Identifying Behavioral Responses in Labor Supply:
Revisiting the New York Taxi Industry

Marc-Antoine Schmidt*

Draft version: October 21, 2016

Abstract

This paper presents empirical evidence that the labor supply elasticity of taxi drivers can be negative in response to temporary positive earnings shocks. The observed pattern is inconsistent with the neoclassical life-cycle model of labor supply and suggests some form of behavioral preferences. To get this result, I decompose unexpected earnings variations into a market wage component and an idiosyncratic component. This differs from previous studies that assume a homogeneous labor supply effect of unexpected earnings shocks. I identify abnormally large tips from the universe of New York medallion taxi trips in 2013 and use them as a source of exogenous variation of the idiosyncratic component. I find that the negative labor supply elasticity is only 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 market wage shock causes their labor supply to increase, consistent with an optimizing rational agent. While the empirical findings of this paper rejects the neoclassical model of labor supply, they cannot distinguish between competing behavioral explanations. Therefore, three broad classes of behavioral models 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, and the CEPA seminar at the University of Toronto for their helpful comments.

1 Introduction

One of the main goals of labor economics is to understand how agents make labor supply decisions. Most importantly, how does an economic agents labor supply respond to variations in wage? New econometric methods and datasets have generated a large amount of research estimating elasticities of labor supply. Some of this research provides conflicting estimates, not only in the magnitude of elasticities, but also in their sign. While the standard view is that the sign depends on the magnitude of the income and the substitution effects, recent findings suggests that the reality is much more complex and often context-dependent. For instance, when we relax the hypothesis that any type of income is fungible with one another, patterns in the data seems to indicate that tips and regular wage have different impact on the labor supply decision of a worker. This paper provides some evidence that may reconcile opposing views in this literature.

The benchmark framework is generally a neoclassical model of labor supply where the sign of the effect of a wage change is ambiguous. This ambiguity depends on the relative magnitude of the substitution and income effects. When wage variation is temporary and small relative to total earnings, the income effect will be negligible, and so the substitution effect will dominate. As a workers wage increases, the opportunity cost of leisure goes up, resulting in a positive labor supply response. This framework is certainly useful to answer many questions but might fail to correctly explain interesting behaviors, especially when we start to disaggregate the data.

In the context of this temporary and small wage variation, some economists have argued that the labor supply elasticity might instead be negative. Advocates of reference-dependent models suggest that an agent might work less in response to a temporary wage increase. An individual with reference-dependent preferences responds to income, 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. This reference point makes the marginal benefit of working discontinuous, which creates incentives to bunch at the target. One possible consequence of this type of reference-dependent preference is that 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 empirical work of Camerer et al. (1997), controlled and natural experiments have generated results that seem to support the presence of reference-dependent preferences.

More recently, Farber (2015) provides evidence that disputes these earlier findings. Using a very large dataset of taxi drivers labor decisions, he is unable to replicate the negative labor supply elasticity found in earlier papers, finding instead an elasticity consistent with the neoclassical model. While studies using controlled and field experiments continue to find patterns indicative of reference-dependence preferences, Farber (2015) casts a shadow on the external validity of those results in the real world.

Farber (2015)s result is based on the assumption that any unexpected earnings shocks will have the same impact on every driver. However, it is easy to imagine shocks that might affect a single driver that would not affect all of the drivers working at one particular time. In this paper, I will explore whether drivers react differently to unexpected shocks that affect a large portion of the market (e.g. a subway closure) and unexpected individual shocks (e.g. getting a large tip because a customer was in a good mood). This distinction is important in order to reconcile the literatures conflicting results. By comparing the predictions of a neoclassical model to estimated labor supply elasticities from both types of shock, I will be able to test whether the neoclassical model explains taxi drivers behavior.

The following 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, no one has used the tips to identify clear idiosyncratic shocks to hourly earnings. The main empirical contribution of this paper will be to show that a driver's labor supply responds negatively to idiosyncratic income shocks, a behavior that cannot be explained by the neoclassical model. In other words, when a taxi driver has a "lucky day" (positive idiosyncratic shock), he tends to reduce its hours worked significantly.

However, contrary to previous studies that assumed any deviation from the neoclassical model was caused by some sort of reference-dependent preferences, I will 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 shocks are able to affect a driver's target will fit 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 recurent shocks. Psychologically, 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 without this bias towards the present.

The remaining of the paper is organized as follows: Section 2 provides a basic review of the literature. Section 3 describes the empirical strategy used to test the neoclassical model. In section 4, I present important characteristics of the dataset and the aggregation methodology. 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 Previous literature

2.1 Reference-dependence and labor supply elasticity

The task of estimating a labor supply elasticity goes back many decades ago. The first credible estimates were done in the 1980's by, for instance, MaCurdy (1981), Browning et al. (1985), or Altonji (1986). However, as noted in later studies, interpreting the results of these studies is challenging. First, wage changes are rarely transitory and dissimilar, causing the income effect to interfere with the substitution effect. Furthermore, many other endogenous factors play a role in both wages and the labor supply decision. The relationship between wages and hours uncovered by these studies is, therefore, less causal than we would like.

At the same time, with the increasing popularity of Kahneman and Tversky (1979)'s prospect theory in many economic fields, it was a matter of time before someone incorporated the idea of loss aversion to labor supply. The first model that clearly incorporated reference-dependent preferences, (i.e. loss-aversion with respect to a target) was developed by Tversky and Kahneman (1991). Over the years, many researchers have tried to empirically demonstrate real-world occurrences of reference-dependence. However, it proved to be a very difficult task in part because the reference-point of an agent is never directly observed in the data.

In general, one of the biggest problems when studying labor supply decisions is that, in developed economies, most industries and occupations have constraints that make the hours-worked decision trivial or its estimation impossible. Fixed hours is probably the most obvious of these constraints. There are two ways of getting around this problem. The first solution is to create a controlled experiment where the researchers define a task that participants must complete to be rewarded. Here, the researcher tries to mimic the real-world labor market at a smaller scale. Although these studies can be informative on the existence of such behaviors, their external validity is often weak. For a good example, see Abeler et al. (2011).

Another solution that economists use is to study worker decisions in industries where the constraints are less restrictive. 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.[1] These studies provided valuable empirical and methodological insights into the study of reference-dependence. However, for multiple reasons, it is the taxi industry that provided the most utilized datasets in the reference-dependent literature.

2.2 Labor supply of taxi drivers

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 duration on the wage rate. The authors, recognizing that this approach would be plagued by a division bias (See Borjas, 1980), instrumented the actual wage by the average wage of the other drivers. The 2SLS estimate of the elasticity, with driver fixed effects, was negative and large (close to -1). This finding stood in stark contrast to the prediction of the neoclassical model. Their results were confirmed 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 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. The main critique came from Farber (2005). Arguing that the econometrics 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.[2]

Using Farber (2005)'s data, Crawford and Meng (2011) presented an empirical test of reference-dependence based on the theoretical work of Kszegi and Rabin (2006). Similar to

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

[2] 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, while Doran (2014) showed that the labor supply elasticity depends on the duration of the wage variation. See also Agarwal et al. (2015).

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. Kszegi and Rabin (2006) were the first to notice that a driver's reaction to unanticipated wage variation could be very different than to anticipated wage variation.

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. 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. Farber estimates the labor supply response of both anticipated and *unanticipated* variation in wage. Instead of decomposing the wage into those two components, he argues that date and time dummies capture the anticipated portion of wage variations. The wage coefficient would thus represent the labor supply elasticity with respect to only unanticipated wage variations. 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, Kszegi 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.[3] The study concludes that reference-dependence is not of first-order importance 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 (2015a) replicated 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. All in all, it seems that the literature did not settle on a clear and exhaustive empirical test of reference-dependence.

3 Empirical strategy

Previous tests of reference-dependence implicitely made the assumption that income sources are fungible for taxi drivers. Consequently, if expectations about future earnings stay con-

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

stant, the labor supply elasticity should be the same no matter what the source of income variation that identifies it.

Taxi drivers face multiple sources of income variation. This allows us to decompose earnings in an intuitive way. Firstly, 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. Similar to previous studies, the estimating equation will ideally control perfectly expected variation in wage. This can be done, albeit imperfectly, by adding time and date fixed effects. The fixed effects will capture any reccuring 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, we are left with the arguably more interesting 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 mecanical division bias that arises when estimating a labor supply elasticity with unprecise 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 a shift, in hours, be H, and the total earnings during a shift be E. The benchmark estimating equation, which is the same as the main estimating equating in Camerer et al. (1997) and Farber (2015), is:

$$\log(H_{is}) = \delta \log\left(\frac{\widehat{E_{is}}}{H_{is}}\right) + \beta X_{is} + \mu_i + \nu_{is} \tag{1}$$

$$\log\left(\frac{\widehat{E_{is}}}{H_{is}}\right) = \log\left(\frac{\sum_{k\in A_s} E_k}{\sum_{k\in A_s} H_k}\right) + \epsilon_{is} \tag{2}$$

The variables are indexed by the driver (i) and the shift (s). Shift specific controls are included in X_{is} and contains the time and date fixed effects. μ_i is a driver fixed effect. Thus, the identifying variation comes from within driver. This will become even more relevant when the idiosyncratic shocks will be added. In the first stage (equation 2), 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[4]. Of course, if we assume that the impact of idiosyncratic shocks and market-level shocks on a shift's duration is the same, then the coefficient δ will represent the labor supply elasticity.

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. Even though 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 length duration. Finding a customer that will give you a large tip or needs a ride to the airport will be the proxy for idiosyncratic shock.

Adding this to the benchmark equation is easy. The estimating equation becomes:

$$\log(H_{is}) = \delta \log\left(\frac{E_{is}}{H_{is}}\right) + \gamma T_{is} + \beta X_{is} + \mu_i + \nu_{is} \tag{3}$$

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 x during the shift s[5]. Furthermore, the tip needs to be received in the first $\tilde{h}$ hours of a shift. The reason why the second condition is needed will become clear later. In section 6, a sensitivity test will show the main coefficient's response to variation in these thresholds.

While the idea of using large tips seems very simple, three issues require our attention. The first one, already discussed in section 4, 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. Finally, the fact that there will be reverse causation is the third issue. 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

[4] 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.

[5] 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

case for reference-dependence harder. Further measurement error will be introduced to solve the reverse causality issue. Indeed, the rule that the tip needs to be received in the first four hours of the shift equalizes the probability of receiving a large tip across observations[6].

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.

4 Data

4.1 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.

[6]Shifts with a duration of fewer than four hours are also dropped.

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 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 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

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

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 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 the outlier data.

4.2 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[8]. 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 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.

[8]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.

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_{st}^{c}), cumulative earnings (E_{st}^{c}), cumulative earnings at the market wage ($\tilde{E}_{st}^{c}$), and cumulative earnings shocks (U_{st}^{c}). 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.

5 Results

Two decisions need to be made before obtaining the regression results. First, a time threshold ($\tilde{h}$) must be decided. I have chosen 4 hours as it seems to keep a large proportion of shifts while giving enough time to have a chance of receiving a large tip. Second, the definition of a large tip requires a threshold in the amount of money (x). I have decided to use 30$. 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).

Results of the regressions are presented in table 1. The first column shows the coefficient of a regression that replicates the specification of Camerer et al. (1997) and Farber (2015) (see equation 1). The second column shows the regression of the log of the shift duration on the log of the wage and the tip indicator without any controls. The third column adds driver fixed effects and date/time controls. These additional controls include hour of the day, day of the week, and month dummies.

Compared with the estimates in Farber (2015), we observe a slightly lower, but qualitatively similar, labor supply elasticity (wage coefficient) in the four specifications. The difference, even in the first specification, is due to the different sample and a slightly different instrument. This positive estimate is indicative 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 his shift length by 3 to 5 percent. As for the strength of the instrument, the massive number of observation produces a very large F statistics higher than 1000.

The second row contains the estimated coefficient of the tip indicator variable. It represents the reaction of drivers to at least one large tip. The estimate of column 3 implies that a large tip, as defined earlier, would reduce the labor supply of the driver by approximatively 5 percent. The coefficients do not precisely represent how a driver would react to any idiosyncratic income shock but can be seen as a lower bound to the reaction's magnitude. Even with a clear bias towards zero, we still observe a significant negative response to receiving an

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

	(1)	(2)	(3)	(4)
log Wage	0.391***	0.400***	0.392***	0.500***
	(0.00495)	(0.00600)	(0.00495)	(0.00395)
Tip indicator		-0.0614***	-0.0515***	
		(0.00385)	(0.00276)	
Tip indicator $(t-1)$				0.000796
				(0.00244)
date/time dummies	yes	no	yes	yes
driver fixed effects	yes	no	yes	yes
obs	6,655,551	6,656,146	6,655,551	6,655,551

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors in parentheses.
The dependent variable is the duration, in hours, of a shift. The "log hourly earnings" regressor is computed as total fare (excluding tips) divided by hours worked. It is instrumented with the average hourly wage of all drivers during the shift (see the appendix for the methodology). "Tip indicator" 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. 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). Column (1) replicates the specification of Camerer et al. (1997) and Farber (2015). Observations with a shift duration of more than 20 hours were dropped.

idiosyncratic increase in hourly earnings.

To test whether the negative elasticity comes from a very large income effect, the fourth column shows the effect of an idiosyncratic shock during the last shift on the duration of the contemporaneous shift. The coefficient is not statistically different from zero, while maintaining a similar standard deviation. Compared with the same coefficient of column 3, it seems that the strong effect present during the same shift vanished when looking at future shifts. This result cannot be explained by the neoclassical life-cycle model of labor supply, which would imply some smoothing of the impact of this shock over a period of time.

6 Robustness Checks

The negative labor supply elasticity with respect to large tips found in table 1 supports the reference-dependent model. However, it does not provide enough evidence to show that the relationship is not caused by something else. In this section, I provide multiple tests to show that the neoclassical model of labor supply cannot explain this behavior.

It could also be that the negative relationship between large tips and hours worked is caused by an omitted variable. 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 an idiosyncratic shock is trips from Manhattan to the JFK airport. Although this variable is arguably less exogenous than the tip, the underlying idea is the same. 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. To a driver, it is clear this represents a positive idiosyncratic shock.

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

	(1)	(2)	(3)
log hourly earnings	0.392***	0.391***	0.383***
	(0.00495)	(0.00495)	(0.00492)
tip indicator	-0.0515***		-0.0491***
	(0.00276)		(0.00272)
JFK indicator		-0.0294***	
		(0.000543)	
date/time dummies	yes	yes	yes
driver fixed effects	yes	yes	yes
neighborhood fixed effects	no	no	yes
obs	6,655,551	6,655,551	6,655,551

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors in parentheses.
The dependent variable is the duration, in hours, of a shift. The "Wage" regressor is computed as total fare (excluding tips) divided by hours worked. It is instrumented with the average hourly wage of all drivers during the shift (see the appendix for the methodology). "Tip indicator" 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. 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). See figure B5 for a map of the geographical division.

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 its hours worked in response to an 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. Of course, both (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 intra-driver. Nevertheless, 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.[9] 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 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 1 shows the coefficient at different threshold levels. As I vary the thresholds, the coefficients remain negative and significantly different from zero, suggesting that our results were not driven by the choice of threshold.

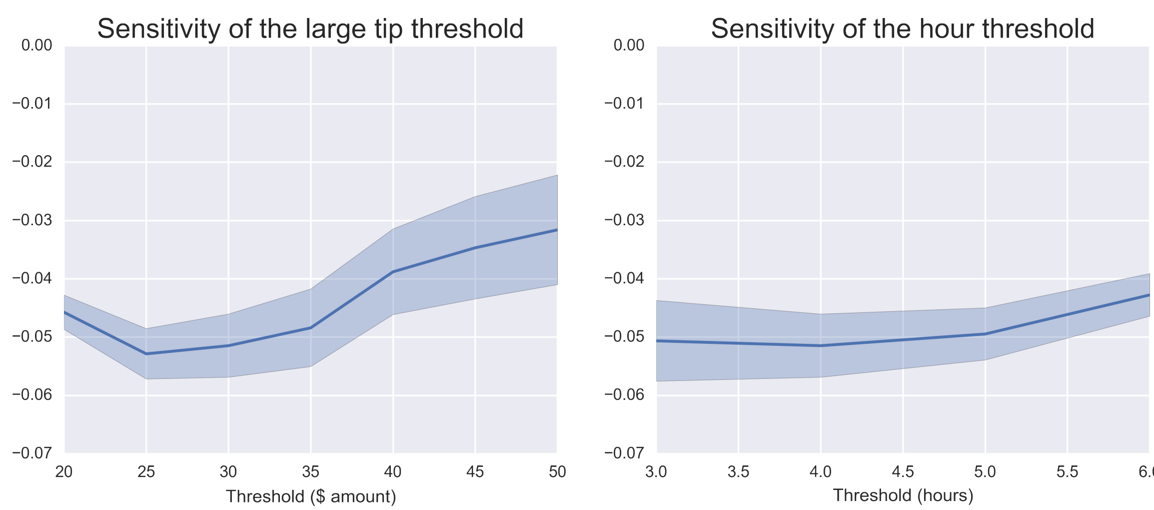

Figure 1: Sensitivity analysis of the coefficient

[9] 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.

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}). If the model presented earlier is correct, and drivers do have reference-dependent preferences, we should see that positive idiosyncratic shocks increase the probability of stopping.

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 are 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:

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

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

Terms in this equation are indexed by the driver (i), the shift (s), and the trip (t). Equation 4 replicates the framework of Farber (2005). Equation 5 decomposes the cumulative earnings. 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 observation (more than 170 million compared to around 8 million shifts), I used a random sample of 1500 drivers[10]. Furthermore, similar to Leah-Martin (2015a), I use a linear probability model instead of a probit for two reasons. First, the linear probability model is computationally easier. Second, it has been argued that non-linear models do not behave normally in the presence of fixed effects (Greene, 2004).

[10] These 1500 drivers represents around 3.5% of the total sample. I ran the same analysis on different samples and got qualitatively the same results.

Table 3: Discrete choice stopping model

	(1)	(2)	(3)	(4)
H^c_{ist}	0.0169***	0.0419***	0.0541***	
	(0.0000872)	(0.000174)	(0.000225)	
E^c_{ist} (100$)	0.00361***			
	(0.000276)			
$\tilde{E}^c_{ist}$ (100$)		-0.0803***	-0.113***	-0.107***
		(0.000556)	(0.000707)	(0.000691)
U^c_{ist} (100$)		0.0175***	0.0238***	0.0232***
		(0.000272)	(0.000298)	(0.000297)
Time dummies	yes	no	yes	yes
Driver fe	yes	no	yes	yes
Cumul. time dummies	no	no	no	yes
obs	6041915	6041915	6041915	6041915

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Clustered standard errors in parentheses. The dependent variable is a binary variable taking the value of 1 if the trip was the last trip of the shift. H^c_{ist} is defined as the cumulative duration of the shift, in hours, up to the end of the trip t. E^c_{ist} is similarly defined as the cumulative earnings (excluding tips) up to the end of the trip. $\tilde{E}^c_{ist}$ is defined as the cumulative income a driver would have made if he had consistently earned the average wage in the market. U^c_{ist} is simply the difference between actual cumulative earnings and market cumulative earnings. 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). Column (1) replicates the specification of Farber (2005) while column (2)-(3) show the coefficient of the same regression with the earnings decomposed into cumulative earnings at the market wage and an idiosyncratic shock components. Computed using a random sample of 1500 drivers.

Table 3 presents the result of the discrete choice stopping model. The first column shows the result without the decomposition (equation 4). We can see that even if the coefficient is positive and significant, it is very close to zero. This supports the view of Farber (2005, 2015) that reference-dependence is not a large factor in the labor supply decisions of taxi drivers. The coefficient indicates that each 100$ of income during a shift would increase the probability of stopping by 0.4 percent.

The second and third column show the regression coefficient when we estimate equation 5. As we can see, this simple decomposition allows us to see that using the wage as a single measure does, in fact, give us an average. In other words, the reference-dependence is masked through 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 11 percent.

The coefficient on the idiosyncratic income is also significant but positive. Increasing a driver's idiosyncratic income should raise his probability of stopping by around 2 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.

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 dummies. This approach is less parametric than the previous one. Column 4 of table 3 present the result. The coefficient is almost identical.

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 (2015b) 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 Kszegi 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).[11] 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 suffering from present bias will acknowledge all the options and consequences, but the strong preference for immediacy will dwarf the utility obtained from future periods.

[11] 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.

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

This paper provided evidence that taxi drivers labor supply reaction to idiosyncratic income shocks supports the presence of suboptimal psychological factors. Furthermore, the positive labor supply elasticity with respect to market wage shocks provide evidence that drivers modify their target in response to generalized demand shocks. The evidence suggests that previous studies of taxi drivers labor supply had a misspecified empirical model in which reference-dependent behaviors were masked and diluted. Indeed, combining idiosyncratic shocks, where reference-dependence seem to be active, and market wage shocks resulted in an insignificant average of the two effects.

The debate concerning the importance of understanding real-world occurrences of reference-dependence is far from over. The results proved that we do not yet understand perfectly what motivates a worker to make its labor supply decision. This paper provided a starting point for further research on the subject. For instance, combining the present findings to a framework that would allow for heterogeneity would be valuable. Furthermore, one could incorporate other targets, such as hours target, to increase its accuracy. Research concerning the labor supply of taxi drivers could be expanded in many other ways. While this paper dealt with the intra-day labor supply decisions (number of hours worked), future research could look at a longer time frames (weekly or monthly) to uncover new patterns. Moreover, with the precision of the dataset at hand, it could be possible to look at the intensive margin (effort) response. Taxi drivers could, instead of modifying their extensive labor supply decision (number of hours worked), decide to change their "strategy" to reduce their effort.

References

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011): "Reference Points and Effort Provision," *American Economic Review*, 101, 470–492.

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

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

Altonji, J. G. (1986): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, 94.

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.

Browning, M., A. Deaton, and M. Irish (1985): "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 53, 503–543.

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*.

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

Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–292.

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

Leah-Martin, V. (2015a): "How Much To Work: Evidence of Reference Dependence in Taxi Drivers," *Working Paper*, 1–38.

——— (2015b): "How Much To Work: Evidence of Reference Dependence in Taxi Drivers," *Working Paper*.

MaCurdy, T. E. (1981): "An empirical model of labor supply in a life cycle setting," *Journal of Political Economy*, 89, 1059–1085.

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

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.

Tversky, A. and D. Kahneman (1991): "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106, 1039–1061.

Appendices

A Creation of the average wage

To create 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

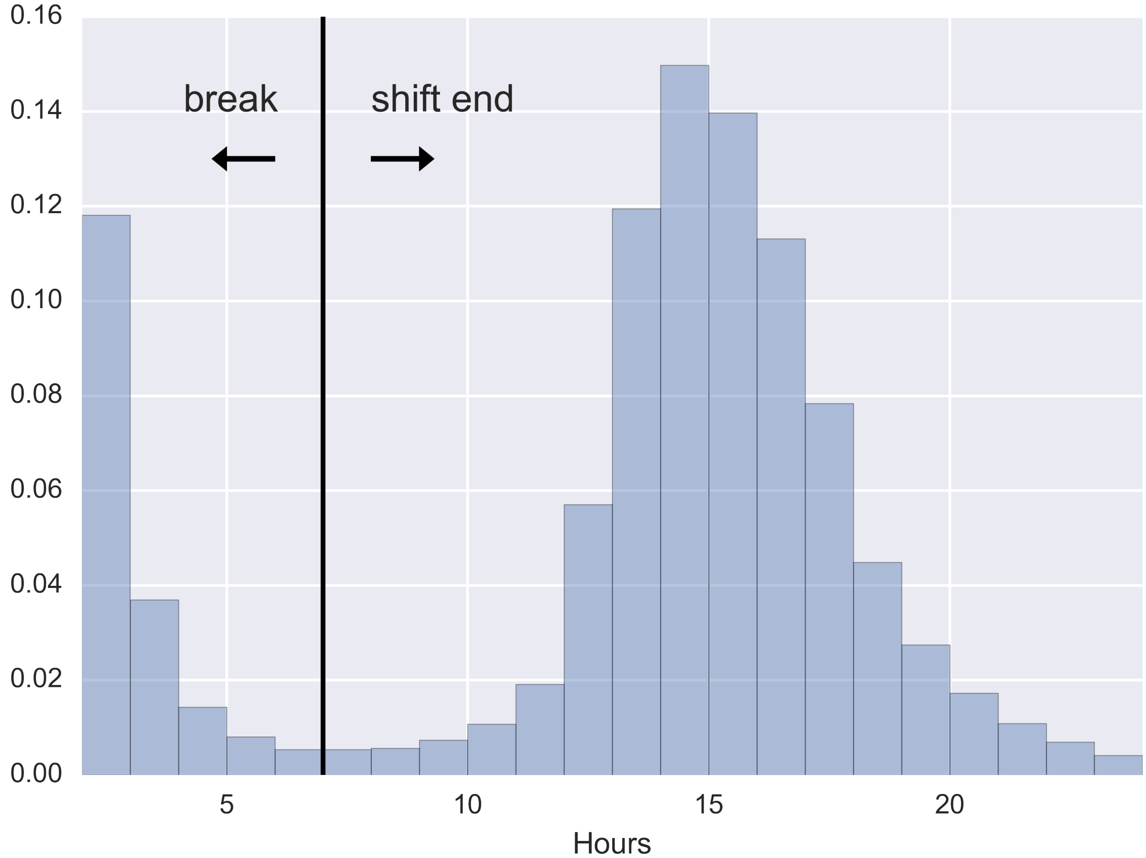

Figure B1: Distribution of the waiting time between two trips, conditional on being more than two hours and less than 24 hours

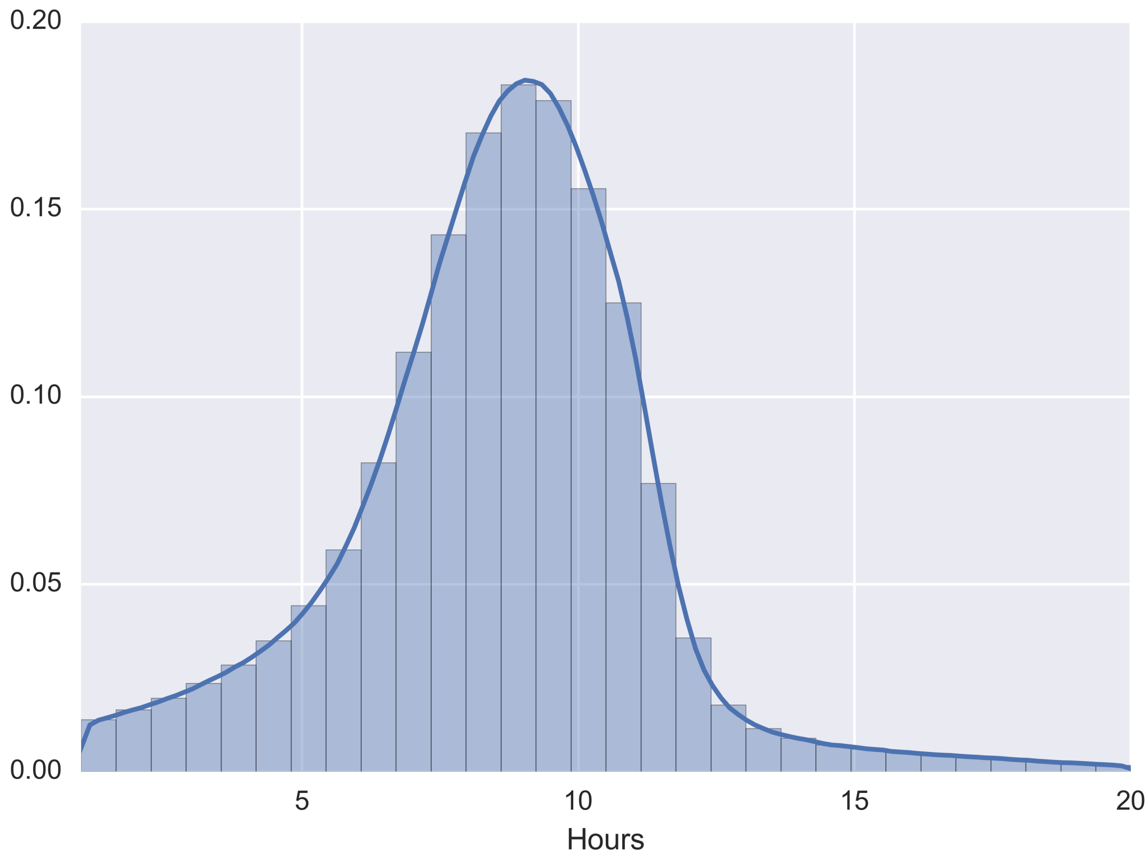

Figure B2: Distribution of the shift duration

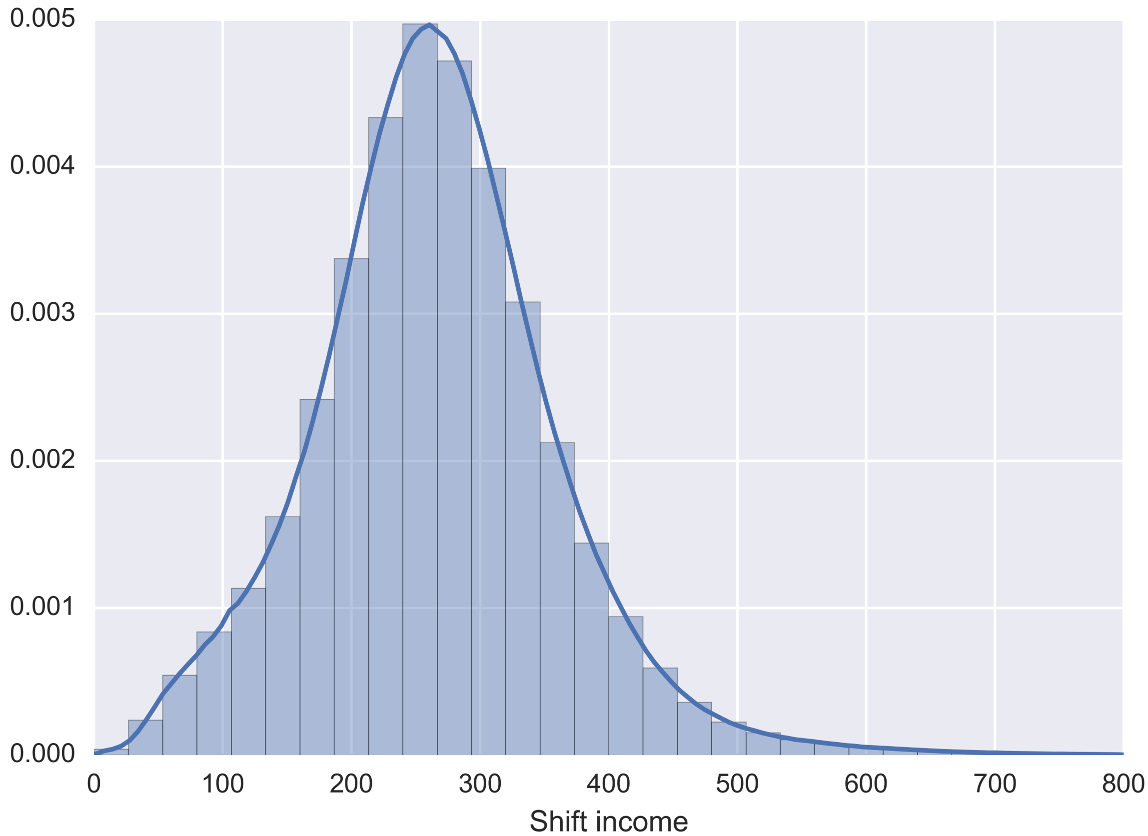

Figure B3: Distribution of earnings during a shift

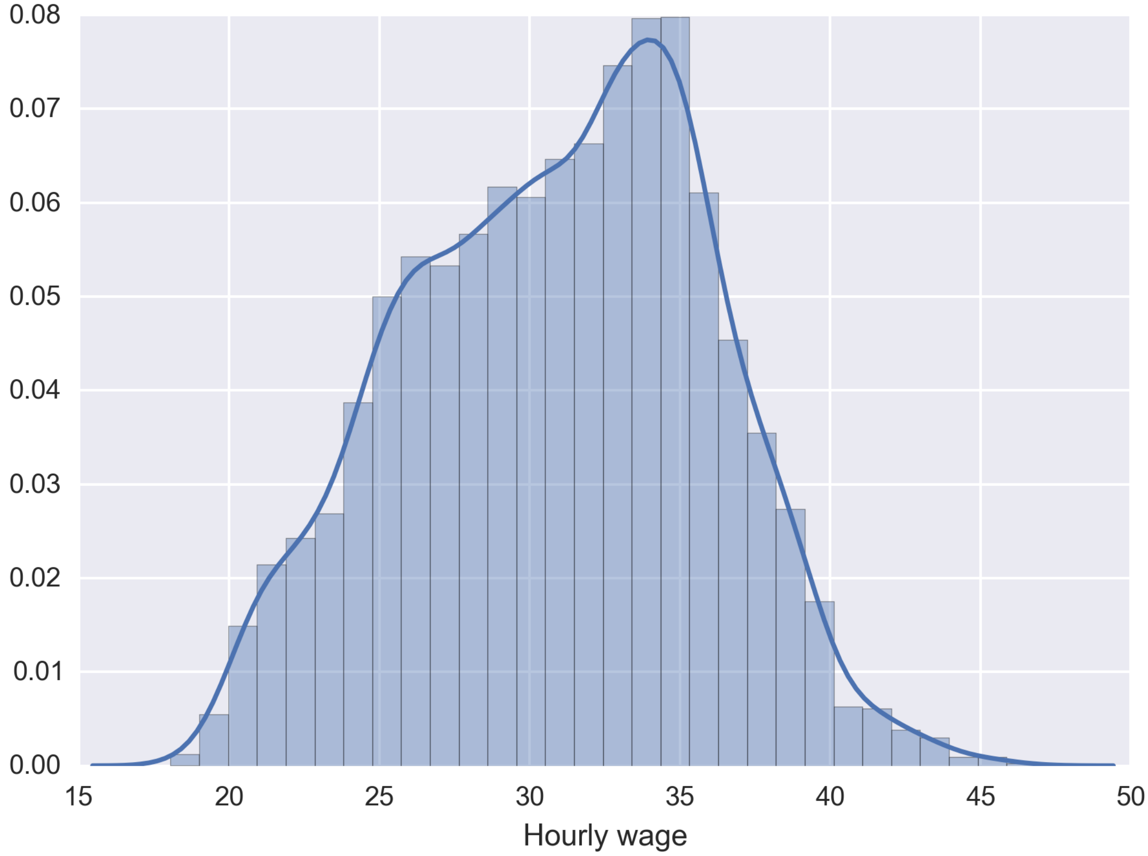

Figure B4: Histogram of the hourly average wage

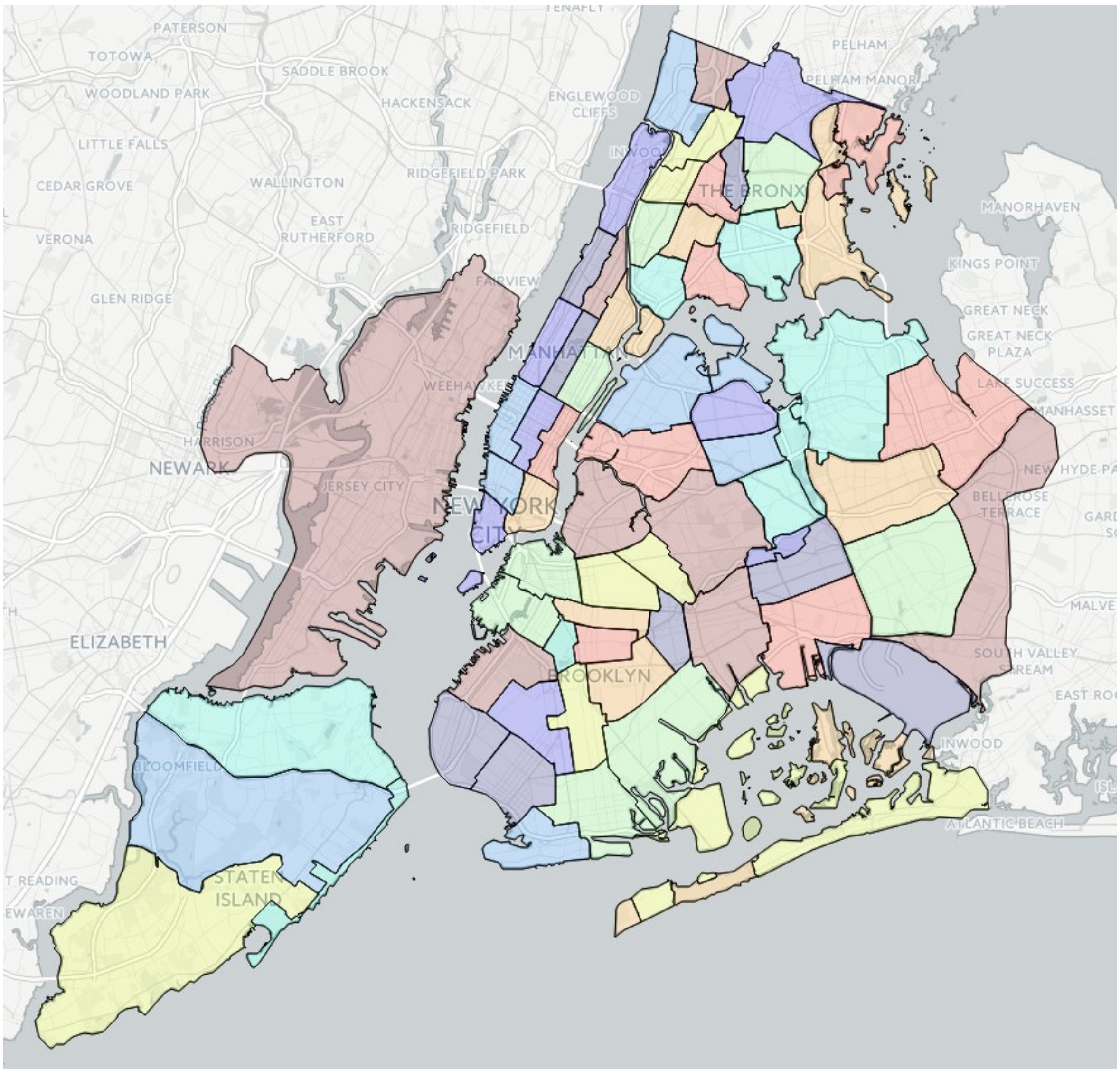

Figure B5: The different neighborhoods used to classify the geospatial fixed effects

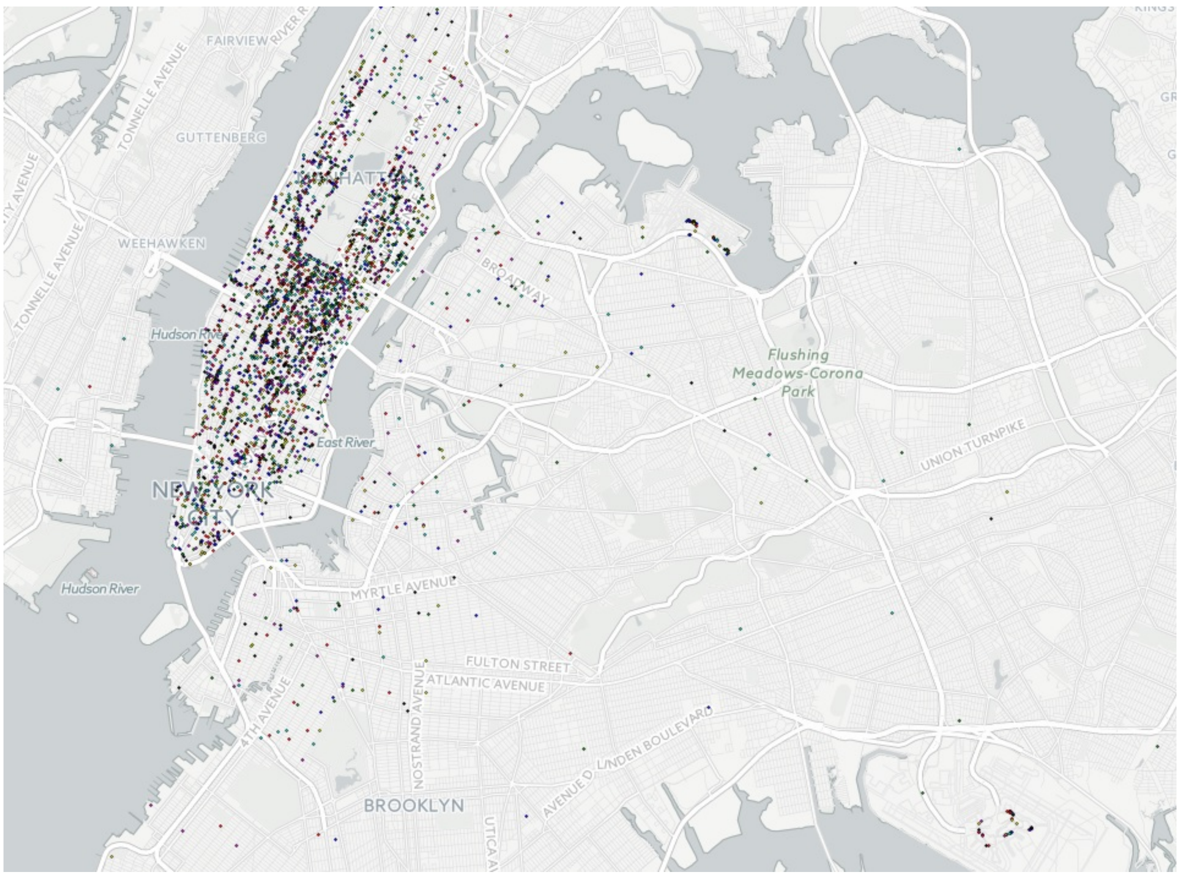

Figure B6: Geographical dispersion of a sample of 3000 pickup locations