The Daily Labor Supply Response to Worker-Specific Earnings Shocks

Marc-Antoine Schmidt*

Draft version: November 2018

Click here for the most up-to-date version

Abstract

This paper presents empirical evidence indicating that the daily labor supply elasticity of workers is large and negative in response to idiosyncratic earnings shocks (e.g. a large tip), contrary to the prediction of the standard neoclassical model. I use microdata covering the universe of New York taxi trips to reconstruct drivers' daily work shifts in 2013. In the main specification, I identify variation in idiosyncratic earnings using large tips received by drivers and find that they respond to these shocks by *decreasing* their labor supply substantially; I obtain similar results when using trips from Manhattan to JFK Airport as idiosyncratic earnings shocks. I also find that these shocks do not affect future labor supply, indicating that standard neoclassical income effects cannot explain this result. In contrast, a positive earnings shock at the *market level* causes drivers' labor supply to increase, consistent with optimizing rational agency. The large and negative response to idiosyncratic earnings shocks indicates that such shocks can have significant effects on labor supply. My results suggest that transferring the share of income earned through tips or bonuses to a more predictable wage structure could be a cost-neutral way to increase labor supply.

JEL-Classification: D03, J22
Keywords: Labor Supply, Behavioral Model, Reference Dependence, Taxi Drivers

*I would like to thank Philip Oreopoulos, Kory Kroft, Gustavo Bobonis, and Robert McMillan for their constructive feedback. I also benefited from helpful discussions with Stephen Ayerst, Jean-William Laliberté, Mathieu Marcoux, Baxter Robinson, and participants at the CEA Meeting in Ottawa and the SCSE Conference in Québec City. Email: ma.schmidt@mail.utoronto.ca.

1 Introduction

A primary goal of labor economics is to understand how agents make labor supply decisions when faced with changes in income. The daily labor supply response to variation in hourly earnings has been the focus of most previous studies, whereas we know little about the daily response to small (relative to lifetime earnings) windfall gains. According to the neoclassical model, the labor supply elasticity with respect to such small windfall gains should be indistinguishable from zero, since the income effect is infinitesimal and there is no substitution effect.

In this paper, I present empirical evidence indicating that the daily labor supply elasticity of workers is large and negative in response to idiosyncratic earnings shocks. I focus on the taxi industry, where workers have the ability to flexibly choose their daily labor supply, and use large tips as the main proxy for idiosyncratic earnings shocks. I also measure the labor supply response to market-level earnings shocks and find, as predicted by the neoclassical model, a positive elasticity. How can we reconcile those results?

The most common theory that has been suggested to explain negative daily labor supply elasticities is reference dependence. Under this model, a worker's utility depends on how far he is from a self-determined income target. Below the target, the individual incurs disutility from not reaching the goal, which creates an additional incentive to work, which disappears when the target is reached. Models of reference dependence predict that in response to small windfall gains, a worker will end her shift earlier since this contributes to reaching the target sooner.

Past studies have tried to distinguish between the two theories by looking at the labor supply response to unexpected variation in hourly earnings. Similar to the response to windfall gains, the two models generate opposing predictions in response to unexpected variation in hourly earnings.[1] The empirical evidence from prior work is mixed: both positive and negative labor supply responses have been estimated using data from the same population.[2]

[1]The neoclassical model predicts that individuals should increase their labor supply if the wage rate increase because the opportunity cost of leisure increased. Standard model of reference dependence predicts that individuals have fixed reference points when they start their day. Increasing the wage rate will help them reach their target faster, and could therefore decrease their labor supply.

[2]Most notably, the New York City taxi industry has been extensively studied using trip sheets manually filled out by drivers. Camerer et al. (1997) and Crawford and Meng (2011) find a negative labor supply

The question arises whether windfall gains can help us better understand how these workers make their daily labor supply decisions.

The medallion taxi industry in New York City offers a great context to tests these predictions. Contrary to most other cities, New York medallion taxi drivers must find their customer only through hailing and cannot be booked in advanced. Therefore, idiosyncratic shocks can make up a large portion of a driver's daily earnings. The empirical analysis is carried out using a recently available dataset—the Taxicab Passenger Enhancement Project (TPEP) dataset—which contains the universe of medallion taxi trips taken in New York City. From the raw data covering an entire year, I construct the taxi drivers' daily work shifts. I aggregate the transaction-level data into approximately 7 million shifts from more than 40,000 drivers. This dataset contains, among other things, the credit card tip (gratuity) given to the driver, which I use as a proxy for idiosyncratic shock. Even though the taxi industry has been perhaps the most studied real-world environment to answer this type of question, to the best of my knowledge, only Thakral and Tô (2017) also use tips to study drivers' labor supply.[3]

In the main specification, I use very large tips (more than $30) to estimate the labor supply response to idiosyncratic shocks. I argue that features of the New York City taxi industry make it likely that very large tips occur quasi-exogenously 'within-driver.' The sign of the labor supply elasticity is estimated from the relationship between hours worked and the occurrence of large tips during a shift. I estimate the labor supply elasticity with respect to market-level shocks from the relationship between the shift duration and hourly earnings (the implicit wage).

I find a large and significantly negative labor supply elasticity in response to idiosyncratic shocks. Moreover, idiosyncratic shocks only affect the labor supply decision during the day they occur, indicating that even with extreme preference parameters, standard income effects cannot be the sole source of this negative elasticity. This result is more easily explained

response to hourly earnings shocks while Farber (2005, 2008) argues that drivers do not end their shift earlier when accumulated earnings are higher.

[3]In their analysis, tips are treated as a measure of effort, whereas I use large tips as an idiosyncratic shock. Even if large tips are likely mostly explained by the customer's characteristics, effort could still be a component, which would bias the estimates. I show that large tips do not predict a higher average tip for the other trips in the same shift, indicating that effort throughout the shift is at most a negligible part of large tips.

by reference dependence. In contrast, I find that the labor supply elasticity is significantly positive in response to market-level shocks and quantitatively similar to the estimates of Farber (2015). The positive elasticity in response to market-level variation indicates that the neoclassical model remains a reasonable approximation when earnings are not subject to idiosyncratic shocks.

I analyze several other specifications to ensure that the results are not driven by an omitted variable correlated with large tips.[4] First, when replacing the 'large tips' variable by the number of trips from Manhattan to JFK Airport, the results are qualitatively similar. The fare for those trips is fixed at $52 and represents about 20 percent of the median shift income, making this variable a suitable candidate to replace large tips. Second, adding neighborhood fixed effects does not change the results significantly. This suggests that the negative relationship between shift duration and large tips cannot be explained by drivers working in locations with a higher likelihood of large tips when they plan to quit earlier. Third, breaks could serve as a margin of labor supply adjustment. I find that adjusting the measure of labor supply to remove breaks increases the magnitude of the response, suggesting that the reduction in labor supply operates both through an earlier stopping time and longer breaks.

Fourth, to address the concern expressed in Farber (2005) that taxi drivers do not use the implicit wage parametrically, I estimate a discrete-choice stopping model to look at how cumulative earnings affect the probability of stopping. Under the neoclassical model, cumulative earnings should not affect the stopping probability.[5] However, under reference dependence, cumulative earnings also measures how near someone is from the target, and should thus positively influence the stopping probability. I decompose cumulative earnings of the drivers into a market-level component and an idiosyncratic component, finding that drivers are more likely to end their shift as their cumulative idiosyncratic earnings get larger. I find the reverse for the marginal effect of cumulative market-level earnings. These results are consistent with the findings of the main empirical framework: positive idiosyncratic shocks

[4]For instance, these robustness checks provide evidence that the estimated negative elasticity with respect to idiosyncratic shocks is not entirely generated by a different level of effort.

[5]Thakral and Tô (2017) argues that under the hypothesis of positive autocorrelation of hourly earnings, this coefficient could even be positive because higher cumulative earnings could indicate a higher continuation value to work.

seem to reduce daily labor supply but drivers respond optimally to variation in market-level shocks.

The remainder of the paper is organized as follows. Section 2 provides a basic overview of recent findings on daily labor supply. In Section 3, I present important characteristics of the dataset as well as the methodology to infer the drivers' working shifts. Section 4 describes the empirical strategy used to test the neoclassical model. The results are presented in Section 5, and multiple robustness checks are performed in Section 6. I discuss different potential behavioral models that could explain the results in Section 7, and Section 8 concludes.

2 Daily labor supply elasticities in the literature

The literature estimating daily labor supply elasticities revolves around testing for the presence of reference-dependent preferences as opposed to a purely neoclassical labor supply model. In this section, I discuss the recent developments in this debate to contextualize the contributions of my paper.

In their seminal contribution, Camerer et al. (1997) presented empirical evidence indicating that New York City taxi drivers had a negative labor supply elasticity. Their basic approach was to regress shift duration on hourly earnings. Their instrumental variables estimate of the elasticity, with the individual hourly earnings instrumented for by moments of the hourly earnings distribution of other drivers on that day, was negative and large. This finding stood in stark contrast to the prediction of the neoclassical model. Their results were replicated by a similar study of taxicabs in Singapore (Chou, 2002). Although it was not a direct test of reference-dependent labor supply, Camerer et al. (1997) discussed how this behavior could be caused by reference-dependent preferences through income targeting.

Many studies then sought to refine and expand our knowledge of daily labor supply decisions. However, there still is no consensus, as some studies find evidence in support of reference dependence while others do not. For instance, Crawford and Meng (2011) find that taxi drivers seem to have reference-dependent preferences along two dimensions: hours and income. Their results support a model of labor supply with expectation-based reference dependence as modeled by Kőszegi and Rabin (2006). Other researchers find evidence of reference dependence among Indian boat owners (Giné et al., 2017), pear packers (Chang and Gross, 2014), bicycle messengers (Fehr and Goette, 2007), and taxi drivers in both developed (Doran, 2014; Agarwal et al., 2015; Leah-Martin, 2017) and developing countries (Dupas et al., 2017). Yet, many studies have cast doubt over these results. Most notably, Farber (2005) finds no evidence of reference dependence for taxi drivers, arguing that the findings of Camerer et al. (1997) suffered from methodological issues.[6] Similarly, reference dependence does not seem to be important for stadium vendors (Oettinger, 1999), Florida fishermen (Stafford, 2015), day laborer in Malawi (Goldberg, 2016), and Uber drivers (Sheldon, 2016).

In a recent paper, Farber (2015) has made important contributions to this literature. Along with Haggag and Paci (2014) and Morgul and Ozbay (2015), his is one of the first studies to use the TPEP dataset and the first to use the universe of taxi trips made during an extended period of time (2009-2013). Farber shows that the original results of Camerer et al. (1997) do not hold in this dataset. More specifically, he finds a positive and significant daily labor supply elasticity. He also looks at the heterogeneity across drivers and finds that the most experienced drivers exhibit more neoclassical behaviors, indicating that reference dependence might be a temporary outcome of inexperience. Conversely, the drivers with the most negative labor supply elasticity tend to exit the market more often. The study concludes that reference dependence is not of first-order importance when considering the taxi industry. While I replicate Farber's main result in this paper, I find a positive labor supply elasticity in response to market-level shocks only.

Two other studies are noteworthy in the context of this paper. First, Andersen et al. (2017) study the daily labor supply decision of Indian vendors in a randomized field experiment setting. They provide some vendors with an expected increase in hourly earnings and others with an unexpected earnings shock, the unexpected earnings shock being very similar to the idiosyncratic earnings shock I use in the remainder of this paper. Selected vendors receive a substantial overpayment for a transaction at the beginning of their day. The experiment design makes this unexpected earnings shock very likely to be idiosyncratic in the view of the vendor. The authors show that, in response to this idiosyncratic shock, vendors immediately consume

[6] While the result of Farber (2008) does not fully support the presence of reference dependence, he finds evidence of a large discontinuity of the stopping probability at the reference point.

more leisure by taking a midday break. Their results seem to confirm that idiosyncratic earnings shocks influence labor supply negatively in the very short-run.

Second, Thakral and Tô (2017), using the TPEP data, study other types of behaviors from taxi drivers to provide evidence of psychological factors in the labor supply decision. Their structural approach complements the empirical findings of this paper. For instance, they find that the timing of the earnings influences the labor supply decision differently, suggesting that earnings are non-fungible. Similarly, my results show non-fungibility in the source of the income (idiosyncratic vs market-level). To show that their results are robust to the possibility that earnings are correlated with effort, they use tips as an instrument for total earnings. Instead of using all tips, I restrict my analysis to large tips, showing that large tips do not seem correlated with the average tips received in the rest of the shifts. This indicates that effort is not a confounding factor in my findings.

3 Data

3.1 Taxicab Passenger Enhancement Project dataset

Up until Morgul and Ozbay (2015) and Farber (2015), previous studies of taxicab labor supply decisions used the storage technology of the time: physical taxi trip-sheets. Data entry was done manually and a dataset of hundreds of shifts was considered large. Recent technological innovations have made it easier for taxi agencies to store this information digitally. More specifically, since 2009, the New York Taxi and Limousine Commission standardized the storage of those data for the medallion taxi. The Taxicab Passenger Enhancements Project allowed the installation of computerized meters able to store a multitude of data: date and time of pickup and drop-off (to the second), distance traveled, number of passengers, fare, method of payment, tip, etc. Although on a much smaller sample size and with uncertain measurement error, this information was already available to the previous literature. In this paper, I use a subsample of the data of one year (2013) for which I have access to unique identifiers for the driver and the medallion, necessary to track individual workers' labor supply.

This dataset is set apart by two new characteristics. First, it contains precise GPS data on the pickup and drop-off locations of each trip. Second, the TPEP dataset contains the universe of trips made with New York City medallion taxis.[7] To put this into perspective, the largest dataset previously used by Farber (2008) and Crawford and Meng (2011) contained less than 600 shifts made by 21 drivers. The TPEP dataset used in this paper covers all of 2013 and contains information on 180 million trips made by more than 40,000 drivers over 6 to 7 million shifts.

There was no fare modification during 2013. The last fare modification, as of writing this paper, was on September 4, 2012. Thus, the fare structure remained constant throughout the studied period. All taxicabs in New York City face the same fare structure: a time-varying starting fare of $2.50 to $3.50 and an increasing function of either distance or time in traffic. A few specific destinations have a different fare structure. For instance, trips between Manhattan and JFK Airport are subject to a flat fare of 52$ and trips to Newark Airport are charged a surcharge of $17.50 on top of the regular metered fare. Because of the fixed nature of the price structure, the taxi industry is an interesting case study since equilibrium effects are constrained. Supply and demand affect only a driver's earnings by changing the driver's probability of finding its next customer.

Medallion taxis have a monopoly on the Manhattan region (except for the northern part), and face competition in the other boroughs by non-medallion taxis. They are the only type of taxis that can pick up a hailing customer. Other types of taxis that operate outside of Manhattan are usually called by a dispatch center. Figure A6 show the geographic location of pickups from a random sample of 3000 trips from the TPEP dataset. It is clear that the main geographical market for medallion taxis is the core of Manhattan, although there seem to be bunching around the LaGuardia and JFK airports.

Although the TPEP dataset is a large improvement over previous datasets, it has some limitations that could lead to estimation biases. The first limitation is directly related to the main estimation strategy of this paper. Data on the tip is truncated. When a trip is paid by cash, the tip is never recorded. It is only when a customer pays with a credit or debit card that the amount of the tip is recorded. Furthermore, a customer paying with a debit or

[7] Medallion taxis can be recognized by their distinctive yellow color.

credit card might give a tip in cash and will not be recorded. Every previous study of taxi drivers' labor supply decided to simply ignore the tip. It was not a part of a driver's earnings. This should not cause any issues unless the average rate of tipping is correlated with earnings excluding tips.

Payment behaviors evolved greatly over time. In 2009, only 20% to 25% of the customers paid by credit or debit card Haggag and Paci (2014). In 2013, according to the data, this proportion went up to around 53%. This generalization in the usage of credit and debit card is strengthening the position that drivers are not selecting the types of customer. Furthermore, refusing to serve a customer because he is paying with a credit or debit card is not allowed by the overseeing agency. An overview of tipping behaviors using the same data source but for the year 2009 can be found in Haggag and Paci (2014).

Finally, even though the data collection process is computerized, many errors still remain. The most common error involves the geospatial data. The latitude and longitude are sometimes erroneous. However, those observations are easy to identify as they are either coded as zero or a value outside the possible range of GPS coordinates (longitude: +/- 180; latitude: +/- 90). I have also identified coding errors in the fares and timestamps variables. To make sure these errors do not drive the results, all the regressions were run both with them and without them (where possible). The results never change by a large amount in response to the inclusion of outliers.

3.2 Aggregation and shift construction

The analysis that will be conducted in Section 5 requires the aggregation of trips into shift units. The smaller datasets used in previous studies had one advantage over the TPEP dataset. Indeed, hand-written trip sheets are a collection of trips made in a single shift. Shifts were defined in a straightforward and objective manner. Because the TPEP dataset does not contain this information, the definition of a shift will require a subjective rule.

To define a shift, I will use a rule similar to Farber (2015). One natural way of defining a shift is to group consecutive trips without a large break between them. Figure A1 shows the distribution of the driver's wait time between trips. It is clear that the smoothness of the distribution makes it hard to implement a perfect threshold. I use the subjective rule that waiting time of more than 6 hours between trips represents a shift delimiter.[8] Figure A2 presents the resulting distribution of shift durations. The median shift duration of 9 hours confirms that the shift delimiter did not introduce an implausible distribution of values.

New variables are created during and after the aggregation process. For instance, total earnings is the sum of fares and surcharges received during a shift. The shift length is defined as the elapsed time between the start of the shift's first trip and the end of the shift's last trip. The instrument used for a driver's hourly earnings will be the average hourly earnings of all drivers during the shift. The methodology to create the instrument for the hourly earnings and the cumulative earnings at the market wage is described in the appendix. Figure A4 shows the distribution of hourly average wage in 2013. We observe a lot of heterogeneity, the 5th percentile being about half that of the 95th percentile of the wage distribution.

Trip-level observations will be used as the basis of a robustness check. For every trip t, this analysis requires the computation of four cumulative variables: cumulative hours (H^c_{st}), cumulative earnings (E^c_{st}), cumulative earnings at the market wage ($\tilde{E}^c_{st}$), and cumulative idiosyncratic earnings (U^c_{st}). Cumulative hours and cumulative earnings are simply the total hours worked and total earnings up to a given trip during a shift. Cumulative earnings at the market wage is the cumulative earnings a driver would have made if he constantly earned the hourly average market wage. The cumulative idiosyncratic earnings is the difference between the cumulative earnings and the cumulative earnings at the market wage.

4 Empirical strategy

The primary empirical strategy I employ uses large tips received by taxi driver to proxy for idiosyncratic shocks. The set of controls used and the driver fixed effects reduce and potentially eliminate the possibility of other channels affecting both the labor supply and the probability of receiving a large tip. The sign of the labor supply elasticities are obtained from a regression of the log shifts' duration on both types of earnings variation. It should be noted that large tips are one of the components of total idiosyncratic income during a shifts and

[8] Farber (2015) used the same threshold of 6 hours. The main regression specifications were also estimated with a threshold of 4 hours (not reported) and the results were almost identical.

other proxies can be used to further solidify the results. For instance, in Section 6, I use trips to JFK Airport as a proxy for idiosyncratic income.

Previous tests of non-standard labor supply decisions implicitly made the assumption that income is fungible relative to its source. In other words, taxi drivers do not distinguish between an unexpected \$30 tip and a \$30 increase in fares due to an unexpected and temporary demand surge. Consequently, if expectations about future earnings stay constant (i.e. no substitution effect), the labor supply elasticity should be the same for both sources of income. This is the case in either standard neoclassical models or expectation-based reference-dependence models. My empirical strategy allows me to relax this assumption with regards to market-level and idiosyncratic income shocks.

It is important to distinguish between expected and unexpected earnings shocks. Theoretically, the labor supply elasticity with respect to expected wage variations is similar in both the neoclassical and expectation-based reference-dependence models. The estimating equation will ideally control for expected variation in wage. This can be done, albeit imperfectly, by adding an array of time and date fixed effects. The fixed effects will capture any recurring difference in average earnings. For example, it can be seen in the data that a weekend night shift will have a different expected earnings than weekday night shift.

I look at two types of unexpected variation: market-level shocks and idiosyncratic shocks. Market-level shocks, while unexpected, affect more than one driver by the same amount. A good example of such shock would be a short subway closure in which a whole area gets a surge in demand. Idiosyncratic shocks, on the other hand, affect directly a single driver and can be seen as a lucky draw. For instance, receiving a very large tip.

The same type of instrument used in Camerer et al. (1997) and Farber (2015) can be used to identify market-level shocks. Instead of directly using hourly earnings in a standard regression setting, the average hourly earnings of all drivers during the shift will be used as an instrument. This instrument removes the mechanical division bias that arises when estimating a labor supply elasticity with imprecise measure of hourly earnings (see Borjas, 1980). In the context of separating idiosyncratic shocks from market-level shocks, this instrumental variable strategy also has the benefit of purging the hourly earnings from idiosyncratic variations unrelated to the average market hourly earnings.

To formalize, let the length of shift s, for a driver i, in hours, be H_{is}, and the total earnings from fares during a shift be E_{is}. The logarithm of the hourly earnings (w_{is}) is defined as $\log(E_{is}/H_{is})$. Similar to Camerer et al. (1997) and Farber (2015), the benchmark estimating equations is:

$$\log(H_{is}) = \delta w_{is} + \mathbf{X}_{is}\beta + \mu_i + \nu_{is} \tag{1}$$

The hourly earnings variable (w_{is}) is instrumented by the average hourly earnings in the market during shift s. Shift specific controls such as time and date fixed effects, precipitation, temperature[9], and major holidays are included in $\mathbf{X}_{is}$. μ_i are driver fixed effects. Thus, the identifying variation comes from within driver. If we assume that a driver's labor supply response to idiosyncratic and market-level earnings shocks are the same, then the coefficient δ represents the labor supply elasticity with respect to any type of unexpected wage variation.

However, if we relax the assumption of similar elasticities with respect the two types of unexpected shocks, the instrumental variable estimate of the coefficient now identifies the labor supply elasticity with respect to market-level shocks only (δ^M). Although it would be almost impossible to correctly capture every idiosyncratic earnings variation since it would require a knowledge of the agent's expectation, we can approximate them. The main strategy will be to use very large tips received by a driver. Because some drivers receive on average higher tips, it is important to focus on within-driver variations by including driver fixed effects. Other similar strategy can also be tested. For instance, taxi drivers usually find trips to airports more profitable than regular rides. Therefore, one can look at the effect of a ride from Manhattan to the airport on the shift's duration. Receiving a large tip and having a ride from Manhattan to the airport will be the proxies for idiosyncratic shock.

To capture the labor supply response to variation in idiosyncratic earnings, I add the

[9] The weather data come from the National Centers for Environmental Information of the National Oceanic and Atmospheric Administration (NOAA) in the United States. The information comes from a weather station in Central Park.

explanatory variable T_{is} to the regression equation.

$$\log(H_{is}) = \delta^M w_{is} + \delta^I T_{is} + \mathbf{X}_{is}\beta + \mu_i + \nu_{is} \tag{2}$$

T_{is} represents the number of idiosyncratic shocks a driver received during a fixed amount of time in his shift. When using large tips as the proxy, T_{is} counts the number of tips larger than an arbitrary threshold the driver received during a fixed portion shift s.[10] The fixed period length is essential to eliminate a clear problem of reverse causality: longer shifts have mechanically a higher probability of receiving idiosyncratic shocks. Fixing the period for which we look at the number of large tips received, every shift has an equal probability of receiving large tips. In Section 6, a sensitivity test shows the main coefficient's response to variation in the fixed period definition and the large tip threshold.

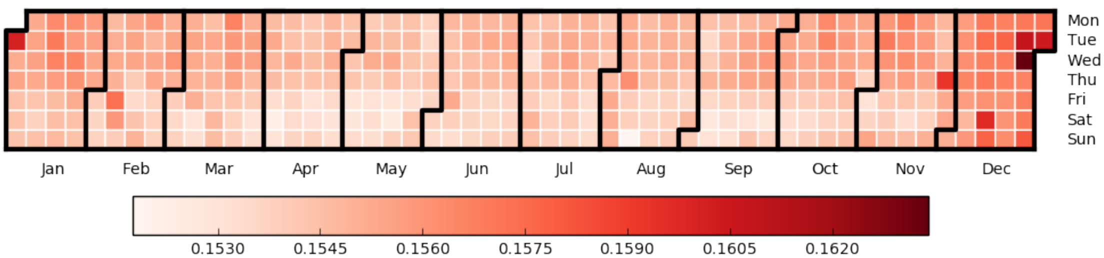
The type of taxi trips and the tipping behaviors vary a lot throughout the year. To understand some of the patterns at play, Fig. 1a shows the daily average tip as a percentage of total fare and Fig. 1b shows the daily average tip amount. Each column represent a week, starting from January 1st on the top left. Many interesting patterns emerge. First, not surprisingly, we can infer that taxi trips made on a Saturday are shorter than the ones made on other days. The low average tip amount is mainly due to lower average fares on that day. We can also see different tipping patterns on major holidays: on Christmas Day or Thanksgiving Day, taxicab riders seem more generous. In general, winter months seems to generate higher tip percentages than summer months. This figure makes clear the need to control for the day of the week and the month of the year in order to eliminate expected tipping patterns that driver could respond to.

It seem likely that large tips are exogenous after controlling for date and time effect as well as the individual drivers' unobservable propensity to receive large tips. Two characteristics of the New York City taxi industry make it extremely hard for drivers to develop a strategy using large tips. First, it was not possible in 2013 to pre-arrange a pickup with a medallion taxi. Medallion taxis can only find customers through street hailing. Therefore, it is impossible for a driver to know the destination before meeting the customer. Second, large tips are very

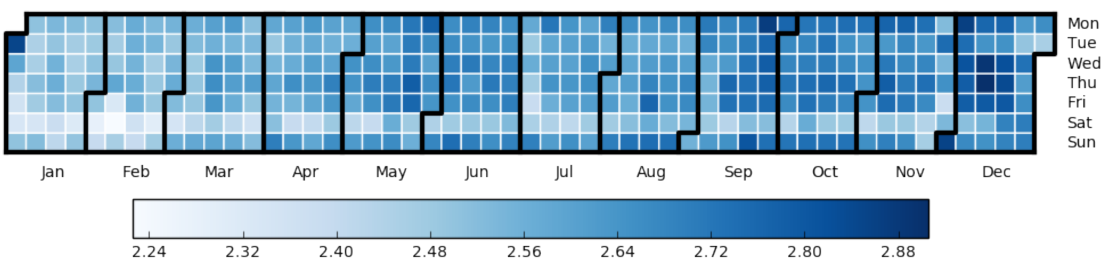
[10]The case where the independent variable is the number of trips to JFK Airport can be defined similarly with a fixed period.

Figure 1: Yearly tipping patterns

(a) Daily average tip as a percentage of total fare



(b) Daily average tip amount



Notes: Each square represents a day. Starting from the top left (January 1st), each column represents a week from Monday to Sunday. Due to data restrictions, the computation are done excluding observations with $0 tip.

rare. Taxi drivers have received a large tip in only 0.35% of shifts. Because of how unlikely very large tips are, it seems improbable that taxicab drivers in New York City search more actively for them in days they expect to work less. Thus, the only plausible channel by which large tips and labor supply are related seems to be the driver's reaction to the income shock.

While the idea of using large tips seems very simple, three potential issues require our attention. The first one, already discussed in Section 3, is that only tips paid with a credit or debit cards are recorded. Second, the rule used to determine whether a tip is large or not is subjective. Those two problems will introduce measurement errors. Third, if we do not standardize the period in which the idiosyncratic shock is received, reverse causality will bias our estimates.

The measurement error introduced by the first two problems is most likely random and brings a downward bias to the estimate. However, since the goal is not to estimate a precise coefficient but rather to uncover the direction of the response, this simply makes the case for non-standard behavioral theories harder. Further measurement error is introduced to solve the reverse causality issue. The rule that the tip needs to be received in the first $\tilde{h}$ hours of the shift equalizes the probability of receiving a large tip across observations[11] but miscategorizes observations which received the idiosyncratic shock at the end of the shift.

Because I am looking at idiosyncratic shocks, and in particular tips, one more test is required to distinguish between the neoclassical and behavioral models. Even if the estimated labor supply elasticity with respect to idiosyncratic shocks is negative, the standard neoclassical model could explain this with a large income effect. By definition, this idiosyncratic shock does not increase a driver's expectation of future wage. Therefore, the substitution effect creating the positive labor supply elasticity is completely shut down. What remains is the income effect which should be negative if we believe leisure to be a normal good. Since these are small and temporary shocks, the income effect should be negligible. Nevertheless, the estimation framework allows us to test for such income effect.

The main idea behind this test is that income effect should be long lasting. The neoclassical model tells us that when a driver earns more money while keeping the expected return of future days constant, he will increase his consumption of a large variety of goods. For leisure,

[11] Shifts with a duration of fewer than $\tilde{h}$ hours are also dropped.

this has the implication that the driver should increase his consumption of contemporaneous leisure as well as future leisure. Taking Eq. (2), we can simply look at the effect an idiosyncratic shocks has on the future labor supply decision. $T_{i,s-1}$ can be used instead of $T_{i,s}$, capturing the labor supply response to idiosyncratic shocks received in the previous shift. If no effect can be detected while the contemporaneous effect is large, this will indicate that some psychological behaviors are needed to explain this and the standard neoclassical model of labor supply can be rejected.

5 Results

Two arbitrary parameters need to be chosen to construct the explanatory variable of interest, T. As explained in section 4, in order to avoid the issue of reverse causality where a longer shift leads to a higher probability of receiving a large tip, I constrain the large tip to occur within the first $\tilde{h}$ hours of the shift. The regressions in this section are computed with $\tilde{h} = 4$. Second, large tips must be larger than an arbitrary amount x. The following regressions will use \$30 as the threshold of large tip. This represents around 10 to 15 percent of the average earnings during a shift (see figure A3 for a distribution of shift income). A sensitivity analysis was made to ensure that the result was not being driven by these thresholds (see section 6).

The first column of Table 1 shows the coefficient of a regression that replicates the specification of Camerer et al. (1997) and Farber (2015) (see Eq. (1)). Compared with the estimates of Farber (2015), we observe a qualitatively similar labor supply elasticity (coefficient on log hourly earnings). This positive estimate supports the presence of a neoclassical response to a variation in hourly earnings. As a driver's hourly earnings increases, his labor supply increases. This elasticity implies that for every 10 percent increase in hourly earnings, a driver increases the duration of his shift by 6.5 percent. As for the strength of the instrument, the massive number of observations produces a very large F statistics (upward of 1000), similar to that of Farber (2015).

Columns 2 to 4 add the "large tips" variable. This explanatory variable is the main addition to the original framework. We are interested in the sign of the coefficient and its statistical significance. A coefficient close to zero would support the neoclassical framework.

Table 1: IV estimates of the wage elasticity and effect of idiosyncratic earnings on shift duration

	(1)	(2)	(3)	(4)	(5)
Large tips		-0.0573***	-0.0671***	-0.0802***	
		(0.00378)	(0.00319)	(0.00327)	
Large tips (t-1)					-0.000466
					(0.00270)
Log hourly earnings	0.653***	0.371***	0.515***	0.665***	0.665***
	(0.00975)	(0.00674)	(0.00409)	(0.00618)	(0.00619)
driver FE	yes	no	yes	yes	yes
date/time FE[a]	m/d/h	no	no	m/d/h	m/d/h
holidays	yes	no	no	yes	yes
weather	yes	no	no	yes	yes
obs	7084914	6664258	6663665	6663665	6626414

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors at the driver-level in parentheses.

The dependent variable is the log of the duration of a shift in hours. The "Log hourly earnings" regressor is computed as total fares divided by hours worked. It is instrumented with the average hourly wage of all drivers during the shift (see the appendix for the methodology). "Large tip" is a dummy variable taking the value of 1 if the driver received at least one tip of over $30 during the first 4 hours of its shift. Observations with a shift duration of more than 20 hours or less than 4 hours were dropped.

[a] m: month of the year; d: day of the week; h: hour of the day.

Indeed, because the idiosyncratic shock should not modify the expected future wage, only the income effect should impact the labor supply decision. Under the standard neoclassical model, the coefficient should be indistinguishable from zero because the income shock is extremely small relative to life-cycle earnings. A positive coefficient could indicate that the drivers expect the probability of receiving another positive income shock to increase after receiving a first positive income shock[12]. Finally, the coefficient could be negative. This would be an evidence for a behavioral effect in the labor supply decision of taxi drivers. Multiple psychological explanations, including the presence of reference-dependent preferences, will be explored in Section 7.

Column 2 report the result of a naive regression without driver fixed effects or other controls. The coefficient of -0.058 is significantly different from zero. It is also economically

[12]Other explanations can explain a positive coefficient, e.g. the model's misspecification.

significant, indicating a 5.8 percent average drop in shift duration in the case of a large tip during the first 4 hours of a shift. Obviously, the coefficient of such regression could be biased for many reasons. A priori, the direction of the bias is not clear. The first tool we can use to eliminate part of the bias is to add driver fixed effects (column 3). This will force the variation to originate within driver. This is especially important if we think that some drivers have a higher probability to receive a large tip than others (e.g. by having better social skills). Adding driver fixed effects slightly increases the magnitude of the coefficient to -0.068.

Simply limiting the variation to originate within driver is not enough to eliminate potential biases. The wage, probability of high tips, and shift duration might all be correlated to certain time of the day or period of the year. Furthermore, the weather could also affect those variables[13]. Column 4 incorporates controls for all these potential sources of bias. The date and time fixed effects include controls for hour of the day, day of the week, and month of the year. Using patterns identified in Fig. 1, I add Christmas Eve and New Years Eve to set of major holidays used in Farber (2015)[14]. When controlling for those potential sources of bias, we see that the coefficient's magnitude increases, giving a stronger support to a negative elasticity with respect to idiosyncratic income shocks. Receiving a large tip in the first 4 hours of a shift seems to decrease on average the shift duration by 8 percent. When looking at the wage, representing the labor supply elasticity with respect to market-level income shocks, the coefficient is almost identical to that of column 1. The main takeaway from these results is clear: the drivers respond positively to a market-level income shock but negatively when the income shock is idiosyncratic.

Because of the censored nature of the tip variable and the resulting bias towards zero, the estimated coefficients of Table 1 should be interpreted as upper bounds (lower bound of the coefficient's absolute value). To roughly translate the coefficients (column 4) into a comparable elasticity, note that the average large tip is \$45.74, and the average income during a shift is \$264.13. This represents a 17% income shock[15]. Thus, a simple approximation suggests a

[13] Farber (2015) finds that rainfall do not impact average earnings. The reason is that the higher demand during rainy weather is completely offset by worst driving conditions that decrease the earnings per minute of drivers.

[14] Major holidays are: New Years Day, Easter Sunday, Memorial Day, Fourth of July, Labor Day, Thanksgiving Day, Christmas Eve, Christmas Day, and New Years Eve.

[15] I am being cautious by assuming that the full amount of the large tip is considered an idiosyncratic shock. The driver probably expected a portion of the large tip. This suggests that the approximated elasticity could

labor supply elasticity of -0.33, -0.39, and -0.47 for columns 2, 3, and 4, respectively. Being an upper bound, a negative elasticity of this magnitude is certainly of economic importance.

To further support the claim that the negative elasticity is not due to a large income effect, I test whether the effect of the large tip remains present during the next shift. If the income effect is indeed the cause of the negative elasticity, we should also observe an effect in the following shifts. The specification use for column 5 contains the same controls than column 4, but the tip indicator is lagged by one shift. The coefficient is not statistically significant from zero and the standard error is small. Using the same strategy as above, we can statistically reject the presence of an elasticity lower than -0.03. This seems to confirm that a psychological factor must be present to explain the large negative elasticity.

6 Robustness Checks

The negative labor supply elasticity with respect to large tips found in Table 1 implies behaviors that are inconsistent with the neoclassical life-cycle model of labor supply. In this section, I provide multiple tests and further evidence to support this claim. I add more controls and I test another source of idiosyncratic shock to show that it is not only an unobservable characteristic of tips that create this relationship. I also estimate a discrete-choice stopping model similar to Farber (2005) which does not depend on tips. This methodology uses the cumulative deviations from the market's average hourly earnings as an explanatory variable of the decision to stop working or not.

In the previous section, I used the same specification for the date and time fixed effects as in Farber (2015). Instead using the day of the week and hour of the day separately, we can use each hour of the week as a distinct fixed effect. This increases the number of date and time fixed effects from 40 to 178[16]. The estimated coefficients from this new specification are found in column 1 of Table 2. While the magnitude of the coefficient of interest slightly decreases, it remains well in the range of a statistically and economically significant effect.

It is also possible that the negative relationship between large tips and hours worked has

be even lower.

[16]Main specification: month of the year (11), day of the week (6), hour of the day (23). Robustness check: month of the year (11), hour of the week (167)

Table 2: IV estimates of wage elasticity and effect of idiosyncratic earnings on shift duration

	(1)	(2)	(3)
Large tips	-0.0558***		-0.0782***
	(0.00282)		(0.00322)
JFK fares		-0.0340***	
		(0.000536)	
Log hourly earnings	0.448***	0.665***	0.656***
	(0.00677)	(0.00618)	(0.00615)
driver FE	yes	yes	yes
date/time FE[a]	m/h*d	m/d/h	m/d/h
holidays	yes	yes	yes
weather	yes	yes	yes
neighborhood	no	no	yes
obs	6663665	6663665	6663264

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors at the driver-level in parentheses.
The dependent variable is the log of the duration, in hours, of a shift. “JFK fares” is defined as the number of trips from Manhattan to the airport during the shift. See the notes of Table 1 for a definition of the other regressors and controls.
[a] m: month of the year; d: day of the week; h: hour of the day.
[b] See Fig. A5 for a map of the geographical division used for the fixed effects.

another causal channel, unrelated to the income earned from the large tip. For instance, it could be that a large tips is sometimes given with the expectation that the driver will spend more leisure time with his family. Even though this explanation would be interesting by itself, it would not easily generalize to any other idiosyncratic shock. To convince ourselves that it is not the case, we can show that the relationship holds for other plausible idiosyncratic shocks. Another potential candidate to proxy for idiosyncratic shocks is the set of trips from Manhattan to JFK Airport.[17] A trip between Manhattan and JFK Airport costs 5 times more in fare ($52) than the average trip ($11). It equals almost one-fifth of the median shift income. From casual discussions with taxi drivers, this type of trip seems considered a positive shock.

Column 2 of Table 2 presents the result of the modified regression model substituting large tips for trips to JFK Airport. The variable is constructed in the same way the large tip

[17]Reverse trips from the airport to Manhattan are not included because the decision to pick up customers at the airport is part of a driver's strategy.

binary variable was: 1 if the driver had a trip from Manhattan to JFK Airport in the first 4 hours of its shift and 0 if he did not. We can see that the coefficient is still negative and significant. While it is difficult to interpret the difference in magnitude of the coefficient, the qualitative results stay the same: a driver reduces his hours worked in response to a positive idiosyncratic shock to income.

Another competing story as to why we would observe a relationship between the length of a shift and idiosyncratic earnings is if we did not control for neighborhood fixed effects. This can be a problem if: (a) drivers can somehow target neighborhoods to receive larger tips (or other idiosyncratic shocks) and (b) the neighborhood strategy was related to the labor supply decision. Both conditions (a) and (b) would need to be satisfied in order for the neighborhood effect to bias the coefficient. Furthermore, with driver fixed effects, the variation would need to be within drivers. It is still possible that a single driver will drive around a different neighborhood on different shift depending on whether he is planning on working a longer or shorter shift. I investigate whether neighborhood targeting is a possibility by adding neighborhood fixed effect.

The methodology to impute neighborhood fixed effects is simple. The modal pickup neighborhood during a shift is used.[18] To get a sense of how finely the neighborhoods are defined and the coverage of the whole city, Fig. A5 shows the neighborhoods' border used to classify the shifts. Because it is not the focus of the paper, the current specification with neighborhood fixed effect is not based on a theoretical model. In fact, it is not clear how one should aggregate the location of the different trips. However, if the baseline specification does suffer from an omitted variable bias due to missing neighborhood information, we should see the estimated coefficient vary when we add these fixed effects. From the results presented in column 3 of Table 2, we see that the change in the coefficient is minimal, suggesting that neighborhood effects cannot explain the negative relationship between variation in idiosyncratic shock and shift duration.

In their analysis of daily labor supply decisions of Indian vendors, Andersen et al. (2017) found that the labor supply response to an unexpected income shock was to reduce labor

[18]For example, assume a driver made ten trips during a shift. Four of those trips originated from neighborhood A, while all the other trips originated from different neighborhoods. The value of the neighborhood fixed effect will be neighborhood A.

supply immediately in the form of a midday break. However, when the vendors came back to work, they left the market slightly later than vendors who did not received the income shock. When ignoring midday breaks, the labor supply elasticity seemed positive in response to an unexpected income shock. This highlights the importance of considering intra-shift breaks when looking at a worker's labor supply.

For taxi drivers, the possibility of using intra-shift breaks as another margin to adjust their labor supply seems reasonable. *Ex-ante*, the effect of using a labor supply measure that captures intra-shift breaks is unclear. If drivers that receive a large tip are more likely to take a midday break, then the "large tips" coefficient should be even more negative. If, in contrast, drivers that receive a large tip are less likely to take a mid-shift break, then the "large tips" coefficient would increase. Table 3 presents new estimates of labor supply elasticities when mid-shift breaks are removed from the hours worked. There is a clear and generalized decrease in the idiosyncratic shock coefficients. This suggests that, similar to Andersen et al. (2017), an unexpected idiosyncratic shock reduces labor supply within the shift. The estimates of the labor supply elasticity with respect to market-level shocks seem to increase. Because breaks can be viewed as another margin to adjust labor supply, this result support a neoclassical response to unexpected market-level wage shocks.

To make sure the arbitrary threshold for large tips did not affect the coefficient, I run a sensitivity analysis. Figure 2 shows the coefficient at different threshold levels. As I vary the thresholds, the coefficient remains negative and significantly different from zero. When moving the minimum dollar amount to be considered a large tip (panel (a)), the resulting function is slightly "U-shaped". A threshold that is too low will incorporate many event that are not considered idiosyncratic to the driver. This explains why the coefficient decreases when the threshold is below \$20. Above \$35, the coefficient becomes less precise, as seen by the widening confidence interval. This could explain the small increase in the coefficient. When changing the time for which large tips are recorded (panel (b)), the coefficient seems to decrease. This is due to the fact that we are removing more and more shifts of smaller duration. However, when looking at both figures, it is clear that the effect is present and strongly different from zero.

Another potential threat to the exogeneity of the "large tips" variable is the presence of

Table 3: IV estimates of the wage elasticity and effect of idiosyncratic earnings on shift duration (net of breaks)

	(1)	(2)	(3)	(4)
Large tips	-0.178***	-0.148***		-0.170***
	(0.00507)	(0.00447)		(0.00483)
JFK fares			-0.135***	
			(0.000750)	
Log hourly earnings	1.153***	0.889***	1.154***	1.122***
	(0.00611)	(0.00663)	(0.00611)	(0.00596)
driver FE	yes	yes	yes	yes
date/time FE[a]	m/d/h	m/h*d	m/d/h	m/d/h
holidays	yes	yes	yes	yes
weather	yes	yes	yes	yes
neighborhood	no	no	no	yes
obs	6661422	6661422	6661422	6661013

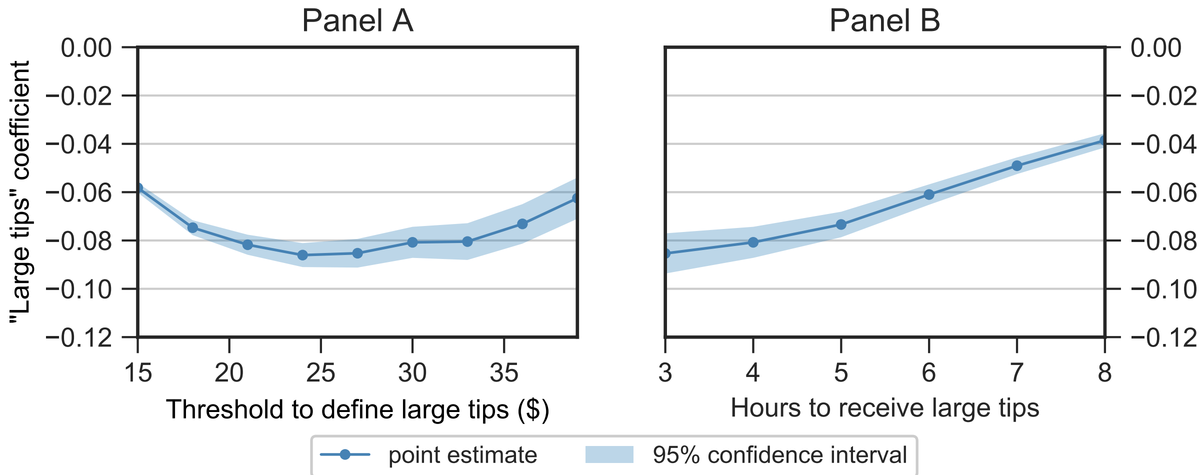
Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors at the driver level in parentheses.
The dependent variable is the log of the duration of the shift, net of breaks that occurred within that shift. Breaks are defined as a period of more than one hour between two trips. See the notes of Table 1 and Table 2 for a definition of the regressors and controls.
[a] m: month of the year; d: day of the week; h: hour of the day.

systematic misreporting. I test for two types of issue by looking at

I pursue the analysis with a discrete-choice stopping model. The model is heavily based on Farber (2005). The main contribution of this new analysis, aside from being done on a much larger dataset, is to decompose the cumulative income (E^c_{ist}) into two components: the cumulative income at the market wage ($\tilde{E}^c_{ist}$) and the cumulative idiosyncratic shock (U^c_{ist}). In the presence of a negative elasticity to labor supply with respect to idiosyncratic shocks, we should observe that positive idiosyncratic shocks increase the probability of stopping. This estimation strategy could further support the claim that, facing a positive idiosyncratic shock to income, drivers reduce their labor supply.

This decomposition is done in a very simple manner. $\tilde{E}^c_{ist}$ is simply the sum of the average hourly earnings of all drivers up until the end of trip t. Therefore, if the driver worked for 8 hours and the average hourly earnings was \$35 during the first 4 hours and \$25 during the last 4 hours, his cumulative income at the market wage would be \$140 after the first 4 hours, \$190 after 6 hours, and \$240 at the end of his shift. The income from idiosyncratic shocks

Figure 2: Coefficient's sensitivity to different thresholds



Notes: In panel (a), the "large tips" coefficient is shown with different threshold above which a regular tip becomes a "large tip". In panel (b), the "large tips" coefficient is shown when different length of time are used to record the large tips.

is computed as the residual (difference between actual cumulative income and cumulative income at the market wage).

Similar to Farber (2005), reduced form equations take the form:

$$\Pr(\text{Stopping}_{ist}) = \beta H^c_{ist} + \delta E^c_{ist} + \mu_i + \epsilon_{ist} \tag{3}$$

$$\Pr(\text{Stopping}_{ist}) = \beta H^c_{ist} + \delta \tilde{E}^c_{ist} + \gamma U^c_{ist} + \mu_i + \epsilon_{ist} \tag{4}$$

Terms in this equation are indexed by the driver (i), the shift (s), and the trip (t). Equation (3) replicates the framework of Farber (2005) while Eq. (4) decomposes the cumulative earnings. Alongside the variables E^c_{ist}, $\tilde{E}^c_{ist}$, and U^c_{ist} that were defined earlier, H^c_{ist} is the cumulative hours up to trip t. μ_i and ϵ_{ist} are respectively driver fixed effects and an error term.

Due to the fact that an analysis at the trip level contains a lot more observations (About 20 times larger[19]), I used a random sample of 1000 drivers for computational ease[20]. I estimate Eq. (3) and Eq. (4) with both a linear probability model and a probit model.

Columns 1 to 3 of Table 4 presents the result of the discrete-choice stopping model estimated with a linear probability model. To avoid the well-known problems of the linear

[19] On average, each working shift contains 21 taxi trips.

[20] These 1000 drivers represents around 2.3% of the total sample. I ran the same analysis on different samples and got qualitatively the same results.

Table 4: Discrete-choice stopping model: marginal effects

	LPM			Probit[a]		
	(1)	(2)	(3)	(4)	(5)	(6)
Hours worked	0.018***	0.045***		0.019***	0.081***	
	(0.001)	(0.002)		(0.001)	(0.004)	
Income (100$)	0.003			0.012***		
	(0.002)			(0.002)		
Market-level income (100$)		-0.081***	-0.080***		-0.117***	-0.120***
		(0.005)	(0.005)		(0.009)	(0.009)
Idiosyncratic income (100$)		0.017***	0.016***		0.045***	0.042***
		(0.002)	(0.002)		(0.005)	(0.004)
date/time dummies[b]	yes	yes	yes	yes	yes	yes
driver fixed effects	yes	yes	yes	yes	yes	yes
Nonparametric time	no	no	yes	no	no	yes
obs	3637537	3637537	3637537	3637528	3637528	3637528

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors at the driver level in parentheses. Hours worked and the three measure of income are all cumulative measures over the working shift of a driver. The dependent variable is a binary variable taking the value of 1 if the trip was the last trip of the shift. Both estimation where made on a random sample of 1000 drivers.

[a] The probit's marginal effects are computed at 8 hours into the shift and at the mean of all other regressors.

[b] The date and time dummies include fixed effects for the month of the year (11), the day of the week (6), and the hours of the day (23).

probability model, the estimation of models with a limited dependent variable is often done with a probit. However, it has been shown that the fixed effects estimator in a probit model is problematic due to the incidental parameter problem (Greene, 2004). This problem is severe in panel dataset when the number of observation per cross-section unit (number of taxi trips per driver) is very small. In the random sample of 1000 drivers used for the estimation, the average number of shifts over the year was 172 and the average number of taxi trips over the year was 3656 (or around 21 taxi trips per shift). The large size of the "T" dimension of the dataset indicate that the bias will be negligible. Results of the probit model are presented in column 4 to 6 of Table 4.

The first and fourth column shows the estimated coefficients of the specification without the decomposition (Eq. (3)). In the linear probability model, the coefficient is extremely close to zero with a small standard error, suggesting a lack of relationship between the accumulation of income during a shift and the decision to end a working shift. This supports the view of Farber (2005, 2015) that reference-dependence (or any other psychological factor that would generate a negative labor supply elasticity) is not a large factor in the labor supply decisions of taxi drivers. When estimated with a probit, the marginal effect of cumulative income increase to a statistically significant level. However, the economic effect is relatively small and the conclusion that the neoclassical model explains a large portion of labor supply behaviors is still consistent with this finding.

Columns 2 and 5 show the regression coefficient when we estimate Eq. (4). The two coefficients representing different portion of the cumulative income have opposite signs. This simple decomposition allows us to see that using the wage as a single measure gives us the average of two opposing effects. In other words, the behaviors inconsistent with the neoclassical model are masked by the effect on the market wage variation. The coefficient on the cumulative earnings at the market wage is negative and significant; that is, increasing the cumulative market earnings, keeping everything else constant, does seem to reduce the probability of stopping. A $100 increase in cumulative earnings at the market wage reduces the driver's probability of stopping by 8 to 12 percent.

The coefficient on the idiosyncratic income is also significant but positive. When estimated using the linear probability model, increasing a driver's idiosyncratic income by $100 raises his probability of stopping by 1.7 percent. While this effect is larger than the simple income coefficient, it is still small. However, one should keep in mind that the decomposition was done in a very naive way. The driver will probably be able to predict a greater proportion of his earnings. Thus, our idiosyncratic income proxy will include a part that is still predictable, having at the same time a neoclassical effect. We should also keep in mind that this effect is an average over the complete shift. To look at the effect of a $100 increase in idiosyncratic income close to the end of an average shift, we can look at the results from the probit model since the marginal effects shown in Table 4 are computed at 8 hours into the shift. The estimated coefficient increases to reach 4.5%, which is a fourfold increase compared to the simple cumulative income.

A possible critique of the discrete-choice stopping model is that if we do not control properly for the time, we might be picking the relationship between the gradual increase of the cumulative income with the increased probability of stopping with time. Therefore, I run the same model in which, instead of controlling linearly for the time, I construct bins of 15 minutes added as fixed effects. This approach is less parametric than the previous one and allows the effect of an additional hour of work when the shift started different than when the driver is into his 10th hour of work. Column 3 and 6 of Table 4 present the result. The coefficients are almost identical, supporting the robustness of the results.

7 Discussion

It is important to note the different identification strategy of this paper. Previously, the focus was put on income variations that generate a substitution effect while arguing that the income effect should be negligible. In this paper, since idiosyncratic shocks do not modify future earnings expectations, substitution effects are by construction absent. Because a large contemporaneous income effect could be the cause of the negative labor supply elasticity observed in the data, we need to consider alternative psychological factors to explain this decision-making anomaly.

This section will consider three broad classes of psychological factors to explain the results: reference-dependence, narrow bracketing, and present bias. In their own way, they are able to explain features of the data. Because the reduced-form results do not allow me to distinguish between them, I will simply discuss how they are consistent with the main results of section 5.

7.1 Reference-Dependence

A discussion on the labor supply decision of taxi drivers would be incomplete without a look at reference-dependence. This is especially true considering the amount of research done on this subject in the past two decades. The most recent pieces of evidence are mixed. On one side, Farber (2015) argues that income reference-dependence does not play an important role, and, on the other, Agarwal et al. (2015), Leah-Martin (2017), and Thakral and Tô (2017) find empirical anomalies that can be explained by reference-dependence. Similarly, my results are also mixed. While I find a negative labor supply elasticity in response to idiosyncratic shocks, the labor supply elasticity with respect to variations in the market-level wage is positive.

This last finding is consistent with the neoclassical life-cycle model of labor supply.

The central idea behind any model with reference-dependent preferences is the presence of a reference point for the driver. When this target is reached, the marginal utility from an additional taxi trip is diminished. What makes reference-dependent models so hard to estimate is the fact that we do not know exactly how agents choose their target. Standard models of reference-dependence based on Kőszegi and Rabin (2006) makes the assumption that the target is determined by the expected value of earnings during that period. For instance, if a taxi driver usually makes $300 during a regular Saturday night, the model assumes this will be their target during other regular Saturday nights. The positive estimate for the labor supply elasticity with respect to an unanticipated market-level shock does not support the presence of this type of reference-dependence.

It does not, however, rule out the possibility that drivers set a target during a shift. Anecdotal evidence from discussions with taxi drivers seem to indicate that they sometimes set a target (income or time) but they also modify those targets when some external shocks occur to optimize their medium-term earnings. To reconcile the results and the reference-dependence hypothesis, a simple solution is to modify the assumption on target setting behaviors. Indeed, a reference-dependence model in which the target adjusts to market-level shocks but not to idiosyncratic shocks would generate the patterns I observed in the data.

A similar type of preferences can be found in Thakral and Tô (2017). Also looking at New York City taxi drivers, they observe that drivers display a stronger income effect in response to more recent earnings. This can be seen as another dimension where labor supply respond differently to earnings shocks. Although their model cannot explain why drivers would react differently to idiosyncratic and market-level shocks if they are received at the same time, a simple modification of the reference-point determination equation that adjust more quickly to market-level earnings shock would be sufficient to nest both empirical patterns.

7.2 Narrow Bracketing

The second class of psychological factors can be grouped under the what behavioral economists call narrow bracketing. Rather than taking into account all the information and consequences, narrow bracketing implies that the agent uses only a subset of the available information.

Generally, this will yield a lower overall utility. Read et al. (1999b) present multiple randomized experiments in which they clearly observe a preference reversal when the framing was modified (e.g. by forcing the subjects to make decisions sequentially instead of simultaneously). These systematic preference reversals are in contradiction to the predictions of the neoclassical model with rational agents.

The negative labor supply elasticity found in section 5 could be explained by narrow bracketing. Indeed, the drivers, by making continuous labor supply decisions during a shift, might simply be unable to compute the complete optimization problem and therefore simplify it by not considering all the consequences. For instance, when receiving a tip of $30, a driver would not re-optimize their life-cycle labor supply decision in the face of this income shock. Instead, he might decide to simply reduce the problem to a daily labor supply decision, and will decide to "purchase" more leisure during the day.

To support this view, it is important to note that both studies that find or reject the presence of behavioral effects in the labor supply of taxi drivers also found that these effects are smaller or nonexistent the more experienced a driver is (Camerer et al., 1997; Farber, 2015; Sheldon, 2016). Combined with evidence on the relationship between cognitive ability and preference anomalies (Benjamin et al., 2013) and on learning by doing processes in the taxi industry(Agarwal et al., 2016; Haggag et al., 2017), it suggests that drivers learn to optimize and act closer to a rationally maximizing agent as they acquire more experience. This supports the theory that continuously making labor supply decisions over the day is a costly process that new driver tend to reduce to a daily decision, while more experienced drivers are able to consider the problem over a longer time horizon.

Moreover, the narrow bracketing hypothesis is not undermined by the positive labor supply elasticity with respect to market-level shocks. This behavior can be explained by a strong substitution effect. Even though the income effect seems economically significant, it is possible that the substitution effect is larger, generating a net positive labor supply response to variations that modify the expected earnings in the short run.

7.3 Present Bias

The third and final category of behavioral models I will discuss is related to the concept of time-inconsistent preferences. More specifically, time-inconsistency refers to a situation in which a person's current preferences are at odds with their future preferences regarding the same decision. This type of behavior has been observed in many environments. For instance, Benartzi and Thaler (1995) show how this type of preference can explain the equity premium puzzle. It has even been observed in surprising contexts such as making the decision between watching a "highbrow" or a "lowbrow" movie (Read et al., 1999a).[21] Time-inconsistent preferences are generally modeled using discounting functions other than the standard exponential discounting. Hyperbolic discounting and beta-delta discounting are two of the most popular functions considered in the literature. In both cases, people tend to excessively weight the present, leading to reward-salient activities being done too soon and cost-salient activities being done too late compared to a perfectly rational agent's decision.

For taxi drivers, a present bias could explain a large income effect I observed during the contemporaneous shift. When choosing how to spend the income generated from the positive idiosyncratic shock, the driver has an overwhelming preference to spend it today. This increased wealth is spent, at least in part, on leisure, thus decreasing the labor supply.

For our purposes, present bias and temporal narrow bracketing are very similar. Conceptually, someone with present-bias preferences will acknowledge all the options and consequences, but the strong preference for immediacy will dwarf the utility obtained from future periods. For narrow bracketing, the latter periods are simply ignored. In our setting, the result is the same: it creates a large daily income effect. The explanation for the positive elasticity with respect to market-level shocks is also the same: the substitution effect brings the overall effect back in positive territory. They are thus empirically indistinguishable with the observational data at hand. On the other side, the way by which reference-dependence generates the negative labor supply elasticity is quite different. With data generated in a controlled environment, the

[21] In the experiment, when the subjects had to choose a movie to watch immediately, the majority chose the "lowbrow" movie. With time-consistent preferences, we would expect the choice not to differ depending on the delay between the decision and the reward (watching the movie). However, "highbrow" movie were significantly more likely to be chosen if the subject had to wait a day or two before the viewing.

presence of a reference point for taxi drivers could possibly be verified or ruled out. However, with the available data, it is not possible to tell the three classes of models apart.

8 Conclusion

In this paper, I used very large tips received by New York City taxi drivers to show that the labor supply elasticity is negative in response to unexpected idiosyncratic shocks. These windfall gains are small relative to yearly or monthly earnings and should not generate any noticeable income effect according to the standard neoclassical model of labor supply. To provide further evidence that this result cannot be generate by the standard neoclassical model with very large income effects, I estimate the effect of receiving a large tip during a shift on the labor supply during the next shift. Income effect should not be extremely short-lived. However, I find that a large tip has no effect on future labor supply. In contrast to Camerer et al. (1997), the labor supply elasticity is not negative for all types of unanticipated variation. When the unanticipated earnings shock is generalized at the market-level, I find a positive labor supply elasticity.

The empirical results contained in this paper are not a test of reference-dependence. Other types of well-known psychological factors could explain the negative elasticity with respect to idiosyncratic income variations. I explored some of the behavioral models that could possibly play a role in the labor supply decision. Aside from the presence of reference-dependent preferences, I discussed the potential for present bias and narrow bracketing to explain the results. Further research should focus on disentangling these psychological factors.

The original findings of Camerer et al. (1997) were important for the development of the literature on psychological factors influencing labor supply decisions. Even if Farber (2005, 2015) presents compelling arguments that the neoclassical model can explain most of the labor supply decisions, a sizable number of researchers generally accept the claim that psychological factors, and in particular reference-dependence, affect labor supply decisions (Ordóñez et al., 2009; Barberis, 2013). In this paper, I attempted to reconcile those two views. On one hand, if the neoclassical model makes the right predictions on most of the anticipated as well as the unanticipated market-level income variations, the policy-relevance of reference-dependent preferences is minor. Indeed, most of the policymaker's tools will impact the anticipated or market-level income variations. Furthermore, the majority of the labor force works under well-defined and stable contracts in which an idiosyncratic component is rarely applicable. Therefore, the predictions of the neoclassical model should not be discarded. On the other hand, the result of this paper demonstrated that psychological factors are playing a role in the labor supply decision of taxi drivers when idiosyncratic shocks are involved. If one goal of labor economists is to lean toward a universal understanding of how humans make labor supply decisions, then capturing how psychological factors affect these decisions should not be taken lightly.

The debate on the importance of psychological factors to aggregate labor supply behavior is far from over. This paper provided a starting point for further research on how different source of income might generate different responses, even when the usual income and substitution effects have been taken into account. For instance, extending the methodology to a framework that would allow for heterogeneity in the presence of psychological factors or their intensity would be valuable. Furthermore, even if previous research quickly pointed toward reference-dependence as the cause of the negative elasticity of labor supply, it is still unclear whether it is the only channel generating the negative labor supply elasticity or if other psychological factors are contributing to the behavior.

References

Agarwal, S., S.-f. Cheng, J. Keppo, and R. Sato (2016): "Learning by Driving: Evidence from Taxi Driver Wages in Singapore," Working Paper, 00002.

Agarwal, S., M. Diao, J. Pan, and T. F. Sing (2015): "Are Singaporean Cabdrivers Target Earners?" Working Paper, 00007.

Andersen, S., A. Brandon, U. Gneezy, and J. A. List (2017): "Toward an Understanding of Reference-Dependent Labor Supply: Theory and Evidence from a Field Experiment," Working Paper, 00008.

Barberis, N. C. (2013): "Thirty Years of Prospect Theory in Economics: A Review and Assessment," *Journal of Economic Perspectives*, 27, 173–195, 00451.

Benartzi, S. and R. H. Thaler (1995): "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, 110, 73–92, 02866.

Benjamin, D. J., S. A. Brown, and J. M. Shapiro (2013): "Who Is 'Behavioral'? Cognitive Ability and Anomalous Preferences," *Journal of the European Economic Association*, 11, 1231–1255, 00268.

Borjas, G. J. (1980): "The Relationship between Wages and Weekly Hours of Work : The Role of Division Bias," *Journal of Human Resources*, 15, 409–423.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997): "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, 112, 407–441.

Chang, T. and T. Gross (2014): "How Many Pears Would a Pear Packer Pack If a Pear Packer Could Pack Pears at Quasi-Exogenously Varying Piece Rates?" *Journal of Economic Behavior & Organization*, 99, 1–17.

Chou, Y. K. (2002): "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore," *Singapore Economic Review*, 47, 17–47.

Crawford, V. P. and J. Meng (2011): "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101, 1912–1932.

Doran, K. (2014): "Are Long-Term Wage Elasticities of Labor Supply More Negative than Short-Term Ones?" *Economics Letters*, 122, 208–210.

Dupas, P., J. Robinson, and S. Saavedra (2017): "The Daily Grind: Cash Needs, Labor Supply and Self-Control," Working Paper, 00028.

Farber, H. S. (2005): "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 113, 46–82.

——— (2008): "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, 98, 1069–1082.

——— (2015): "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers," *Quarterly Journal of Economics*, 130, 1975–2026.

Fehr, E. and L. Goette (2007): "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 97, 298–317.

Giné, X., M. Martinez-Bravo, and M. Vidal-Fernández (2017): "Are Labor Supply Decisions Consistent with Neoclassical Preferences? Evidence from Indian Boat Owners," Working Paper.

Goldberg, J. (2016): "Kwacha Gonna Do? Experimental Evidence about Labor Supply in Rural Malawi," *American Economic Journal: Applied Economics*, 8, 129–149.

Greene, W. (2004): "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects," *Econometrics Journal*, 7, 98–119.

Haggag, K., B. McManus, and G. Paci (2017): "Learning by Driving: Productivity Improvements by New York City Taxi Drivers," *American Economic Journal: Applied Economics*, 9, 70–95.

Haggag, K. and G. Paci (2014): "Default Tips," *American Economic Journal: Applied Economics*, 6, 1–19.

Kőszegi, B. and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 121, 1133–1165, 01922.

Leah-Martin, V. (2017): "When to Quit: Narrow Bracketing and Reference Dependence in Taxi Drivers," Working Paper.

Morgul, E. F. and K. Ozbay (2015): "Revisiting Labor Supply of New York City Taxi Drivers: Empirical Evidence from Large-Scale Taxi Data," Working Paper.

Oettinger, G. S. (1999): "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors," *Journal of Political Economy*, 107, 360–392.

Ordóñez, L. D., M. E. Schweitzer, A. D. Galinsky, and M. H. Bazerman (2009): "Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting." *Academy of Management Perspectives*, 23, 6–16.

Read, D., G. Loewenstein, and S. Kalyanaraman (1999a): "Mixing Virtue and Vice: Combining the Immediacy Effect and the Diversification Heuristic," *Journal of Behavioral Decision Making*, 12, 257–273.

Read, D., G. Loewenstein, and M. Rabin (1999b): "Choice Bracketing," *Journal of Risk and Uncertainty*, 19, 171–197.

Sheldon, M. (2016): "Income Targeting and the Ridesharing Market," Working Paper.

Stafford, T. M. (2015): "What Do Fishermen Tell Us That Taxi Drivers Do Not? An Empirical Investigation of Labor Supply," *Journal of Labor Economics*, 33, 683–710.

Thakral, N. and L. T. Tô (2017): "Daily Labor Supply and Adaptive Reference Points," Working Paper.

Appendix A Creation of the hourly market wage

To eliminate the division bias when the labor supply is present both as the dependent variable and in the denominator of the hourly earnings, I use an instrumental variable strategy. Camerer et al. (1997) and Farber (2015) use measures of the distribution of the wage of other drivers starting a shift on the same calendar day. I improve this instrument by computing the average market wage *only* when the driver was working.

Define E_k^{all} as the sum of fares of trips starting in hour k.Similarly, define D_k as the number of drivers working during hour k, weighted by the fraction of the hour worked. The instrument is constructed as follow:

1. Compute the average earnings during hour k: $W_k^{all} = \frac{E_k^{all}}{D_k}$

2. For each shifts, compute the weighted average of W_k^{all} using the fraction of the hour k worked as the weights. This is the instrument $\widehat{W}_{is}$ for shift s of driver i.

For the discrete-choice stopping model presented in Section 6, the cumulative earnings at the market wage is computed as the sum of W_k^{all} for which k is part of the cumulative hours worked. I adjust incomplete hours accordingly.

Appendix B Additional Tables and Figures

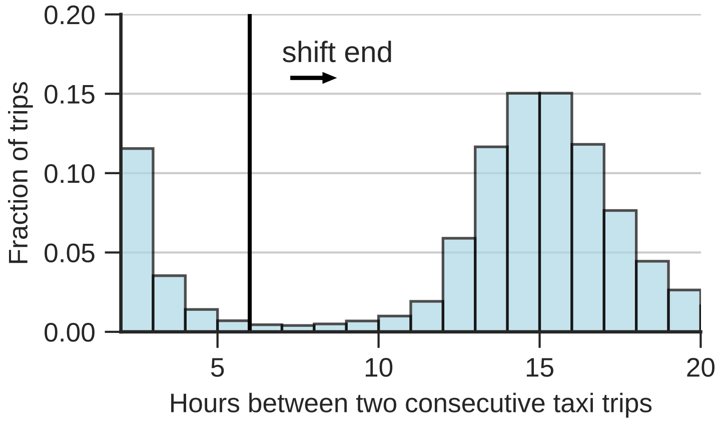
Table A1: IV estimates of the wage elasticity and effect of idiosyncratic earnings on shift duration (alternative large tips)

	(1)	(2)	(3)	(4)	(5)
Large tips	-0.0651***	-0.0262***	-0.0135***	-0.0800***	-0.0843***
	(0.00313)	(0.00298)	(0.00310)	(0.00330)	(0.00346)
Log hourly earnings	0.615***	0.615***	0.615***	0.615***	0.615***
	(0.00600)	(0.00600)	(0.00600)	(0.00600)	(0.00600)
driver FE	yes	yes	yes	yes	yes
date/time FE[a]	m/d/h	m/d/h	m/d/h	m/d/h	m/d/h
holidays	yes	yes	yes	yes	yes
weather	yes	yes	yes	yes	yes
minimum tip percentage	0.2	0.3	0.5	0	0
minimum fare amount	0	0	0	5	10
obs	6661422	6661422	6661422	6661422	6661422

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors at the driver level in parentheses.

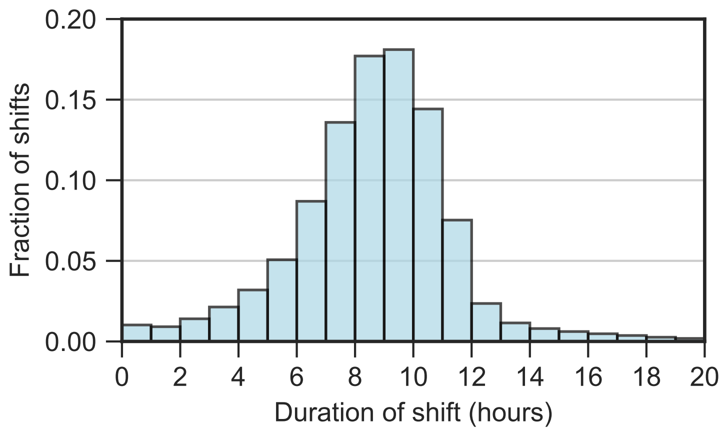
See the notes of Table 1 for a definition of the regressors and controls. Columns (1) and (2) exclude tips that account for less than 20% and 50%, respectively, of the fare. Columns (3) and (4) exclude tips with an associated fare of less than \$5 and \$10, respectively.

Figure A1: Distribution of the waiting time between two trips



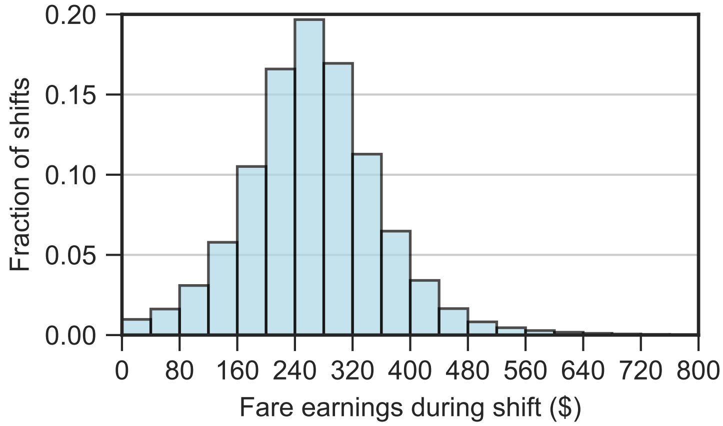
Notes: Wait time is defined as the difference between the ending time of a trip and the start time of the same driver's next trip. Wait times below 2 hours and above 20 hours are excluded from this figure.

Figure A2: Distribution of shift durations



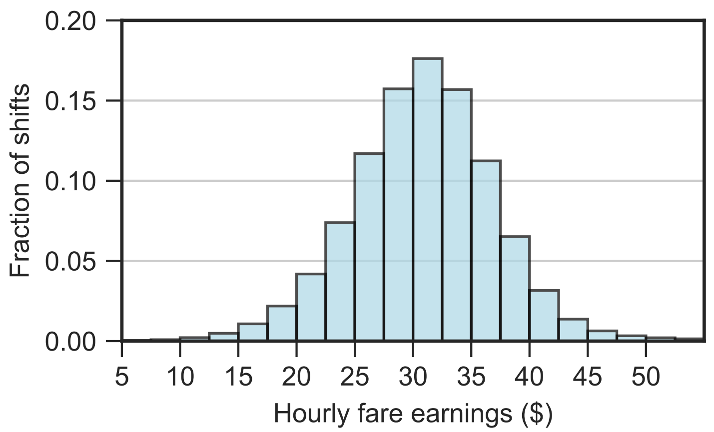
Notes: A shift starts at the first trip after a wait time of at least 6 hours and ends after the last trip before another period of at least 6 hours of wait time. A shift's duration is the difference between the ending time of the last trip and the starting time of the first trip in a shift.

Figure A3: Distribution of earnings during a shift



Notes: A shift starts at the first trip after a wait time of at least 6 hours and ends after the last trip before another period of at least 6 hours of wait time. A shift's fare earnings is the sum of all fares received during a shift, including surcharge but excluding tips.

Figure A4: Distribution of hourly earnings

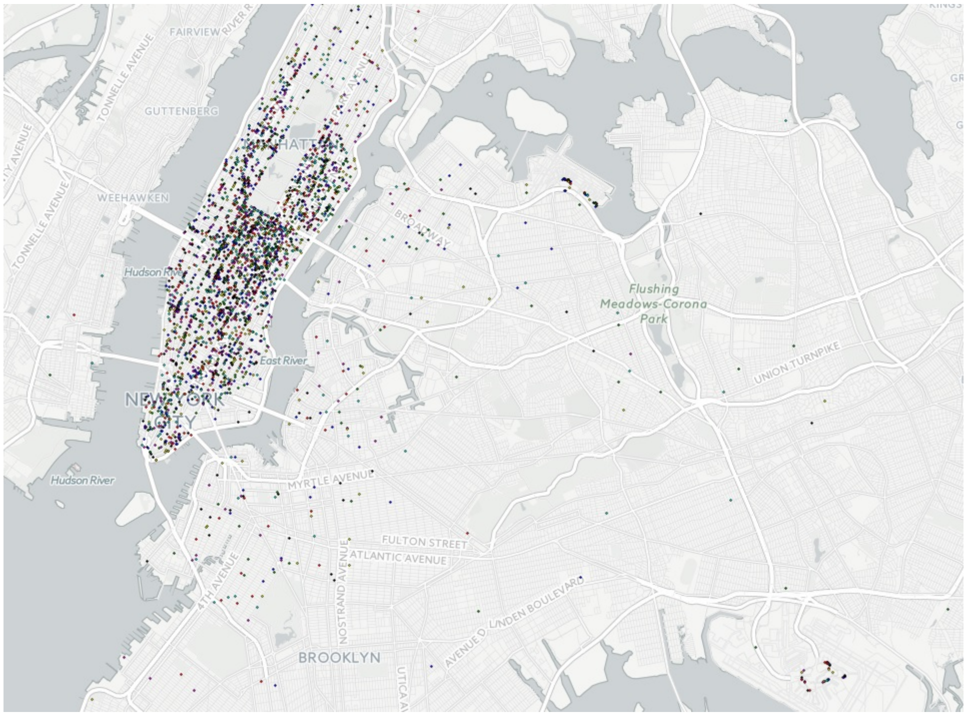


Notes: A shift starts at the first trip after a wait time of at least 6 hours and ends after the last trip before another period of at least 6 hours of wait time. The hourly earnings of a shift is computed as the ratio of earnings from fares and the duration of the shift in hours.

Figure A5: New York City districts

Notes: This map presents the limits of the neighborhoods used to add geospatial fixed effects. There are 71 New York districts (59 community districts and 12 other non-residential districts). Furthermore, I added Jersey City as a region to include Newark Airport.

Figure A6: Pickup locations



Notes: A random sample of 3000 taxi trips is used.