The Impact of Behavioral and Economic Drivers on Gig Economy Workers

Gad Allon
The Wharton School, University of Pennsylvania, gadallon@wharton.upenn.edu

Maxime C. Cohen
Desautels Faculty of Management, McGill University, Montreal, Canada, maxime.cohen@mcgill.ca

Wichinpong Park Sinchaisri
The Wharton School, University of Pennsylvania, swich@wharton.upenn.edu

Problem Definition: Gig economy firms benefit from labor flexibility by hiring independent self-scheduling workers. Such flexibility poses a great challenge in terms of planning and committing to a service capacity. Understanding the motivations that drive gig economy workers is thus of great importance. In collaboration with a ride-hailing platform, we study how on-demand workers make labor decisions: specifically, when to work and for how long.

Academic/Practical Relevance: Our model offers a way to potentially reconcile competing theories of labor supply regarding the impact of income shocks on labor decisions. We are interested not only in improving the prediction of the number of active workers but also in understanding how to design better financial incentives for workers.

Methodology: Using a large comprehensive dataset, we analyze workers' decisions and responses to incentives to work while accounting for sample selection, simultaneity, and endogeneity biases.

Results: We find that financial incentives have a significant positive influence on the decision to work and on the number of hours worked—confirming the positive income elasticity posited by the standard income effect. We also find support for a behavioral theory such as income-targeting behavior (working less when reaching an earning goal) and inertia (continuing to work more after working for a longer period).

Managerial Implications: We show via numerical experiments that incentive optimization based on our model can increase service capacity by 22% without incurring additional cost, or maintain the same capacity at a 30% lower cost. Ignoring behavioral factors could lead to understaffing by 10–17% below the optimal capacity level. Lastly, inertia could be a potential sign of workers' loyalty to the platform.

Key words: empirical and behavioral operations, gig economy, incentives, sample selection

1. Introduction

Gig economy is a labor-sharing market system where workers engage in short-term projects or freelance work as opposed to permanent jobs. In 2017, 57.3 million Americans or 36% of the U.S. workforce engaged in gig work (Upwork 2017), providing a wide range of services, from ride-hailing (e.g., Uber, Lyft, and Didi) to food delivery (e.g., DoorDash and Caviar) to web development (e.g., Upwork and Fiverr). By 2025, it is estimated that the majority of the workforce will participate in the gig economy—leading to a global GDP boost of $2.7 trillion (Manyika et al. 2015). The unique and novel feature of this business model relates to the nature of employment: independent workers can freely choose when to work as well as seamlessly switch between multiple platforms. For example, as of 2017, nearly 70% of on-demand U.S. drivers work for both Uber and Lyft, and 25% drive for more than just those two (Campbell 2017). In other words, gig workers are in control of their work schedules: when to work, how long to work, and for which platforms. Such control attracts many workers to the gig economy.

Companies also greatly benefit from increased labor flexibility as they can hire workers with different skill levels to work at different times, while paying them only for the work they perform. Although the core of the gig economy success lies in the perfect matching of supply with demand, firms need to ensure that their services appeal not only to customers (demand) but also to service providers (supply). This poses an enormous challenge in planning and committing to a service capacity, during both peak hours, when demand is high, and during off-peak times, when only a handful of workers are needed. At the same time, policymakers have been increasingly passing new regulations in response to the growth of the gig economy. As part of its Vision Zero initiative, New York City (NYC) passed fatigued driving prevention rules in 2017. These rules limit the number of daily and weekly hours a ride-hailing driver can work with the goal of reducing driver fatigue and enhancing road safety. In 2019, the European Parliament has approved new EU rules that provide minimum rights and enforce better job transparency and compensation for gig workers.

To examine how firms can staff the right number of on-demand workers at the right time and how policymakers can develop effective regulations, it is important to first understand how gig workers make labor decisions. We address this question by collaborating with a ride-hailing platform with the goal of not only improving the way of predicting the number of active gig economy workers, but also understanding how to design better incentives for flexible workers.

For decades, economists have studied how labor supply is affected by financial incentives and wage variation. The standard income effect predicts that workers are lifetime-utility maximizers. As a result, when the wage increases, individuals are more likely to work or will work for longer. While several observational studies find evidence for this theory (e.g., Carrington 1996, Oettinger 1999, Sheldon 2016), other studies suggest the opposite prediction; that is, workers may actually work less when offered a higher wage due to their psychological reaction. For example, taxi drivers may work for fewer hours on a high-paying day because they can meet their earning goal faster; that is, they exhibit *income-targeting* behavior (e.g., Camerer et al. 1997, Thakral and Tô 2017). Providing further support for the behavioral theory of labor supply, Crawford and Meng (2011) and Farber (2015) suggest that workers' behavior may also be influenced by their duration goal or *time-targeting* behavior.

Our paper aims, in part, to reconcile this on-going debate in labor economics by proposing an operational framework to predict labor supply using both behavioral and economic incentives. Most recent studies in the context of gig economy operations has focused on the system equilibrium or on the welfare of customers and service providers (e.g., Ibrahim 2018, Taylor 2018, Kabra et al. 2017). To our knowledge, among the papers that focus on the supply side (e.g., Gurvich et al. 2016, Dong and Ibrahim 2017, Benjaafar et al. 2018, Hall et al. 2018), our work is the first to study the causal effect of behavioral and economic factors on workers' decisions and to incorporate workers' behavior into financial incentives decisions.

Research questions and methodology. Our key research questions are: (i) *How do gig economy workers make labor decisions?* How do they react to incentives? What are the factors that shape each worker's decision? Are their decisions rational or do they exhibit behavioral biases? (ii) *How can firms design incentives to entice workers?* How can they meet their desired service level by offering the right incentives?

We answer these questions by estimating an econometric model of workers' labor decisions and conducting numerical experiments on financial incentives. Prior empirical studies on the relationship between wage and labor decisions have not distinguished between the decision of whether to work and for how long to work, and treated them essentially as a single decision due to data limitations. We overcome this challenge by leveraging our rich dataset which includes real-time financial incentives and work decisions for every driver registered with our ride-hailing partner. Accordingly, we gain clearer insight into work decisions by using data on financial incentives and historical work experience of the drivers who ended up not driving at a particular period. In our empirical model, we address econometric challenges such as sample selection bias and simultaneity, and account for several factors (e.g., weather and past work frequency)—allowing us to test the impact of potential behavioral biases. Finally, we conduct simulations to compute improved incentive allocations and to quantify the loss incurred when the platform ignores workers' behavioral factors.

Contributions. Our paper contributes to the economics and operations literatures in four ways. First, we offer a potential way to reconcile the two competing theories of labor supply by showing that workers respond to wage variation in the same way as suggested by the standard income effect, while also exhibiting targeting behaviors with respect to recent cumulative earnings and hours worked. We find that financial incentives have a positive influence on the decision to work and on the work duration. However, cumulative earnings from earlier hours or days have a negative impact on both decisions. This can be explained by an income-targeting behavior: the closer a worker is to his/her earning goals, the shorter amount of time s/he is likely to work. Second, we unravel a new behavioral bias, which we refer to as *inertia*. Our results indicate that the number of hours worked earlier in the day or in the week has a positive influence on the decision to continue working and on the future work duration. We refer to this phenomenon as inertia as it captures the tendency of workers to continue working after having worked for some time. It can potentially reflect workers' loyalty to the service platform. We find this effect to be consistent over time. Third, we demonstrate that behavioral biases play an important role in workers' labor decisions. Both in-sample and out-of-sample analyses suggest that workers' reaction to cumulative earnings and past work hours is a key driver of their labor decisions. We then demonstrate via simulations that not accounting for these behavioral factors will result in understaffing by 10–17%. Finally, we use our econometric models to prescribe operational decisions. Specifically, we show that, if the company re-allocates incentives while accounting for its workers' behavior, it can increase the service capacity by 22% without incurring additional cost, or maintain the same service level at a 30% lower cost.

Structure of the paper. This paper is organized as follows. Section 2 introduces the behavioral model of how gig economy workers make labor decisions and develops several hypotheses. We then describe our dataset and present descriptive statistics in Section 3. We discuss our empirical approach, including sample selection bias, the control function method, and instrumental variables, in Section 4, and present our estimation results in Section 5. Managerial insights are then drawn through discussions and numerical experiments in Section 6. Finally, we conclude and discuss future research directions in Section 7.

2. Labor Supply Theories and Hypotheses Development

Researchers across several disciplines have studied labor supply decisions for decades. Economists have offered two different perspectives centered around the elasticity of labor supply. The traditional approach follows a lifecycle model where individuals maximize their lifetime utility. Accordingly, when facing a wage increase, workers should work more, hence exhibiting positive income elasticity. On the other hand, empirical studies, notably in the context of taxi drivers, suggest that income elasticity could be negative if workers are loss averse and benchmark their outcome relative to a reference point. Then, a wage increase would translate into a shorter time to reach earning goals, and ultimately in working less. Although many previous studies focused on workers who have some discretion over their work schedule, it is unclear whether existing findings can apply to gig economy workers who freely decide their schedule and can easily switch between providers. In this section, we review in greater detail the two contrasting models of labor supply and develop hypotheses for gig economy workers' behavior.

2.1. Traditional Model of Labor Supply

In the neoclassical microeconomics tradition, each worker is a rational agent who maximizes lifetime utility. A positive wage shock should then lead to a larger group of workers joining the platform or to a higher level of activity from the workers. In other words, workers are expected to exhibit a positive wage elasticity. This perspective seems plausible but finding evidence in the field has been challenging as in reality, workers cannot easily adjust their work hours. However, positive elasticities have been observed among workers who have some level of discretion over their schedule, such as Trans-Alaskan pipeline workers (Carrington 1996), vendors in a baseball stadium (Oettinger 1999), and Florida lobster fishermen (Stafford 2015). These studies found that wage shocks, which are typically driven by temporary demand variation, have a positive effect on labor supply—both the number of workers and work hours. Sheldon (2016) investigated driving partners of a ride-hailing platform and found that drivers' elasticities are positive and significant, suggesting that their behavior can also be explained by the traditional model of labor supply.

2.2. Behavioral Model of Labor Supply

The behavioral perspective on labor supply offers an explanation to the opposite sign of elasticity observed for taxi drivers. Taxi markets provide a suitable setting to study labor supply decisions given that taxi drivers' wages are influenced by a variety of transitory shocks with low within-day variance and high across-day variance. Camerer et al. (1997) studied NYC taxi drivers' and found substantial negative elasticities. Their seminal work suggested that drivers' daily decisions on work hours were influenced by their individual targets (known as the income-targeting effect), which sparked a lot of interest among researchers. Using data from a different set of NYC taxi drivers, Farber (2005) found that the probability to stop working is closely related to the realized income earned in the same day only when drivers' fixed effects are not controlled for. Farber (2008) further introduced a structural model based on daily income targeting and found that the probability to stop working increases once the income target is reached. However, the author concluded from the targets' large standard error that the reference-dependent model of labor supply might not be useful. Reconsidering the possibility of reference-dependent preferences, Crawford and Meng (2011) adapted the econometric strategies used in Farber (2005) and Farber (2008) to estimate models based on the influential theory of Kőszegi and Rabin (2006), which allows for consumption and gain-loss utilities. More recently, Thakral and Tô (2017) estimated a structural model of labor supply using a NYC taxi drivers dataset and confirmed that the income-targeting effect exists, when controlling for the number of work hours.

These findings offer a realistic behavioral explanation and align well with insights from behavioral economics; however, support for the behavioral theory has been lacking outside the taxi industry. Sheldon (2016) noted that the income elasticity of ride-hailing drivers is positive and seems to increase over time. This result suggests that if income targeting exists, it would be only temporary.

2.3. Labor Supply in the Gig Economy

The gig economy offers workers a flexible, part-time, work schedule without providing traditional benefits such as medical insurance and retirement savings. In 2016, 72% of Americans have used some type of shared or on-demand online services, while 8% of Americans have participated as gig workers across industries, such as ride-hailing, online tasks, and cleaning services (Pew-Research Center 2016). Some workers fully rely on gig work as a primary source of income, while others keep their full-time job and earn additional income via the gig economy. As gig platforms appeal to a broad range of workers with different backgrounds and preferences, predicting the turnout or service capacity at any point in time is remarkably challenging. A common way to incentivize new workers to join and to keep existing workers active on the platform is to offer financial incentives. For example, a new Lyft driver will earn a one-time sign-on bonus when joining and will receive a weekly guaranteed earning rate for the first few weeks of driving.[1] Lyft drivers can also earn more through a Power Driver Bonus program which grants bonuses based on the number of completed trips in a given week.[2] Real-time bonuses, such as Uber's surge prices and DoorDash's Busy Pay, reward workers who work during rush hours or periods with high demand. For platforms such as the food delivery service Caviar, workers are notified in advance of time-specific bonuses so they can plan their schedule accordingly. Beyond direct monetary rewards, several companies go one step further by using a combination of gamification and psychology. For example, Uber drivers can earn badges for achievements such as excellent service and entertaining ride and are constantly reminded of how close they are to their earning goals. While these incentive strategies are prevalent in practice, less is known in academic research about their influence on workers' labor decisions.

Our paper belongs to the fast-growing research trend that examines operational and pricing decisions in the context of the gig economy (see, e.g., Allon et al. 2012, Jiang and Tian 2016, Taylor 2018). Most relevant to our work are studies that examine how dynamic wages affect supply, especially when independent providers can decide whether and when to work. In Gurvich et al. (2016), the platform determines the number of agents, a cap on the number of agents allowed to work, and the market-condition-contingent price and wage. Then, the agents decide whether or not to work after observing the market condition. The authors find that agent independence reduces the number of workers and increases the optimal price. Chen and Sheldon (2016) analyze data from 25 million ride-hailing trips and show that a dynamic wage due to surge pricing can entice drivers to work longer. Hall et al. (2018) use differences in timing and size of the city as well

[1] For example, new drivers in Los Angeles, CA are eligible for $1,000 weekly guarantees in their first four weeks: https://www.lyft.com/terms/incentives/5542d91a1703f12a950b7c1d

[2] The program is only available in selected regions and its requirements include a 2011 or newer car, 90% acceptance rate for the week, and a minimum number of trips (during and outside rush hours).

as platform-initiated fare changes to identify the impact of the fare on market equilibrium. The authors observe that hourly earnings rise immediately following a fare increase, but then decline shortly thereafter due to a drop in utilization. Several studies consider the problem of designing the right incentives on prices and wages to coordinate supply with demand for on-demand service platforms. In Hu and Zhou (2017), the authors study the pricing of an on-demand platform and show the good performance of a flat-commission contract. Taylor (2018) studies an on-demand service platform that decides both a per-service price for customers and a wage posted to agents. The author shows that the uncertainty in delay-sensitive customers' valuations or in the agents' opportunity costs can lead the intermediary to raise the price during congestion. In Cohen and Zhang (2018), the authors study a model where two-sided platforms engage simultaneously in wage and price competition. Assuming that each side of the market (i.e., customers and workers) follow a multinomial logit model, the authors show the existence and uniqueness of equilibrium.

While there is a large body of research on flexible or self-scheduling workforce in operations management, there are relatively few studies that investigate the supply side behavior and its impact on the platform's operational decisions. Moreover, most of the studies are of a theoretical nature and focus on the equilibrium of matching supply with demand. Ibrahim (2018) examines a queueing system with a random number of servers and characterizes the optimal staffing policy and the resulting cost. Dong and Ibrahim (2017) further study optimal staffing decisions when the workforce is composed of both contingent and permanent workers. The authors show that staffing decisions depend on the uncertainty of the flexible workers, the operating costs, and customer demand. Benjaafar et al. (2018) study an equilibrium model of labor welfare that accounts for interactions between supply and demand and investigate the possible alignment and misalignment of the platform's and workers' interests. The authors find that labor welfare first increases with the labor pool size and then decreases.

The only empirical studies that incorporate workers' behaviors to our knowledge, besides Sheldon (2016), are Chen et al. (2017) and Kabra et al. (2017). Chen et al. (2017) documented how Uber drivers value real-time flexibility and estimated the driver surplus from allowing a flexible schedule. The authors found that drivers earn higher surplus from Uber's flexible nature relative to less flexible arrangements. Kabra et al. (2017) investigated the impact of incentives using data from a Singaporean taxi company. Their structural estimation results suggest that offering incentives to drivers is more effective than passengers' incentives. While both of these papers rigorously captured how drivers and passengers responded to incentives and controlled for endogeneity, their models did not consider potential behavioral biases and did not causally explain workers' behavior. This is due to data limitations given that most datasets record trips only when they happen. In our dataset, however, we observe the information available to drivers even when they decided not to work. We focus on the behavior of gig economy workers and on how the platform can improve its operational decisions by understanding such behavior. We next present hypotheses on labor decisions of gig economy workers.

2.4. Hypotheses Development

We are interested in studying how gig economy workers make labor decisions, specifically whether they will work at a particular time and, if they will, for how long. Labor decisions typically depend on multiple factors such as weather and external commitments. Yet, these are not controlled by the platform and thus, while we attempt to control for such factors, we focus on the impact of economic drivers such as financial incentives and behavioral factors (income and time targeting). Given that financial incentives are the main lever companies can use to stimulate workers, their impact on the labor supply of traditional employees has been extensively studied in the literature. Following the structure of our data, we divide the incentive into two parts: an hourly wage rate and a temporary promotion. As is common in empirical studies of taxi drivers, we include income and time targeting effects in our model. Several companies have exploited workers' tendency to set goals by helping workers track their progress and nudging them to work for longer. Since individuals' targets cannot be observed, we model the amount of income each worker has already earned and the number of hours worked so far as proxies for income and time targets, respectively. We next present our hypotheses regarding the impact of each factor on gig economy workers' labor decisions.

H1: Higher wage increases the probability of working and the number of work hours. Following the standard income effect, we expect that a higher hourly wage will increase the probability of working. Empirical studies of workers who have discretion over their work hours, such as Oettinger (1999) and Stafford (2015) suggest that workers adjust labor decisions in the same direction as wage. Such positive income elasticity can be explained by two reasons. First, unlike traditional employees, gig economy workers tend to work on smaller tasks (e.g., assembling furniture, driving within a neighborhood) that require less time to complete, especially for workers who keep their full-time job. Consequently, decisions relate to short timeframes so that the main objective is likely to maximize utility (or the monetary outcome) in the following period. Second, we still believe that there exists a behavioral explanation that drives workers to stop working. This effect can be driven by the income earned so far and the time worked so far (see H2 and H3 below). Past studies that found an income targeting effect only modeled the relationship between the number of work hours and the average daily wage. We postulate that the negative impact on the number of work hours will only be apparent during specific times of day (resp. days of week), when workers might be closer to their daily (resp. weekly) income targets. Thus, when controlling for both cumulative income and work hours, we should observe a positive income elasticity. Further, it may become less positive later in the day (or week) when workers are closer to their targets.

H2: Higher income earned so far decreases the probability of working and the number of work hours. Past studies on labor supply assumed that workers will participate in the workforce if the offered wage is higher than their reservation price (see, e.g., Heckman 1974). Studies of taxi drivers including Camerer et al. (1997), Farber (2008), and Thakral and Tô (2017) provide support for an income-targeting behavior: the probability to stop working increases once the income target is reached. Thakral and Tô (2017) demonstrate that drivers' decisions are highly influenced by recent earnings. Note that none of the previous work attempted to model both the decision to work and the work duration, as we do in this paper. We build on the previous finding that the more workers have earned so far, the more likely they will stop working. An alternative explanation of the negative impact is related to *fatigue*. Specifically, a higher cumulative income could indicate a greater level of effort. Consequently, workers experience fatigue and thus work for a shorter time. As a result, we expect to see a negative impact of the cumulative income on both the probability of working and on the number of work hours.

H3: Longer time worked so far decreases the probability of working and the number of work hours. Previous work in labor economics suggest another type of targeting behavior: time targeting. Using the dataset of Farber (2005, 2008), Crawford and Meng (2011) develop a structural stopping estimation model that allows for reference points in both daily income and daily hours. The authors conclude that drivers are income loss averse relative to a reference point based on income and time. Agarwal et al. (2015) estimate hours and income targets using ex-post data of Singaporean taxi drivers and find that drivers with high cumulative hours and/or cumulative income for the day are more likely to stop working earlier when they reach their targets. Farber (2015) uses a discrete choice stopping model to analyze the complete record of all trips taken in NYC taxis from 2009 to 2013 and finds that the probability of ending a shift is positively related to work hours rather than income earned. Moreover, longer work hours could lead to fatigue. This phenomenon is aligned with recent findings suggesting that work performance deteriorates toward the end of long shifts among paramedics (Brachet et al. 2012) and part-time call center agents (Collewet and Sauermann 2017). Thus, one may expect that the longer the workers have worked, the less likely they will continue working and, if so, for a shorter time.

3. Data: Ride-hailing Platform in New York City

To answer our research questions, we collaborate with an on-demand ride-hailing company (referred to as "the company" or "the platform") and analyze a large comprehensive dataset of driving activities and financial incentives in New York City over a period of 358 days (from October 2016 to September 2017). Our data includes: each driver's vehicle type, experience with the platform, number of hours driven, and financial incentives offered and earned. The key advantage of our data is that we observe the incentives that were offered to *every* driver regardless of the decision to drive. In other words, even for drivers who decided not to drive for a particular time period, we still know their offered wage and promotions for that period. In total, we have several million driver-shift observations and several thousand unique drivers.[3] We next present an overview of the platform and report descriptive statistics of working shifts, financial incentives, and vehicle types.

3.1. Platform Overview

The company is a ride-hailing online platform that offers services in several cities worldwide. The users (riders) may request rides in real-time through a smartphone application. Then, the platform will match riders with available drivers. This platform offers a sharing service (i.e., several passengers heading in the same direction may share the same vehicle). To make the service more efficient, passengers can be picked up and dropped off at an optimized location near the exact requested locations. Finally, drivers usually own larger vehicles and the vast majority are compensated according to a guaranteed hourly rate regardless of the number of completed rides (this compensation model is different from several other ride-hailing platforms). We focus on drivers who are paid by the hour as this scheme resembles the traditional wage model but with more flexibility on the drivers' side. This allows us to investigate how drivers' work decisions are influenced by variations in financial incentives.

Figure 1 Breakdown of shifts for each operating day

3.2. Shifts and Work Schedule

Each operating day is divided into six shifts (see an illustration in Figure 1): morning non-rush hours from midnight to 7am (*AM Off-peak*), morning rush hours from 7 to 9am (*AM Peak*), midday from 9am to 5pm (*Midday*), afternoon rush hours from 5 to 8pm (*PM Peak*), evening non-rush hours from 8 to 9pm (*PM Off-peak*), and late night from 9pm to midnight (*Late night*). The largest volume of activities happen during PM Off-peak, followed by PM Peak, and Midday, while AM Off-peak hours are the least busy. In our data, a driver works on average 11 weeks (per year), 2.23 days per week, and 3.27 hours per day.

In this paper, we analyze drivers' behavior both at the shift and day levels. We control for the day of the week to account for demand and supply variation. In our data, 49.46% of all completed trips occurred between Tuesday and Thursday, potentially confirming the popularity of the service

[3] We cannot reveal the exact number of drivers and the size of our dataset due to confidentiality. However, these exact numbers do not affect any of our results or findings.

among city commuters. Monday and Friday trips account for 30.91% of all trips, while weekend trips account for 19.62%. Fewer trips on these days can represent either lower demand from customers or increased competition from other ride-hailing platforms. While the flexible work schedule allows drivers to choose when to work, they often stick to their "regular" times. For example, 30.41% of drivers never worked on weekends. We therefore include past work schedules in our models to control for such heterogeneity among drivers.

3.3. Earnings and Incentives

Drivers receive an hourly rate for the time they are *active* on the platform, which we refer to as *offer* in the sequel. They are considered active when they log on to the application on their mobile device and report to their designated start location. This compensation scheme can be considered as a guaranteed payment, in contrast to a commission-based contract that compensates drivers for each completed trip—commonly used by several platforms. It is possible under this scheme that drivers could be paid even if there are no ride requests for the entire hour. Similar schemes are used by other gig economy companies such as DoorDash, a U.S. on-demand delivery service.

The guaranteed hourly offer comprises two components: a base rate and a promotional rate. These two variables vary over time (shifts and days of week) and across different drivers. The base rate for each driver is decided when the driver joins the platform for the first time. For the same driver, the base rate may be different for different shifts and different days of the week, but typically remains the same across weeks. In addition to the base rate, drivers are frequently offered three types of promotional incentives: rate-based, minimum-hour, and combo promotions. Drivers can only earn these bonuses if they drive during particular shifts as indicated in the promotion terms and conditions. Rate-based promotions are offered as an additional bonus to the hourly base rate during specific times. In our data, 32.71% of shifts include rate-based promotions and the average promotion rate is an additional 50.36% of the base rate. Minimum-hour promotions provide a lump sum if the driver works continuously for at least some pre-determined number of hours during a specific time. Similarly, a combo promotion is unlocked once a driver meets certain requirements.

At the time of our data, promotions were decided as follows: First, the platform sets a number of promotional rates as benchmarks. Then, an algorithm uses these rates to assign the final rate for each driver based on the past work behavior and the vehicle type (see Section 3.4). Ultimately, the promotion mainly depends on the past work behavior and on the vehicle type. The platform then sends text messages to drivers every evening to communicate the promotional rates for the next day. This suggests that drivers are likely to plan their upcoming work schedule ahead of time. Occasionally, drivers may receive real-time adjustments to their rates but will never experience lower rates than initially informed. All rates are pro-rated to the actual amount of time worked for a given shift. Earnings are cumulative until the end of the week when drivers have the option to transfer them to their bank account.

3.4. Drivers and Vehicle Types

Drivers are identified by a unique ID. For each shift, we observe the decision to work (i.e., to become active) for every driver registered in the system. For drivers who started working after the first day of our dataset (October 7, 2016), we record their first day to control for their experience with the platform. All drivers were present in the dataset for every day after their first day, but will only drive in some of the days. For the analysis conducted in this paper, we only consider the drivers who own a single vehicle (89.9% of all drivers). There are six types of vehicles: a 3-passenger SUV, a 4-passenger SUV, a 5-passenger SUV, a 5-passenger van, a 6-passenger van, and a 3-passenger sedan. We exclude from our analysis the van drivers who lease their vehicle from the company (rather than owning their vehicle or leasing it from an external third party) as they may have a completely different utility from driving, leaving us with 84.4% of the original pool. For our main analysis, we focus on two types of drivers: sedan and 5-passenger SUV, which are 33.2% of the entire driver pool. These two vehicle types potentially represent drivers who may behave differently. Sedan vehicles are generally less expensive to maintain and the owners can use their vehicles for multiple purposes. From our data, we observe that SUV drivers typically work more frequently and for longer hours relative to sedan drivers. We obtained similar qualitative results for other vehicle types; but omit the results for conciseness.

3.5. Supplementary Data

To better capture the market conditions, we incorporate trip records for both yellow taxis and for-hire vehicles (FHVs) collected by the New York City Taxi and Limousine Commission (TLC).[4] In particular, we analyze 101,487,565 yellow taxi trips and 129,868,077 ride-hailing or FHV trips that occurred between October 2016 and September 2017. Yellow taxi trip records include dates/times/locations of pick-ups and drop-offs, itemized fares, and driver-reported passenger counts. FHV trip records consist of dates/times/locations of pick-ups and drop-offs and the dispatching base license number that is associated with a ride-hailing platform.

We also retrieve weather data from the Dark Sky API, which provides minute-level weather information for a specific location (NYC in our case). Such information includes humidity, precipitation probability and intensity, precipitation type (rain, snow, or sleet), temperature, apparent or "feel-like" temperature, visibility, and wind speed. For shift-level analysis, we use averages of these measures for each hour in the shift. Similarly, we use day averages for the day-level analysis.

[4] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

4. Empirical Approach

To test the hypotheses developed in Section 2, we estimate the impact of financial incentives, income and time targets, and other covariates on two labor decisions: (1) driving or not (captured by the binary variable $Drive_{i,t}$) and (2) how long to drive ($Hours_{i,t}$, a positive number). Specifically, $Drive_{i,t} = 1$ if Driver i drives at least five minutes in Shift t and $Drive_{i,t} = 0$ otherwise. $Hours_{i,t}$ represents the number of hours that Driver i is active in Shift t. Given the long tails in $Hours_{i,t}$, we apply a Box-Cox transformation to normalize its distribution, conditional on the covariates. We obtain consistent results under different transformations (e.g., logarithm and square root). We exclude outliers defined as drivers for which the number of work hours in a given shift or day exceeds the 1.5 interquartile ranges (IQRs). We also exclude public holidays from our analysis.

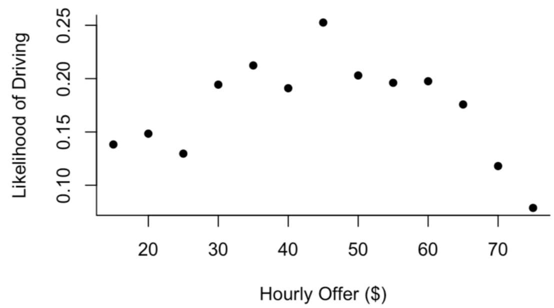
We estimate our model both at the shift (i.e., within day) and day levels, separately for each vehicle type. This allows us to capture type-specific heterogeneity: drivers operating different vehicle types may have different preferences, costs, and utility functions, and thus make their labor decisions differently. In this section, we first discuss the empirical challenges that arise from our observational data—particularly simultaneity and sample selection bias—and our econometric approach to overcome these challenges. We then provide a detailed description of our two-stage method.

4.1. Simultaneity Bias Correction: Instrumental Variables

Simultaneity Bias. As discussed, the standard income effect suggests that financial incentives encourage workers by increasing their likelihood of working or their number of work hours. Nevertheless, quantifying the effect of incentives by regressing the labor decision on financial incentives can lead to misleading results. In our dataset, we observe that a smaller fraction of drivers who received an hourly offer of \$65 decided to work relative to those who received \$45 per hour (see Figure 2). One possible implication is that financial incentives are not effective in inducing some drivers. Alternatively, these appealing promotions might have been strategically offered to engage inactive drivers. Consequently, regressing the driving decision (or the number of work hours) on financial incentives can lead to a *simultaneity bias* as the incentives are likely to be determined by the state of the supply or the number of active drivers. Overlooking this issue may yield to a bias estimate of the effect of financial incentives. A common solution is to use instrumental variables (IVs) that are highly correlated with the financial incentives, but affect the work decision only through the incentives (Levinsohn and Petrin 2003).

Instrumental Variables. The main endogenous variables in our data are the hourly financial incentives, $w_{i,t}$, and the hourly earnings, $\tilde{w}_{i,j}$. Our ideal instrument is one that is highly correlated with each endogenous variable and affects the dependent variable (the decision to drive or the work hours) only through the endogenous variable. In other words, we are looking for instruments that are not correlated with the unobserved variables in the error terms. Our industry partner confirmed that the financial incentives were endogenously determined with respect to supply decisions.

Figure 2 **Example of a nonlinear relationship between the likelihood of working and the hourly offer**



Specifically, the company sets the financial incentives based on past work history, level of inactivity, and vehicle type. This insight motivated us to focus on instruments that categorize drivers based on these three factors. We propose to use three types of IVs:

1. Our first type of IV is based on the notion of *co-workers*. For each driver who is available to work at a particular time (i.e., has not terminated his/her partnership with the platform), we define his/her co-workers as the drivers who meet the following conditions: (i) available to work at the same time, (ii) drive a different vehicle type, and (iii) have made the same work decision in the past (i.e., the same shift in the previous week or previous month). Work decisions of interest are driving, not driving, or not yet registered. Assuming that random shocks are not correlated across drivers, we propose to use the average hourly offers received by co-workers for the focal period as an IV. This IV satisfies the *relevance condition:* since both the focal driver and his/her co-workers made the same work decision in the past, their incentives should be highly correlated. From the first stage of our IV estimation, the estimate for the instrument is consistently signifiant and F-statistics for all models are larger than the conventional threshold of 10. This IV also satisfies the *exclusion restriction:* current incentives for co-workers should not directly influence the focal driver's work decision because (i) they drive different vehicle types, and (ii) the focal driver does not have access to co-workers' incentives information.
2. Our second type of IV follows a similar idea. Instead of matching drivers based on their work decisions at a specific time in the past, we now match drivers based on the level of past inactivity. For every day in our data, we categorize drivers into four groups based on each quartile of the number of consecutive days they have been inactive. We call the drivers of a different vehicle type who belong to the same group *co-skippers*.
3. Finally, to further test the robustness of our results, we also consider a Hausman-type IV. This instrument is based on the same strategy as in previous literature (e.g., Hausman et al. 1994, Sheldon 2016) and uses the average hourly offer rate received by all other registered drivers during the same shift on the same day as an instrument for the offer rate.

We obtain consistent insights under all three specifications. Further discussion and our estimation results using these last two IVs are deferred to Appendix B.1. Lastly, since incentives are updated on a shift basis, we assume that drivers have to make a decision at the beginning of each shift whether or not to work and for how long. Cumulative earnings and hours worked, which are related to work decisions in previous shifts, should not be related to random shocks in the new shift. As a robustness check, we consider a model in which we use lagged values of cumulative earnings and worked hours as instruments for the current values. The results remain qualitatively the same.

4.2. Two-Stage Estimation

Sample Selection Bias. Previous studies such as Camerer et al. (1997) and Sheldon (2016) investigated the relationship between the number of work hours and the hourly wage rate conditional on drivers who worked on a given day. This would not be a concern if drivers randomly decide whether or not to work. In reality, however, it is more plausible that they make such decisions based on factors which are not observed by the researcher. In other words, the selection of drivers who choose to work at a given time is not random. Consequently, this approach may yield a biased estimate of the sensitivity to incentives (i.e., income elasticity). Fortunately, the comprehensiveness of our data offers an opportunity to address this challenge. Since we observe both active and inactive drivers, we can directly estimate the selection problem.

Modified Heckman Two-Stage Estimation. Our dataset provides a unique advantage as we observe financial incentives offered to every driver, including those who chose not to drive at any given time. To account for the sample selection bias, we use the modified two-stage Heckman method (Heckman 1979) to first estimate the driving decision across the entire population of drivers using a probit regression, and then estimate the number of work hours for drivers who chose to work. The key to connect the two estimation stages is the inverse Mills ratio (IMR), which is computed from the predicted probability of driving and included as a regressor in the second stage.

Let $w_{i,t}$ and $\tilde{w}_{i,t}$ be the hourly offer and hourly earning rate earned by Driver i at time t, respectively. $Hour_{i,t}$ is the observed number of hours that Driver i worked in shift or day t. $\boldsymbol{X_{i,t}}$ and $\boldsymbol{Z_{i,t}}$ are the relevant set of covariates that affect the decision to work and the work hours, respectively. We model the two stages as follows:

$$Hour_{i,t} = \begin{cases} Hour^*_{i,t} & \text{if} \quad Drive_{i,t} = 1 \\ \text{unobserved} & \text{otherwise} \end{cases} \tag{1}$$

$$Drove_{i,t} = \begin{cases} 1 & \text{if} \quad Drive^*_{i,t} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$Hour^*_{i,t} = \beta_{0,i} + \beta_{\tilde{w}} \tilde{w}_{i,t} + \boldsymbol{\beta} \boldsymbol{Z_{i,t}} + u_{i,t} \tag{3}$$

$$Drive^*_{i,t} = \alpha_0 + \alpha_w w_{i,t} + \boldsymbol{\alpha} \boldsymbol{X_{i,t}} + v_{i,t} \tag{4}$$

$$\begin{bmatrix} \sigma_v^2 \\ \sigma_u^2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_u \\ \rho\sigma_u & \sigma_u^2 \end{bmatrix} \right). \tag{5}$$

The two stages that we estimate are given by:

$$P(Drive_{i,t} = 1|\boldsymbol{X_{i,t}}) = \Phi(\alpha_{0,i} + \alpha_{0,t} + \alpha_w w_{i,t} + \boldsymbol{\alpha X_{i,t}}), \tag{6}$$

$$f(Hour_{i,t}) = \beta_{0,i} + \beta_{\tilde{w}} \tilde{w}_{i,t} + \boldsymbol{\beta Z_{i,t}} + \theta\lambda_{i,t} + u_{i,t}, \tag{7}$$

where $\Phi(\cdot)$ is the normal c.d.f. and $\lambda_{i,t}$ is the IMR calculated from the predicted probability in Equation (6) (*"Choice Equation"*). Thus, we essentially estimate a probit model for the driving decision in Equation (6) and compute the IMR for each observation. We then fit an OLS model of the (transformed) number of hours that depends on all covariates and the IMR (Equation (7)) while controlling for the drivers who drove (*"Level Equation"*). The estimated coefficient $\theta = \rho\sigma_u$ will potentially confirm the existence of a sample selection bias.

While Heckman-type selection model has been widely used in several applications, it has also been criticized on its potential pitfalls, particularly the weak nonlinearity of the IMR and the multicollinearity of regressors in both stages (Puhani 2000). To address these concerns, we carefully choose the sets of regressors for both stages ($\boldsymbol{X_{i,t}}$ and $\boldsymbol{Z_{i,t}}$) to be different (see more details in Section 4.3) and we check for collinearity by regressing the IMR on the regressors of the second stage. On average, the standard deviation of the errors is 44.52% less than the standard deviation of the IMR, which suggests a substantial difference. We also perform a different approach suggested by Puhani (2000) by estimating a subsample OLS or a two-part model. In the two-part model, a binary choice model is estimated for the probability of observing a positive-versus-zero outcome (e.g., the number of work hours). This is essentially the same as the first stage of our main approach. Then, conditional on a positive outcome (e.g., drivers who worked during a particular shift), an OLS regression model is estimated for the number of work hours (Cragg 1971, Madden 2008, Farewell et al. 2017). This is the same as the second stage of our main approach excluding the IMR. We report the estimates from both the two-part model and our main approach in Section 5.

For robustness purposes, we also consider the Dahl's approach for sample selection correction. Dahl (2002) argued that, when the choice equation is multinomial, semi-parametric estimation of this model may face the "curse of dimensionality." As the number of alternatives grows, it requires the estimation of a large number of parameters, which rapidly makes it intractable. Dahl's approach relies on using a basis spline to approximate the choice probability for only a subset of relevant choices. For more details, we refer the reader to Bourguignon et al. (2007) that provides Monte Carlo comparisons across different selection models and to Bray et al. (2019) that implements this correction to model proximity-based supplier selection. In our context, the choice for each driver is binary: whether to work or not. Instead of using an IMR to correct for sample selection, we use a basis spline to approximate the choice probability. Our results remain consistent under both sample selection correction approaches and can be found in Appendix B.2.

4.3. Estimation Details

Relevant Covariates. The covariates for the choice equation include financial incentives, income and time targets, driving habits, as well as other controls. We use the hourly offer rate (i.e., hourly base rate plus promotions, if any) as a proxy for incentives. We use the total number of driving hours since the beginning of the day until right before the focal shift, or "hours so far" HSF, as a proxy for a *time target*. For the day level analysis, we use the number of hours the driver drove since Monday until the focal day as HSF. A higher HSF can be viewed as a signal that the driver is getting closer to his/her time target. Similarly, we use the cumulative earnings since the beginning of the day (or week) until the focal period, or "income so far" ISF, as a proxy for an *income target*. Both of these variables represent how close drivers are to their privately known targets. To further identify potential nonlinear effects of targets, we also consider specifications that include quadratic terms of both variables: HSF^2 and ISF^2.

A potential concern of including both HSF and ISF in the same specification is the multicollinearity issue. This issue does not significantly affect our results because of three reasons. First, despite a positive correlation, HSF and ISF are not a direct transformation of each other, hence there is no perfect correlation. Intuitively, HSF increases linearly with time as it denotes the exact amount of time the driver has been working, while ISF evolves dynamically as it depends on time-varying financial incentives. Second, multicollinearity generally makes causal inference difficult because the variance of each estimate would be inflated, leading to statistical insignificance. Our results (see Section 5) show that this is not the case for our model as both coefficients are statistically significant in most cases. Third, potential problems from high collinearity can be largely offset with sufficient power (Mason and Perreault Jr 1991). Our dataset consists of large enough number of observations to provide sufficient statistical power even when we separately estimate our model by vehicle type, day of the week, and shift of the day. When controlling for drivers with similar HSF (ISF), we can exclusively focus on the effect of ISF (HSF) and confirm its direction and significance. To further reduce the collinearity, we also consider alternative specifications that replace ISF with a natural logarithm of ISF and obtain consistent insights.

To capture aggregate demand variation for ride-hailing services, we include hourly weather information such as humidity, apparent temperature, and precipitation probability, the day of week, and month-year fixed effects. Furthermore, we compute the numbers of trips operated on other ride-hailing platforms using the TLC trips record data. For the choice equation, we include the trips initiated in the previous period (shift or day) to reflect the market condition observed by the drivers on our platform. For the level equation, we use the trips initiated in the same period to reflect the subsequent decision of work hours (i.e., when to quit). We also include other variables that capture drivers' short- and long-term habits. Short-term habits include the number of hours each driver worked on the same day and shift of the previous week and the number of hours worked during the previous week. Long-term habits are captured by the driver's experience (i.e., whether s/he is new to the platform) and drivers' fixed effects. We randomly split our observations into 65% training, 30% testing, and 5% validation sets for model selection. The final sets of covariates are then decided based on in-sample fitting, comparing AICs through stepwise method and the mean squared errors through LASSO regression, and out-of-sample prediction accuracy. The final sets of regressors in our main model are:

- **Choice:** hourly offer, ISF, HSF, number of hours worked last week, new driver indicator, humidity, apparent temperature, precipitation probability, number of other ride-hailing trips in the previous shift/day (in thousands).
- **Level:** hourly earning rate, ISF, HSF, number of hours worked on the same shift of last week, humidity, apparent temperature, precipitation probability, number of other ride-hailing trips during the same shift/day (in thousands).

4.3.1. *Choice:* Control Function Probit. The first stage (choice estimation) is based on a probit model of labor decisions. In our context, we estimate whether each driver would work during a given shift (or day). As discussed, financial incentives and work decisions are simultaneously related so that we need to use instrumental variables. However, we cannot perform a two-stage least squares (2SLS) to account for endogeneity in our probit model as this will lead to inconsistent estimates because certain properties of the expectation and linear projection operators do not carry over for nonlinear models (Newey 1987). Instead, we implement the control function method to account for endogeneity for our nonlinear probability model (Imbens and Wooldridge 2007, Wooldridge 2015). The first step is identical to the first step of 2SLS, that is, we estimate an OLS regression of the endogenous variable on the exogenous covariates and instrumental variables. We can then keep the endogenous variable (hourly offer rate) in the model and include the residuals from the first stage as an additional regressor. The intuition behind this method relies on using the instrument to split the unmeasured confounders into two parts, one that is correlated with the endogenous regressor and one that is not. We correct for the standard errors using the standard deviation of the residuals following Imbens and Wooldridge (2007). We can then compute the IMR for each observation using the fitted probability.

We also allow for drivers and time fixed effects throughout our estimation. Adding fixed effects to the nonlinear choice equation is known to suffer from the incidental parameters problem. More precisely, the usual asymptotic properties of the maximum likelihood estimator are not guaranteed, thus leading to a biased and inconsistent estimator (Greene 2004). Fortunately, recent developments in bias correction, such as the jackknife estimation method (see Hahn and Newey 2004, Dhaene and Jochmans 2015 for more details on this method), allow us to obtain asymptotically unbiased estimates and alleviate the incidental parameters problem. The estimates from this modified choice equation remain similar both in terms of sign and statistical significance. The resulting level equation also yields the same qualitative results.

4.3.2. ***Level:*** **Fixed Effects 2SLS.** The second stage aims to estimate the number of work hours conditional on the fact that the driver works during the focal shift (or day). Incorporating the IV approach to the level equation is straightforward as we can simply perform a 2SLS regression in which we first obtain the predicted value of $\tilde{w}_{i,t}$ based on exogenous covariates and the IV. As we include the IMR as one of the regressors in the second stage, we bootstrap the standard errors by repeating our analysis on resampled datasets. Furthermore, even if we already perform separate analyses for drivers with different vehicle types, we still include a driver-specific intercept in the level equation to control for unobserved heterogeneity among drivers. Time (month-year and day of week) fixed effects are also included to capture seasonal trends.

5. Empirical Results

We first conduct our analysis at the shift level. We compare the results obtained for SUV and sedan drivers during the Midday shift. We then draw several insights that help us rigorously verify the hypotheses developed in Section 2. We next consider the problem at the day level and contrast the results relative to the shift-level analysis.

5.1. Within-Day Analysis

We examine drivers' labor decisions by considering the decision at the beginning of each of the six operating shifts (see Figure 1). As 97% of the drivers in our data did not drive overnight, we assume that the first shift of the day is AM Off-peak. Using the first two shifts (AM Off-peak and AM Peak) as baselines, we analyze the remaining four shifts (Midday to Late Night) to investigate how labor decisions are influenced by financial incentives ($Offer$) as well as by cumulative earnings (ISF) and work hours (HSF) since the beginning of the day.

For each shift, we first estimate the choice equation in which the outcome variable is a binary decision of whether to work for the focal shift. We then estimate the level equation that relates the number of hours worked for the shift to the hourly earning rate, cumulative earnings and hours previously worked, and other covariates. We compare three model specifications: baseline OLS, 2SLS without correction for sample selection bias ("two-part model"), and our main model which is a 2SLS with sample selection correction. Tables 1 and 2 display our estimates for the Midday shift of SUV and sedan drivers, respectively. The first column in both tables reports the estimates from the control function probit of the choice equation. The second column reports the estimates from a baseline OLS for the level equation in which we replicated the model implemented in previous work (Camerer et al. 1997, Sheldon 2016). We included the following covariates: log hourly wage, temperature, rain indicator, day of week, and month dummies. We also used the same IV as in previous work, that is, the average of other drivers' hourly wages and estimated the model both with and without drivers' fixed effects. We report the estimates for the model with the best in-sample and out-of-sample fit. We then present the estimates from the level equation of the two-part model (i.e., our main model without sample selection correction) in the third column and our main model for the level equation in the fourth column.

Table 1 Estimates of two-stage selection models of SUV drivers' decisions during Midday shifts

	Choice Eq	*Level Eq Baseline*	*Level Eq Two-Part*	*Level Eq Main Model*
Incentives/targets				
Offer/Earnings	0.002^{***} (0.0006)	-0.083^{***} (0.019)	0.001 (0.001)	0.001 (0.001)
Income so far	-0.017^{***} (0.004)	-	-0.009^{***} (0.002)	-0.008^{***} (0.002)
Hours so far	2.904^{***} (0.163)	-	1.690^{***} (0.068)	1.826^{***} (0.070)
Hours last week				
Total	0.017^{***} (0.0003)	-	-	-
Same shift	-	-	0.056^{***} (0.002)	0.059^{***} (0.002)
New driver	0.590^{***} (0.060)	-	-	-
IMR	-	-	-	0.271^{***} (0.029))
Driver's FE	Yes	No	Yes	Yes
Month's FE	Yes	Yes	Yes	Yes
Weather	Yes	Yes	Yes	Yes
Observations	124,769	45,330	45,329	45,329
R^2	-	0.378	0.552	0.552

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

SUV drivers. For the choice equation, we find that hourly offer rate and cumulative hours worked so far have a positive impact on the decision to work, while cumulative earnings have a negative impact. The positive impact of hourly offer rate suggests that drivers respond positively to financial incentives (e.g., are more likely to work as the incentive increases) as predicted by the standard income effect. The positive effect of HSF suggests that drivers who have been active for a longer period will be more likely to continue working for a new shift. We refer to this behavior as *inertia*, which we will discuss further as it becomes more prevalent across different shifts. In contrast, the negative effect of ISF reflects a potential income-targeting behavior: drivers become more likely to quit (i.e., not continue working in a new shift) if they have earned more income or become closer to their unobserved earning goals. We also find that the number of hours each driver worked in the previous week has a significant positive impact on the decision to work. This could suggest that drivers tend to stick to their habits and hold regular work schedules. In other words, their past work frequency could play an important role in how their income and time targets are formed. Furthermore, we observe that newer drivers who recently joined the platform are significantly more likely to work.

We next consider the level equation. Interestingly, under the baseline model, we observe that SUV drivers exhibit a negative income elasticity, similar to full-time cab drivers investigated in Camerer et al. (1997) and Thakral and Tô (2017), rather than a positive income elasticity observed in ride-hailing drivers (Sheldon 2016). Controlling for drivers' fixed effects and sample selection bias conditional on drivers who worked, the estimates for the level equation are relatively consistent for all models. We observe a positive impact of hourly earnings on the number of hours worked, providing additional evidence that drivers exhibit positive income elasticity. The impact of ISF is significantly negative, suggesting that income-targeting behavior negatively influences both a decision to work and work duration. On the other hand, the impact of HSF or inertia behavior is positive across all specifications. Another evidence that drivers might stick to their schedules is the positive correlation between the number of hours worked during the same shift in the previous week and the number of hours worked during the current shift. In addition, the estimated coefficient of our sample selection correction variable (IMR) is statistically significant, confirming that selection into working is not random. Overall, we observe that the positive effects of hourly earnings and HSF dominate the negative impact of ISF on the number of hours worked. For an average SUV driver during a Midday shift, increasing the hourly earning rate by \$10 boosts the work probability by 0.82% and extends the work duration by 30 seconds. A \$10 increase in ISF reduces the work probability by 5.73% and shortens the work duration by 4.87 minutes. Lastly, an additional hour of HSF increases the work probability by 57.21% and triggers almost two hours of extra work (109.53 minutes).

Table 2 Estimates of two-stage selection models of sedan drivers' decisions during Midday shifts

	Choice Eq	*Level Eq Baseline*	*Level Eq Two-Part*	*Level Eq Main Model*
Incentives/targets				
Offer/Earnings	0.007^{***} (0.0008)	0.080^{***} (0.028)	0.001 (0.001)	0.001 (0.001)
Income so far	-0.031^{***} (0.006)	-	-0.007^{***} (0.002)	-0.007^{***} (0.002)
Hours so far	3.243^{***} (0.192)	-	1.073^{***} (0.058)	1.058^{***} (0.061)
Hours last week				
Total	0.022^{***} (0.0004)	-	-	-
Same shift	-	-	0.079^{***} (0.003)	0.078^{***} (0.003)
New driver	0.660^{***} (0.042)	-	-	-
IMR	-	-	-	-0.029 (0.029)
Driver's FE	Yes	Yes	Yes	Yes
Month's FE	Yes	Yes	Yes	Yes
Weather	Yes	Yes	Yes	Yes
Observations	113,444	20,307	20,297	20,297
R^2	-	0.389	0.580	0.580
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$			

Sedan drivers. We perform the same estimation and obtain similar results for sedan drivers: hourly offer/earning rate and HSF have a positive impact on the decision to drive and the number of hours worked. Under the baseline approach, we observe that, for sedan drivers, (log) hourly earnings rate positively affects the number of hours worked. Such positive income elasticity is in line with findings from Sheldon (2016) which also investigates the behavior of ride-hailing drivers. This suggests that SUV and sedan drivers are potentially different types of workers: SUV drivers' behaviors are similar to full-time professional taxi drivers, whereas sedan drivers' behaviors are similar to average drivers on ride-hailing platforms. While descriptive statistics suggest that SUV drivers tend to drive more often and for longer periods relative to sedan drivers, both types of drivers respond to financial incentives and behavioral biases in the same way. Note that the estimate for IMR is not statistically significant (at p=0.05) for this shift, suggesting that the evidence of selection of bias is weak. Nevertheless, our insights remain valid as the estimates are consistent across several models. Furthermore, IMR estimates are statistically significant for all the other shifts (See Appendix A). For an average sedan drivers during a Midday shift, a \$10 increase in hourly offer increases the work probability by 1.28% and extends the work duration by 45 seconds. A \$10 increase in ISF dampens the work probability by 4.70% and shortens the work duration by 4.21 minutes. Lastly, an additional hour to HSF boosts the work probability by 84.74% and increases the work duration by an hour (63.46 minutes).

Figure 3 Signs and statistical significance for estimates of two-stage models of drivers' shift-level decisions

	Choice (Work or not)						**Level** (How long)						
SUV	Mean	IV-F	Offer	ISF	HSF	N	Mean	IV-F	Earn	ISF	HSF	R^2	N
Midday	0.343	372.9	+	-	+	124,769	4.987	180.2	+	-	+	0.552	45,329
PM-Peak	0.277	345.1	-	-	+	131,910	2.421	58.5	+	-	+	0.244	39,592
PM-OPeak	0.182	320.6	+	-	+	130,651	0.731	50.5	+	-	+	0.281	26,699
Late Night	0.117	379.0	+	-	+	125,382	1.996	39.91	+	-	+	0.296	17,137
Sedan	Mean	IV-F	Offer	ISF	HSF	N	Mean	IV-F	Earn	ISF	HSF	R^2	N
Midday	0.137	224.9	+	-	+	113,444	4.186	78.0	+	-	+	0.580	20,297
PM-Peak	0.123	254.7	-	-	+	117,152	2.327	32.6	+	-	+	0.273	19,613
PM-OPeak	0.099	298.8	+	-	+	124,611	0.803	29.9	+	-	+	0.252	17,025
Late Night	0.071	299.5	+	-	+	124,280	2.167	32.9	+	-	+	0.304	15,623

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.

Estimates and effect sizes. Figure 3 summarizes the signs and statistical significance of the key estimates (hourly offer/earnings, ISF, and HSF) for each vehicle type and each shift. Each cell in the main three columns is color-coded as follows: a green cell with a bolded plus sign indicates a significant positive estimate ($p < 0.05$), a yellow cell with a bolded minus sign indicates a significant negative estimate ($p < 0.05$), and a white cell corresponds to a non-significant effect (the sign indicates the directional effect). In addition, mean work probability, F-statistics for IV estimation, mean number of hours worked conditional on working, adjusted total R^2, and number of observations are displayed alongside the estimates. Figure 4 provides the effect sizes for an average driver when one of the following conditions happens: (i) a \$10 increase in hourly offer or earning rate, (ii) a \$10 increase in ISF, and (iii) an additional hour to HSF.

Figure 4 **Effect sizes of changes in hourly financial offer/earnings, ISF, and HSF on drivers' shift-level decisions**

	% Change in P(Work)					Change in Minutes Worked				
SUV	Mean	+$10 Offer	+$10 ISF	+1h HSF	N	Mean	+$10 Earn	+$10 ISF	+1h HSF	N
Midday	0.343	0.82	-5.73	57.21	124,769	4.987	0.51	-4.87	109.53	45,329
PM-Peak	0.277	-4.39	-0.57	15.27	131,910	2.421	13.64	-0.24	18.96	39,592
PM-OPeak	0.182	0.27	-0.36	6.43	130,651	0.731	1.72	-0.08	1.18	26,699
Late Night	0.117	0.34	-0.22	3.32	125,382	1.996	14.87	-0.11	1.35	17,137
Sedan	Mean	+$10 Offer	+$10 ISF	+1h HSF	N	Mean	+$10 Earn	+$10 ISF	+1h HSF	N
Midday	0.137	1.28	-4.70	84.74	113,444	4.186	0.75	-4.21	63.46	20,297
PM-Peak	0.123	-1.45	-0.18	9.08	117,152	2.327	11.99	-0.50	6.95	19,613
PM-OPeak	0.099	0.31	-0.28	5.59	124,611	0.803	1.62	-0.11	0.33	17,025
Late Night	0.071	0.18	-0.20	4.12	124,280	2.167	21.65	-0.83	3.81	15,623

Note: Green: significantly positive, yellow: significantly negative, white: non-significant at p = 0.05.

We observe that the estimates for both types of drivers are substantially similar for most shifts. Hourly offers have a consistent positive impact on both choice (driving decisions) and level (hours worked) for both types of drivers. This result is consistent with the standard income effect that predicts a positive income elasticity and confirms our first hypothesis, that is, financial incentives encourage the decision to work and the work duration. However, we also observe behavioral biases with regards to cumulative earnings and hours. The impact of ISF on both stages is significantly negative, suggesting that drivers of either type become less likely to drive and drive for a shorter period when they have earned a higher income. This phenomenon reflects an income-targeting behavior among drivers and provides support for the behavioral theory of labor supply, that is, labor decisions are negatively influenced by income targeting. We thus find support for our second hypothesis. Lastly, we derive a new insight from the consistently positive impact of HSF for both types of drivers. Specifically, drivers who have been active for a longer period are more likely to be active in a new shift and work for longer. We refer to this phenomenon as *inertia*. Our third hypothesis is therefore rejected in the sense that, when controlling for both income- and time-targeting behaviors, drivers do not exhibit an aversion to working too many hours.

On average, the impact of adding an additional hour to HSF on the work probability is much stronger than the impact of additional hourly offer or ISF, but as the day proceeds this difference in magnitude decreases. We observe an interesting pattern for the marginal effects on work duration. Early in the day, the marginal effect of HSF is larger than that of hourly earnings, but in later shifts, the marginal effect of hourly earnings becomes larger than that of HSF. Fatigue or deteriorated performance discussed in Section 3 seems to be explained by income-targeting behavior rather than by its time counterpart. Putting these together, we conclude that *drivers exhibit positive income elasticity as predicted by the standard income effect but are also influenced by behavioral factors such as income targeting and inertia.*

5.2. Across-Day Analysis

We now aggregate drivers' decisions, incentives, and other covariates at the day level to investigate daily labor decisions. We assume that the week starts on Monday so that ISF and HSF capture cumulative income and hours worked since Monday. In this analysis, ISF and HSF are considered as a proxy for weekly income and time targets. The covariates in both stages are nearly identical to the ones used in Section 5.1 except that we replace the number of hours worked on the same *shift* of the previous week by the number of hours worked on the same *day* of the previous week. Figure 5 displays the estimates from our model for both vehicles types.

Figure 5 Signs and statistical significance for estimates of two-stage models of drivers' day-level decisions

	Choice (Work or not)						**Level** (How long)						
SUV	Mean	IV-F	Offer	ISF	HSF	N	Mean	IV-F	Earn	ISF	HSF	R^2	N
Tuesday	0.409	43.6	+	+	+	28,883	8.696	18.3	-	-	+	0.422	9,482
Wednesday	0.418	55.9	+	+	+	21,965	8.964	26.2	-	-	+	0.422	10,120
Thursday	0.426	73.4	+	+	+	29,233	9.053	34.6	-	-	+	0.412	9,894
Friday	0.412	74.0	+	+	+	20,294	8.915	33.7	+	-	+	0.436	9,283
Saturday	0.203	98.1	-	-	+	15,788	8.435	19.1	-	+	-	0.398	4,372
Sunday	0.162	82.2	-	-	+	13,025	7.927	15.1	+	+	-	0.390	3,240
Sedan	Mean	IV-F	Offer	ISF	HSF	N	Mean	IV-F	Earn	ISF	HSF	R^2	N
Tuesday	0.169	31.1	+	+	+	21,283	7.687	7.3	-	-	+	0.564	4,681
Wednesday	0.182	37.3	+	+	+	23,280	7.680	9.8	+	-	+	0.567	5,278
Thursday	0.179	47.5	+	+	+	19,982	7.724	11.6	-	-	+	0.542	5,081
Friday	0.171	46.7	+	+	+	18,418	7.568	11.2	-	-	+	0.533	4,666
Saturday	0.148	53.3	+	-	+	15,762	8.022	11.7	-	-	+	0.514	3,817
Sunday	0.129	45.5	-	-	+	12,602	7.708	11.4	+	-	+	0.560	3,065

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.

Analyzing the results across days, we draw considerably different conclusions from our shift-level analysis. While the positive impact of HSF on a decision to work remains consistent, the impact of hourly offer and ISF appear to be time-dependent. Prior to the weekend, both hourly offer and ISF positively encourage drivers to work. Drivers may perceive a recent high income as an indicator of a high demand and form an optimistic outlook on future market conditions. However, both effects become negative for Saturday and Sunday, resembling weaker income-targeting behavior. The results for the level equation shed another interesting insight. We do not find significant effects from the three main drivers in most specifications, except a consistent inertia observed among sedan drivers. Note that the estimates for the IMR are significant across all models, suggesting that there is indeed a sample selection bias in the daily work decision. One potential explanation is that, while gig economy workers make strategic decisions of whether to work on a daily basis, they do not seem to decide ahead of time how many hours they would work in the day. Instead, they are likely to make such decisions at the shift or hourly level as observed in our shift-level analysis.

5.3. Discussion

Our results offer an improved explanation of how gig economy workers make labor decisions and, in part, reconcile the debate between neoclassical and behavioral theories of labor supply. Table 3 summarizes our hypotheses and results. We find that, as predicted by the standard income effect, drivers respond positively to financial incentives. While we do not observe the strong negative income elasticity from the literature (such as Camerer et al. 1997), we find empirical evidence of an income-targeting behavior among drivers, suggesting that their labor decisions are influenced by recent earnings or income goals. Several gig economy platforms provide an application that features a real-time dashboard, making it simple for workers to track their progress and work history. In other words, information surrounding past earnings and work activities have become much more salient relative to traditional settings. By separating cumulative income from financial incentives, we show that the negative impact of income targeting stems from cumulative income rather than the hourly wage. Thakral and Tô (2017) draw a similar insight by showing the existence of income targeting among taxi drivers and identifying the recently earned cumulative income as a key factor in the decision to quit. In addition, we establish a new behavioral bias: workers with greater cumulative work hours are more likely to work even more, instead of exhibiting a time-targeting behavior. We refer to this phenomenon as inertia to reflect the tendency of workers with longer work hours to continue working and stay active for longer. With a better understanding of how gig workers make labor decisions (i.e., identifying behavioral factors such as income targeting and inertia), companies can design effective incentives and personalize these incentives based on individual workers' behaviors. Lastly, we find that gig workers make working decisions at both shift and day levels, whereas the work duration decision seems to be done at a more granular shift basis. The latter potentially highlights the unique flexibility of gig jobs.

Table 3 Summary of hypotheses and results

		Shift-level		Day-level	
	Statement	SUV	Sedan	SUV	Sedan
H1a	Higher wage increases P(work)	✓	✓	✓→ ✗	✓→ ✗
H1b	Higher wage increases work hours	✓	✓	✗	✗
H2a	Higher income so far decreases P(work)	✓	✓	✗→ ✓	✗→ ✓
H2a	Higher income so far decreases work hours	✓	✓	✗	✗
H3a	Longer work hours so far decreases P(work)	✗	✗	✗	✗
H3b	Longer work hours so far decreases work hours	✗	✗	✗	✗

Note: P(work): likelihood of working, ✓: fail to reject, ✗: reject, →: result differs later on in the day or week.

6. Managerial Implications: Incentive Optimization

In this section, we illustrate how gig economy firms can use our insights on workers' behavior to enhance their operations. We first investigate the benefit of improved incentive allocation based on two perspectives: increasing service capacity while keeping a fixed budget, and maintaining the same service capacity at a lower cost. We then further highlight the potential pitfalls of ignoring behavioral factors and quantify the resulting capacity loss.

6.1. Targeted Incentives

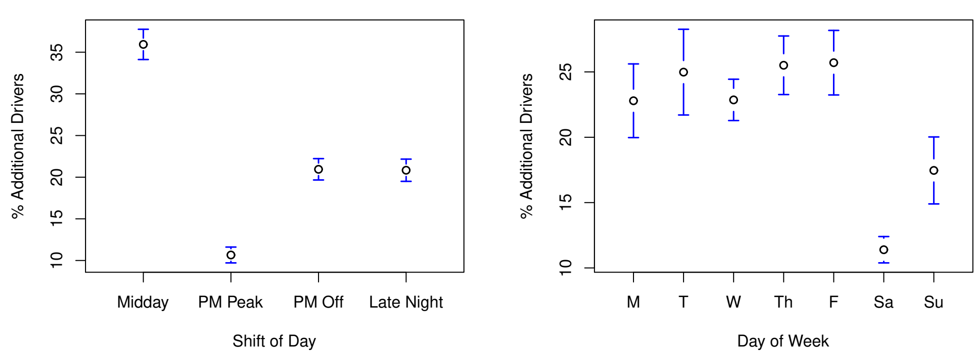
As mentioned, the impact of incentives on the number of active drivers may be nonlinear. This suggests that targeting specific drivers with different incentives can be beneficial. Instead of sending a universal incentive, we examine how the platform can improve its operational performance by offering personalized incentives based on drivers' attributes. As a benchmark, we compute the platform's actual (implemented) allocation of promotions. We then re-allocate incentives more efficiently using the following two perspectives: (i) increasing the service capacity (i.e., staffing more drivers) using the same budget, and (ii) maintaining the same service capacity at a lower cost. We propose to use a heuristic that ranks the drivers depending on the incentive level they need to receive in order to start working. We first compute the average projected and actual promotion expenses incurred by the platform for each shift (and day of the week) in our training data. The sum of all drivers' offered promotions is called the *projected bonus*. Recall that not every driver who received a promotion decided to drive. We thus also compute the average *actual bonus*.

We compare the service capacity and cost of our model relative to the actual allocation between January 1 and September 29, 2017. For each shift in each day, we train our model using all observations (in the same shift and day of the week) prior to the focal shift. Across 1,012 day-shifts, we observe that 94.59% of active drivers were sent a promotion but only 18.4% of these drivers claimed the offer and ended up driving. Moreover, 94% of the drivers who drove were not provided a promotion. These observations suggest that there is an opportunity to improve the current allocation of financial incentives.

We first compute the average proportion of drivers who drove during a given shift in a given weekday (using all past data). We denote this quantity by $\bar{D}$. We then compute the inverse c.d.f. evaluated at $\bar{D}$: $\tilde{D} = \Phi^{-1}(\bar{D})$, that is, $\tilde{D}$ represents the argument of $\Phi(\cdot)$ in the right hand-side of Equation (6). In other words, $\tilde{D}$ corresponds to the combination of drivers' attributes that will induce a probability of driving equal to $\bar{D}$. For each driver, we use all the covariates' values with the base rate (instead of the offered rate) in our fitted model. This will predict the probability of driving when offered only the base rate, $\hat{p}_i^{base}$. If $\hat{p}_i^{base} \geq \tilde{D}$, we label the driver as "driving without promotion." For other drivers, we compute the difference, $\Delta_i = \tilde{D} - \hat{p}_i^{base} > 0$, to determine the level of additional incentive needed for Driver i to start driving.

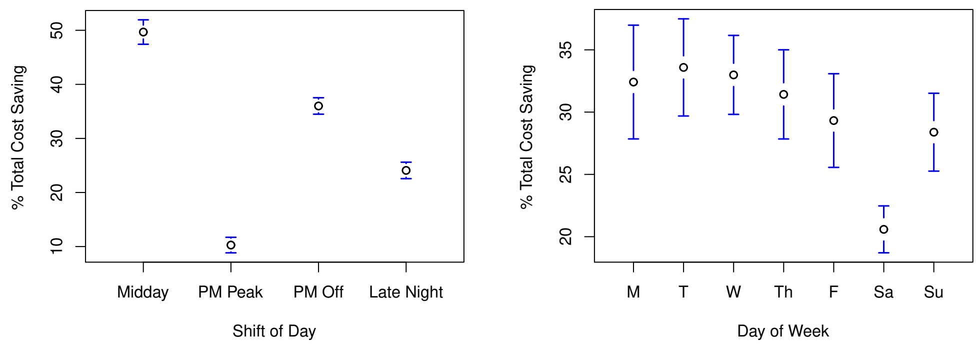
Improving service capacity while keeping the same budget. Using the actual promotion offered to drivers and the total cost observed in the data as a baseline, we can compute the platform's budget for promotions for any given day and shift, assuming that such budget is separate from the budget for base rates. It is fairly common in the retail and service industries that the initial pricing and subsequent promotions are decided separately. Thus, we first determine the number of drivers who would drive regardless of promotions (i.e., their base rates are appealing enough to trigger their decisions to work), and then rank the remaining drivers by increasing values of Δ_i. We compute the *minimum driving-inducing promotion level* by dividing Δ_i by the estimated coefficient $\hat{\beta}_{offer}$. We call this value $\tilde{\Delta}_i$. Then, a desired strategy can be to allocate the promotion budget to drivers with the smallest $\tilde{\Delta}_i$ until we exhaust the budget (or we can no longer encourage additional drivers). On average, our re-allocation procedure sends promotions to 6.27% of all available drivers. The 95% interval for the fraction of drivers who should receive a promotion is [0.44%, 19.92%]; these fractions are substantially lower than the current practice of the company. As a result, each targeted driver will receive an attractive deal. Under the allocation observed in the data, drivers were offered an average promotion of 0.58× relative to their base rate. In our proposed allocation, however, targeted drivers will receive an average promotion of 2.09×. Consequently, the proposed allocation will target a much smaller number of drivers but will offer a significantly more attractive promotion. Ultimately, using the same promotion budget, our approach can staff 22.1% additional drivers on average with a 95% interval of [2.46%, 50.50%]. Figure 6 reports the percentage increase in the number of drivers for each shift and weekday.

Figure 6 **Number of additional drivers using our allocation strategy**



Maintaining service capacity at a lower cost. An alternative strategy is to re-allocate the promotions so as to maintain the same number of drivers. Similar to the previous case, we rank all drivers by increasing values of promotion levels that trigger a driving decision (i.e., $\tilde{\Delta}_i$). We subtract the number of drivers who are predicted to drive without receiving any promotion from the desired service capacity. Instead of having a budget constraint, we now send promotions to drivers who require the smallest incentive $\tilde{\Delta}_i$ until we reach the desired service capacity. On average, our allocation costs 30.10% less relative to current practice with a 95% interval of [0.75%, 63.54%]. Figure 7 shows the percentage of cost savings for each shift and weekday.

Figure 7 **Simulated cost savings while maintaining the same service capacity**



6.2. Ignoring Behavioral Factors

In this section, we quantify the impact of capturing the main behavioral factors obtained in our estimation results. To this end, we investigate how much prediction accuracy the company would lose and how many workers it would fail to attract if it did not incorporate income targeting and inertia. We compare the following three scenarios to our model:

(a) ISF Only: The firm assumes that work decisions are influenced by ISF but not HSF.

(b) HSF Only: The firm assumes that work decisions are influenced by HSF but not ISF.

(c) *Base*: The firm completely ignores targeting and inertia behaviors.

Our analysis is at the day-shift level and reports out-of-sample predictions. The testing set consists of each day-shift between January 1, 2017 and September 30, 2017. For each day-shift in the testing set, we train four separate choice equations—one for each model (a)-(c) above and one for our model (i.e., drivers' decisions are influenced by both income so far and hours so far)—using all historical observations of the same day-shift from October 7, 2016 to the week prior to the focal date. Each of the four choice equations represents the predicted outcome depending on the assumption on workers' behavior. We first compute the fraction of drivers' work decisions that each model predicts correctly out-of-sample relative to the actual realization in the data. On average, our model outperforms the other three models in prediction accuracy both at the shift and day levels. Specifically, when the company ignores behavioral factors, it loses 8.6% in prediction accuracy on average. Following the same procedure as in Section 6.1, we compute the incentive allocation under each model. More precisely, we assume that each model is the true state of the world and optimize the financial incentives for each driver given the promotion budget observed in the data. Once the incentives are allocated to all available drivers, we estimate the expected number of drivers assuming that the true state of the world is actually governed by our model. Note that by construction, our model will always outperform the other models in terms of expected number of drivers. Our main goal is to quantify the magnitude of capacity reduction when the company uses different assumptions on workers' behavior.

Figure 8 **Impact of ignoring behavioral factors on the expected number of active drivers**

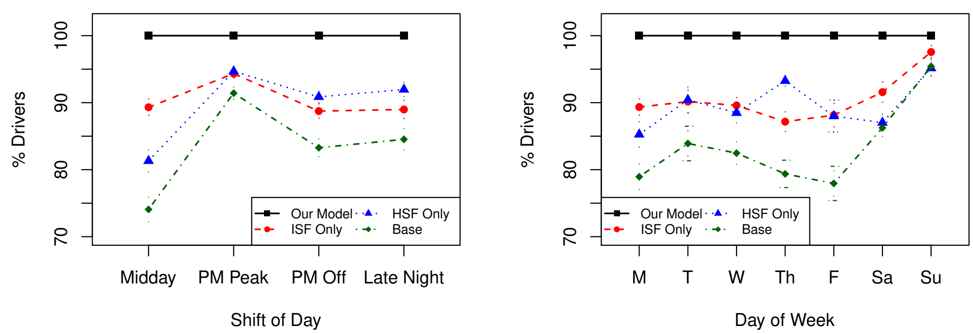


Figure 8 shows that ignoring behavioral factors can lead to a significant decrease in the number of active drivers. Specifically, the Base model leads to an average reduction of 16.70% in active drivers relative to our model, with a standard deviation of 13.06%. The ISF Only and HSF Only models suggest an average reduction of 9.63% and 10.32% in the expected number of active drivers, respectively, with standard deviations of 9.10% and 10.20%.

In summary, the above results suggest that it is important for gig platforms to account for income targeting and inertia. Ignoring these behavioral factors can decrease prediction accuracy, and more importantly, induce misleading incentive decisions that may result in suboptimal capacity levels.

7. Concluding Remarks

The recent rise of the gig economy has changed the way people think about employment. Unlike traditional employees who work under a fixed schedule, gig economy workers are free to choose their own schedule: when to work and for how long. Such flexibility poses a great challenge to gig platforms in terms of planning and committing to a service capacity. In this paper, we propose a framework to investigate how gig economy workers make labor decisions. Using data from a ride-hailing platform, we develop an econometric model that accounts for sample selection, simultaneity, and endogeneity bias. We find that financial incentives have a positive effect on the decision to work and on the numbers of hours worked, confirming the positive income elasticity from the standard income effect. We also observe the influence of behavioral factors through the cumulative income and number of hours worked. The dominating effect, inertia, suggests that, the longer workers have been active so far, the more likely they will continue working and the longer period they will work. Our results also reflect a unique feature of gig economy work. While workers decide whether to work on both shift and day levels, they decide on the duration of their service on a shift basis. Finally, our numerical experiments demonstrate that gig economy platforms can benefit from incorporating our insights on labor decisions into their incentive optimization.

One of the important phenomena that emerge from this paper is the existence of inertia among drivers. While we cannot conclude that all gig economy workers exhibit such a behavior, we believe that it has important implications that go beyond this study. Indeed, we believe the findings are generalizable for three reasons. First, there is nothing specific nor exclusive about the platform studied in this paper. Second, drivers working for this platform are often also working for competing platforms. Third, policies used by our industry partner are quite common in the industry. Therefore, there is a lesson to be learned about the fundamental impact of such policies. Amidst intensifying competition among providers of similar on-demand services, companies are making every effort to win over a mutual pool of workers. This paper empirically identifies several key behavioral factors that affect gig economy workers' decisions. Our research can be used to sharpen platforms' understanding on how gig economy workers make labor decisions, and ultimately improve platforms' operational decisions (e.g., sending the right offer to the right worker at the right time).

This paper opens several avenues for future research. It could be interesting to validate our findings by running a controlled field experiment. Given that online platforms routinely run experiments to confirm insights, testing the income targeting and inertia effects can be of interest. A second direction is to further investigate how workers construct their reference points or targets in both financial and time dimensions, and how these targets are updated over time. This will allow companies to gain insights about the (dis)utility of working as well as understanding how workers switch between service providers. Finally, our incentive allocation is based on simple ranking arguments. Developing a more comprehensive optimization framework to optimize incentives for each driver in each shift under further operational constraints is also an interesting extension. The main goals of this research stream are to refine our understanding of gig economy workers and develop data-driven methods that can be used by gig platforms to efficiently motivate and strengthen their relationships with their flexible workforce.

References

Agarwal S, Diao M, Pan J, Sing TF (2015) Are singaporean cabdrivers target earners?, working paper.

Allon G, Bassamboo A, Çil EB (2012) Large-scale service marketplaces: The role of the moderating firm. *Management Science* 58(10):1854–1872.

Benjaafar S, Ding JY, Kong G, Taylor T (2018) Labor welfare in on-demand service platforms, working paper.

Bourguignon F, Fournier M, Gurgand M (2007) Selection bias corrections based on the multinomial logit model: Monte carlo comparisons. *Journal of Economic Surveys* 21(1):174–205.

Brachet T, David G, Drechsler AM (2012) The effect of shift structure on performance. *American Economic Journal: Applied Economics* 4(2):219–46.

Bray RL, Serpa JC, Colak A (2019) Supply chain proximity and product quality. *Management Science* .

Camerer C, Babcock L, Loewenstein G, Thaler R (1997) Labor supply of new york city cabdrivers: One day at a time. *The Quarterly Journal of Economics* 112(2):407–441.

Campbell H (2017) The rideshare guy - 2017 reader survey. `https://docs.google.com/document/d/1QSUFSqasfjM9b9UsqBwZlpa8EgqNj6EBfWybFBSHj3o/edit`, Last accessed on 2018-08-31.

Carrington WJ (1996) The alaskan labor market during the pipeline era. *Journal of Political Economy* 104(1):186–218.

Chen MK, Chevalier JA, Rossi PE, Oehlsen E (2017) The value of flexible work: Evidence from uber drivers, working paper.

Chen MK, Sheldon M (2016) Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. *EC*, 455.

Cohen MC, Zhang R (2018) Competition and coopetition for two-sided platforms, working paper.

Collewet M, Sauermann J (2017) Working hours and productivity. *Labour Economics* 47:96–106.

Cragg JG (1971) Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica (pre-1986)* 39(5):829.

Crawford VP, Meng J (2011) New york city cab drivers' labor supply revisited: Reference-dependent preferences with rationalexpectations targets for hours and income. *The American Economic Review* 101(5):1912–1932.

Dahl GB (2002) Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica* 70(6):2367–2420.

Dhaene G, Jochmans K (2015) Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies* 82(3):991–1030.

Dong J, Ibrahim R (2017) Flexible workers or full-time employees? on staffing service systems with a blended workforce. Technical report, Northwestern University, working paper.

Farber HS (2005) Is tomorrow another day? the labor supply of new york city cabdrivers. *Journal of political Economy* 113(1):46–82.

Farber HS (2008) Reference-dependent preferences and labor supply: The case of new york city taxi drivers. *The American Economic Review* 98(3):1069–1082.

Farber HS (2015) Why you cant find a taxi in the rain and other labor supply lessons from cab drivers. *The Quarterly Journal of Economics* 130(4):1975–2026.

Farewell V, Long D, Tom B, Yiu S, Su L (2017) Two-part and related regression models for longitudinal data. *Annual review of statistics and its application* 4:283–315.

Greene W (2004) Fixed effects and bias due to the incidental parameters problem in the tobit model. *Econometric reviews* 23(2):125–147.

Gurvich I, Lariviere M, Moreno A (2016) Operations in the on-demand economy: Staffing services with self-scheduling capacity, working paper.

Hahn J, Newey W (2004) Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72(4):1295–1319.

Hall JV, Horton JJ, Knoepfle DT (2018) Pricing efficiently in designed markets: Evidence from ride-sharing, working paper.

Hausman J, Leonard G, Zona JD (1994) Competitive analysis with differenciated products. *Annales d'Economie et de Statistique* 159–180.

Heckman J (1974) Shadow prices, market wages, and labor supply. *Econometrica: journal of the econometric society* 679–694.

Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica: Journal of the econometric society* 153–161.

Hu M, Zhou Y (2017) Price, wage and fixed commission in on-demand matching, working paper.

Ibrahim R (2018) Managing queueing systems where capacity is random and customers are impatient. *Production and Operations Management* 27(2):234–250.

Imbens G, Wooldridge J (2007) Control function and related methods, what's new in Econometrics, National Bureau of Economic Research.

Jiang B, Tian L (2016) Collaborative consumption: Strategic and economic implications of product sharing, forthcoming in Management Science.

Kabra A, Elena B, Karan G (2017) The efficacy of incentives in scaling marketplaces, working paper.

Kőszegi B, Rabin M (2006) A model of reference-dependent preferences. *The Quarterly Journal of Economics* 121(4):1133–1165.

Levinsohn J, Petrin A (2003) Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies* 70(2):317–341.

Madden D (2008) Sample selection versus two-part models revisited: The case of female smoking and drinking. *Journal of health economics* 27(2):300–307.

Manyika J, Lund S, Robinson K, Valentino J, Dobbs R (2015) A labor market that works: Connecting talent with opportunity in the digital age. *June. McKinsey Global Institute. http://www. mckinsey. com/˜/media/McKinsey/dotcom/Insights/Employment% 20and% 20growth/Connecting* 20.

Mason CH, Perreault Jr WD (1991) Collinearity, power, and interpretation of multiple regression analysis. *Journal of marketing research* 28(3):268–280.

Newey WK (1987) Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 36(3):231–250.

Oettinger GS (1999) An empirical analysis of the daily labor supply of stadium venors. *Journal of political Economy* 107(2):360–392.

Pew-Research Center (2016) Shared, collaborative and on demand: The new digital economy.

Puhani P (2000) The heckman correction for sample selection and its critique. *Journal of economic surveys* 14(1):53–68.

Sheldon M (2016) Income targeting and the ridesharing market, working paper.

Stafford TM (2015) What do fishermen tell us that taxi drivers do not? an empirical investigation of labor supply. *Journal of Labor Economics* 33(3):683–710.

Taylor TA (2018) On-demand service platforms. *Manufacturing & Service Operations Management* .

Thakral N, Tô LT (2017) Daily labor supply and adaptive reference points, working paper.

Upwork (2017) Freelancing in america: 2017. *Upwork and the Freelancers Union* .

Wooldridge JM (2015) Control function methods in applied econometrics. *Journal of Human Resources* 50(2):420–445.

Appendix A: Additional Details of the Main Results

Figures A.1 and A.2 provide additional details of the main results from our two-stage models of drivers' decisions on a shift level and a day level, respectively. For each of the key variables, we provide an estimated coefficient and a standard error in parenthesis. Within each model, we also report an estimated coefficient and a standard error for IMR and two R-squared's, total R-squared (top) and within R-squared (bottom, italicized). We acknowledge that a few of the IMR estimates are not statistically significant, suggesting that the selection bias is weak in some cases. However, our insights regarding the impact of financial incentives, cumulative income, and cumulative work hours on the decisions of both stages are consistent across different model specifications and selection approaches (e.g., two-part model and Dahl's correction). In addition, we observe that SUV and sedan drivers may differ in terms of selecting into working for the shift or day.

Figure A.1 Estimates of two-stage models of drivers' shift-level decisions

	Choice (Work or not)				**Level** (How long)					
SUV	Offer	ISF	HSF	N	Earn	ISF	HSF	IMR	R^2	N
Midday	**0.0024 (0.0006)**	**-0.0173 (0.0036)**	**2.9044 (0.1632)**	124,769	0.001 (0.001)	**-0.008 (0.002)**	**1.826 (0.070)**	**0.271 (0.029)**	0.552 *0.239*	45,329
PM-Peak	**-0.0177 (0.0013)**	**-0.0022 (0.0002)**	**0.5020 (0.0082)**	131,910	**0.023 (0.005)**	**-0.0004 (0.0001)**	**0.316 (0.009)**	**0.627 (0.043)**	0.244 *0.092*	39,592
PM-OPeak	**0.0018 (0.0008)**	**-0.0024 (0.0001)**	**0.3436 (0.0048)**	130,651	**0.003 (0.001)**	**-0.0001 (0.00003)**	**0.020 (0.002)**	0.009 (0.011)	0.281 *0.029*	26,699
Late Night	**0.0035 (0.0010)**	**-0.0024 (0.0001)**	**0.2817 (0.0047)**	125,382	**0.025 (0.002)**	-0.0002 (0.0001)	**0.022 (0.011)**	-0.088 (0.054)	0.296 *0.027*	17,137
Sedan	Offer	ISF	HSF	N	Earn	ISF	HSF	IMR	R^2	N
Midday	**0.0068 (0.0008)**	**-0.0309 (0.0056)**	**3.2429 (0.1916)**	113,444	0.001 (0.001)	**-0.007 (0.002)**	**1.058 (0.061)**	-0.029 (0.029)	0.580 *0.206*	20,297
PM-Peak	**-0.0109 (0.0016)**	**-0.0013 (0.0004)**	**0.4787 (0.0133)**	117,152	**0.020 (0.004)**	**-0.001 (0.0002)**	**0.116 (0.009)**	**-0.120 (0.034)**	0.273 *0.014*	19,613
PM-OPeak	**0.0031 (0.0010)**	**-0.0028 (0.0003)**	**0.4133 (0.0090)**	124,611	**0.003 (0.0005)**	**-0.0002 (0.00004)**	**0.005 (0.002)**	**-0.098 (0.007)**	0.252 *0.029*	17,025
Late Night	0.0018 (0.0014)	**-0.0021 (0.0002)**	**0.3356 (0.0082)**	124,280	**0.036 (0.004)**	**-0.001 (0.0002)**	**0.063 (0.011)**	**-0.378 (0.048)**	0.304 *0.026*	15,623

Note: Green: significantly positive, yellow: significantly negative, white: non-significant at p = 0.05.

Figure A.3 provides the effect sizes for an average driver of each type when one of the following conditions happens: (i) a \$10 increase in hourly offer or earning rate, (ii) a \$10 increase in ISF, and (iii) an additional hour to HSF.

Appendix B: Alternative Model Specifications

B.1. Instrumental Variables

B.1.1. Co-skippers IV. This IV follows a similar idea to our main IV, but instead of matching drivers based on their past work decisions at a specific time in the past, we now match drivers based on the level of past inactivity. For every day in our data, we categorize drivers into four groups based on each quartile of the number of consecutive days they have been inactive. We call the drivers of a different vehicle type who belong to the same group *co-skippers*. This IV satisfies the *relevance condition:* Since both the focal driver and his/her co-skippers have been inactive for approximately the same time, their incentives should be highly correlated. From the first stage of our IV estimation, the estimate for the instrument is consistently

Figure A.2 Estimates of two-stage models of drivers' day-level decisions

	Choice (Work or not)				**Level** (How long)					
SUV	Offer	ISF	HSF	N	Earn	ISF	HSF	IMR	R^2	N
Tuesday	0.0039 (0.0021)	0.0006 (0.0003)	**0.0581 (0.0137)**	28,883	-0.003 (0.010)	-0.001 (0.001)	0.027 (0.029)	**-1.711 (0.184)**	0.422 *0.037*	9,482
Wednesday	0.0036 (0.0020)	**0.0005 (0.0002)**	**0.0461 (0.0087)**	21,965	-0.001 (0.008)	-0.0003 (0.0005)	0.028 (0.021)	**-1.274 (0.192)**	0.422 *0.040*	10,120
Thursday	**0.0087 (0.0019)**	**0.0005 (0.0001)**	**0.0358 (0.0061)**	29,233	-0.006 (0.008)	-0.0004 (0.0003)	**0.042 (0.014)**	**-0.973 (0.217)**	0.412 *0.046*	9,894
Friday	**0.0069 (0.0019)**	0.00001 (0.0001)	**0.0506 (0.0046)**	20,294	0.013 (0.008)	-0.0004 (0.0002)	**0.055 (0.012)**	0.007 (0.229)	0.436 *0.031*	9,283
Saturday	**-0.0246 (0.0036)**	-0.0002 (0.0001)	**0.0292 (0.0038)**	15,788	-0.002 (0.030)	0.0001 (0.0003)	-0.013 (0.017)	**-2.149 (0.640)**	0.398 *0.045*	4,372
Sunday	**-0.0216 (0.0034)**	**-0.0006 (0.0001)**	**0.0504 (0.0040)**	13,025	**0.049 (0.024)**	0.00005 (0.0004)	-0.032 (0.021)	**-3.102 (0.580)**	0.390 *0.040*	3,240
Sedan	Offer	ISF	HSF	N	Earn	ISF	HSF	IMR	R^2	N
Tuesday	**0.0216 (0.0028)**	0.0008 (0.0007)	**0.0766 (0.0221)**	21,283	**-0.040 (0.015)**	-0.002 (0.002)	**0.070 (0.035)**	**-0.940 (0.141)**	0.564 *0.097*	4,681
Wednesday	**0.0128 (0.0027)**	**0.0016 (0.0004)**	**0.0435 (0.0142)**	23,280	0.015 (0.012)	**-0.002 (0.001)**	**0.122 (0.023)**	**-0.657 (0.150)**	0.567 *0.114*	5,278
Thursday	**0.0115 (0.0026)**	**0.0010 (0.0003)**	**0.0351 (0.0095)**	19,982	-0.002 (0.011)	-0.00004 (0.0005)	**0.052 (0.016)**	-0.254 (0.164)	0.542 *0.100*	5,081
Friday	**0.0173 (0.0024)**	**0.0004 (0.0002)**	**0.0375 (0.0068)**	18,418	-0.009 (0.011)	-0.00002 (0.0004)	**0.026 (0.013)**	-0.321 (0.209)	0.533 *0.067*	4,666
Saturday	0.0035 (0.0049)	-0.0003 (0.0002)	**0.0502 (0.0062)**	15,762	-0.006 (0.028)	-0.0002 (0.0004)	**0.038 (0.014)**	-0.066 (0.311)	0.514 *0.067*	3,817
Sunday	-0.0081 (0.0046)	**-0.0007 (0.0002)**	**0.0626 (0.0063)**	12,602	**0.058 (0.022)**	**-0.001 (0.0004)**	**0.062 (0.015)**	-0.317 (0.342)	0.560 *0.101*	3,065

Note: Green: significantly positive, yellow: significantly negative, white: non-significant at p = 0.05.

Figure A.3 Effect sizes of changes in hourly financial offer/earnings, ISF, and HSF on drivers' day-level decisions

	% Change in P(Work)					Change in Minutes Worked				
SUV	Mean	+\$10 Offer	+\$10 ISF	+1h HSF	N	Mean	+\$10 Earn	+\$10 ISF	+1h HSF	N
Tuesday	0.409	1.53	0.24	**2.26**	28,883	8.696	-2.03	-0.45	1.65	9,482
Wednesday	0.418	1.39	**0.19**	**1.81**	21,965	8.964	-0.46	-0.18	1.70	10,120
Thursday	0.426	**3.36**	**0.20**	**1.38**	29,233	9.053	-3.71	-0.26	**2.54**	9,894
Friday	0.412	**2.70**	0.01	**1.96**	20,294	8.915	7.51	-0.27	**3.31**	9,283
Saturday	0.203	**-6.23**	-0.05	**0.83**	15,788	8.435	-1.42	0.03	-0.76	4,372
Sunday	0.162	**-5.04**	**-0.16**	**1.33**	13,025	7.927	**29.32**	0.03	-1.90	3,240
Sedan	Mean	+\$10 Offer	+\$10 ISF	+1h HSF	N	Mean	+\$10 Earn	+\$10 ISF	+1h HSF	N
Tuesday	0.169	**6.19**	0.21	**2.08**	21,283	7.687	**-23.80**	-0.92	**4.23**	4,681
Wednesday	0.182	**3.71**	**0.44**	**1.22**	23,280	7.680	8.81	**-1.38**	**7.31**	5,278
Thursday	0.179	**3.19**	**0.26**	**0.95**	19,982	7.724	-1.26	-0.02	**3.10**	5,081
Friday	0.171	**4.85**	**0.11**	**1.00**	18,418	7.568	-5.56	-0.01	**1.55**	4,666
Saturday	0.148	0.89	-0.08	**1.29**	15,762	8.022	-3.57	-0.12	**2.27**	3,817
Sunday	0.129	-1.98	**-0.18**	**1.63**	12,602	7.708	**34.52**	**-0.63**	**3.72**	3,065

Note: Green: significantly positive, yellow: significantly negative, white: non-significant at p = 0.05.

signifiant and F-statistics across all models except one are larger than the conventional threshold of 10. This IV also satisfies the *exclusion restriction:* Current incentives for co-skippers should not directly influence the focal driver's work decision because (i) they drive different vehicle types and (ii) the focal driver does not have access to co-skippers' incentives information.

The estimates from shift- and day-level analyses are consistent with our main results. Figure A.4 presents the signs and statistical significance (at p=0.05) of the estimates across shifts and days. However, these models are outperformed by our main models in both in-sample and out-of-sample prediction accuracy.

Figure A.4 Estimates across shifts and days using the co-skippers IV

	Choice (Work or not)				**Level** (How long)			
SUV	IV-F	Offer	ISF	HSF	IV-F	Earning	ISF	HSF
Midday	433.1	+	+	+	266.8	+	+	+
PM-Peak	289.5	-	-	+	58.7	+	-	+
PM-OPeak	260.2	+	-	+	45.1	+	-	+
Late Night	329.9	+	-	+	36.4	+	-	+
Sedan	IV-F	Offer	ISF	HSF	IV-F	Earning	ISF	HSF
Midday	229.2	+	+	+	104.4	+	+	+
PM-Peak	231.3	+	-	+	31.8	+	-	+
PM-OPeak	255.9	+	-	+	24.3	+	-	+
Late Night	270.0	+	-	+	30.9	+	-	+

	Choice (Work or not)				**Level** (How long)			
SUV	IV-F	Offer	ISF	HSF	IV-F	Earning	ISF	HSF
Tuesday	37.9	-	+	+	21.1	-	-	+
Wednesday	41.3	+	+	+	25.5	-	-	+
Thursday	67.9	+	-	+	43.9	-	+	+
Friday	67.5	+	+	+	41.9	-	-	+
Saturday	89.1	+	+	+	19.4	-	+	-
Sunday	82.1	-	-	+	16.0	+	+	-
Sedan	IV-F	Offer	ISF	HSF	IV-F	Earning	ISF	HSF
Tuesday	25.3	+	+	+	8.8	-	-	+
Wednesday	24.9	+	+	+	10.6	+	-	+
Thursday	43.9	+	+	+	15.8	+	-	+
Friday	39.9	+	+	+	13.6	-	+	-
Saturday	58.3	+	+	+	12.8	-	+	-
Sunday	48.7	-	-	+	12.5	+	-	+

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.

B.1.2. Hausman-type IV. Inspired by previous studies such as Sheldon (2016), we use the average hourly offer rate received by all other registered drivers during the same shift on the same day as an instrument for the offer rate. Similarly, we use the average hourly earnings rate earned by all other active drivers during the same shift on the same day as an instrument for the hourly earnings rate. These instruments can be thought of as a mutual offer or earning rate for eligible drivers in New York City at a particular time. In addition, the incentives offered to other drivers should not directly influence the focal driver's decision to work. Controlling for weather and market conditions using the TLC data, we rule out potential confounders that affect both the variation in incentives and the labor decisions. Unlike other ride-hailing platforms, drivers on our platform do not compete with other drivers for promotions as both the base and promotional rates are decided and announced ahead of time. Moreover, promotions are not offered as a way to relocate drivers to high-demand areas (see Section 3.3 for more details). Thus, it suggests that this IV satisfies the exclusion restriction. The results we obtained using this IV are qualitatively similar as illustrated in Figure A.5. While this type of IV appears to be valid for the choice equation, low F-statistics suggest that it is a relatively weaker IV relative to both the co-workers and co-skippers IVs.

Figure A.5 Estimates across shifts and days using Hausman-type IV

	Choice (Work or not)				Level (How long)			
SUV	IV-F	Offer	ISF	HSF	IV-F	Earning	ISF	HSF
Midday	365.4	+	+	+	183.7	+	+	+
PM-Peak	318.8	-	-	+	56.9	+	-	+
PM-OPeak	301.7	+	-	+	49.1	+	-	+
Late Night	371.7	+	-	+	38.1	+	-	+
Sedan	IV-F	Offer	ISF	HSF	IV-F	Earning	ISF	HSF
Midday	223.2	+	+	+	81.1	+	+	+
PM-Peak	255.6	-	-	+	32.9	+	-	+
PM-OPeak	274.9	+	-	+	27.3	+	-	+
Late Night	278.6	+	-	+	29.8	+	-	+

	Choice (Work or not)				Level (How long)			
SUV	IV-F	Offer	ISF	HSF	IV-F	Earning	ISF	HSF
Tuesday	41.4	+	+	+	18.1	-	-	+
Wednesday	52.2	+	+	+	24.5	+	-	+
Thursday	61.6	+	-	+	33.2	-	+	+
Friday	62.3	+	+	+	29.8	-	-	+
Saturday	83.1	+	+	+	15.2	-	+	-
Sunday	70.9	+	-	+	11.5	+	+	-
Sedan	IV-F	Offer	ISF	HSF	IV-F	Earning	ISF	HSF
Tuesday	30.3	+	+	+	8.0	-	-	+
Wednesday	35.3	+	+	+	10.3	+	-	+
Thursday	42.6	+	+	+	12.9	+	+	+
Friday	39.5	+	+	+	10.6	-	+	-
Saturday	44.4	+	-	+	9.8	-	+	-
Sunday	40.7	+	-	+	7.9	+	-	+

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.

B.1.3. Instrumenting for ISF and HSF. Lastly, we perform a robustness check for our assumption that ISF and HSF are not endogenous. Here, we allow them to be endogenous and use their lagged values from the previous hour or day as their instruments. We make one assumption that the drivers' fixed effects are not correlated with the error terms [5]. Figures A.6 and A.7 illustrate the signs and statistical significance (at $p = 0.05$) for the estimates across shifts and days, respectively. Again, insights remain qualitatively consistent. We note that the impact of cumulative income on the work duration becomes statistically insignificant even at a shift level. This suggests that, if cumulative earnings and work hours are endogenous, we still observe the consistently positive impact of cumulative work hours on both the decision to work and work duration.

Figure A.6 Estimates across shifts when also instrumenting for ISF and HSF

	Choice (Work or not)					Level (How long)				
SUV	Mean	Offer	ISF	HSF	N	Mean	Earn	ISF	HSF	N
Midday	0.343	+	-	+	124,769	4.987	-	-	+	45,329
PM-Peak	0.277	-	-	+	131,910	2.421	+	+	-	39,592
PM-OPeak	0.182	+	-	+	130,651	0.731	+	-	+	26,699
Late Night	0.117	+	-	+	125,382	1.996	+	-	-	17,137
Sedan	Mean	Offer	ISF	HSF	N	Mean	Earn	ISF	HSF	N
Midday	0.137	+	-	+	113,444	4.186	+	-	+	20,297
PM-Peak	0.123	-	-	+	117,152	2.327	+	-	+	19,613
PM-OPeak	0.099	+	-	+	124,611	0.803	+	-	-	17,025
Late Night	0.071	+	-	+	124,280	2.167	+	-	+	15,623

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.

[5] Future researchers may consider an alternative approach which is to use Mundlak- or Chamberlain-type correlated random effects regression models that include the driver-level mean of the time-varying covariate.

Figure A.7 **Estimates across days when also instrumenting for ISF and HSF**

	Choice (Work or not)					Level (How long)				
SUV	Mean	Offer	ISF	HSF	N	Mean	Earn	ISF	HSF	N
Tuesday	0.409	+	+	+	28,883	8.696	+	-	+	9,482
Wednesday	0.418	+	+	+	21,965	8.964	+	-	+	10,120
Thursday	0.426	+	+	+	29,233	9.053	-	+	-	9,894
Friday	0.412	+	+	+	20,294	8.915	+	-	+	9,283
Saturday	0.203	-	-	+	15,788	8.435	+	-	+	4,372
Sunday	0.162	-	-	+	13,025	7.927	+	+	+	3,240
Sedan	Mean	Offer	ISF	HSF	N	Mean	Earn	ISF	HSF	N
Tuesday	0.169	+	+	+	21,283	7.687	-	-	-	4,681
Wednesday	0.182	+	+	+	23,280	7.680	-	-	-	5,278
Thursday	0.179	+	+	+	19,982	7.724	+	-	+	5,081
Friday	0.171	+	+	+	18,418	7.568	+	-	+	4,666
Saturday	0.148	+	-	+	15,762	8.022	-	+	+	3,817
Sunday	0.129	-	-	+	12,602	7.708	+	-	+	3,065

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.

B.2. Sample Selection Bias Correction

B.2.1. Dahl's Correction. Following Dahl (2002) and Bray et al. (2019), we use the selection probability as a sufficient statistic for the selection bias. Since, in our context, the choice for each driver is only binary: to work or not, we do not suffer from a curse of dimensionality. Revisiting our level equation (Equation (7)),

$$f(Hour_{i,t}) = \beta_{0,i} + \beta_{\tilde{w}}\tilde{w}_{i,t} + \beta\mathbf{Z_{i,t}} + \theta\lambda_{i,t} + u_{i,t},$$

we can substitute IMR (λ) with all basis functions of a B-spline with using the quantiles of work probabilities for all drivers, $\mathbf{P_{work}} = [P(Drive_{i,t} = 1|\mathbf{X_{i,t}}), \forall i]$ as interior knots. Let $\mathfrak{B}(\mathbf{P_{work}}, j)$ be the j^{th} basis function of a degree n B-spline with the quantiles of $\mathbf{P_{work}}$ as m interior knots. Also, define $\eta_{i,t} = u_{i,t} - \sum_{j=0}^{m+n}\gamma_j\mathfrak{B}(\mathbf{P_{work}}, j)$ to maintain the orthogonality of the error term and the expected hours worked. Thus, our level equation under this approach becomes:

$$f(Hour_{i,t}) = \beta_{0,i} + \beta_{\tilde{w}}\tilde{w}_{i,t} + \beta\mathbf{Z_{i,t}} + \sum_{j=0}^{m+n}\gamma_j\mathfrak{B}(\mathbf{P_{work}}, j) + \eta_{i,t}. \tag{A.1}$$

In Figure A.8, we present the estimates for the level equation when choosing $m = n = 3$. Our results remain consistent under both approaches for sample selection correction. Note that, for all but sedan drivers' decisions on Friday and Saturday, the selection variables are significant at p = 0.05, confirming that there exists a selection bias in the decision to work.

B.3. Additional Model Specifications.

As a further robustness test, we consider several other model specifications: allowing for non-linear effects of the targeting behavior, allowing for an interaction between financial incentives and the targets, isolating income targeting and inertia effects, using an OLS approach in the choice equation rather than a probit, and including an indicator for vehicle type rather than separately estimating models by vehicle type. Our main insights remain robust. We present the results for the first specification below.

Figure A.8 Estimates for the level equation using Dahl's correction

	SUV Drivers			Sedan Drivers		
Shift	Earning	ISF	HSF	Earning	ISF	HSF
Midday	+	+	+	+	+	+
PM-Peak	+	-	+	+	-	+
PM-OPeak	+	-	+	+	-	+
Late Night	+	-	+	+	-	+
Day	Earning	ISF	HSF	Earning	ISF	HSF
Tuesday	-	-	+	-	-	+
Wednesday	+	-	+	+	-	+
Thursday	-	-	+	+	+	+
Friday	-	-	+	-	+	+
Saturday	-	+	-	-	+	-
Sunday	-	+	-	+	-	+

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.

Non-Linearity of targeting effects. Tables A.9 and A.10 display estimates when we include the quadratic terms, ISF^2 and HSF^2, in all our equations. This allows us to investigate whether there is potentially a non-linearity effect of each target on work decisions. We find that estimates for the linear terms remain consistent with our main insights. The estimates for the quadratic terms are generally either having the opposite sign from the linear counterpart or are statistically insignificant. We can interpret this result as follows: in general, both income targeting and inertia effects have a diminishing marginal impact. One interesting observation is that the level estimates for HSF and HSF^2 are both significantly positive in some shifts. This suggests that inertia becomes much more apparent as workers' work duration increases.

Figure A.9 Estimates across shifts when including quadratic terms

	Choice (Work or not)						**Level** (How long)					
SUV	IV-F	Offer	ISF	I^2	HSF	H^2	IV-F	Earning	ISF	I^2	HSF	H^2
Midday	379.1	+	+	-	+	-	188.6	+	-	+	-	+
PM-Peak	346.7	-	-	+	+	-	61.9	+	-	+	+	+
PM-OPeak	310.4	+	-	+	+	-	52.4	+	-	+	+	+
Late Night	368.9	+	-	+	+	-	40.7	+	-	+	+	+
Sedan	IV-F	Offer	ISF	I^2	HSF	H^2	IV-F	Earning	ISF	I^2	HSF	H^2
Midday	230.0	+	+	-	+	-	82.9	+	-	+	-	+
PM-Peak	259.6	+	-	+	+	-	33.3	+	-	-	+	+
PM-OPeak	282.5	+	-	+	+	-	29.1	+	-	+	+	+
Late Night	281.9	+	-	+	+	-	31.3	+	-	+	+	+

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.

Figure A.10 **Estimates across days when including quadratic terms**

	Choice (Work or not)						**Level** (How long)					
SUV	IV-F	Off	ISF	I^2	HSF	H^2	IV-F	Earn	ISF	I^2	HSF	H^2
Tues	43.5	+	+	-	+	-	21.9	-	-	+	+	-
Wed	55.0	+	+	-	+	-	29.8	+	-	+	+	-
Thu	67.9	+	+	-	+	-	40.6	-	-	+	-	+
Fri	67.1	+	+	-	+	-	38.1	-	-	+	+	+
Sat	90.4	+	-	+	+	-	18.1	-	-	+	+	-
Sun	75.7	+	-	+	+	-	14.9	+	-	+	-	-
Sedan	IV-F	Off	ISF	I^2	HSF	H^2	IV-F	Earn	ISF	I^2	HSF	H^2
Tues	31.1	+	+	-	+	-	9.0	-	-	+	+	-
Wed	36.2	-	+	-	+	-	11.8	+	-	+	+	+
Thu	45.2	+	+	-	+	-	14.6	+	-	+	-	+
Fri	41.9	+	+	-	+	-	13.2	-	-	+	-	+
Sat	46.5	+	+	-	+	-	11.7	-	+	+	-	+
Sun	43.3	+	-	+	+	-	11.7	+	+	-	-	+

Note: Green with "+": significantly positive, yellow with "-": significantly negative, white: non-significant at p = 0.05.