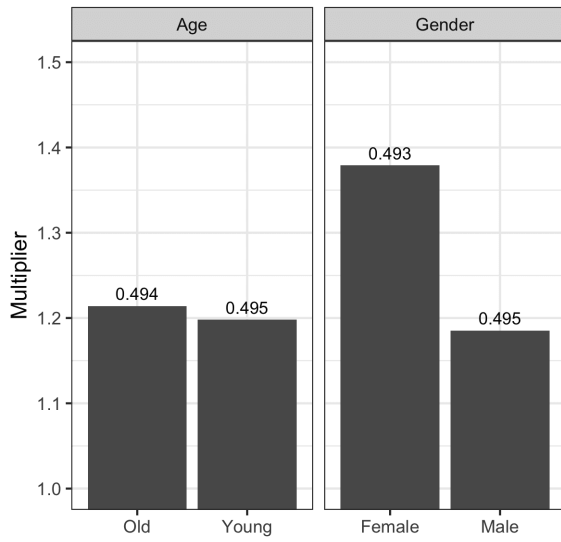
Notes: The left panel removes each drivers' favorite 5-hour block from their choice sets. The "Best hours" indicates the favorite 5-hour block of the week for the drivers. The middle panel removes weekdays or weekends from drivers' choice sets. The right panel removes certain hour blocks across the entire week from drivers' choice sets. The numbers in the parentheses indicate the fraction of each type of drivers.

5.3 The Value of the Ability to Adjust Work Schedules

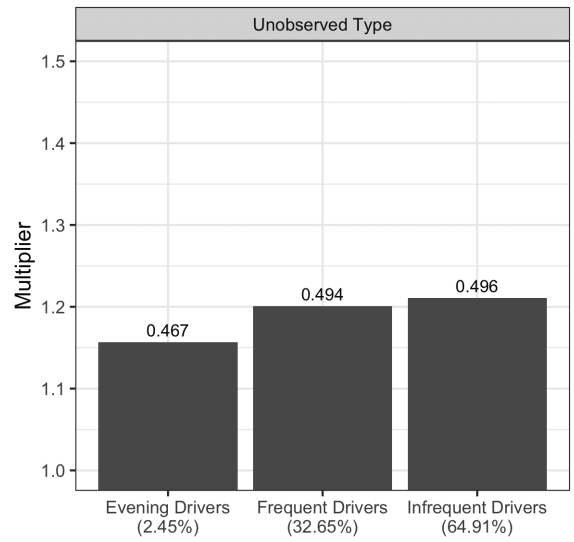
Another source of job flexibility comes from the drivers' ability to adjust work schedules from day to day or even hour to hour in response to unanticipated changes to market wages or the costs of driving. We quantify the value of this job flexibility by forcing the driver to commit to a work schedule before observing the realizations of the innovations to wages and preferences. In particular, we compare the expected values of the two distinct types of work arrangements. The first is a commitment scheme where each driver commits to working the 5-hour block that gives her the highest expected utility. The alternative is a flexible scheme in which each driver is allowed to adjust their choice of 5-hour block once she observes the realizations of the innovations to wages and preferences. By comparing a driver's utility in these two cases, we can calculate the wage multiplier that she would need to accept the commitment scheme instead of the flexible scheme.

In Figure 26, we report the average wage multipliers that the drivers would demand to accept the commitment scheme. For now, each driver is only allowed to adjust her work schedule once per week in the flexible scheme. On average, about half of the drivers would use this flexibility and change their work schedule due to preference or wage shocks. The wage multiplier needed for drivers to accept the commitment scheme is 1.21. In other words, the average worker would require 21 percent higher wages to prefer the commitment scheme over the flexible scheme. Female drivers place a higher value on the flexibility of adjustment to shocks than male drivers; they need a multiplier of 1.38, while male drivers only require 1.19.

Next, we examine how the wage multipliers change as we increase the number of times drivers are allowed to adjust to shocks. In this analysis, we compare the commitment scheme to flexible schemes where drivers are allowed to adjust once, or twice, or three times within the next week. The results are presented in Figure 27. The wage multiplier approaches 1.4 when we compare the commitment scheme to a flexible scheme where each driver can make three adjustments to her weekly work schedule as she observes the realizations of her wage and preference shocks.



(a) Demographic Groups



(b) Unobserved Types

Figure 26: Wage Multipliers to Accept the Commitment Scheme Instead of the Flexible Scheme

Notes: In Figure (a), we compute the weighted average of the wage multipliers needed for drivers in each demographic group to be indifferent between the flexible scheme and the commitment scheme. In Figure (b), we compute the wage multipliers needed for drivers of each unobserved type to be indifferent between the flexible scheme and the commitment scheme. The estimates on the bars show the fraction of drivers that will switch to the second-best 5-hour block.

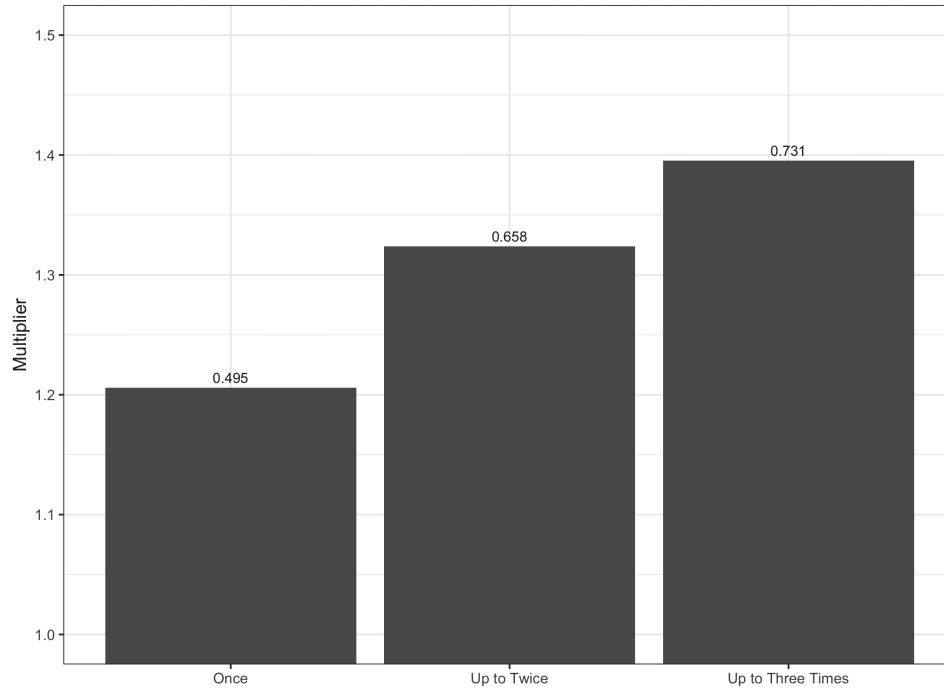


Figure 27: Wage Multipliers to Accept the Commitment Scheme as Compared to Flexible Schemes where Drivers may Adjust Work Schedules Once, Twice or Three Times

Notes: In this figure, we compute the weighted average of the wage multipliers needed for drivers to be indifferent between the commitment scheme and the three types of flexible schemes (where drivers are allowed to adjust to shocks once, twice, and three times). The estimates on the bars show the fraction of drivers that will switch away from the first best window.

5.4 The Value of the Ability to Both Customize and Adjust Work Schedules

We now quantify the value of the ability to both customize and adjust work schedules. To this end, we consider three types of work schedules: (1) a fully fixed work schedule where drivers are assigned a particular work schedule and cannot adjust, (2) a commitment scheme where workers can customize but not adjust work schedules, and (3) a flexible scheme where workers can both customize and adjust work schedules. Comparing (3) to (1) allows us to infer the total value of the ability to both customize and adjust work schedules, while the comparison between (3) and (2) isolates the value of the ability to adjust work schedules. In each case, every person drives 5 consecutive hours per week. For now, we only let the drivers make one adjustment per week in (2) and (3). In (1), we fix the work schedule of each driver to be Friday evening, 7:00 -11:59 p.m.

Figure 28 presents the results. The total value of the ability to both customize and adjust work schedules corresponds to a 43 percent increase in wages. About half of this wage increase can be attributed to the value of the ability to adjust work schedules. In Figure 29, we compare the wage multiplier needed by gender, age, and type of driver. Most heterogeneity arises from the gender and the type of drivers. Female and evening drivers place a particularly high value on work flexibility.

Next, we examine how the wage multipliers change as we increase the number of times drivers are allowed to adjust in (3). The results are presented in Figure 30. The wage multiplier exceeds 1.45 when we compare the commitment scheme to a flexible scheme where each driver can make many adjustments to her weekly work schedule as she observes the realizations of her wage and preference shocks.

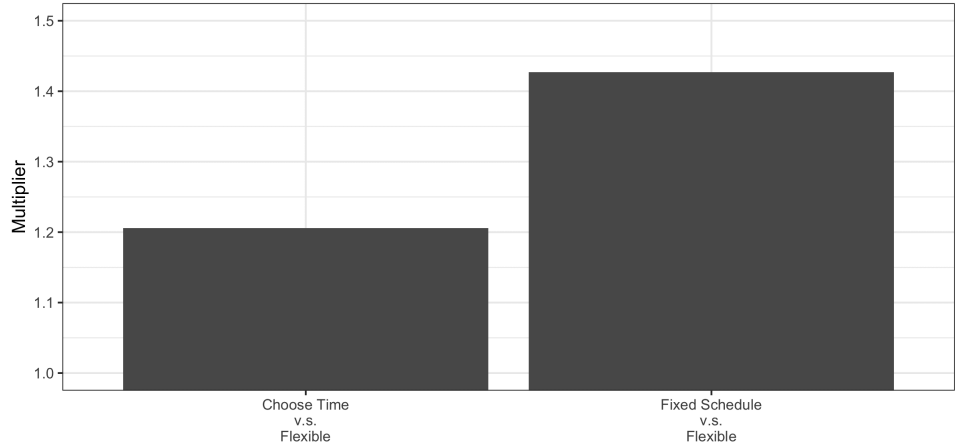
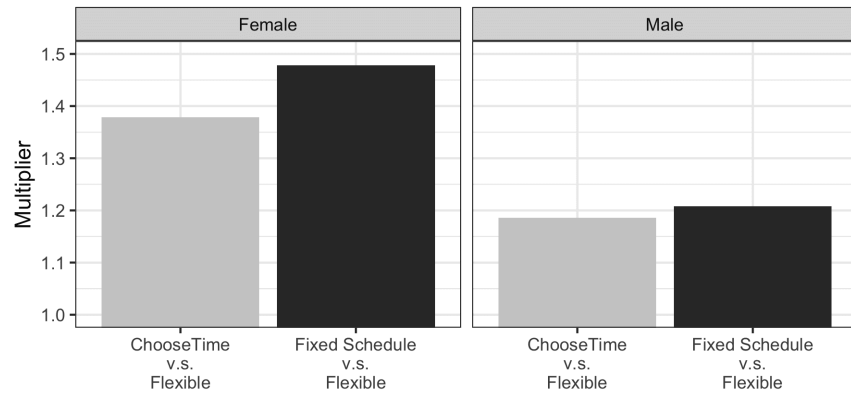
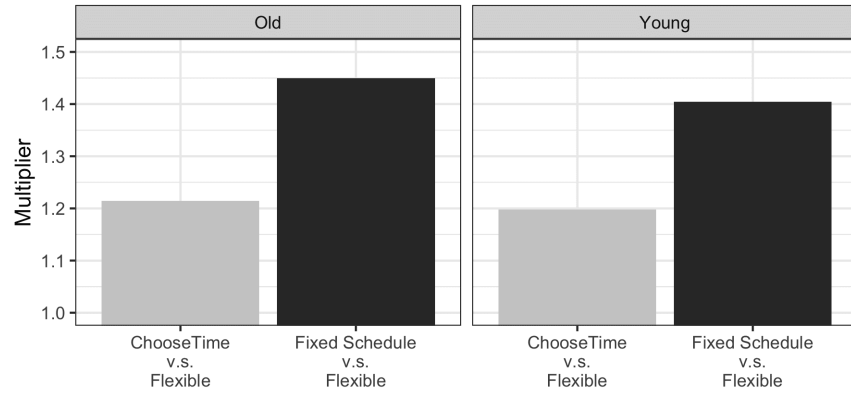


Figure 28: Wage Multipliers to Accept Restrictions on the Choice Set and Work Schedule Adjustments

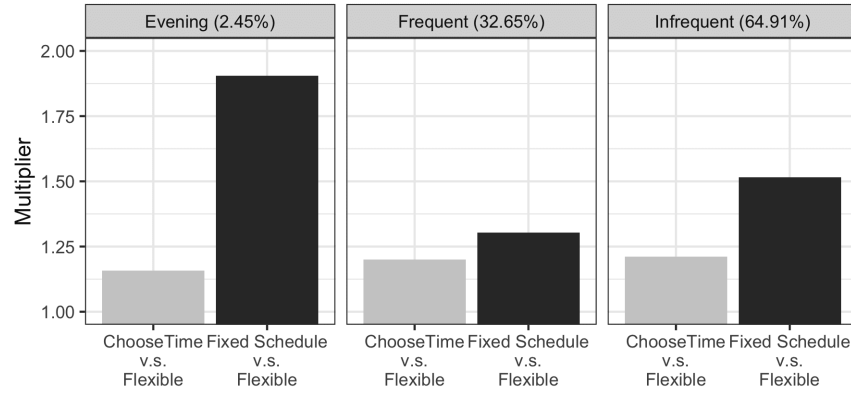
Notes: In this figure, we compute the weighted average of the wage multipliers needed for drivers to be indifferent between the different types of work schedules. The fixed work schedule starts from 7:00 p.m. on Friday and ends at 11:59 p.m. on Friday.



(a) Gender



(b) Age



(c) Unobserved Types

Figure 29: Heterogeneity in Wage Multipliers to Accept Restrictions on the Choice Set and Work Schedule Adjustments

Notes: In these figures, we compute the wage multipliers needed for drivers to be indifferent between the different types of work schedules. The fixed work schedule starts from 7:00 p.m. on Friday and ends at 11:59 p.m. on Friday.

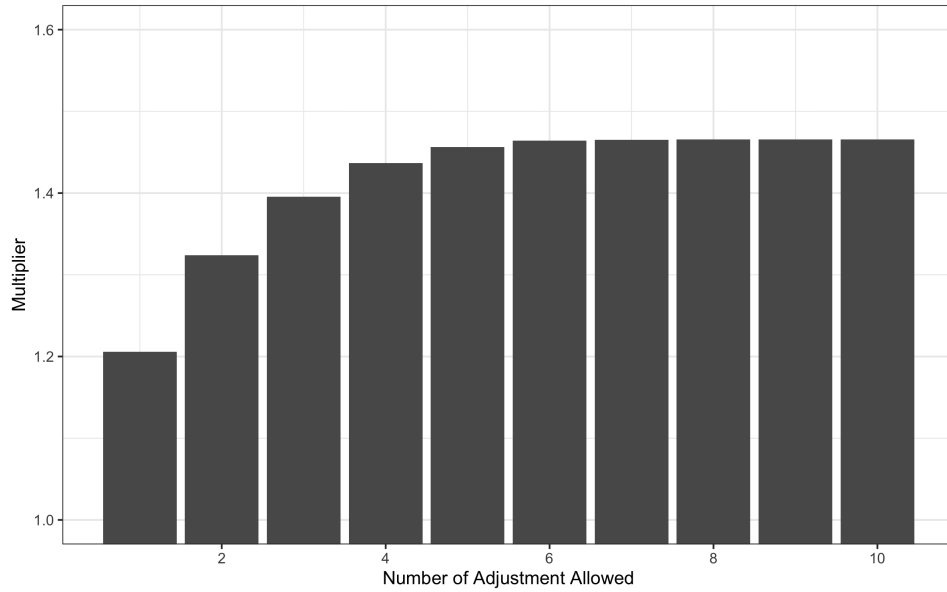


Figure 30: Wage Multiplier with Increasing Number of Work Schedule Adjustments

Notes: In this figure, we compute the wage multipliers needed for drivers to be indifferent between the commitment scheme and the flexible scheme with increasing numbers of adjustments.

5.5 Effectiveness of Work Incentives

The last set of counterfactuals that we perform is aimed at quantifying the effectiveness of work incentives to drivers. Ride-sharing companies regularly use driver incentives to encourage drivers to work. The presence of preference heterogeneity and adjustment costs can be important to accurately predicting the effectiveness of such work incentives. Given the preference heterogeneity over time, it is more costly to incentivize drivers to drive late at nights than during the daytime and during weekdays than weekends. Due to the adjustment costs, it is costly for drivers to begin to drive, and therefore cheaper to encourage drivers to continue driving than to induce drivers to start to drive.

In Figure 31, we report the wage increase that would be needed to increase employment to a pre-specified amount (from a baseline probability of working of 0.042 per hour). We decompose the increases in employment into the part that is due to additional drivers starting to drive (new drivers) and the part that is due to more drivers continuing to drive (continuing drivers). It takes a \$9.55 increase in hourly wage (38% increase as compared to the average wage) to increase employment per hour by 0.5 percentage point. Continuing drivers are more responsive since they do not have to pay the adjustment cost. However, relatively few drivers are working at any given point in time. Thus, most of the increase in employment comes from additional drivers starting to drive.

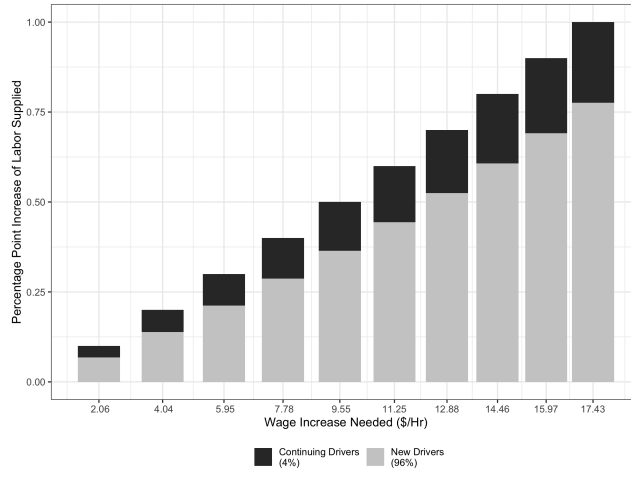


Figure 31: Wage Increases Needed to Incentivize Drivers

Notes: The y-axis denotes the target percentage point increase in labor supply, and the x-axis denotes the increase in hourly wages needed. We decompose the wage increase needed based on the lagged choice. The numbers in the parentheses indicate the fraction of the drivers in each group.

In Figure 32-34, we plot the increase in hourly wages that is needed to achieve a 0.5 percentage point increase in employment in a given hour of the week, hour of the day, and day of the week. The results mirror the findings for reservation wages. It is cheaper to incentivize drivers during the daytime than late at nights. It is also cheaper to incentivize drivers during the weekend than weekdays, although these differences are relatively modest.

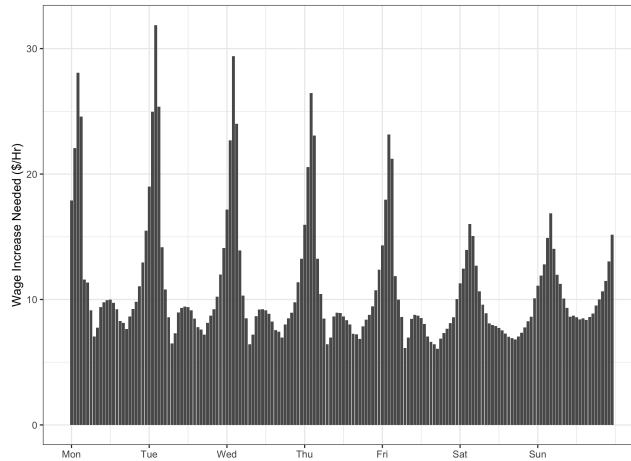


Figure 32: Wage Increases Needed to Incentivize Drivers by Hours of the Week

Notes: The y-axis denotes the increase in hourly wage needed for a 0.5 percentage point increase in labor supply.

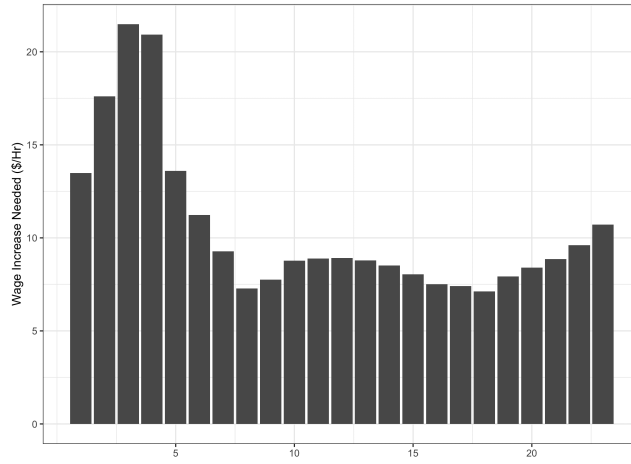


Figure 33: Wage Increases Needed to Incentivize Drivers by Hours of the Day

Notes: The y-axis denotes the increase in hourly wages needed for a 0.5 percentage point increase in labor supply.

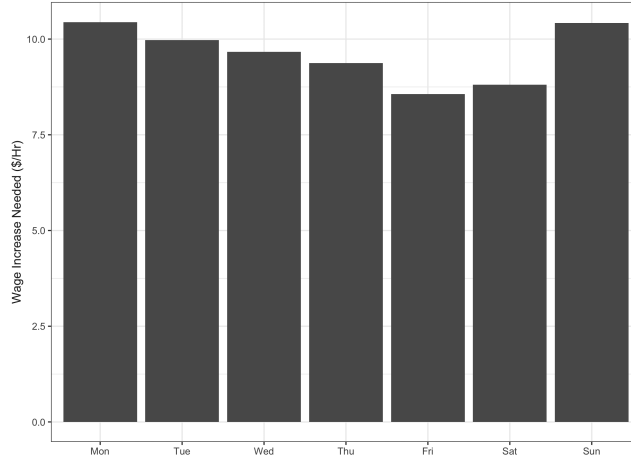


Figure 34: Wage Increases Needed to Incentivize Drivers by Days of the Week

Notes: The y-axis denotes the increase in hourly wages needed for a 0.5 percentage point increase in labor supply.

5.6 Substitution Possibility to Lyft

In the period and cities we consider, Uber has a large majority of the U.S. consumer ride-sharing market. Nevertheless, one might be worried that our estimates are affected by substitution between Lyft and Uber.

To assess this, we perform two sets of analyses. First, we examine if our findings differ systematically in cities or in periods in which Lyft has relatively high or low market share. Second, we compare the results for all drivers to those we obtain from a subsample of drivers who are ineligible to drive for Lyft. The results are reported in the Appendix. In each set of analyses, we find that neither the experimental estimates nor the estimated model parameters depend materially on the possibility of substitution between Lyft and Uber.

6 Conclusion

Over the years labor markets have varied dramatically in both their flexibility for where and when the agent works. Just considering the U.S. in the past 75 years, World War II pushed more than 9 million women

who previously worked at home to enter factory and other office positions. Work life was a fixed daily and hourly routine completing inflexible tasks. The past decade has witnessed a shift back, with the Gig Economy providing much higher levels of flexibility, autonomy, and task variety. We leverage a natural field experiment at Uber to quantify how reservation wages and labor supply elasticities vary between people and over time, and to infer workers' valuation of flexibility in their choice of work hours. Economists and policymakers are keenly interested in these quantities, especially lately with the growth in jobs that offer flexible work schedules. Combining the experiment with high frequency panel data on wages and work decisions, we documented how labor supply responds to exogenous changes in offered wages in a setting with no restrictions on hours choices stemming from the demand side of the market. We found evidence of systematic heterogeneity in labor supply responses between people and over time, significant fixed costs to starting to drive, and high demand when it is costly for drivers to work.

These experimental findings motivated a dynamic model of labor supply with flexible heterogeneity in preferences over work schedules, start up costs, and the correlation between offered wages and cost of driving in a given period. The primitives of the model were recovered from a combination of the experimental estimates and other data moments. We used the estimated model to compute how labor supply elasticities and reservation wages vary between people and over time, and to perform counterfactual analyses. These analyses allowed us to infer drivers' willingness to pay for the ability to customize and adjust their work schedule, and to examine how preference heterogeneity and adjustment costs influence the effectiveness of wage incentives that Uber can offer.

References

- Angrist, J. D., S. Caldwell, and J. V. Hall (2017). Uber vs. taxi: A driver’s eye view. Technical report, National Bureau of Economic Research.
- Attanasio, O., P. Levell, H. Low, and V. Sánchez-Marcos (2018). Aggregating elasticities: intensive and extensive margins of female labour supply. Technical report, National Bureau of Economic Research.
- Beffy, M., R. Blundell, A. Bozio, G. Laroque, and M. To (2019). Labour supply and taxation with restricted choices. *Journal of Econometrics* 211(1), 16–46.
- Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approaches. In *Handbook of labor economics*, Volume 3, pp. 1559–1695. Elsevier.
- Chen, M. K., J. A. Chevalier, P. E. Rossi, and E. Oehlsen (2019). The value of flexible work: Evidence from uber drivers. *Journal of Political Economy*.
- Chernozhukov, V., I. Fernández-Val, and Y. Luo (2018). The sorted effects method: discovering heterogeneous effects beyond their averages. *Econometrica* 86(6), 1911–1938.
- Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review* 101(3), 471–75.
- Cohen, P., R. Hahn, J. Hall, S. Levitt, and R. Metcalfe (2016). Using big data to estimate consumer surplus: The case of uber. Technical report, National Bureau of Economic Research.
- Cook, C., R. Diamond, J. Hall, J. A. List, and P. Oyer (forthcoming). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *Review of Economics Studies*.
- Dagsvik, J. K. (2020). Marginal compensated effects and the slusky equation for discrete choice models.
- DellaVigna, S., J. A. List, and U. Malmendier (2012, 01). Testing for Altruism and Social Pressure in Charitable Giving *. *The Quarterly Journal of Economics* 127(1), 1–56.
- DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2016, February). Estimating social preferences and gift exchange at work. Working Paper 22043, National Bureau of Economic Research.
- Eriksson, T. and N. Kristensen (2014). Wages or fringes? some evidence on trade-offs and sorting. *Journal of Labor Economics* 32(4), 899–928.
- Farber, H. S. (2005). Is tomorrow another day? the labor supply of new york city cabdrivers. *Journal of political Economy* 113(1), 46–82.
- Farber, H. S. (2015). Why you can’t find a taxi in the rain and other labor supply lessons from cab drivers. *The Quarterly Journal of Economics* 130(4), 1975–2026.

- Hall, J. V. and A. B. Krueger (2018). An analysis of the labor market for uber’s driver-partners in the united states. *ILR Review* 71(3), 705–732.
- Hamermesh, D. S. (1999). Changing inequality in markets for workplace amenities. *The Quarterly Journal of Economics* 114(4), 1085–1123.
- Harrison, G. W. and J. A. List (2004). Field experiments. *Journal of Economic literature* 42(4), 1009–1055.
- He, H., D. Neumark, and Q. Weng (2019). Do workers value flexible jobs? a field experiment on compensating differentials. Technical report, National Bureau of Economic Research.
- Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies* 60(3), 497–529.
- Hunnicutt, B. K. (1996). *Kellogg’s six-hour day*, Volume 71. Temple University Press.
- Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* 77(1), 135–175.
- Katz, L. F. and A. B. Krueger (2019). The rise and nature of alternative work arrangements in the united states, 1995-2015. *ILR Review* 72(2), 382–416.
- Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature* 49(4), 961–1075.
- Killingsworth, M. R. and J. J. Heckman (1986). Female labor supply: A survey. *Handbook of labor economics* 1, 103–204.
- Laibson, D. and J. A. List (2015, May). Principles of (behavioral) economics. *American Economic Review* 105(5), 385–90.
- Lamadon, T., M. Mogstad, and B. Setzler (2019). Imperfect competition, compensating differentials and rent sharing in the us labor market. Technical report, National Bureau of Economic Research.
- Maestas, N., K. J. Mullen, D. Powell, T. Von Wachter, and J. B. Wenger (2018). The value of working conditions in the united states and implications for the structure of wages. Technical report, National Bureau of Economic Research.
- Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete decision processes. *Econometrica* 70(2), 801–816.
- Mas, A. and A. Pallais (2017). Valuing alternative work arrangements. *American Economic Review* 107(12), 3722–59.
- Oyer, P. (2016). The independent workforce in america.
- Pencavel, J. (1986). Labor supply of men: a survey. *Handbook of labor economics* 1, 3–102.
- Pierce, B. (2001). Compensation inequality. *The Quarterly Journal of Economics* 116(4), 1493–1525.

- Rosen, S. (1986). The theory of equalizing differences. *Handbook of labor economics* 1, 641–692.
- Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.
- Santovec, M. L. (2014). Creating flexible work policies: One size does not fit all. *Women in Higher Education* 22(8), 1–2.
- Wiswall, M. and B. Zafar (2017). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics* 133(1), 457–507.

Appendix

A. Wage Prediction Model

We now describe how we construct the expected market wages. As described in Section 2.3, a driver’s expected hourly wage is the product of two components: a wage multiplier and the pre-multiplier wage. In the subsections below, we describe each component in detail.

A1. Wage Multipliers The wage multiplier augments a driver’s baseline compensation for driving. It is the maximum of the surge level and the GSL experimental wage multiplier, i.e. $\max(\text{GSL}, \text{Surge})$. However, since both GSL and surge levels vary over time and across places, this complicates how we construct the wage multipliers.

We illustrate how GSLs vary at the time×place level and how surge levels could limit the experimental variation in GSLs with Table 7. The table provides a simple example of a GSL menu for drivers in Chicago which was divided into five regions during most of our sample period: North, West, Loop, South, and Others.

Across all five regions, wherever the GSL experiment is switched on, the treated drivers receive 0.1 higher GSL multipliers than the control group drivers. However, the treatment effect on the wage multiplier is not always 0.1 for two reasons. First, the wage multiplier takes on the maximum value of GSL and the surge level for a given region and time. In regions where the surge levels are higher than the treatment group’s GSL level, all drivers in the region receive the same wage multiplier (i.e., the surge level) regardless of their treatment status. Second, there is spatial variation in GSLs and surge levels so GSLs may be higher than surge levels in some regions but not in the others. As a result, we compute the average of the maximum of GSL and surge levels for each driver, weight them by the driver’s fraction of time driving in each region in the past, and use this weighted average to predict the driver-specific wage multiplier at every hour.

	Region in Chicago				
	South	North	Loop	West	Others
Treatment	1.5×	1.3×	1.0×	1.1×	1.0×
Control	1.4×	1.2×	1.0×	1.0×	1.0×

Table 8: GSL Example in Chicago Across Regions

Notes: Within each region, the treatment group has a GSL multiplier 0.1× higher than the control group.

A2. Pre-multiplier Wages

We calculate the pre-multiplier wages as the observed hourly wages divided by the calculated wage multipliers. We then fit the following regression model to the panel data on the pre-multiplier wages:

$$\tilde{W}_{it} = \alpha_i + \kappa_{h(t)} + \epsilon_{it}$$

where t is an hour, $h(t)$ is the hour of the week at t , i denotes a driver, α_i and $\kappa_{h(t)}$ are driver and hour-of-week fixed effects, and \tilde{W} is the pre-multiplier hourly wage. For each hour t , we fit the model with the panel data

up to $t - 1$, and use the estimated fixed effects to compute a predicted value for \tilde{W} for every worker at every hour. The predicted hourly wage, \hat{w}_{it} , is then constructed as the product of the wage multiplier and the predicted pre-multiplier wage.

A3. Other Prediction Model and Cross-Validation

To assess how well our prediction model performs, we compare it to alternative approaches using a cross-validation procedure. This procedure repeatedly divides samples into a training sample and a testing sample. For each approach, we use the training samples to estimate a prediction model, and then we use the estimated model to form a prediction and calculate the out-of-sample mean squared errors (MSE) on the testing samples.

Formally, MSE is defined as the squared distance between the predicted hourly wage rate, \hat{w}_{it} , and the observed hourly wage rate w_{it} :

$$MSE = \frac{1}{N_1} \sum_{t=1}^T \sum_i^N (\hat{w}_{it} - w_{it})^2$$

where N_1 is the number of observed hourly wage w_{it} . The smaller the MSE is, the better the prediction model performs.

In addition to our current prediction model, we consider several alternative approaches with matching models. In these models, we consider two sets of observables for matching. The first set of observables consists of demographic variables, including gender and age. The second set of observables includes drivers' past driving histories.

For each set of observables in the matching models, we consider two ways to form the prediction. The first way uses the mean of the observed hourly wages in the previous week. For example, we predict young male drivers' hourly wages at 9 a.m. on a Monday by their mean hourly wages on the Monday 9 a.m. in the previous week. The second way of prediction uses the rolling average of the observed hourly wages. For example, we take average of young male drivers' observed hourly wages at 9 a.m. on all past Mondays to form the wage prediction.

One challenge in this procedure is that a driver's driving history is high-dimensional. Therefore, we conduct dimensionality reduction on driving histories. We consider two types of drivers' work histories, the histories on observed wages and the histories on minutes worked. We first obtain a 168-dimensional vector for each driver by averaging observed wages (or minutes worked across) the past weeks. Each element in the 168-dimensional vector corresponds to a driver's average wages (or minutes worked) in an hour of the week. Next, we cluster drivers based on these 168-dimensional vectors by the K-means clustering algorithm. Formally, let $\vec{a}_{it} = \frac{1}{t} \sum_{s=1}^t \vec{a}_{is}$ be a vector representing the average observed wages (or minutes worked) for each hour of the week up to t for driver i . Driver groups are obtained by solving the following optimization problem:

$$\arg \min_{k_1, \dots, k_N} \sum_{i=1}^N (a_{it-1} - \bar{a}_{k_{it-1}})^2$$

where $i \in \{1, \dots, N\}$ denotes a driver, $k_i \in \{1, \dots, K\}$ denotes driver i 's group, and $\bar{a}_{k_{it-1}}$ denotes the mean of a_{it-1} of the group k_i .

We present the results in Table 9. The demographic groups we consider include age and female dummies.

Specification		MSE
Fixed Effect Model		
Fixed Effects		0.067
Matching Model		
<i>Var. Matched On</i>	<i>Prediction From</i>	
Age \times Gender	Last Week	0.077
Age \times Gender \times History of Wage	Up to Last Week	0.088
Age \times Gender \times History of Wage	Last Week	0.067
Age \times Gender \times History of Minutes Worked	Up to Last Week	0.077
Age \times Gender \times History of Minutes Worked	Last Week	0.088

Table 9: Cross-Validation

Notes: MSE refers to the mean squared error of each prediction model. “Prediction From” indicates the periods used to form wage predictions.

The driving history includes drivers’ histories on observed hourly wages and histories on minutes worked. Our fixed effect wage prediction model outperforms most matching models in terms of out-of-sample mean squared errors. Given its performance and simplicity, we adopt the fixed effect model as our preferred wage prediction model.

B. Model Identification and Estimation

B1. Model Identification

A driver’s flow payoff associated with action a_{it} is given by:

$$U_{it} = \begin{cases} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \xi_{it} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{cases}$$

whereas the equation for the predicted hourly wage is specified as:

$$w_{it} = \delta_{h(t)} + \delta_0 m_t + \delta_1 m_t \times z_{it} + u_{it}$$

By Assumption 1, we can rewrite ξ_{it} as

$$\xi_{it} = \frac{\rho_{u\xi}}{\sigma} u_{it} + \psi_{it} = \frac{\rho_{u\xi}}{\sigma} (w_{it} - \delta_{h(t)} - \delta_0 m_t - \delta_1 m_t \times z_{it}) + \psi_{it}$$

where $\psi_{it} \sim \mathcal{N}(0, 1 - \rho_{u\xi}^2)$ and $\psi_{it} \perp u_{it}$ by construction. We define a new state variable $\phi_{it} \equiv w_{it} - \delta_{h(t)} - \delta_0 m_t - \delta_1 m_t \times z_{it}$. Thus, the flow payoffs of the problem become

$$U_{it} = \begin{cases} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \frac{\rho_{u\xi}}{\sigma} \phi_{it} + \nu_{it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{cases}$$

where $\nu_{it} = \psi_{it} + \epsilon_{1it}$. By further substituting w_{it} in the flow payoff with the wage equation, we have

$$U_{it} = \left\{ \begin{array}{ll} \gamma\delta_0 m_t + \gamma\delta_1 m_t \times z_{it} + (\beta_{h(t)} + \gamma\delta_{h(t)}) + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + (\frac{\rho_{u\xi}}{\sigma} + \gamma)u_{it} + \nu_{it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

It is useful to observe that the state variables can be either expressed as

$$(w_t, h(t), a_{it-1}, \phi_{it}, j(i)),$$

or equivalently,

$$(m_t, z_{it}, h(t), a_{it-1}, u_{it}, j(i)).$$

Note also that we can recover u_{it} from the wage equation, so u_{it} becomes an observed state in the driver's decision problem. Under Assumption 2, the only unobservables in the flow payoffs are ν_{it} , ϵ_{0it} , and $\eta_{j(i),h(t)}$. In the case where unobserved type $j(i)$ were observed, we can identify $(\gamma\delta_0, \beta_{h(t)} + \gamma\delta_{h(t)}, \mu, \frac{\rho_{u\xi}}{\sigma} + \gamma)$ given the observed state $(m_t, z_{it}, h(t), a_{it-1}, u_{it}, j(i))$. Since $\delta_{h(t)}$, δ_0 and δ_1 are identified from the wage equation, the structural parameters $(\gamma, \beta_{h(t)}, \mu, \frac{\rho_{u\xi}}{\sigma})$ are thus identified. In the case where $j(i)$ is unobserved, we apply (REACH) and (NTYPE) of Assumption 2 as in Kasahara and Shimotsu (2009) to jointly identify $\eta_{j(i),h(t)}$ and $(\gamma, \beta_{h(t)}, \mu, \frac{\rho_{u\xi}}{\sigma})$.

B2. Model Estimation

In this section, we provide details on the estimation procedure.

Wage Equation.

We start with estimating the following wage equation using OLS to obtain \hat{u}_{it} and the estimates of $\delta_{h(t)}$, δ_0 , δ_1 :

$$w_{it} = \delta_{h(t)} + \delta_0 m_t + \delta_1 m_t \times z_{it} + u_{it}.$$

Estimation of State Transition.

Next, we estimate the joint distribution of the state variables. Recall that we assume the state transition follows a first order Markov process, where the state vector contains hours of the week $h(t)$, treatment status z_{it} , wage component u_{it} , experiment hours m_t , lagged action a_{it-1} , and unobserved type $j(i)$. We also assume that drivers with different unobserved types share the same state transition probability.

Ideally, we can discretize the state space¹⁴ and nonparametrically estimate the transition probabilities. However, the experiment menus vary across blocks and cities, resulting in around 260 distinct experiment menus. Including all of the experiment menus in the estimation of our model greatly enlarges the dimension of the state space. For computational feasibility, we group together similar experiment menus for each city by the K-means clustering method. Specifically, for each experiment menu, we construct a vector of binary indicators for whether an hour has the GSL experiment switched on. We then apply the K-means clustering

¹⁴Since u_{it} is the only continuous component in this model, we discretize the distribution of \hat{u}_{it} into deciles.

algorithm on the indicator vectors. We arrive at 5 average experiment menus for each block type and each city. These average experiment menus are taken as inputs in the estimation of the model.

Full Model Estimation.

We use the EM algorithm to obtain the maximum likelihood estimates based on the following objective function:

$$\{\hat{\theta}, \hat{\pi}\} = \arg \max_{\theta, \pi} \sum_{i=1}^N \ln \left[\sum_{j=1}^J \pi_j(s_{i1}) \prod_{t=1}^{T_i} l(a_{it} | s_{it}, j, \hat{p}, \theta) \right]$$

where $s_{it} = (u_{it}, z_{it}, a_{it-1}, h(t), m_t)$ is the vector of the observed states, \hat{p} is a vector of empirical conditional choice probabilities, $\pi(s_{i1})$ is the population probability of type j conditional on the initial state s_{i1} . The number of unobserved types, J , is assumed to be known and equal to 3. We assume that $\hat{\pi}_j(s_1) = \hat{\pi}_j(w_1, a_0)$ so that a driver's unobserved type depends on her initial wage and lagged actions observed in the data.

To obtain the starting values in the EM algorithm, we start with initializing $(\eta_{1,Day}, \eta_{1,Night}, \eta_{2,Day}, \eta_{2,Night})$. The initialization of the remaining structural parameters is set to the estimates of a structural model without any unobserved type. We set the tolerance level to be $1e - 7$.

C. Full Comparison with Alternative Model Specifications

In this section, we illustrate the potential bias that may arise from more restrictive model specifications. We progressively build up from the simplest possible discrete choice model to highlight how several of our modeling choices – including adjustment costs, permanent observed and unobserved heterogeneity, and the field experiment to address wage endogeneity – are key not only to match the data but also for the estimates of the reservation wages and for the results from the counterfactual analyses. The estimates of all the model specifications discussed in this section are reported in Table 13.

C1. Static Model with Constant Time Preference and Exogenous Wage

Setup.

At the beginning of each period, a driver observes the wage w_{it} and the idiosyncratic preference shocks $(\epsilon_{0it}, \epsilon_{1it})$, and she chooses between work ($a_{it} = 1$) and rest ($a_{it} = 0$):

$$U_{it} = \begin{cases} \gamma w_{it} + \beta + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{cases}$$

A driver's labor supply decision depends only on the wage, idiosyncratic shocks, and a constant utility of work β . Once the driver observes the realization of the wage and the preference shocks, she chooses to work if and only if $\gamma w_{it} + \beta \geq \epsilon_{0it} - \epsilon_{1it}$. We maintain the assumptions (IID) and (EXOG) as in Section 4.2. We further assume that ϵ_{0it} and ϵ_{1it} are T1EV and independent: $\epsilon_{1it} \perp \epsilon_{0it}$. In this model, we specify the wage equation as $w_{it} = \delta_{h(t)} + u_{it}$. We also show the results of a nonparametric wage process as a robustness check in the later section.

Model Fit.

Figure 35 compares the model implied probability of working across hours of a week against the data counterparts. While the data reveals a clear picture of systematic variation in the probabilities of working overtime, the simple static model does not predict enough variation. Since the preference shocks are IID across time and drivers in our model, the variability in the probability of working comes only from the variability in wages. Therefore, we conclude that the variation in wages is not enough to generate the large variation in the probabilities of working that we observe in the data.

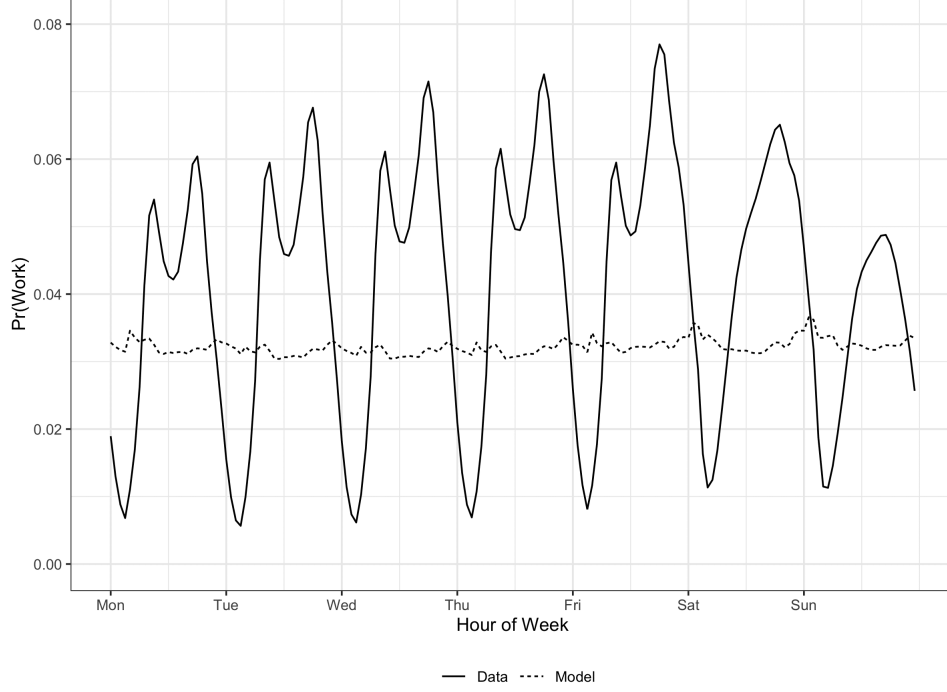


Figure 35: Model Fit of Probability of Working Across Hours of the Week

Notes: The solid line plots the data and the dashed line plots the prediction from the model.

C2. Static Model with Varying Time Preference and Exogenous Wage

Since the variation in wages is not sufficient to generate the observed variation in the probability of working over time, we allow the utility of work to vary across hours of the week in the model.

Setup.

At the beginning of each period, a driver observes the wage w_{it} and the idiosyncratic preference shocks $(\epsilon_{0it}, \epsilon_{1it})$, and she chooses between work ($a_{it} = 1$) and rest ($a_{it} = 0$):

$$U_{it} = \left\{ \begin{array}{ll} \gamma w_{it} + \beta_{h(t)} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

where $h(t)$ is the hour of a week at t , and $\beta_{h(t)}$ is the time-specific shifter of the cost of driving at a given hour of the week. We maintain the above assumptions and the parametric specification on the wage process.

Model Fit.

Figure 36a compares the predicted probability of working across hours of a week against the data counterparts. With time-varying time preferences, the model now successfully captures drivers' working patterns across hours of a week.

C3. Dynamic Model with Varying Time Preference and Exogenous Wage

Although the static model specification in the previous section fits the working patterns across hours of a week excellently, it fails to capture the dynamics of work that we observe in the data. In Figure 36b we compare the probabilities of working conditional on lagged employment choices. The static model predicts no difference in the probabilities of working between those who worked in the previous period and those who rested. Yet we observe in the data a much higher probability of working if one worked in the previous period. This state-dependency emphasizes the importance of incorporating the dynamic component in the driver's decision problem to capture the connection between the decision to drive in the current period and future utility. We now consider a dynamic model with adjustments costs and time-varying preferences.

Setup.

At the beginning of each period, a forward-looking driver observes the wage w_{it} and the idiosyncratic preference shocks $(\epsilon_{0it}, \epsilon_{1it})$, and forms expectations on future wages and preference shocks. The driver then chooses between work ($a_{it} = 1$) and rest ($a_{it} = 0$) based on the following choice-specific flow payoffs, taking into account how her current choice will affect the expected payoffs in the future:

$$U_{it} = \begin{cases} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{cases}$$

Note that the flow payoff from work depends on whether the driver worked in the previous period; if the driver did not work at $t - 1$, she needs to pay an adjustment cost μ to start to work. Both the adjustment cost and the wage transition are linking together periods, since a driver who has paid the adjustment cost in t has a higher continuation value in $t + 1$. Assuming a fixed costs, the model now becomes a dynamic discrete choice problem. We maintain the above assumptions and the parametric specification of the wage process. In addition, we assume (CI-X) and (DISCOUNT) with the following modification:

(CI-X') State transition probability F satisfies

$$F(w_{it+1}|a_{it}, a_{it-1}, w_{it}, h(t), \epsilon_{1it}) = F(w_{it+1}|a_{it}, a_{it-1}, w_{it}, h(t))$$

Model Fit.

Figure 37a plots the actual and model implied probabilities of working conditional on lagged employment choice. The dynamic model with the adjustment cost fits the probabilities of working conditional on the lagged choice very well. We, therefore, conclude that (some form of) adjustment costs need to be included to meaningfully represent the driver's decision problem.

C4. Observed Heterogeneity

So far, we have assumed no systematic preference heterogeneity among drivers. Conditional on wages, the hour of a week, and the lagged choice, the only source that generates differences in drivers' choices is the idiosyncratic preference shocks. Figure 37b plots the probability of working across the four subgroups of drivers partitioned by gender and age. This figure shows substantial heterogeneity across the four demographic groups in the probabilities of working. On average, male drivers and older drivers work more than other drivers. This large difference in the probabilities of working across the demographic groups suggests that it is key to let preferences of the drivers to vary by observables such as gender and age.

Estimation and Model Fit.

In order to capture the observed heterogeneity along the demographic (gender \times age) dimension, we estimate the dynamic model subgroup by subgroup. The dynamic model fits excellently each demographic group's working patterns. Figure 38a plots the conditional choice probabilities for each subgroup against the data counterparts.

C5. Time-Invariant Unobserved Heterogeneity

We have shown that there exists a lot of observed heterogeneity across drivers. However, there may also be important unobserved heterogeneity across drivers in the costs or non-wage benefits of working. To illustrate this, Figure 38b shows the difference in the average probabilities of working of what we refer to as full-time drivers and infrequent drivers. Full-time drivers have average hours worked above the median. Infrequent drivers have average hours worked below the median. The rank of average hours worked is calculated by pooling all drivers from the four subgroups. As is shown in Figure 38b, there is a large difference in the average work probabilities between the top percent and the bottom 50 percent of the drivers, pointing out that there is a great deal of heterogeneity even conditional on observables. We now introduce persistence in the costs or non-wage benefits of working to the model by including unobserved driver types in the flow payoff of work.

Setup of Time-Invariant Unobserved Types.

At the beginning of each period, a forward-looking driver observes the wage w_{it} and the idiosyncratic preference shocks $(\epsilon_{0it}, \epsilon_{1it})$, and forms expectations on future wages and preference shocks. The driver then chooses between work ($a_{it} = 1$) and rest ($a_{it} = 0$) based on the following flow payoffs, taking into account how the current choice will affect her payoffs in the future:

$$U_{it} = \left\{ \begin{array}{ll} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i)} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{array} \right\}$$

where $j(i)$ is driver i 's unobserved type, and $\eta_{j(i)}$ is a shifter of the cost of driving, e.g., full-time drivers have higher $\eta_{j(i)}$ and are thus more likely to drive. We maintain all the assumptions above, use the parametric specification of the wage equation, and invoke the assumption (REACH) and (NTYPE) in Section 2. For now, we assume there are two unobserved types of drivers, full-time drivers and infrequent drivers.

Model Fit.

Figure 39a plots the difference in the average probabilities of working between drivers whose average hours worked rank above the median and those below the median. There is a significant improvement in the model fit once we include the two unobserved types to the model. Capturing such unobserved persistence in driving is important to model the labor supply of Uber drivers.

C6. Time-Varying Unobserved Types

The assumption of time-invariant unobserved types implies that observationally equivalent workers may systematically differ in how much they work, but not in when they work a lot. In the Uber setting, however, it is likely to be a subset of drivers who work mostly in the evenings due to daytime jobs other than driving for Uber. To illustrate this, Figure 39b plots the average probabilities of working within a day of three groups of drivers: (1) infrequent drivers whose average hours worked in the daytime and in the nighttime both rank among the bottom 80 percent,¹⁵ (2) evening drivers whose average hours worked in the daytime rank among the bottom 80 percent but hours worked in the nighttime rank among the top 20 percent, and (3) the remaining frequent drivers. Figure 39b reveals a large amount of heterogeneity in the persistence of working within a day across the three types of drivers. By comparison, the model with time-invariant unobserved types is not able to generate such time-specific persistence in driving, even conditional on observables. Motivated by this finding, we introduce time-varying unobserved types to the previous model.

Setup of Time-Varying Unobserved Types.

At the beginning of each period, a forward-looking driver observes the wage w_{it} and the idiosyncratic preference shocks $(\epsilon_{0it}, \epsilon_{1it})$, and forms expectations on future wages and preference shocks. The driver then chooses between work ($a_{it} = 1$) and rest ($a_{it} = 0$) based on the following flow payoffs, taking into account how the current choice will affect the payoffs in the future:

$$U_{it} = \begin{cases} \gamma w_{it} + \beta_{h(t)} + \mu \mathbf{1}\{a_{it-1} = 0\} + \eta_{j(i),h(t)} + \epsilon_{1it} & , \text{ if } a_{it} = 1 \\ \epsilon_{0it} & , \text{ if } a_{it} = 0 \end{cases}$$

where $\eta_{j(i),h(t)}$ is a shifter in the cost of driving for driver i with unobserved type j at the hour of week $h(t)$. For example, evening drivers have lower $\eta_{j(i),h(t)}$ for $t \in \text{Morning}$ but higher $\eta_{j(i),h(t)}$ for $t \in \text{Evening}$, so they are more likely to drive only in the evenings. We maintain all the assumptions as above and assume there are three unobserved types of drivers.

Model Fit.

Figure 40 shows the average probabilities of working of the three groups of drivers (infrequent drivers, full-time drivers, and evening drivers) in the data and in the simulated data from the model. The model captures reasonably well the three unobserved types of drivers after taking into account the observed heterogeneity.

¹⁵We define the daytime to be 6 a.m.-9 p.m. and the nighttime to be 10 p.m.-5 a.m. next day.

			Non-parametric Wage Model				Parametric Wage Model			
			Old Male	Young Male	Old Female	Young Female	Old Male	Young Male	Old Female	Young Female
Preference for Wage	γ		0.004	0.007	-0.009	-0.006	0.004	0.007	-0.009	-0.005
Time Preferences	β	$E[\beta_{h(t)}]$	-0.666	-0.757	-0.643	-0.724	-0.673	-0.761	-0.645	-0.751
		$Sd(\beta_{h(t)})$	0.615	0.506	0.797	0.694	0.614	0.506	0.794	0.697
		$Median(\beta_{h(t)})$	-0.400	-0.619	-0.258	-0.427	-0.410	-0.616	-0.282	-0.453
		$q_{10}(\beta_{h(t)})$	-1.642	-1.417	-1.825	-1.793	-1.641	-1.429	-1.840	-1.822
		$q_{90}(\beta_{h(t)})$	-0.093	-0.205	0.052	-0.066	-0.100	-0.208	0.042	-0.091
Adjustment Cost	μ		-5.515	-5.651	-6.303	-6.256	-5.517	-5.653	-6.306	-6.212
Unobserved Types	η	$\eta_{(1,Night)}$	1.313	1.130	1.715	1.513	1.312	1.131	1.713	1.529
		$\eta_{(1,Day)}$	-0.074	-0.009	-0.047	0.007	-0.074	-0.010	-0.500	0.006
		$\eta_{(2,Night)}$	1.213	1.106	1.683	1.575	1.213	1.106	1.678	1.581
		$\eta_{(2,Day)}$	0.339	0.407	0.457	0.498	0.339	0.406	0.458	0.502

Table 10: Estimates of Non-Parametric and Parametric Wage Models

Notes: “Young” is defined as those whose ages are less than or equal to the median age.

The patterns of work vary distinctly over time across the three types of drivers. This is true both in the actual and the simulated data.

C7. Model of Wages

Before allowing for endogeneity of wages, we show that our estimation results from the previous model do materially change if we maintain the assumption of exogenous wages but relax the restriction on the process for how wages evolve over time. We relax the wage equation by allowing for a nonparametric evolution of wages as follows. The wage transition is assumed to follow a first-order Markov process. The wage distribution nonparametrically depends on lagged wages, lagged choices, and the hour of the week.¹⁶ We compare the parameter estimates from the models with and without the parametric assumptions on wages in Table 10. The estimates of the structural parameters are very similar between the types of models.

Until now, we have assumed exogeneity of wages, ruling out the possibility that market wages may co-move with the cost of driving. As we point out in Section 3, the downward bias in the OLS estimates of labor supply responses to wage changes is consistent with demand being high when it is costly or difficult to drive. The inclusion of fixed effects of workers and hours of a week helps reduce but not eliminate the bias, suggesting that market wages and the cost of driving may co-move due to idiosyncratic factors, such as weather conditions or entertainment events. This motivates the use of the experiment to address concerns about wage endogeneity.

¹⁶We first discretize the wage distribution into decile grids, and then nonparametrically estimate the transition probability of the wage process conditional on lagged wages, lagged choices, and the hour of the week.

			Time-Variant Unobs. Types					Full Model with Control Function				
			Weighted Average	Old Male	Young Male	Old Female	Young Female	Weighted Average	Old Male	Young Male	Old Female	Young Female
Preference for Wage	γ		0.002	0.004	0.007	-0.009	-0.006	0.036	0.040	0.042	0.024	0.021
Time Preferences	β	$E[\beta_{h(t)}]$	-0.704	-0.666	-0.757	-0.643	-0.724	-1.544	-1.612	-1.760	-1.257	-0.977
		$Sd(\beta_{h(t)})$	0.607	0.615	0.506	0.797	0.694	0.513	0.516	0.462	0.597	0.570
		$Median(\beta_{h(t)})$	-0.468	-0.400	-0.619	-0.258	-0.427	-1.328	-1.375	-1.598	-0.972	-0.734
		$q_{10}(\beta_{h(t)})$	-1.599	-1.642	-1.417	-1.825	-1.793	-2.295	-2.428	-2.351	-2.164	-1.861
		$q_{90}(\beta_{h(t)})$	-0.113	-0.093	-0.205	0.052	-0.066	-1.055	-1.151	-1.301	-0.653	-0.432
Adjustment Cost	μ		-5.757	-5.515	-5.651	-6.303	-6.256	-6.377	-6.191	-6.269	-6.771	-6.864
Unobserved Types	η											
		$\eta_{(1,Night)}$	1.320	1.313	1.130	1.715	1.513	1.857	1.847	1.901	2.039	1.571
		$\eta_{(1,Day)}$	-0.036	-0.074	-0.009	-0.047	0.007	0.601	0.589	0.669	0.547	0.483
		$\eta_{(2,Night)}$	1.277	1.213	1.106	1.683	1.575	1.178	1.120	1.237	1.288	1.067
		$\eta_{(2,Day)}$	0.399	0.339	0.407	0.457	0.498	0.794	0.773	0.802	0.838	0.788
Selection Term	$\frac{\rho_{u\xi}}{\sigma}$							-0.039	-0.043	-0.044	-0.028	-0.024

Table 11: Estimates of the Exogenous Wage Model and the Full Model

Notes: “Weighted average” is calculated by averaging the estimates of the four demographic groups weighted by the share of the drivers. “Young” is defined as those whose ages are less than or equal to the median age.

Table 11 compares the point estimates of the model assuming exogenous wages and the one allowing for endogenous wages. Notably, the point estimates of the preference for wage parameter, γ , which captures drivers’ sensitivity to wage changes, increase substantially for all of the four subgroups. In the case of female drivers, the sign of γ estimates flips and becomes positive. The estimate of the correction term $\frac{\rho_{u\xi}}{\sigma}$ is negative in all four subgroups. Recall that $\rho_{u\xi}$ is the correlation coefficient between the preference shock ξ_{it} and the wage component u_{it} . Our estimation results indicate that the costs of working tend to co-move with the market wages. Therefore, it is important to take this endogeneity into account to obtain reliable estimates of the preference parameters.

C8. Economic Implications of the Alternative Modeling Choices

We have demonstrated that the fit of the models improves as we gradually build up from the simple static model up to the full specification. The inclusion of the key components in our model also affects the estimates of the key parameters of interest. We now examine how these parameter estimates, such as preference heterogeneity and adjustment costs, influence the implied reservation wages and the effectiveness of driver incentives that Uber may offer.

	Reservation Wage				Wage Increase Needed
	Daytime	Evening	Weekday	Weekend	
Constant Time Preference	1.80	1.80	1.80	1.80	7.03
Varying Time Preference	0.81	0.99	0.87	0.88	3.14
Adjustment Cost	6.10	7.99	6.70	6.81	71.98
Obs. Heterogeneity	3.16	4.18	3.48	3.56	75.76
Time-Invariant Unobs. Types	4.87	6.25	5.31	5.38	90.48
Obs. & Unobs. Heterogeneity	3.27	4.64	3.70	3.79	88.16
Full Model	1.34	1.65	1.43	1.46	9.55

Table 12: Comparison of Counterfactual Analyses by Model Specifications

Notes: “Reservation Wage” is calculated as the average minimal wages required to work. We normalize the reservation wages to the old male drivers’ reservation wage in the daytime in the full model. “Wage Increase Needed” is computed as the increase in the hourly wage in order to achieve a 0.5 percentage point increase in the average probability of working.

In Column 1-4 in Table 12, we report the model implied reservation wages (normalized by old males’ reservation wages in the daytime). Clearly, the model specification matters for the estimates of reservation wages. Under the static model with a constant preference for time, the model implied expected reservation wages do not vary over time. Once we allow for time-varying preferences, the expected reservation wages exhibits variation over time. Compared with the full model, the exogenous wage models have much higher reservation wages.

In Column 5, we present the hourly wage increase needed to achieve a 0.5 percentage point increase the probability of working. The inclusion of adjustment costs is crucial to the calculation of wage incentives. The wage increase needed in the static model is much smaller compared to the adjustment cost model. There is also a large decline in the wage needed when we move from exogenous wage models to the full model.

Specification	Preference for Wage	Time Preferences				Adjustment Cost		Unobserved Types				Selection Term	
	γ	$\frac{E[\beta_{h(t)}]}{Sd(\beta_{h(t)})}$	β	$Median(\beta_{h(t)})$	$q_{10}(\beta_{h(t)})$	$q_{90}(\beta_{h(t)})$	μ	η	$\eta(1, Night)$	$\eta(1, Day)$	$\eta(2, Night)$	$\eta(2, Day)$	$\frac{\rho_{\Delta}}{\sigma}$
Constant Time Preference	0.015	-3.692											
Varying Time Preference	0.034	-4.059	0.672	-3.778	-5.259	-3.441							
Adjustment Cost	0.005	-0.376	0.401	-0.283	-0.812	0.021	-6.397						
Obs. Heterogeneity													
Time-Invariant Unobs. Types													
Time-Variant Unobs. Types													
Time-Variant Unobs. Types (Nonparam)													
Full Model with Control Function													

Table 13: Estimates of All Specifications

D. Robustness Checks with Regards to Substitution to Lyft

We now present the results from two sets of robustness checks. First, we examine if our findings differ systematically in cities or in periods in which Lyft has relatively high or low market shares. Second, we compare the results for all drivers to those we obtain from a subsample of drivers who are ineligible to drive for Lyft.

D1. Variation in Lyft Share Across Markets

Experimental Estimates by Market. Our sample consists of 3 cities (Boston, Chicago, and San Francisco), each of which experienced differential growth in the market share of Lyft during the sample period. We estimate wage elasticities for each city-month pair, using the GSL experimental variation in wages, and present the elasticities against Lyft market share in Figure 41. As the figure suggests, there is no evidence for correlation between wage elasticities and Lyft market shares.

Structural Model Estimates by Market. We now show how much the estimated model parameters vary with the Lyft market share over time and across cities. For each city \times time pair,¹⁷ we estimate the model and present the estimates of the value of time parameter, $\beta_{h(t)}$, in Figure 42.¹⁸ We see limited changes in $\beta_{h(t)}$ estimates across markets for each demographic subgroup.

D2. Variation in Lyft Eligibility

Experimental Estimates with Lyft Ineligible Drivers. Drivers whose vehicle’s model year falls between 2001 and 2003 can only drive for Uber but not for Lyft. In our sample, 3 percent of drivers have vehicles in this range. We define these drivers as Lyft ineligible drivers and estimate the labor supply elasticities on this Lyft ineligible sample. Similar to the labor supply elasticity estimate 0.53 from the sample of all drivers, the Lyft ineligible drivers’ elasticity is estimated to be 0.56.

Structural Model Estimates with Lyft Ineligible Drivers. Taking advantage of the variation in the the Lyft driving eligibility, we separately estimate the structural model on the full sample and the Lyft ineligible sample for male drivers. We allow for heterogeneity by age but no unobserved types, due to the much smaller sample size of Lyft ineligible drivers. In Figure 43, we compare the parameter estimates of $\beta_{h(t)}$ for all drivers to those we obtain from the subsample of drivers who are ineligible to drive for Lyft. The estimated value of time parameters, $\beta_{h(t)}$, for Lyft ineligible drivers exhibit a very similar pattern as compared to all drivers. We report the structural parameter estimates from the full sample and the Lyft ineligible sample in Table 14. Reassuringly, the structural estimates from the two samples mirror each other.

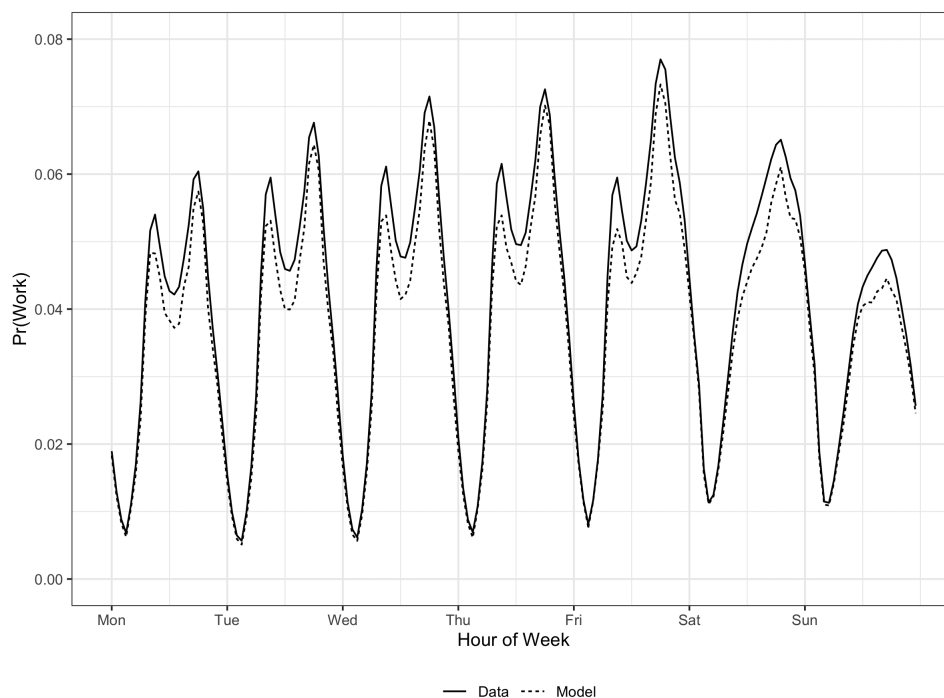
¹⁷We define time as a 5-month period.

¹⁸To avoid empty cells due to sample size restrictions, we estimate a model with endogenous wage but no unobserved heterogeneity.

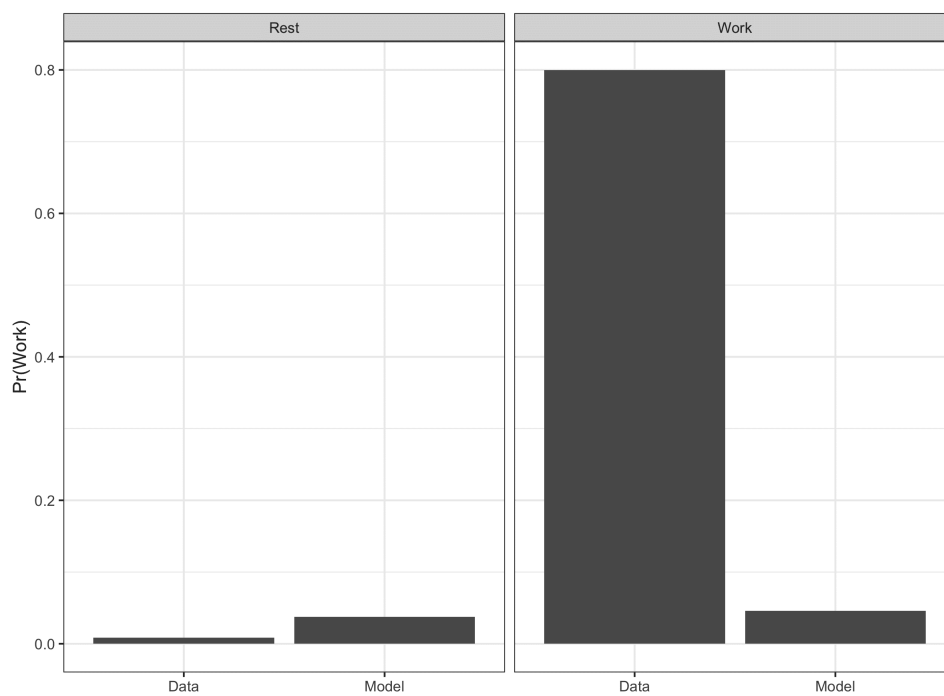
			Full		Lyft Ineligible	
			Old Male	Young Male	Old Male	Young Male
Preference for Wage	γ		0.042	0.039	0.044	0.043
Time Preferences	β	$E[\beta_{h(t)}]$	-1.152	-0.986	-1.140	-1.082
		$Sd(\beta_{h(t)})$	0.418	0.409	0.592	0.596
		$\text{Median}(\beta_{h(t)})$	-1.068	-0.872	-0.995	-0.976
		$q_{10}(\beta_{h(t)})$	-1.641	-1.515	-1.776	-1.691
		$q_{90}(\beta_{h(t)})$	-0.754	-0.607	-0.633	-0.521
Adjustment Cost	μ		-6.479	-7.115	-7.331	-8.176
Selection Term	$\frac{\rho_{u\xi}}{\sigma}$		-0.034	-0.030	-0.037	-0.035

Table 14: Estimates by Lyft Eligibility

Notes: Column 1 reports the estimates using the full sample and Column 2 shows the estimates using the Lyft ineligible sample.



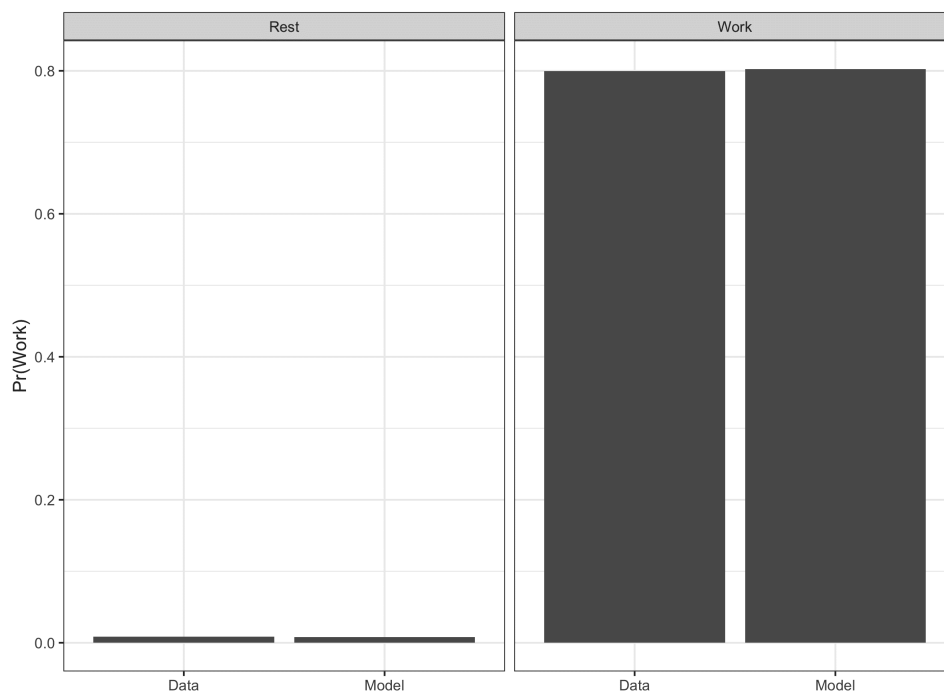
(a) Hours of the Week



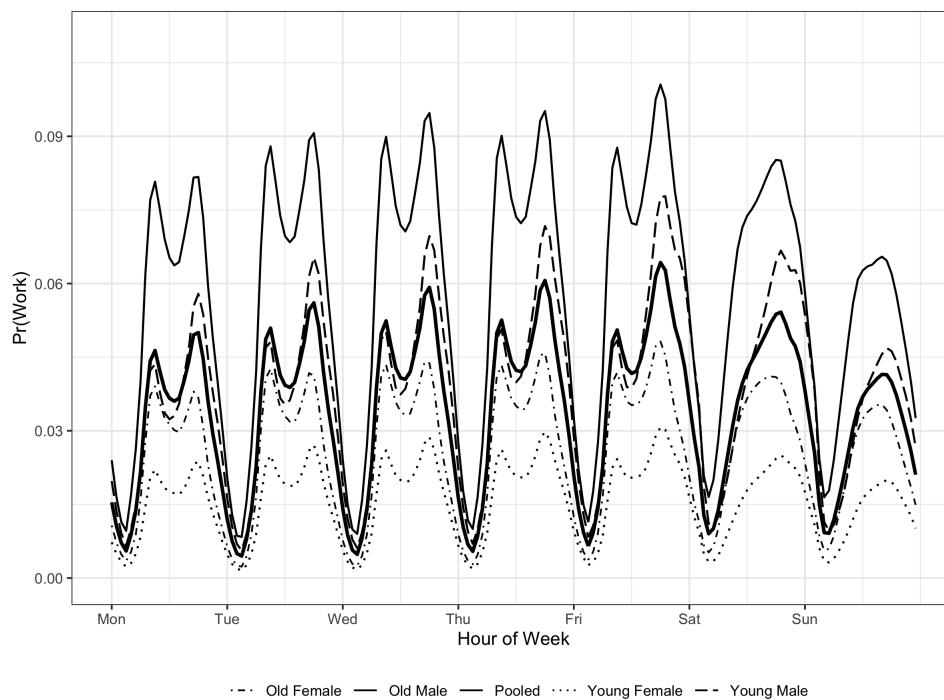
(b) Lagged Choices

Figure 36: Model Fit of Probability of Working

Notes: In Figure (a), the solid line plots the data and the dashed line plots the prediction from the model. In Figure (b), the "Rest" panel shows the probability of working conditional on drivers not working in the previous hour. The "Work" panel shows the probability of working conditional on drivers working in the previous hour.



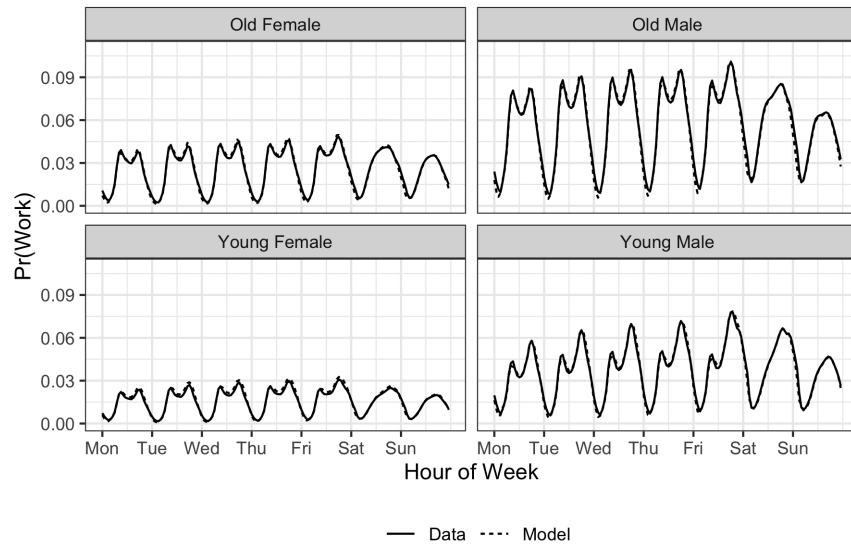
(a) Lagged Choices



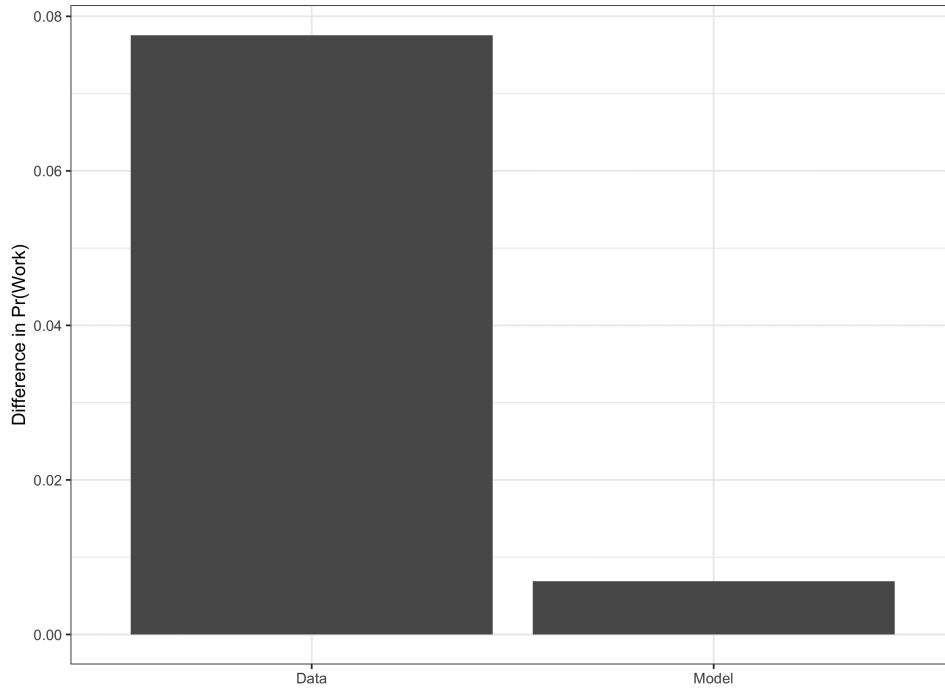
(b) Demographic Groups

Figure 37: Model Fit of Probability of Working

Notes: In Figure (a), the “Rest” panel shows the probability of working conditional on drivers not working in the previous hour. The “Work” panel shows the probability of working conditional on drivers working in the previous hour. In Figure (b), we estimate the probability of working across hours of the week for each demographic group.



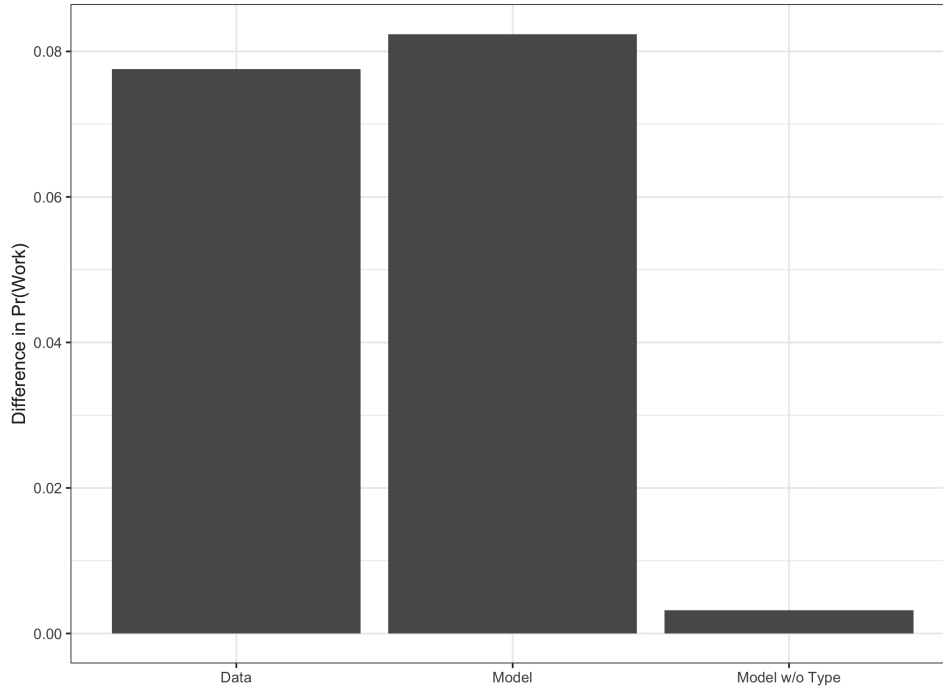
(a) Hours of the Week Across Demographic Groups



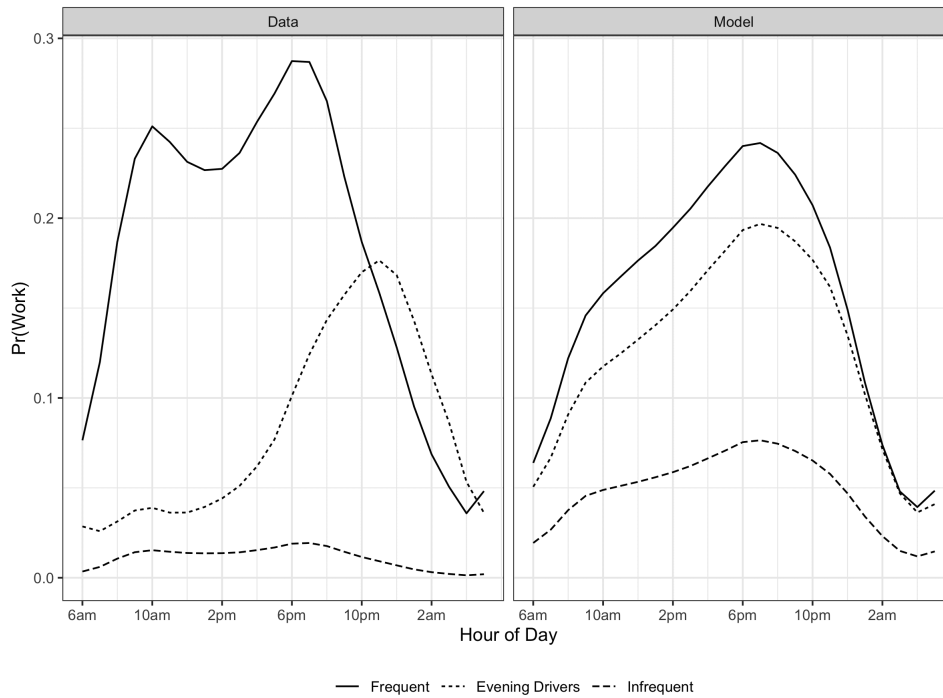
(b) Bottom 50 Percent v.s. Top 50 Percent

Figure 38: Model Fit of Probability of Working

Notes: In Figure (a), the solid line plots the data and the dashed line plots the prediction from the model. We estimate the probability of working across hours of a week for each demographic group in the data and compare it with the probability predicted by the model. In Figure (b), we estimate the difference in the average probability of working between drivers in the top 50 percent and those in the bottom 50 percent, where drivers are ranked by the average hours worked per week.



(a) Bottom 50 Percent v.s. Top 50 Percent



(b) Hours of the Day by Driver Types

Figure 39: Model Fit of Probability of Working

Notes: In Figure (a), we estimate the difference in the average probability of working between drivers above and below the median in the distribution of the average hours worked per week. We conduct the estimation on the data, the simulated data from the model without unobserved types, and the simulated data from the model with two time-invariant unobserved types. In Figure (b), we plot the average probabilities of working within a day of three groups of drivers: (1) infrequent drivers whose average hours worked in the daytime and in the nighttime both rank in the bottom 80 percent, (2) evening drivers whose average hours worked in the daytime rank in the bottom 80 percentile but hours worked in the nighttime rank in the top 20 percent, and (3) the remaining frequent drivers. The ranks are calculated by pooling all four observed subgroups. The daytime is defined to be 6 a.m.-9 p.m., and the nighttime to be 10 p.m.-5 a.m. next day.

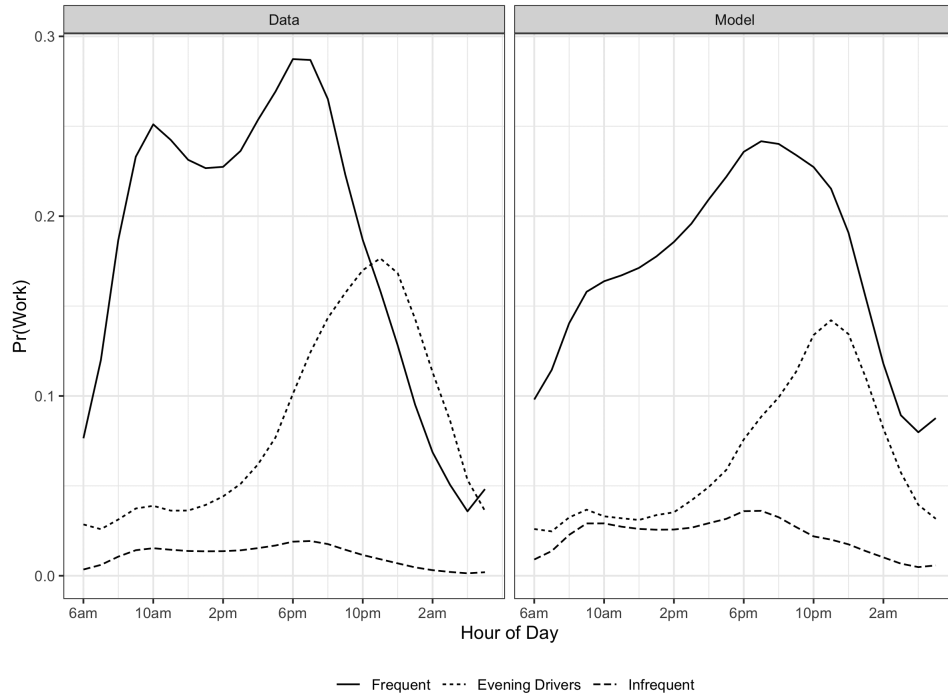


Figure 40: Model Fit of Probability of Working Across Hours of the Day by Driver Types

Notes: We plot the average probabilities of working within a day of three groups of drivers: (1) infrequent drivers whose average hours worked in the daytime and in the nighttime both rank in the bottom 80 percent, (2) evening drivers whose average hours worked in the daytime rank in the bottom 80 percent but hours worked in the nighttime rank in the top 20 percent, and (3) the remaining frequent drivers. The ranks are calculated by pooling all four observed subgroups. The daytime is defined to be 6 a.m.-9 p.m., and the nighttime to be 10 p.m.-5 a.m. next day.

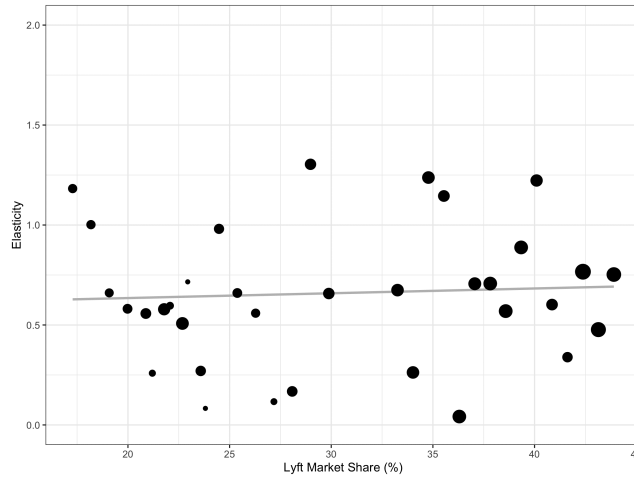


Figure 41: Wage Elasticities by Lyft Market Share

Notes: Each dot represents a city \times month wage elasticity estimate. The x-axis denotes the Lyft market share. The size of each dot indicates the number of observations in each city \times month. The solid line plots the linear regression fit weighted by the number of observations in city \times month.

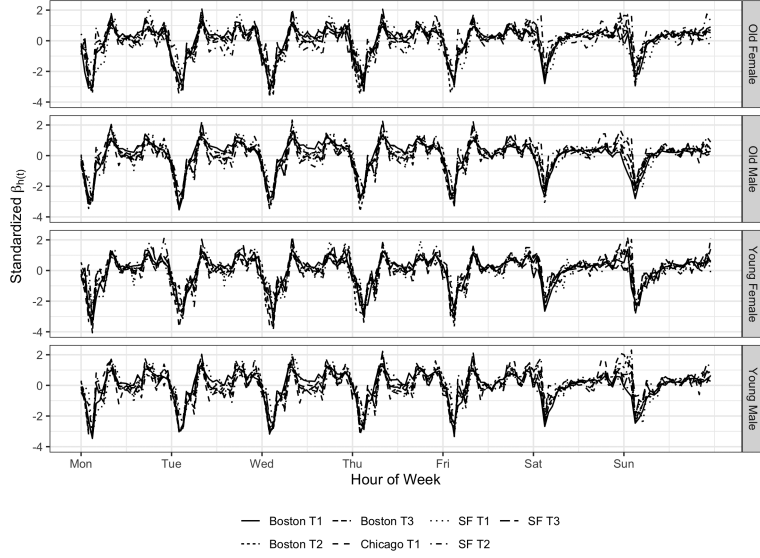


Figure 42: Standardized $\beta_{h(t)}$ by Markets and Demographic Groups

Notes: We define a market as a 5-month \times city pair. We estimate the model for each market. T1 represents the period from 2016/10/21 to 2017/03/31. T2 represents the period from 2017/04/01 to 2017/09/31. T3 represents the period from 2017/10/01 to 2018/03/01. We standardize $\beta_{h(t)}$ by de-meaning the raw estimates and dividing them by the standard deviation.

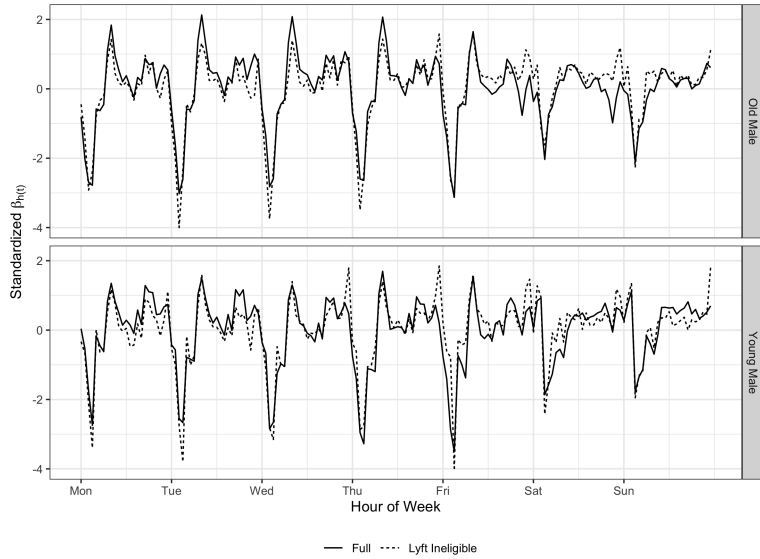


Figure 43: Standardized $\beta_{h(t)}$ by Lyft Eligibility

Notes: We standardize $\beta_{h(t)}$ by de-meaning the raw estimates and dividing them by the standard deviation.