Identifying Behavioral Responses in Labor Supply: Idiosyncratic Shocks vs. Market-Level Shocks

Marc-Antoine Schmidt*

Draft version: August 2017

Abstract

This paper presents empirical evidence that the labor supply elasticity of taxi drivers has a different sign depending on the source of the earnings shocks. The observed pattern cannot be explained only by different magnitude of the income or substitution effects and is thus inconsistent with the neoclassical life-cycle model of labor supply. To get this result, I decompose unexpected earnings variations into a market-wide component and an idiosyncratic component. I use microdata on the universe of New York medallion taxi trips in 2013 to identify abnormally large tips. These large tips act as an exogenous proxy for the idiosyncratic component of the drivers' earnings. I find that a negative labor supply elasticity is observed for the idiosyncratic component; taxi drivers respond to a positive idiosyncratic shock by decreasing their labor supply by an economically significant amount. On the other hand, a positive earnings shock *at the market level* causes their labor supply to increase, consistent with an optimizing rational agent. This differs from previous studies which assume a homogeneous labor supply effect of unexpected earnings shocks. Three types of psychological factor consistent with my empirical findings are compared and discussed.

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

*email: ma.schmidt@mail.utoronto.ca. I would like to thank Philip Oreopoulos, Kory Kroft, Gustavo Bobonis, Stephen Ayerst, Jean-William Laliberté, Mathieu Marcoux, and Baxter Robinson for their helpful advice and suggestions. I also benefited from comments and discussions at the CEA meeting in Ottawa and the SCSE conference in Québec city.

1 Introduction

One of the main goals of labor economics is to understand how agents make labor supply decisions. In particular, how does an economic agent's labor supply respond to variation in their wage? On this question, an important focus of research in the last two decades has been to understand which model would best predict the labor supply response of a representative variation in hourly earnings.[1] In this paper, I investigate whether the labor supply reactions to two different types of hourly earnings variations are similar. More precisely, I decompose a worker's unexpected labor income into two components depending on its origin: one from market-level sources and another one from idiosyncratic sources. If the labor supply reactions to these two types of earnings variation are not similar, are the common models of labor supply able to explain this difference?

In this paper, I explore whether taxi drivers react differently to unexpected earnings shocks that affect a portion of the market (market-level shocks) and unexpected earnings shocks that affect only a single driver (idiosyncratic shocks). This distinction could be important in order to reconcile the literature's conflicting results. Indeed, the labor supply elasticity estimates will depend on the identification strategy's reliance on one type of earnings shock or the other. By comparing the predictions of a neoclassical model to the estimated labor supply elasticities from both types of shock, I test whether the neoclassical model explains the behavior of taxi drivers. I find that in response to market-level shocks, the labor supply elasticity estimate is positive. However, when looking at the response to idiosyncratic shocks, proxied by large tips, the elasticity is negative. Moreover, idiosyncratic shocks affects only the labor supply decision during the day they occur, indicating that income effects cannot be the sole source of this negative elasticity.

The benchmark framework is a neoclassical model of labor supply where the sign of the effect of a variation in wage is ambiguous. This ambiguity depends on the relative magnitude of the substitution and income effects. When the wage variation is temporary and small relative to total earnings, the income effect will be negligible. Furthermore, while the substitution effect impacts the immediate labor supply, the income effect is spread over multiple periods. In this setting, estimates of the labor supply elasticity will only measure the substitution effect. If a worker's wage temporarily increases, the opportunity cost of leisure goes up, resulting in a positive labor supply response. This simple framework is certainly useful to answer many questions but might fail to correctly explain interesting

[1] For instance, papers that explicitly estimate a labor supply elasticity (Camerer et al., 1997; Farber, 2015) assume that it is representative of the response to all unanticipated earnings shocks. Papers that estimate the marginal effect of cumulative earnings on the stopping probability with a reduced form or a structural model (Farber, 2005, 2008; Crawford and Meng, 2011; Agarwal et al., 2015) assume that all sources of earnings are treated equally by taxi drivers. In other words,

behaviors, especially when we start to look at the daily supply decision.

A competing explanation is that workers have reference-dependent preferences or income targeting. Advocates of reference-dependent models argue that an agent might work less in response to a temporary and unanticipated wage increase. An individual with reference-dependent preferences responds to earnings, just as in the neoclassical model, but also responds to how far this individual is from some target, often a level of income in a single day. Below the target, the individual incurs additional losses of utility. On days when the wage is higher, reaching an income target requires less time. This generates a negative relationship between wages and hours worked. Following the seminal work of Camerer, Babcock, Loewenstein, and Thaler (1997), controlled and natural experiments have generated results that seem to support the presence of reference-dependent preferences and income targeting.

The empirical analysis rests on a dataset that became recently available: the taxicab passenger enhancement project dataset. This dataset will allow me to look at the universe of taxi driver shifts in New York City during 2013. It is multiple orders of magnitude larger than datasets used in studies prior to Farber (2015). Even though the taxi industry has been the most studied real-world environment to answer this type of question, only Thakral and Tô (2017) also used tips to study the taxi drivers' labor supply. It is worth noting that because some drivers consistently receive higher tips than others, I will only use variation within individuals to identify the elasticity.

Many robustness checks are conducted to ensure that the results are not driven by an unobservable characteristic of large tips. Furthermore, applying the framework of Farber (2005) and using a discrete choice stopping model, the results are consistent with the main empirical framework. Decomposing the cumulative earnings of the drivers into a market-level component and an idiosyncratic component, I find that drivers are more likely to stop their shift as the cumulative idiosyncratic earnings gets larger. Similar to Farber (2015), I find an opposite effect for the marginal effect of cumulative market-level earnings. Altogether, these results clearly support the presence of psychological factors in labor supply decision, while showing that the neoclassical model remains a good approximation when earnings are not subjects to idiosyncratic shocks.

Previous studies on the labor supply of taxi drivers assumed that any deviation from the neoclassical model was caused by reference-dependent preferences. In this paper, I also explore competing alternative explanations. While the neoclassical model is rejected by the test, the standard reference-dependence model is also not fully able to explain the drivers' behavior. Nevertheless, a simple extension of the standard reference-dependence model in which market-level shocks are able to affect a driver's target fits the pattern observed in the data.

Aside from reference-dependence, two other broad classes of model with psychological factors will be analyzed in lights of the empirical findings. First, it is possible that drivers do not consider all the consequences of small but recurrent shocks. This can be explained by models with narrow bracketing or models of bounded rationality. Second, a bias towards the present might create time-inconsistent preferences. Under this hypothesis, a driver will spend an idiosyncratic income shock earlier then would be optimal with time-consistent preferences.

The remaining of the paper is organized as follows. Section 2 provides a basic review of the literature. In section 3, I present important characteristics of the dataset and the aggregation methodology. Section 4 describes the empirical strategy used to test the neoclassical model. The results are presented in section 5 and multiple robustness checks are explored in section 6. Finally, section 7 will discuss the potential different behavioral models that could explain the results.

2 Competing models of labor supply

In a seminal paper, Camerer, Babcock, Loewenstein, and Thaler (1997) showed evidence of reference-dependent labor supply among taxicab drivers. Their simple approach was to regress the shift's duration on the hourly earnings. The 2SLS estimate of the elasticity, with the hourly earnings instrumented by moments of the hourly earnings distribution of other drivers on that day, was negative and large (close to -1). 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, they discuss how this behavior could be caused by reference-dependent preferences through income targeting.

The basic idea of reference-dependent preferences in a labor supply setting is that drivers set a target, usually in terms of daily earnings. During shifts in which the wage is high, the driver's cumulative income grows faster, reaching his target earlier. When the target is reached, the marginal benefit of further earnings decreases. This means that increasing the wage of drivers with reference-dependent preferences should decrease their labor supply around their target. This is the opposite consequence of a wage variation in a neoclassical world. Indeed, the neoclassical model would suggest that increasing the wage will raise the opportunity cost of leisure, thus increasing labor supply if the utility function is smooth, hours are somewhat flexible, and the income effect is not too large.

These two studies were followed by many papers that presented evidence for and against the usefulness of reference-dependent modeling of labor supply decisions. For instance, Oettinger (1999) studied the labor supply decisions of stadium vendors. In this setting, stadium vendors were free to choose whether they wanted to work or not on a particular day. This is referred to as the extensive margin. They estimate the elasticity on the extensive margin to be positive and large. This was in part corroborated by Fehr and Goette (2007) who looked at bicycle messengers. Their main finding is that the labor supply elasticity on the extensive margin is positive, but that the elasticity on the effort margin is negative. This finding is consistent with the behavior of reference-dependent agents.[2] These studies provided valuable empirical and methodological insights into the study of reference-dependence.

The first critique of Camerer et al. (1997) came from Farber (2005). Arguing that the econometric method used in Camerer et al. (1997) was flawed, he presents a discrete-choice stopping model in which the driver decides whether to stop or not after the end of each trip. If reference-dependent preferences were really important in the decision of the driver, then we would observe a significant effect of cumulative earnings on the probability of stopping. Instead, he finds that cumulative income does not appear to affect the probability of stopping in any significant way[3].

Using the same dataset as Farber (2005), Crawford and Meng (2011) presented an empirical test of reference-dependence based on the theoretical work of Kőszegi and Rabin (2006). Similar to Farber's study, they employ a discrete-choice stopping model. The crucial difference is that a driver has both an income target and a shift duration target. This seemed to fit the data better than the simple income-targeting model of Farber.

More recently, Farber (2015) is the first to use the TPEP dataset, comprising the universe of taxi trips made during an extensive period (2009-2013) in New York City, to look at the labor supply decision. The data used in this paper come from the same source. With this new dataset, the original result of Camerer et al. (1997) does not hold. More specifically, he finds a positive and significant labor supply elasticity. He also looks at the heterogeneity across drivers of the labor supply elasticity. He finds that the most experienced drivers exhibit more neoclassical behaviors. Conversely, the drivers with the most negative labor supply elasticity tend to exit the market more often. Finally, Kőszegi and Rabin (2006)'s model predicts that reference-dependent behaviors are only caused by unanticipated variation in wage. Farber conducts a variance decomposition exercise and shows that more than 85% of the variation in average wage can be categorized as anticipated.[4] The study concludes that reference-dependence is not of first-order importance

[2] See also Stafford (2015) for a study of labor supply of fishermen.

[3] Since then, a number of papers have improved our knowledge of reference-dependence among taxi drivers. Farber (2008) was among the first to estimate a structural model with explicit reference points, while Doran (2014) showed that the labor supply elasticity mostly depends on the duration of the wage variation. See also Agarwal et al. (2015).

[4] This exercise assumes that anticipated wage variation can be explained by time and date fixed effects.

when considering the taxi industry.

Although Farber (2015)'s results suggest that the negative relationship between hours worked and the wage was either spurious or changed over the last 20 years, the debate is not over. For example, Leah-Martin (2015) replicates Farber's positive labor supply elasticity for taxi drivers in San Francisco. However, using a modified version of Farber (2005)'s discrete-choice stopping model, he presents evidence that a significant portion of taxi drivers in both New York City and San Francisco seem to exhibit reference-dependent labor supply. Furthermore, while also finding that most of the average wage variation can be explained by time and date fixed effects, variation in individual wage tend to be a lot less anticipated.

Also using the TPEP dataset, Thakral and Tô (2017) look at 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 and most of their results are similar. For instance, they find that the timing of the earnings impacts the labor supply decision differently, suggesting that earnings are non-fungible. In this paper, I find this non-fungibility in the source of the income (idiosyncratic vs market-level). To show that their result are robust to the possibility that earnings are correlated with effort, they use tips as an instrument for total earnings. This approach is fundamentally different than what is use in the remaining of this paper. Instead of using all tips, I restrict my analysis to abnormally large tips and argue that these tips must be unexpected. Because they use the tip as an instrument for earnings, Thakral and Tô (2017) only uses the part of the variation in the tip that is correlated with earnings, excluding unexpectedly large tips.

3 Data

3.1 Taxicab passenger enhancement project dataset

Up until 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 (TPEP) allowed the installation of computerized meters able to store a multitude of data: unique identifiers for the driver and the medallion (taxicab), 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.

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.[5] 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 B6 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 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

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

paid by credit or debit card Haggag et al. (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. For an overview of tipping behaviors using the same data source but for the year 2009, see Haggag et al. (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 the outlier data.

3.2 Aggregation and variable creation

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 B1 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 four hours between trips represents a shift delimiter[6]. Figure B2 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

[6]Farber (2015) used a threshold of 6 hours. Every regression were run with both thresholds and the results did not change. The advantage of using a lower threshold is to reduce the number of extremely long shifts. Indeed, some drivers in the dataset work for more than three days consecutively without taking a 4-hour break.

hourly earnings of all drivers during the shift. The methodology to create the instrument for the wage and the cumulative earnings at the market wage is described in the appendix. Figure B4 shows the distribution of hourly average wage in 2013. We observe a lot of variation, the 5th percentile being close to half of the wage in the 95th percentile of the 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 earnings shocks (U^c_{st}). Cumulative hours and cumulative earnings are simply the 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 wage in the market. The cumulative earnings shocks is the difference between the actual cumulative earnings and the cumulative earnings at the market wage.

4 Empirical strategy

The primary empirical strategy I employ is to use large tips received by taxi driver to proxy for idiosyncratic shocks. The set of controls used and the driver fixed effects will reduce the possibility of other channels affecting both the labor supply and the probability of receiving a large tip. The large tips are only a component of the total idiosyncratic income during a shifts and other methods can be used to futher solidify my claim. As a robustness check, I use trips to the JFK airport as the proxy for idiosyncratic income.

Previous tests of non-standard labor supply decisions implicitly made the assumption that income sources are fungible for taxi drivers up to the implied change in expectation for future earnings. In other words, they assumed that when keeping the income and substitution effects constant, taxi drivers did not distinguish between an unexpected $20 tip and a $20 increase in fares because of an unexpected demand surge. Consequently, if expectations about future earnings stay constant, the labor supply elasticity should be the same no matter the source of income variation it originated from. My empirical strategy allows me to relax this assumption with regards to market-level and idiosyncratic income shocks.

Taxi drivers face multiple sources of income variation. This allows me to decompose earnings in an intuitive way. 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 behavioral 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.

Once the expected income variations are controlled for, the coefficient on the hourly wage will identify the effect of unexpected income variations. Two types of unexpected variations will be looked at: market-level shocks and idiosyncratic shocks. Market-level shocks, while still 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.

The same 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, and the total earnings during a shift be E. The logarithm of the hourly earnings (w_{is}) is defined as $\log\left(E_{is}/H_{is}\right)$. Similar to Camerer et al. (1997) and Farber (2015), the benchmark estimating equations of the instrumental variable framework are:

$$w_{is} = \gamma_1 \log\left(\frac{\sum_{k\in A_s} E_k}{\sum_{k\in A_s} H_k}\right) + \gamma_2 X_{is} + \mu_i' + \epsilon_{is} \tag{1}$$

$$\log(H_{is}) = \delta \widehat{w}_{is} + \beta X_{is} + \mu_i + \nu_{is} \tag{2}$$

The variables are indexed by the driver (i) and the shift (s). Shift specific controls are included in X_{is} and can contain time and date fixed effects, rain, cold weather[7], and major holidays. μ_i and μ_i' are driver fixed effects. Thus, the identifying variation comes from within driver. In the first stage (equation 1), total earnings and shift length are summed over shifts in A_s, the set of all shifts worked between the starting time and the ending time of s[8]. If we assume that a driver's income is fungible, then the coefficient δ represents the

[7]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. I define cold weather as anything below 0°C.

[8]A shift in A_s might only be a portion of a shift. For example, if A_s contains all shifts between 5 a.m. and 10 a.m., then only the rides that occur between those hours will be counted towards the sum.

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, δ now represents the labor supply elasticity with respect to market-level shocks only. Although it would be almost impossible to correctly capture every idiosyncratic shock since it would require a knowledge of the agent's expectation, we can approximate them. The main strategy will be to use large tips received by a driver. Because some drivers receive on average higher tips, it will be 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 airport rides 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.

Adding this to the benchmark equation is easy. The first stage equation remains the same and the second stage becomes:

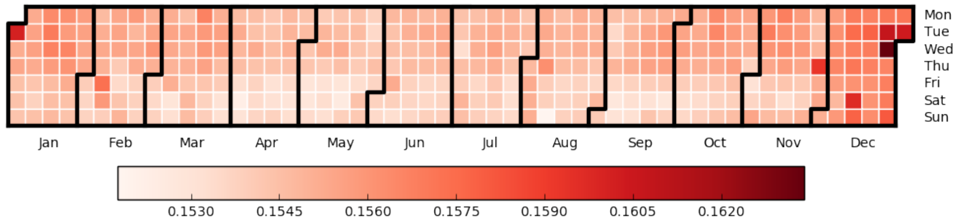
$$\log(H_{is}) = \delta^M \widehat{w}_{is} + \delta^I T_{is} + \beta X_{is} + \mu_i + \nu_{is} \tag{3}$$

$$T_{is} = \begin{cases} 1, & \text{if tip received before } \tilde{h} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

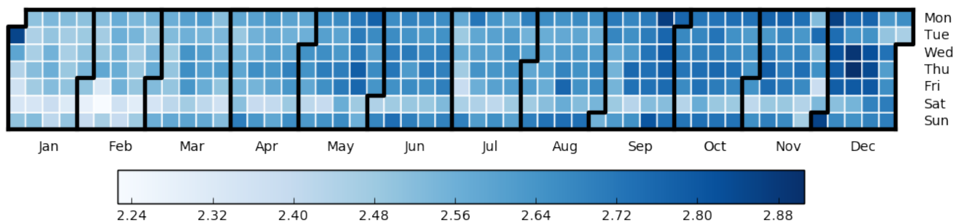
The idiosyncratic shock variable (T) is binary, taking the values of 1 or 0. It takes the value of 1 if the driver receives a recorded tip larger than an arbitrary threshold during the shift s[9]. Furthermore, the tip needs to be received in the first $\tilde{h}$ hours of a shift. This condition is essential to eliminate a clear problem of reverse causality: longer shifts have mechanically a higher probability of receiving at least one idiosyncratic shock. By setting the threshold $\tilde{h}$, all the shifts have an equal probability of receiving a large tip. In section 6, a sensitivity test will show the main coefficient's response to variation in these thresholds.

The type of taxi trips and the tipping behaviors vary a lot throughout the year. To understand some of the patterns at play, figure 1 plots the daily average of summary statistics. Each column represent a week, starting from the top left. Many interesting patterns emerge. First, not surprisingly, taxi trips made on a Saturday are shorter than the ones made on other days. The low average tip is mainly due to lower average fares on that day. We can also see some major holiday. 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 it clear that we need to control for the day of the week and the month of the year in order to eliminate expected tipping

[9]In the case where the independent variable is trip to the airport, T takes the value of 1 if at least a taxi trip from Manhattan to the airport was recorded during the shift



(a) Average percentage tip when tip is recorded to be at least $0.01



(b) Average tip amount when tip is recorded to be at least $0.01

Figure 1: Daily average summary statistics of tipping behaviors in 2013

patterns that driver could respond to.

Controlling for observable aggregate differences in the tipping behavior should be enough to ensure that the variation of interest is exogenous. Because of how unlikely very large tips are, it seems very improbable that taxicab drivers in New York City base their working behavior around them differently in days they expect to work less. In other words, the only plausible channel by which they 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 causation will bias our estimates. Indeed, the probability of receiving a large tip increases with the length of a shift.

The measurement error introduced by the first two problems will most likely be random. This 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 will be 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[10] 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 will be essential 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 construction, 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, this has the implication that the driver should increase his consumption of contemporaneous leisure as well as future leisure. Taking equation 3, we can simply look at the effect an idiosyncratic shocks has on his future labor supply decision. $T_{i,s-1}$ can be used instead of $T_{i,s}$. 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 $\tilde{h}$ hour 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 B3 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 equation 2). Compared with the estimates of Farber (2015), we observe a qualitatively similar labor supply elasticity (wage coefficient). This positive estimate is indicative of a neoclassical response to a

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

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

	(1)	(2)	(3)	(4)	(5)
Log hourly earnings	0.653***	0.371***	0.515***	0.665***	0.665***
	(0.00975)	(0.00674)	(0.00409)	(0.00618)	(0.00618)
Large tip		-0.0578***	-0.0676***	-0.0808***	
		(0.00380)	(0.00320)	(0.00328)	
Large tip (t-1)					-0.000907
					(0.00276)
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	6626417

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors at the driver-level in parentheses.
The dependent variable is the duration of a shift in hours. The "Log hourly earnings" regressor is computed as 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; dh: hour of the week

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 his shift length 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 tip indicator variable. This explanatory variable is the main addition to the original framework. We are interested in the sign of the coefficient. A coefficient close to zero would support the neoclassical framework. Indeed, because the idiosyncratic shock should not modify the expected future wage, only the income effect should impact the labor supply decision. Because I look at extremely small income shocks relative to life-cycle earnings, the coefficient should be indistinguishable from zero. 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[11]. Finally, the coefficient could be negative. This would be an evidence for a behavioral effect

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

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 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[12]. 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 figure 1, I add Christmas Eve and New Years Eve to set of major holidays used in Farber (2015)[13]. 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.

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. This seems to confirm that a psychological factor must be present to explain the large negative elasticity.

[12] 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.

[13] 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.

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. To roughly translate the preferred specification's coefficient (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[14]. Thus, a simple approximation suggests an elasticity of around -0.47. Being an upper bound, a negative elasticity of this magnitude is certainly of economic importance.

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[15]. 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 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

[14] 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 be even lower.

[15] 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)

set of trips from Manhattan to the JFK airport[16]. A trip between Manhattan and JFK 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.

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

	(1)	(2)	(3)
Log hourly earnings	0.448***	0.665***	0.654***
	(0.00677)	(0.00618)	(0.00617)
Large tip	-0.0563***		-0.0761***
	(0.00283)		(0.00319)
JFK airport		-0.0342***	
		(0.000538)	
driver FE	yes	yes	yes
date/time FE[a]	m/d*h	m/d/h	m/d/h
holidays	yes	yes	yes
weather	yes	yes	yes
neighborhood[b]	no	no	yes
obs	6663665	6663665	6648579

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors at the driver-level in parentheses.
The dependent variable is the duration, in hours, of a shift. The "Log hourly earnings" regressor is computed as 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. Similarly "JFK airport" takes the value of 1 if the driver had a ride from Manhattan to the airport during his 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; dh: hour of the week.
[b] See figure B5 for a map of the geographical division used for the fixed effects.

Column 2 of table 2 presents the result of the modified regression model substituting large tips for trips to the JFK airport. The variable is constructed in the same way the large tip binary variable was: 1 if the driver had a trip from Manhattan to the 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

[16]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

response to a positive idiosyncratic shock to income.

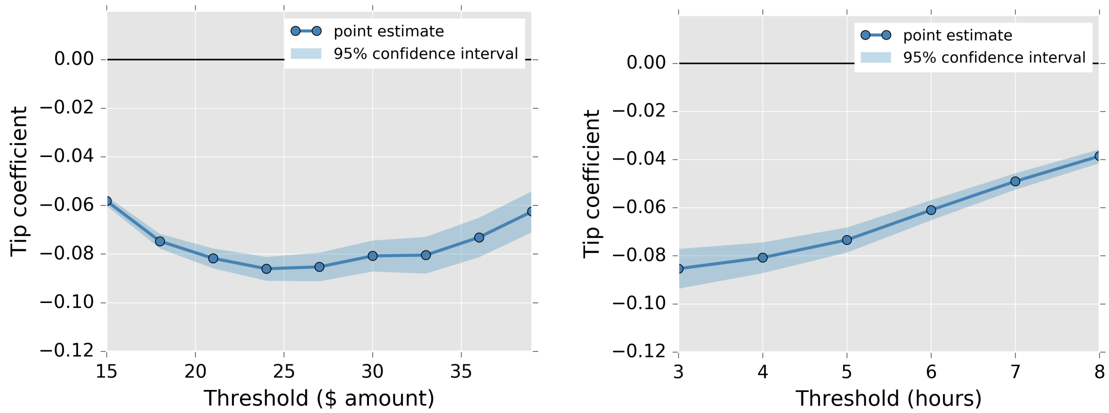
Another competing story as to why we would observe a relationship between the length of a shift and idiosyncratic earnings is if we did not control for neighborhood fixed effects. This can be a problem if: (a) drivers can somehow target neighborhoods to receive larger tips (or other idiosyncratic shocks) 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.[17] To get a sense of how finely the neighborhoods are defined and the coverage of the whole city, Figure B5 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.

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 of the threshold (figure 2a), 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 moving the hour threshold (figure 2b), 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.

I pursue the analysis with a discrete-choice stopping model. The model is heavily based

[17] 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.



(a) Coefficient's sensitivity when changing the minimum dollar amount of large tips

(b) Coefficient's sensitivity when changing the hour threshold of large tips

Figure 2: Daily average summary statistics of tipping behaviors in 2013

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 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:

$$\Pr(\text{Stopping}_{ist}) = \beta H^c_{ist} + \delta E^c_{ist} + \mu_i + \epsilon_{ist} \tag{5}$$

$$\Pr(\text{Stopping}_{ist}) = \beta H^c_{ist} + \delta \tilde{E}^c_{ist} + \gamma U^c_{ist} + \mu_i + \epsilon_{ist} \tag{6}$$

Terms in this equation are indexed by the driver (i), the shift (s), and the trip (t). Equation 5 replicates the framework of Farber (2005) while equation 6 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

Table 3: 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).

and an error term.

Due to the fact that an analysis at the trip level contains a lot more observations (About 20 times larger[18]), I used a random sample of 1000 drivers for computational ease[19]. I estimate equations 5 and 6 with both a linear probability model and a probit model.

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

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

[19] 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.

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 3.

The first and fourth column shows the estimated coefficients of the specification without the decomposition (equation 5). 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 equation 6. 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 3 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 3 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 a 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, Leah-Martin (2015) and Agarwal et al. (2015) find empirical anomalies that could 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 certain 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.

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 reoptimize 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, 2015). Combined with evidence on the relationship between cognitive ability and preference anomalies (Benjamin et al., 2013) and on learning by doing processes

(Haggag et al., 2014; Agarwal et al., 2016), 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).[20] 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

[20] 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.

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 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

The starting point of this paper was to test the hypothesis that taxi drivers have a different labor supply reaction depending on the type of income they receive. Abstracting from the expected difference in substitution and income effects, could one type of income shock generate patterns consistent with the neoclassical model while another type of shock lead to behaviors that support the presence of psychological factors? Focusing on unanticipated earnings, I show that the response to unanticipated idiosyncratic shocks (e.g. receiving a large tips) is different than the response to unanticipated market-level shocks (e.g. a subway closure).

This paper provides evidence that taxi drivers' labor supply reaction to idiosyncratic income shocks supports the presence of psychological factors incompatible with the standard model of labor supply. However, in contrast to Camerer et al. (1997), the labor supply elasticity is not negative for all types of unanticipated variation. When the income shocks are generalized at the market-level, I find a positive labor supply elasticity that the standard neoclassical model is able to explain. The evidence also suggests that previous studies of taxi drivers labor supply had a misspecified empirical model in which one type of unanticipated income variation was masked and diluted.

The simple empirical result 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 preference, I discussed the potential for present bias and narrow bracketing to explain the results. Further research should focus on disentangling these three 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, 2008, 2015) presented 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 variations in income, 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.

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

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

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

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.

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.

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.

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.

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 (2014): "Learning by Driving: Productivity Improvements by New York City Taxi Drivers," Working paper.

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

Leah-Martin, V. (2015): "How Much To Work: Evidence of Reference Dependence in Taxi Drivers," 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. (2015): "Income Targeting and the Ridesharing Market," *Mimeo*, 1–37.

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.

Appendices

A Creation of the average wage

The instrument used for the hourly earnings of a shift (W_s) is the average hourly earnings of all drivers ($\overline{W}_s$) during the shift.

For a given hour k:

1. Compute the total earnings received by all drivers during hour k.
2. Next, compute the number of drivers working during hour k. If a shift started in the middle of hour k, add the fraction of time worked.
3. Dividing the result of step 1 by the result of step 2 gives the hourly average earnings per driver during hour k, ($\hat{W}_k$).

Repeat this process for every hour of the year.

To compute the average market wage during a shift, take the average of $\hat{W}$ for every hour overlapping with the shift.

The cumulative earnings at the market wage is simply the product of the cumulative hours and the average market wage between the start of the shift and the current trip.

B Figures and tables

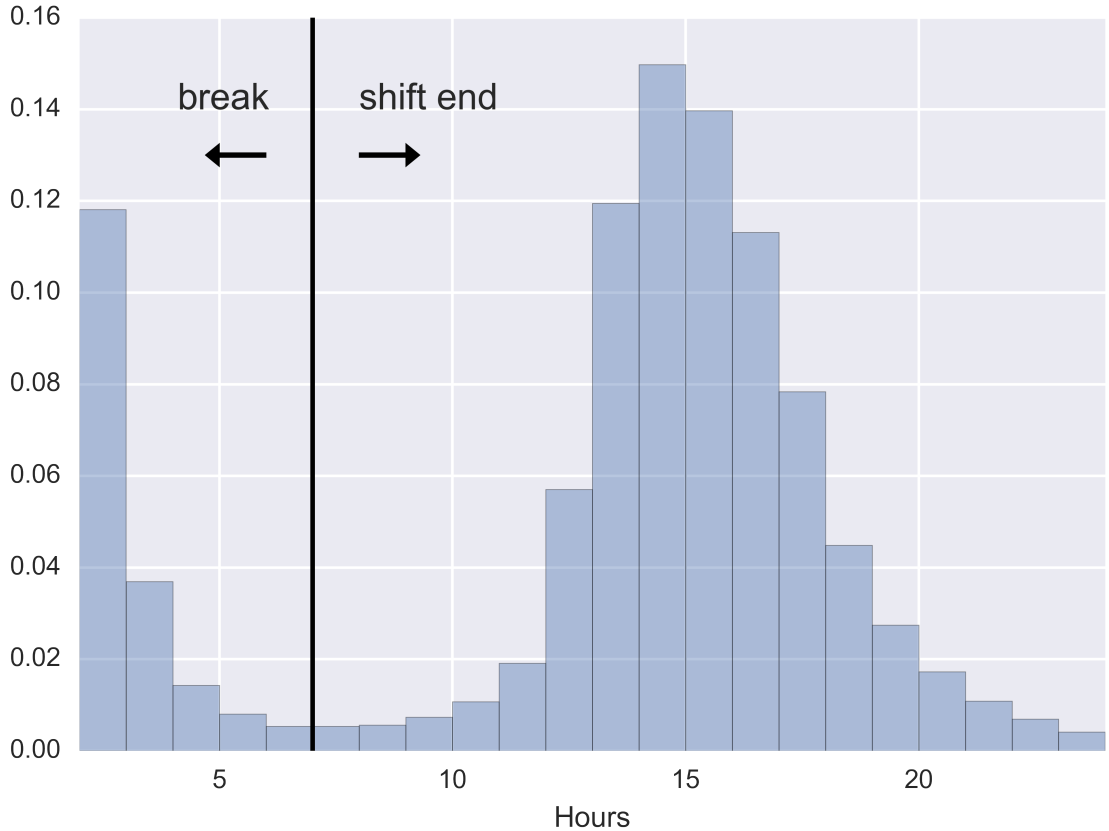


Figure B1: Distribution of the waiting time between two trips, conditional on being more than two hours and less than 24 hours

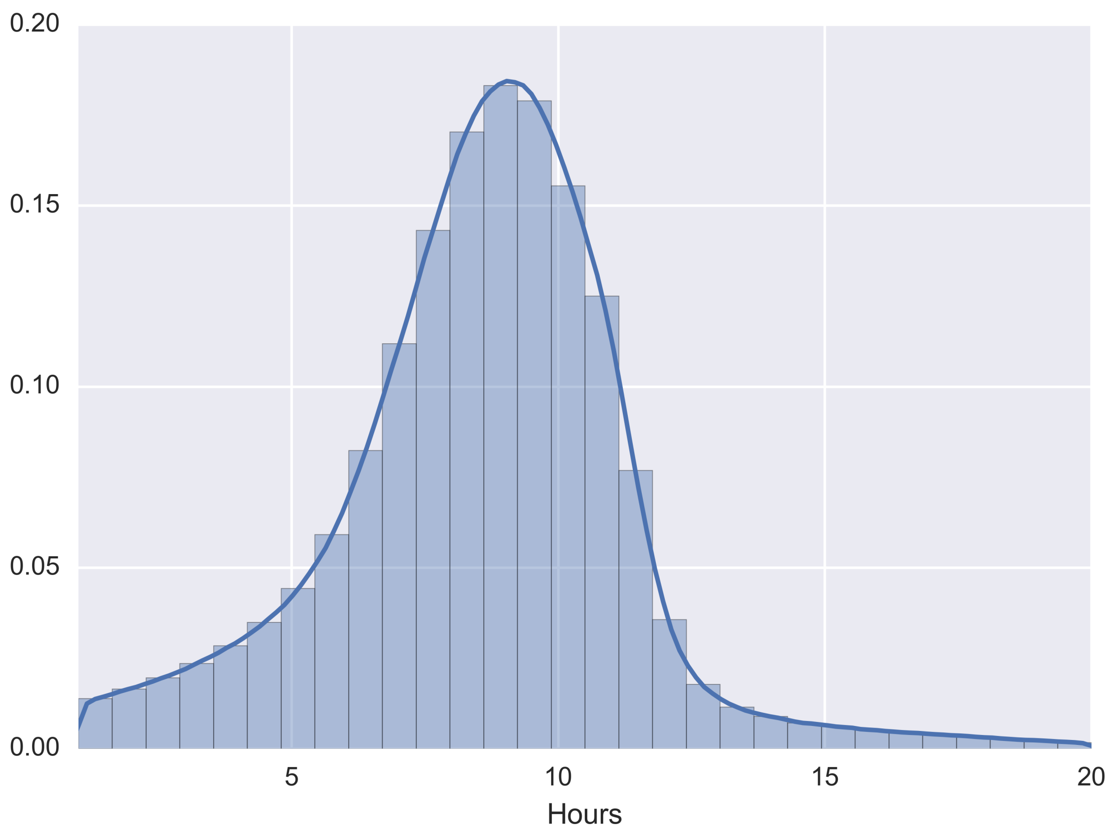


Figure B2: Distribution of the shift duration

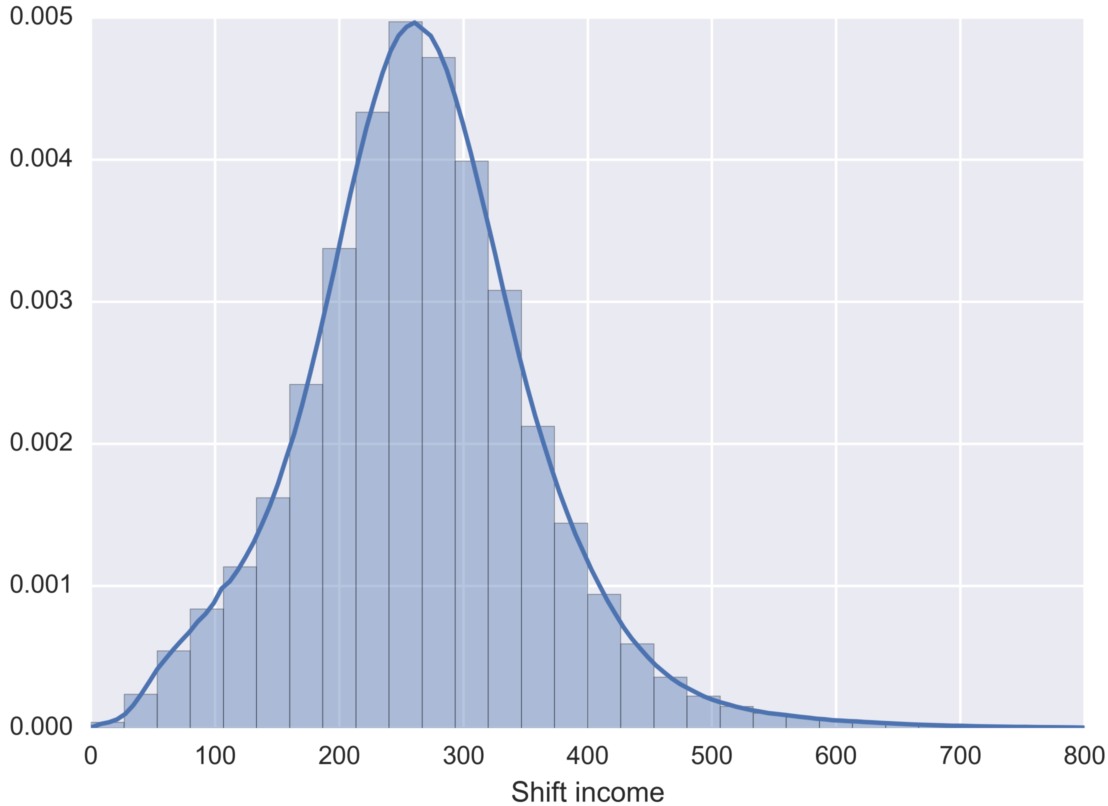


Figure B3: Distribution of earnings during a shift

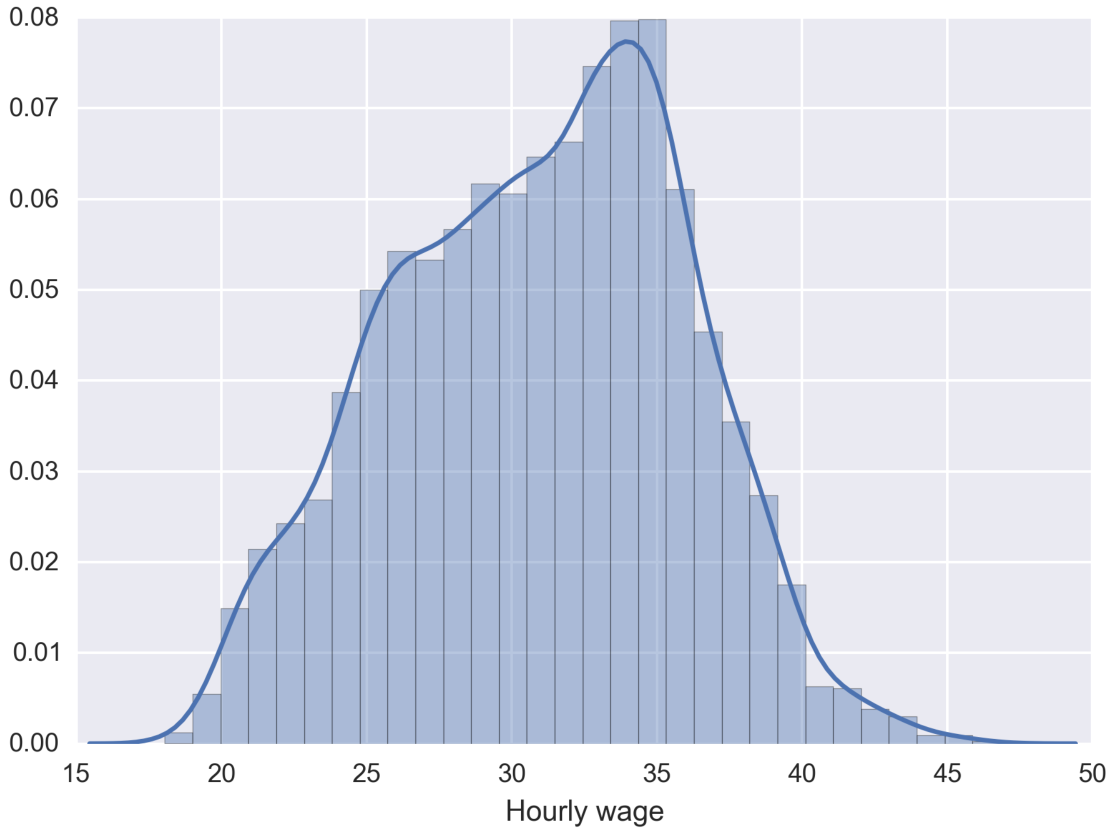


Figure B4: Histogram of the hourly average wage

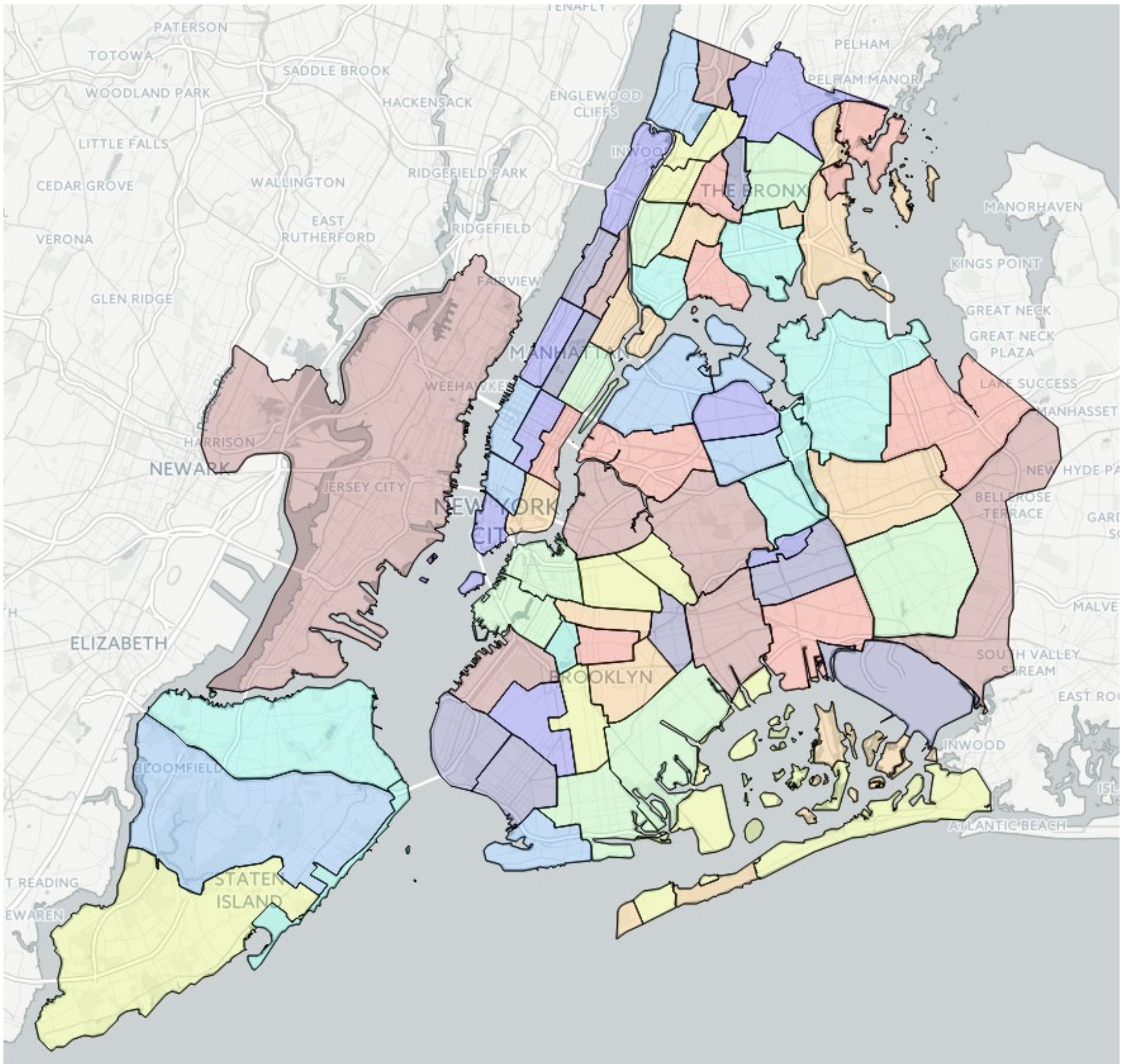


Figure B5: The different neighborhoods used to classify the geospatial fixed effects

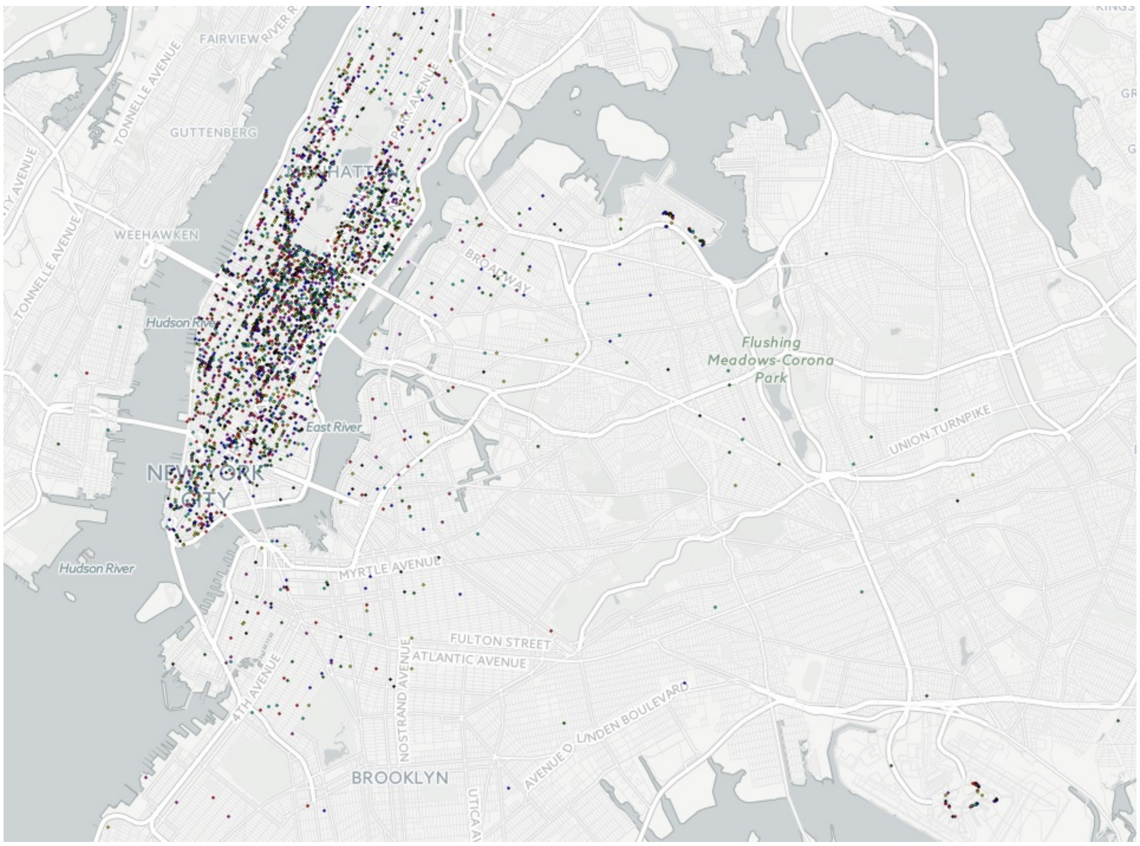


Figure B6: Geographical dispersion of a sample of 3000 pickup locations