



Trouble Underground: Demand Shocks and the Labor Supply Behavior of New York City Taxi Drivers

Alessandro Saia¹

Received: 8 December 2020 / Accepted: 7 June 2021 / Published online: 8 July 2021
© The Author(s) 2021

Abstract

We investigate the relevance of targeting behavior in the labor supply decisions of New York City Taxi drivers using exogenous and transitory positive changes in labor demand. Exploiting high-frequency variations in taxi demand due to subway service disruptions, we show that drivers work more when earnings opportunities are greater both when they are above and when they are below their income target. Surpassing the target, however, significantly reduces drivers' labor supply. Estimates imply that drivers' response to demand shocks is 40% smaller once they have reached their daily income target. These results suggest that, while drivers' behavior seems largely consistent with the prediction of a standard model of labor supply, targeting behavior does nevertheless play an essential role in determining drivers' decisions.

Keywords Labor supply · Reference-dependent preferences · New York City taxi drivers

JEL Classification D01 · J03 · J22

1 Introduction

In recent years, the extensive body of literature within economics that focuses on labor supply decisions has once again become central to the academic debate. Instead of

A. Saia thanks the Editor Luca De Benedictis and three anonymous referees for their helpful comments and suggestions.

✉ Alessandro Saia
alessandro.saia@unibo.it

¹ Department of Economics, University of Bologna, Bologna, Italy

investigating the labor supply behavior of standard workers,¹ several influential papers have been devoted to the analysis of a specific category of workers: New York City taxi drivers. Taxi drivers have proven to be ideal when investigating labor supply decisions. On the one hand, they operate in settings where they are faced with temporary changes in their earning opportunities (in other words, the income effects on the labor supply are likely to be negligible). On the other hand, they are free to decide whether to work, and to what degree, and this is also easy to measure.

In this paper, we exploit positive changes in labor demand driven by subway disruptions to quantify the relevance of behavioral biases in the labor supply decisions of New York City taxi drivers. By creating more favorable demand conditions, subway disruptions have a positive impact on drivers' potential earnings. When subway disruptions commence, commuters will be forced to rely on alternative forms of transportations and, due to the higher the number of commuters on the street looking for a cab, it will be easier for drivers to find a new passenger. To the best of our knowledge, this is the first work to exploit exogenous and unanticipated demand shocks to gauge the importance of behavioral on drivers' labor supply behavior of taxi drivers.

Using web crawling and text recognition techniques, we collected high-frequency information on subway disruptions from different Online resources. In the empirical analysis, we show that subway disruptions largely affect drivers' labor supply behavior. On average, when underground breakdowns are observed the probability of a driver of ending her shift is around 15% smaller, even when accounting for driver fixed-effects, hour fixed effects and calendar day fixed-effects. While this finding is in line with the prediction of the standard model of labor supply (i.e., labor supply increases in response to short, positive changes in earnings opportunities), this result doesn't necessarily rule out the presence of targeting behavior.

To assess the extent to which drivers' responses are affected by income-targeting behavior, we investigate whether they respond differently to demand shocks once they have reached their target. Two main results emerge. On the one hand, our estimates show that drivers' responses to positive changes in labor demand are large and economically relevant both when they are below and when they are above their income goal. On the other hand, they also provide clear evidence of the relevance of behavioral biases in labor supply decisions of taxi drivers: when earning opportunities are temporarily higher, the fact of surpassing the target significantly reduces drivers' labor supply. Drivers' responses to demand shocks, while always positive and statistically significant, are around 40% smaller when drivers are above their income target.

We complement these findings by investigating how drivers' response varies as a function of distance to their target. Results show that drivers' response increases when they are close to the target: the effect of subway disruptions on drivers' stopping behavior is stronger when they are closer to the target. As soon as drivers move away

¹ Until recently, empirical research into the relationship between workers' efforts and their earning opportunities was conducted using individual panel data available for standard workers. However, using standard workers to test the predictions of labor supply models is problematic for two main reasons. First, these workers are faced with variations in wages that are of a permanent nature and are thus correlated with significant income effects. Second, they usually operate within settings in which (at the intensive margin) it is not possible to establish their efforts. Therefore, while the external validity of these results is beyond any doubt, what is difficult to identify is the underlying labor supply behavior.

from the target the effect of an increase in taxi demand on stopping probability, while being always negative, decreases in the magnitude.

Overall, results presented in this paper are in line with the predictions of a model of labor supply where both the standard component and behavioral elements coexist. While drivers' behavior seems largely consistent with the prediction of a standard model of labor supply, the large difference between below-target and above-target responses suggests that targeting behavior nevertheless plays a non-secondary role in drivers' decisions.

2 Related Literature

This work mainly relates, and contributes, to the growing list of papers that investigate the labor supply behavior of workers and, in particular, with the series of influential works on taxi drivers' behavior.

The first attempt to investigate the labor supply behavior of taxi drivers was that of Camerer et al. (1997), who estimated a negative relationship between daily hours worked and average hourly earnings. Negative wage elasticities are not consistent with the prediction of the standard model of labor supply but are in line with the idea that drivers set a daily income target and stop working once they reach that goal. Recently, Farber (2015) replicated and extended the analysis by Camerer et al. (1997) using a larger set of observations. Corresponding results show that, unlike previous findings, drivers' labor supply elasticities are generally positive. This result is in contrast with the idea that behavioral biases play a role in drivers' decisions and suggests that the labor supply behavior of taxi drivers is consistent with the conventional neoclassical inter-temporal labor supply model.² Crawford and Meng (2011) studies the daily labor supply of New York City taxi drivers using an alternative empirical approach based on a discrete-choice stopping model. Their results show that behavioral biases are a key component of drivers' labor supply decisions: drivers are more likely to end their shift once they reached their goal. These results are in line with the results provided by a recent paper by Thakral and Tô (2021). In their paper, the authors provide support for a model of reference dependence in which the target adjusts by documenting reductions in drivers' labor supply in response to more recent earnings within the day. Martin (2017) explores the relationship between income and hazard of stopping. Corresponding results show that the relationship is non-monotonic with decreasing hazard of stopping as NYC drivers earn between 100 and 250 \$. Schmidt (2018) finds that large tips (more than \$30) have a negative effect on the labor supply behavior of New York City taxi drivers and argues that standard model of labor supply cannot explain this finding. More recently, Hai Long et al. (2019) shows how Singaporean taxi drivers

² Evidence of positive inter-temporal substitution elasticity are also obtained using data from Uber drivers. For example, Angrist et al. (2017), Ashenfelter et al. (2010) and Doran (2014) focus on fare increases to investigate the effect of long-term wage changes on drivers' labor supply decisions. Chen and Sheldon (2015) shows that Uber partners drive more at times when earnings are high exploiting a sharp discontinuity created by the pricing algorithm of the platform. While Uber drivers and cab drivers operate in similar settings, there are empirical concerns related to the fact that Uber adjusts its prices based on demand-side and supply-side conditions. This concern is not relevant for taxi drivers, whose fares are regulated and they are not affected by supply conditions.

prolong their shifts after unexpected negative events (i.e., booking cancellations) and they indicate income targeting as main explanation.

To assess the extent to which behavioral biases affect drivers' labor supply decisions, it is of crucial importance to explore whether drivers respond differently to demand shocks once they have reached their income target. If reference-dependent preferences play an important role in determining labor supply decisions, once drivers pass their income target, they should be less responsive to an increase in demand for taxi services. Conversely, the lack of meaningful difference between below-target and above-target responses can be interpreted as evidence of the fact that behavioral biases play no (relevant) role in drivers' labor supply decisions.³ The idea of targeting behavior is that drivers' decisions are not only affected by the *absolute* level of income but also by a reference point (target). Building on Kőszegi and Rabin (2006)' theory of reference-dependent preferences, targets reflect the driver' rational expectations in terms of income and represent the driver's belief concerning possible outcomes.

In this paper, we empirically investigate labor supply behavior using a short and unexpected shift in the general level of earnings opportunities. Both the temporal and non-anticipated nature of shocks are of crucial importance if we are interested in understanding the importance of behavioral anomalies (see, among others, Kőszegi and Rabin 2006). By exploiting an identifiable source of variation in taxi demand approach allows us to provide a precise estimate of the role of behavioral biases in drivers' decisions. Moreover, unlike most of the previous works, our empirical approach does not require any ex-ante assumptions about the autocorrelation of drivers' wages.⁴

This paper also relates to a list of recent papers studying the labor supply of different types of workers. Overall, the empirical evidence presented so far has been mixed. For example, Oettinger (1999) analyzes the daily labor supply behavior of stadium vendors and provides evidence of positive labor supply on the extensive margin. Fehr and Goette (2007) conduct a field on bicycle messengers documenting large positive elasticity of labor supply. Stafford (2015) studies both the intensive (daily hours) and extensive margins (daily participation) of labor supply of spiny lobster fishermen in Florida by providing evidence of positive labor supply elasticities. Similar evidence is found using daily labor participation decisions of South Indian boat owners by Giné et al. (2017). Andersen et al. (2018) conduct a field experiment on vendors that operate in a marketplace in India and their findings suggest a limited role for income targeting in vendor labor supply decisions. Abeler et al. (2011) test the preference of reference-dependent preferences in a laboratory-real effort task and they show that reference points, generated via rational expectations, influence effort provision. Chang and Gross (2014) find strong evidence for reference-dependent preferences using observational data for fruit packers. Dupas et al. (2018) conduct a field experiment on bicycle-taxi drivers in rural Kenya and their results are in line with the predictions of a labor supply

³ In Online Appendix H we present a simple model of labor supply behavior that incorporates demand shocks.

⁴ The ability of researchers to investigate autocorrelation of drivers' wages has been somewhat problematic, and results have been characterized by their sensitiveness to the specific sample composition and empirical approach. For example, while Camerer et al. (1997) and Thakral and Tô (2021) show substantial positive autocorrelations in the hourly wage available within a given day, Farber (2005) does not find significant autocorrelations of the hourly wage within a shift.

model in which people have reference-dependent preferences and form income target. In our paper, we investigate the labor supply decisions that occur within a day. This is especially important because, as we will show in the subsequent paragraphs, positive daily labor supply elasticities are not necessarily inconsistent with the presence of target behavior. Moreover, by focusing on the most debated set of workers, we can show that our findings are driven by the different econometric approach rather than to differences in the dataset used.

The remainder of the paper is organized as follows: Section 3 introduces the data used in the paper. Section 4 is devoted to the presentation of the relevance of subway delays for taxi demand. Section 5 displays the baseline results and Sect. 5.1 displays the main robustness checks. Section 6 contains all results related to the role of behavioral biases in the labor supply behavior of taxi drivers. Finally, Sect. 7 concludes.

3 Data

The main dataset for the analysis presented in the paper is obtained by combining electronically recorded taxi trip-log data and high-frequency information on subway disruptions obtained from web resources.

3.1 Taxi Data

The data used in this paper consists of all rides by licensed taxi drivers in NYC from January 1st, 2013 to July 1st, 2013. The raw dataset included 89,286,340 observations and over 37,000 licensed drivers.⁵

Taxi data were provided by the Taxi and Limousine Commission (TLC) of New York City. The TLC requires all medallion taxicabs to be equipped with a Taxicab Passenger Enhancements Project System (*T-PEP*), which processes payments and allows the TLC to collect electronic trip sheet data. Before *T-PEP*, drivers were required to maintain a trip-log that details every fare that they served. The trip sheet was filled out by hand and stored in paper form. In 2008, the TLC had automated this process and mandated that taximeters in all taxicabs be equipped with the *T-PEP*. Paper trip sheets disappeared as the TLC received the data electronically. For each trip, the System records an (anonymized) unique identifier for the driver, the medallion and all trips' details (e.g., length of the ride, the fare and any surcharges, and GPS coordinates of all pickup and drop-off locations). Though TLC receives information electronically, the database can still contain some glitches (e.g., zero-valued duration and implausible or impossible distances). To clean the data, we follow the steps adopted by Haggag and Paci (2014) which we describe in detail in Online Appendix A.

Our dataset is very similar to the dataset used in Farber (2015) (and Thakral and Tô 2021) but is very different from the data exploited in the earlier papers on the labor supply of taxi drivers (e.g., Crawford and Meng 2011).

⁵ As explained in the following sections, the main results presented in the paper are obtained using a sub-sample of the full dataset where we randomly select 1/5 of the drivers. We conduct several sensitivity checks to show that our results are robust when alternative samples are used.

The main difference between our dataset and the dataset used by Farber (2015) is the length of the period of interest. In his paper, Farber used *T-PEP* data for all trips taken in NYC taxi cabs for the 5 years from 2009–2013. Unfortunately, we were able to retrieve data for a shorter period since only the raw dataset of taxi trips for the year 2013 is still freely available [on-line](#).^{6,7} TLC previously provided trip record data with anonymized hack licenses and medallion numbers, but a party that requested *T-PEP* data was able to use an algorithm to identify drivers and their income. As a consequence, in 2014 the TLC has discontinued that practice to protect both passenger and driver privacy.^{8,9}

Our data, however, remains very different from the dataset used in earlier works. For example, Crawford and Meng (2011) relies on data collected by Farber (2005), that was retrieved before the release of electronic trip sheet data. This dataset was based on trip sheets that drivers filled out by hand during each shift and covered a total of 593 trip sheets for 21 drivers observed over the period from June 1999 through May 2001.¹⁰

3.2 Subway Disruptions Data

The New York City Subway is one of the most extensive rapid transit systems in the world. The network consists of 25 lines with a total network length of around 400 km. There are over 470 stations located throughout the boroughs of Manhattan, Brooklyn, Queens, and the Bronx. The annual total ridership in 2013 was of 1,707,555,714 passengers, with an average daily ridership of 5,465,034.¹¹

Like other rapid transit systems across the world, service disruptions are not uncommon.¹² Electrical problems, switch outages, police activity, and medical emergencies generate delays and increase waiting time for passengers.

⁶ Our decision to focus on the first 6 months of the year is due to the fact that we were not able to retrieve some of the files related to trips realized in subsequent months.

⁷ It is also possible to recover the raw dataset of taxi trips for the year 2009 using the replication files of Haggag and Paci (2014) and Haggag et al. (2017). We refrain from using 2009 taxi data because it was not possible to retrieve subway disruptions events occurred over this period. As explained in the following section, we mainly use Twitter accounts to construct high-frequency information on subway disruptions. One Twitter account has been activated only in September 2009 while the other one has been active since September 2012.

⁸ Our Freedom of Information Law (FOIL) request to obtain the full set of raw trip data over the period 2009–2018 to the Taxi and Limousine Commission (TLC) has been denied due to privacy concerns (FOIL Request Number #28979).

⁹ More information on the usage of NYC's improperly anonymized taxi logs are provided [here \[URL\]](#).

¹⁰ We replicate both the analysis of Farber (2015) and Crawford and Meng (2011) using our data. Corresponding results (presented and discussed in detail in Online Appendix C) suggest that, despite the differences between our data and the data used in the two papers, we can replicate the findings presented in both studies. Obtaining similar findings is of crucial importance since it implies that our results are due to the different econometric approach rather than to differences in the dataset used.

¹¹ Source: [MTA-NYC Subway Ridership](#).

¹² From here on we use the terms “service disruptions”, “service turbulences”, “subway disruptions”, “subway breakdowns”, “subway problems” and “subway delays” interchangeably.

In this context, there is a variety of channels that provide timely information on subway service disruptions and delays. We collected detailed high-frequency information on subway disruptions from different sources using web scraping tools.

Our primary source of information is the Twitter account of MTA-NYC Transit Subway Service (@NYCTSubway). This Twitter account is the official source for news and information for subway service in NYC. It is monitored 24/7 and provides timely information to passengers when a disruption starts and when subway service is restored.

Our approach to constructing a novel high-frequency dataset of subway disruptions events in NYC consists of three steps.

The *first step* consisted in downloading all tweets published by the account @NYCTSubway between January and June 2013 (over 6000 tweets). For each tweet, we collected detailed information about its content and the time each message was created.

After having collected all tweets, the *second step* was to analyze tweets' text to distinguish between (i) messages related to the start of service disruptions and (ii) communications announcing that problems have been solved and subway service has been restored. Since tweets' text and their structure are fairly homogeneous, we can rely on a small set of keywords to distinguish between the two types of message.¹³

While analyzing the content of the message, we also extracted information about which line service among the 25 of the New York City Subway system was being interrupted as well as information related to the nature of the problem.¹⁴ Using tweets' text, we use a keyword-based approach to classify disruptions in three different groups: (i) non-technical, (ii) technical and (iii) other outages. Non-technical problems refer to disturbances due to medical emergencies and fire or police activities. Technical disruptions are infrastructure-related problems (e.g., electrical, switch or signal issues). The third group contains all tweets for which it was not possible to identify the nature of the disruption.

In the *third (and final) step* we link starting disruptions tweets with the corresponding ending disruptions messages using subway line(s) information. After creating "couples" of tweets, we compute the elapsed time between the two using tweets' timestamp (i.e., when tweets were created) to identify the exact portion of the day in which underground breakdowns were observed.

For the sake of clarity, let's consider the two tweets reported in Fig. 1. The tweet in the left panel was published on Thursday, May 30, 2013 at 10:36 p.m. (EDT-GMT-4) and provided information on sick passenger-related delay that affected line 1. Using an automated algorithm, we identify the first tweet that announced the service restoration of line 1. In this case, the tweet (reported in the right panel of Fig. 1) announcing that problem on line 1 had been solved and subway service had been restored was published at 11:56 p.m. (80 min later).¹⁵

¹³ For example, tweets related to the second category contains one of the following keywords: "resumed", "restored", "has ended" or "are operating with residual details".

¹⁴ Due to the fact that messages have a standard format, it is relatively easy to retrieve this information since subway lines are preceded by the symbol # in the text.

¹⁵ We also explore whether our results hold when we adopt an alternative method to construct a dataset of subway disruptions events. With this second approach, we only consider the first 5 min after the tweet

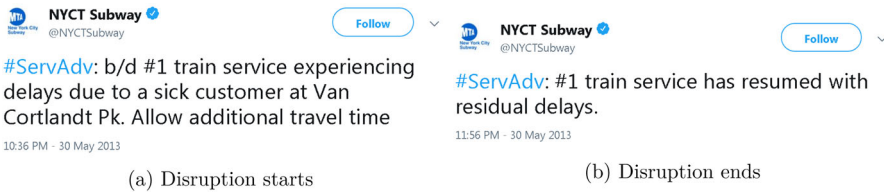


Fig. 1 Tweets on service disruptions from @NYCTSubway

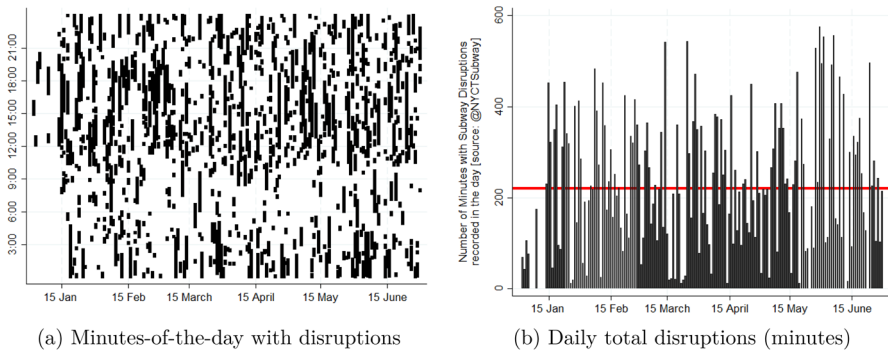


Fig. 2 Subway service disruptions recorded using @NYCTSubway tweets. Note: Author's computations from own subway disruption data. Subway disruptions data was constructed using the official Twitter account of the MTA-NYC Transit Subway Service (@NYCTSubway), following the approach described in Sect. 3.2

Figure 2 plots disruptions data created from @NYCTSubway tweets obtained by performing the steps described above. Panel A displays minutes-of-the-day where subway delays have been recorded. Black dots show portions-of-the-day with measurable disruptions for each calendar day in our period of interest. Total daily disruptions are reported in panel B. From January 2013 to June 2013 there were around 40,000 min during which service disruptions were recorded. The daily average across the whole sample amounts to 220 min. Additional summary statistics are presented in Online Appendix B.

4 Subway Disruptions and Taxi Demand

Demand for taxi fluctuates daily due to demand shocks caused by day-of-the-week effects, holidays, etc. Among others, subways service disruptions are an important shifter of demand for taxi services.

Subway delays can have severe effects on New Yorkers' personal life as well as job and financial security. According to a survey of commuters conducted by the Office of Comptroller of New York City in 2017, 74% of commuters reported being late to a

Footnote 15 continued

announcing the start of a problem as period affected by the delay. Corresponding findings are similar to the estimates obtained with the method presented in the main text. Additional details are provided in Sect. 5.1.

work meeting due to subway delays, while 65% reported being late to pick up or drop off a child and 13% reported losing wages. When facing subway delays, New Yorkers are left to rely on alternative forms of transportations and, according to the survey, 50 percent of commuters were forced to take a taxi.¹⁶ The rise in the number of New York commuters on the street looking for a cab makes it easier for drivers to find a new passenger: metro breakdowns increase drivers' potential earnings opportunities by creating more favorable demand conditions.

We provide support for this claim using actual trip-level data.¹⁷ Using taxi data we compute (i) waiting time and (ii) traveling distance between consecutive trips. Then, we divide our period of interest in 5-min brackets and, for each time window, we compute the median waiting time and the median distance between fares.¹⁸ Then, we combine aggregate taxi market outcomes with data on subway service disruptions observed in the corresponding 5-min window.¹⁹ If it is true that subway delays make less difficult for a driver to find a new passenger then we should observe a negative relationship between subway disruptions and aggregate waiting time (and travelled distance) between two consecutive rides.

Table 1 presents OLS estimates where aggregate taxi outcomes are regressed on service disruptions variables. In order to filter out potential confounding factors, all models include a wide battery of controls for weather conditions observed in the corresponding hour, clock-hour FEs and calendar day FEs in order to take into account any systematic differences between hours of the day and to absorb any calendar day-specific demand (and supply) shifter (e.g., national holiday, etc.).

Result presented in the first column shows how the (log) median of waiting time vary as a function of subway delays. OLS estimates reported in the first panel imply that subway service disruptions cause a drop by around 4.4% in the waiting time observed in the 5-min bracket. A similar pattern emerges from the estimates obtained using the (log) median distance between two consecutive rides as the outcome of interest: estimates reported in column (2) suggest that subway delays also cause a non-negligible reduction in the distance travelled by drivers without passengers. These results provide support for the idea that during subway service disruptions it is easier for a driver to find a new passenger. By increasing the number of available customers on the streets, subway delays shorten both the waiting time and the traveling distance between fares for cab drivers.

Favorable demand conditions should also be observed in other aggregate statistics. For example, if it is true that subway delays create more favorable conditions for taxi drivers, then we should also find a positive link between disruptions and the number of rides. In column (3) we present the results of the model where the (log) total number of trips observed in the 5-min bracket is used as the outcome of interest. Corresponding estimates are highly statistically significant and economically relevant: the point

¹⁶ Source: *The Human Cost of Subway Delays: A Survey of New York City Riders* (2017).

¹⁷ We also exploit hourly-Online search data to provide additional empirical evidence of the existence of a positive relationship between subways service disruptions and taxi demand. Corresponding results are presented in Online Appendix D.

¹⁸ Individual trips are grouped in time brackets using the drop-off time of the ride.

¹⁹ Similar results are obtained using different time windows or using average values instead of median values.

Table 1 Subway disruptions and aggregate taxi market outcomes

	(1) (log) Median Waiting Time _b	(2) (log) Median Dist. Two Rides _b	(3) (log) Total Rides _b	(4) (log) Sum Fares	(5) (log) Median Speed
<i>Subway Disruptions in the 5-min brackets_b</i>	-0.0441*** (0.00416)	-0.0458*** (0.00390)	0.0679*** (0.00485)	0.0693*** (0.00471)	-0.00521*** (0.00126)
Observations	52,116	52,116	52,116	52,116	52,116
R-squared	0.654	0.749	0.677	0.658	0.793
Sample SD	0.598	0.661	0.754	0.701	0.232

Columns report OLS estimates. The dependent variables of interest are reported at the top of the table and are obtained by aggregating taxi trips observed during the period January–June 2013 in 5-min brackets [6]. The variable *Subway Disruptions in the 5-min bracket (dummy)_b* is a dummy that takes value 1 when subway disruptions were observed in the corresponding time bracket. Subway disruptions data was constructed using the official Twitter account of the MTA-NYC Transit Subway Service (@NYCTSubway), following the approach described in Sect. 3.2. All estimates reported in the columns were obtained using the full set of hour fixed effects, calendar day fixed effects and an hourly indicator for observed rainfall in Central Park. Robust standard errors are reported in parentheses. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

estimates of the coefficient of interest imply that the number of rides increases by around 7% if subway problems are observed. Higher demand should also be reflected in higher earnings. In column (4) we present estimates obtained using total earnings²⁰ as the outcome of interest. Corresponding coefficients of interest are once again positive and imply a positive and statistically significant effect of metro breakdowns on the total amount earned by drivers.

While results presented so far reflect a general pattern and cannot be directly related to the labor supply behavior of individual taxi drivers, the estimates presented in this section indicate that subway delays positively affect demand for taxi services. During service disruptions, commuters are forced to use alternative forms of transportations. It is, therefore, easier for drivers to find new customers and, as a consequence, overall taxi market outcomes improve due to the surge in demand.

Before continuing with our analysis, we shall discuss three potential concerns.

Firstly, we shall discuss the possibility that subway delays might also directly affect the supply of taxi services by altering the dis-utility of driving.

For example, metro breakdowns might cause a deterioration in driving conditions. If subway disruptions have a direct impact on drivers' conditions (e.g., cars are moving more slowly) then, it is more difficult to link drivers' decision to end their shift to the presence of behavioral biases: drivers have no incentive in continuing to drive if driving is harder, more difficult and less enjoyable. Following Farber (2015), we use speed as a proxy for road conditions. Corresponding results are presented in the column (5) and show that the effect of subway delays on traffic speed is minimal (-0.5%) (i.e., a decline of 0.02 standard deviations) suggesting that service disruptions don't have a meaningful impact on roads and driving conditions. This result doesn't come as a surprise: factors unrelated to driving conditions usually cause subway disruptions (e.g., sick customers or police emergencies) and they occur underground with limited effect on how pleasant is to drive for taxi drivers.

Secondly, drivers might be able to predict subway breakdowns and, as a consequence, they could adjust their labor supply accordingly. Because behavioral biases in workers' decisions mainly relate to unanticipated changes in labor demand (Kőszegi and Rabin 2006) subway disruptions cannot be used to investigate the role of behavioral biases if drivers can anticipate when breakdowns start. Our approach to mitigate this concern is twofold.

On the one hand, we use our data to assess whether this concern is relevant or not. For example, if drivers can anticipate subway breakdowns, they might decide to start their shifts only when disruptions emerge. However, if this is the case, we should observe a positive correlation between subway disruptions and the share of drivers that have just started their shifts. Using actual data, however, we were not able to find empirical evidence in support for this claim (corresponding tables are presented and discussed in Online Appendix G).

On the other hand, we have also explored whether results presented in the paper hold when we restrict the analysis to a sub-set of disruptive events. It can be argued that delays attributed to medical emergencies, to police activities or fire emergencies are, by nature, more unpredictable than electrical problems or other types of technical

²⁰ We consider total earnings as the sum of fares collected during the shift.

disruptions.²¹ Results obtained with this sub-set of disruptive events are in line with our main empirical findings and provide further support for the validity of our conclusions. Overall, we were not able to find any empirical evidence to support the idea that workers can predict when disruptions start.

A final concern is related to the fact that subway disruptions might also affect the labor supply decisions of taxi drivers if, for example, they use subway services to go back home after their shift. If this is the case, service disruptions make it more difficult for drivers to use the subway and, as a consequence, they will be less prone to end their shift. In the following sections, we will present a wide set of empirical exercises targeted at addressing additional concerns about the potential violation of the exclusion restriction. While we are unable to definitively rule out the possibility that subway disruptions could have some impact on drivers' labor supply beyond its effect working through an increase in taxi demand, the empirical evidence suggests that these other effects, if present, are likely to be not very relevant for our analysis.

5 Demand Shocks and the Labor Supply Behavior of Taxi Drivers

In this section, we explore the extent to which the surge in demand caused by subway disruptions affects drivers' labor supply behavior. Using trip-level data, we randomly select 1/5 of the drivers²² and we estimate a linear model of the probability^{23,24} that driver i will stop driving in her shift s at time t after trip x using the following equation:²⁵

$$\text{Last trip}_{istx} = \beta_0 + \beta_1 \text{Subway disruptions}_t + \mathbf{Z}\boldsymbol{\beta} + \epsilon_{istx} \quad (1)$$

where Last Trip_{istx} is a variable that takes value 1 if the driver i ends her shift s at time t after trip x . $\text{Subway Disruptions}_t$ is a dummy indicating whether subway disruptions were observed at time t and \mathbf{Z} is a vector of baseline control variables that includes the full set of time and locations controls (i.e., hour of the day, day-of-the-week and

²¹ From January 2013 to June 2013 there were around 6200 min during which service disruptions due to non-technical emergencies were recorded. The daily average across the whole sample amounts to 35 min. Additional summary statistics are presented in Online Appendix B.

²² We conduct several sensitivity checks to show that our results are robust when alternative samples are used. Corresponding results are presented in Online Appendix I.

²³ This approach is consistent with the methodology adopted in previous works (e.g., Farber 2008; Crawford and Meng 2011). Due to computational limitations (mainly related to a large number of observations and the extensive set of fixed effects included in the analysis) results presented in the paper are obtained using linear probability models.

²⁴ In Online Appendixes E and F we show that, using alternative empirical approaches, subway disruptions have a positive effect on overall labor supply of drivers.

²⁵ Following Haggag and Paci (2014), we define any gap between trips of more than 5 h as the end of one shift and the start of the next. Time t corresponds to the drop-off time of the trip x (recorded at the day-hour-minute level).

drop-off census track FEs) and drivers' FEs. Standard errors are clustered at the driver level.^{26,27}

Estimation results are reported in Table 2. The first column reports results from the constrained regression, where the variable indicating subway troubles and the baseline structure of FEs are used to explain the probability of ending the shift. The estimated coefficient of the variable *Subway Disruptions* is negative and significant at the 1% level suggesting that positive variations in demand driven by subway service disruptions cause an increase in drivers' effort (i.e., drivers are less likely to end their shift). The estimated effect of demand shocks remains negative and highly significant when we control for adverse weather conditions observed in the corresponding hour (column 2) and the full set of calendar day fixed effects to take into account any systematic difference across days (column 3). These results suggest that when earning opportunities are temporarily higher, drivers adjust their labor supply behavior accordingly.

The effect of subway breakdowns appears to be both statistically significant and relevant: coefficient's estimate reported in column (3) implies that during subway disruptions the likelihood of a driver ending her shift decreases by around 0.6 percentage point. This effect appears to be substantial given that the average stopping probability observed in the sample is approximately 4.5 p.p. Estimates presented in Table 2 imply that drivers' labor supply behavior is largely affected by subway delays, and the probability of ending their shift is around 12% lower due to the increase in demand for taxi rides (to easily understand the underlying magnitude of subway disruptions on drivers' stopping behavior, at the bottom of the table we report the ratio between the coefficient of the variable *Subway Disruptions* and the average probability of stopping observed in the corresponding sample).

In column (4) we include total earnings of the driver in the day and cumulative shift hours (observed after trip x) as additional explanatory variables.²⁸ The inclusion of this additional set of covariates leaves our estimates substantially unchanged, and our coefficient of interest remains negative and highly statistically significant.

We explore whether results hold when we replicate our analysis using only specific types of service disruptions. Using the information contained in tweets, we divide disruptions into three different groups: (i) non-technical, (ii) technical and (iii) other disruptions. The first group contains all disruptions caused by medical, police and fire-related emergencies. The second group includes delays where the nature of the problem is related to technical elements (e.g., switch malfunction, signal or electrical problem, etc.). Corresponding results are reported in Table 3. Coefficients' estimates are always negative and highly statistically significant. Overall, these results confirm the positive effect of (unanticipated) demand shocks on labor supply behavior of taxi drivers and show that this finding is not driven by the type of disruptions considered.

²⁶ Results are very similar when alternative clustering levels are adopted. Corresponding tables are reported in Online Appendix J.

²⁷ For the sake of clarity, all tables obtained using taxi data report t-statistics in parenthesis below coefficients' estimates (rather than coefficients' standard errors).

²⁸ Similar results are obtained when, rather than just a linear term for cumulative shift hours and income, we partition of hours and income as in Farber (2015). Corresponding tables are confined to Online Appendix K.

Table 2 Baseline results of the subway disruptions on drivers' labor supply behavior

<i>Dep. variable: Last trip_{istx}</i>	(1)	(2)	(3)	(4)	(5)	(6)
<i>Subway disruptions_{it}</i>	-0.00492*** (-31.28)	-0.00490*** (-31.21)	-0.00568*** (-32.31)	-0.00684*** (-45.36)	-0.00681*** (-45.27)	-0.00724*** (-43.42)
Observations	12,454,428	12,454,428	12,454,428	12,454,428	12,454,428	12,454,428
R-squared	0.117	0.117	0.118	0.151	0.151	0.152
Driver FEs	Yes	Yes	Yes	Yes	Yes	Yes
Hour FEs	Yes	Yes	Yes	Yes	Yes	Yes
Census track FEs	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-the week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Hourly precipitations	No	Yes	Yes	No	Yes	Yes
Calendar day FEs	No	No	Yes	No	No	Yes
Cum. earnings _{istx}	No	No	No	Yes	Yes	Yes
Cum. hours worked _{istx}	No	No	No	Yes	Yes	Yes
<i>Last trip</i>	0.0459	0.0459	0.0459	0.0459	0.0459	0.0459
Percent change	-10.72	-10.67	-12.37	-14.9	-14.84	-15.78

The dependent variable is a dummy that takes a value of 1 if the driver *i* ends her shift *s* after trip *x* at time *t*. LPM estimates are reported in all columns. The variable *Subway Disruptions_{it}* is a dummy indicating whether subway service disruptions were recorded at time *t*. The service disruption data was constructed using the official Twitter account of the MTA-NYC Transit Subway Service (@NYCTSubway), following the approach described in Sect. 3.2. In order to easily understand the underlying magnitude of the estimates reported in the table, in the last line, we report the ratio between the coefficient of the variable *Subway Disruptions* and the average probability of stopping observed in the corresponding sample (i.e., *Percent change*). The variable *Cum. earnings_{istx}* represents the total earnings of the driver *i* observed after trip *x*. The variable *Cum. hours worked_{istx}* represents the cumulative number of hours worked in the shift by driver *i* after trip *x*. Robust standard error clustered at the driver level are assumed. T-statistics are shown below the estimates in parentheses. Statistical significance is represented by **p* < 0.10, ***p* < 0.05, ****p* < 0.01

Table 3 Types of subway disruptions and drivers' labor supply behavior

Dep. variable: $Last\ trip_{i,s,t,x}$	(1)	(2)	(3)	(4)
<i>Subway disruptions</i> —ALL TYPES	-0.00724*** (-43.42)			
<i>Subway disruptions</i> ₁ —NON-TECHNICAL DISRUPTIONS		-0.00757*** (-20.80)		
<i>Subway disruptions</i> ₂ —TECHNICAL DISRUPTIONS			-0.00812*** (-38.17)	
<i>Subway disruptions</i> ₃ —OTHER DISRUPTIONS				-0.00748*** (-23.24)
Observations	12,454,428	10,949,533	11,884,502	10,886,524
R-squared	0.152	0.155	0.153	0.156
Controls	Yes	Yes	Yes	Yes
Sample mean <i>Last trip</i>	0.0459	0.0461	0.0464	0.0455
Percent change	-15.78	-16.43	-17.51	-16.44

The dependent variable is a dummy that takes a value of 1 if the driver i ends her shift s after trip x at time t . LPM estimates are reported in all columns. The variable $All\ Disruptions_{i,t}$ is a dummy indicating whether subway service disruptions were recorded at time t . The service disruption data was constructed using the official Twitter account of the MTA-NYC Transit Subway Service (@NYCTSubway), following the approach described in Sect. 3.2. Results reported in the first column are obtained using the full set of subway disruptions and correspond to the estimates presented in column (6) of Table 2. Results displayed in column (2) are obtained using only disruptions related to non-technical elements, such as disruptions due to medical emergencies and fire or police activities. Results presented in column (3) refer to disruption due to infrastructure related problems (e.g. electrical, switch or signal problems). The last column presents results obtained all delays for which was not possible to identify the cause of the problem. Further details on disruptions classification are provided in Sect. 3. Estimates are obtained using the full set of controls presented in column (6) of Table 2. All estimates reported in the columns were obtained using the full set of driver fixed effects, hour fixed effects, calendar day fixed effects, census-track fixed effects and an hourly indicator for observed rainfall observed in Central Park plus total earnings of the driver in the day ($Cum. earnings_{i,s,t,x}$) and cumulative shift hours ($Cum. hours\ worked_{i,s,t,x}$) as additional explanatory variables. In order to easily understand the underlying magnitude of the estimates reported in the table, in the last line, we report the ratio between the coefficient of the variable *Subway Disruptions* and the average probability of stopping observed in the corresponding sample (i.e., *Percent change*). Robust standard error clustered at the driver level are assumed. T-statistics are shown below the estimates in parentheses. In column 2 (3) |4] the sample includes all observations with no subway disruptions and non-technical disruptions (technical disruptions) [other disruptions]. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results presented in this section indicate that subway disruptions play a pivotal role in explaining the stopping behavior of drivers: taxi workers adjust their labor supply decisions as a consequence of the surge in taxi demand. Estimates obtained with the most demanding specifications imply that the probability of ending their shift is around 15% lower during subway delays.

5.1 Robustness and Placebo Exercises

Several robustness checks and placebo exercises further validated results presented in the previous section. In the current section we shall limit ourselves to a short account of the robustness analysis, with all robustness tables and further details being relegated to the Online Appendix.

5.1.1 Alternative Data Sources

We investigate whether previous results hold when different sources used to identify subway delays are adopted. To this end, in Online Appendix L we replicate the analysis presented in the previous section using four alternative approaches. We construct a second dataset of metro disruptions that contains information on subway status retrieved from a second verified Twitter account (*@SubwayStats*). This second account provides information about subway status, delays, and closures. It is not affiliated with the MTA of New York, and its only purpose is to share current subway statuses and statistics about the subway train service in New York City.²⁹ Coefficient estimates are in line with the ones presented in the previous pages and confirm the negative relationship between drivers' stopping decisions and subway turbulences. Similar results are also obtained when we combine the two sources or when we focus only on significant disruptions events.

5.1.2 Intensive Margin of Subway Disruptions

The variable of interest used in the previous tables was a dummy indicating whether any disruption was recorded or not. In doing so, we neglected the possibility that (i) several disruptions might co-occur or that (ii) the intensity of two (non-simultaneous) disruptions might differ. For example, it is easy to imagine that the higher is the number of simultaneous delays observed, the larger should be the number of passengers in the street looking for a cab. Similarly, the higher is the number of lines affected by a problem the larger might be the number of individuals affected. In other words, it

²⁹ We also assembled a third dataset using information extracted from the *MTA Alert Archive* (<http://www.mymtaalerts.com>). MTA Alert is a free-of-charge subscription service that provides real-time alerts related to service incidents that impact NYC Transit subways. One crucial problem encountered when using this third source is related to the fact that the website updates date-time fields for no discernible reason. In other words, the date-time field reported in the Archive doesn't always correspond to the original date-time of the alert. Since this concern is not relevant to Twitter messages, we use Twitter-based data as the primary source of information on service breakdowns. Nevertheless, in the Online Appendix M we present results obtained (i) using MTA Alert data and (ii) combining all three sources of service disruptions. Our results remain qualitatively similar.

is plausible to imagine that taxi demand will be positively related to the intensity of the disruption(s). To this end, we replicate the analysis presented in the previous pages by taking into account the *intensive*-margin of subway problems. Corresponding results are presented in Online Appendix N. While these results do not provide any precise general predictions, and are only meant to offer supplementary evidence, it is reassuring to observe that all estimates presented in Table A15 are in line with our hypothesis and provide further support for the existence of a (strong) positive causal relationship between labor demand and labor supply behavior of cab drivers.

5.1.3 Alternative Empirical Specification: Using Drop-Off Locations

So far we have assumed that the surge in demand caused by subway disruptions affects all drivers equally. However, the effect of subway disruptions is likely to be influenced by the location of the driver. Let's imagine two drivers. The first driver ends a trip in a location where there are no subway stations nearby. The drop-off location of the second driver is a place that is surrounded by many subway entrances. If a subway disruption occurs, it will be easier for the second driver to find a new passenger. The idea is the surge in demand caused by subway disruptions will be more relevant for a driver that is located nearby a subway station than for a driver that ends a given trip far away from any given stations. Corresponding results are presented in Online Appendix O and provide further evidence of the relevance of subway disruptions for the labor supply decision of taxi drivers.

5.1.4 Alternative Interpretations of Previous Results

Previous results suggest that drivers are less likely to end their shift when subway disruptions are observed and we argued that this pattern is consistent with the idea that drivers work more when earnings opportunities are greater. An alternative interpretation of previous findings is that drivers are less likely to quit working because they go on a break instead. If this is the case, subway disruptions may actually result in a reduction in drivers' labor supply. To this end, we perform a second alternative exercise using breaks as the dependent variable and we explore the extent to which the surge in demand caused by subway disruptions influences drivers' break-behavior tendencies. Corresponding results are confined in Online Appendix P and suggest that drivers are less likely to go on a break when subway disruptions are observed. This additional set of results is in line with our previous findings, and it is consistent with the idea that taxi drivers' labor supply increases in response to positive changes in earnings opportunities.

5.1.5 Additional Robustness Exercises

Additional sensitivity tests where, for example, (i) we investigate whether results hold when we only consider substantial service disruptions, (ii) we use alternative subsamples created by removing hour-of-the-day from the analysis, (iii) we only consider the first 5 min after the tweet announcing the start of a problem as period affected by

the delay, and (iv) we explore whether subway disruption might also affect the labor supply decisions of taxi drivers beyond its effect working through an increase in taxi demand are confined to the Online Appendixes Q, R, S and T respectively.

6 Drivers' Responses and Targeting Behavior

Estimates presented in the previous section provide evidence of the positive relationship between labor demand and labor supply of taxi drivers: drivers work more when earning opportunities are unexpectedly higher.

While important, this result is not sufficient to reject the hypothesis that behavioral biases influence drivers' labor supply decisions. Drivers fail to respond to a positive change in labor demand if, and only if, their labor supply decisions are exclusively driven by targeting behavior. Therefore, positive response to demand shocks can be compatible with the presence of targeting behavior, if neoclassical elements of labor supply behavior co-exist with behavioral components.

Following Crawford and Meng (2011), we compute sample proxies for drivers' rational expectations using driver-specific sample average income prior to the current shift (i.e., the target for driver i in shift s corresponds to the average earnings of driver i in the previous shifts up to but not including the shift in question).³⁰

After computing drivers' target, we estimate the effect of demand shocks caused by subway disruptions on drivers' stopping behavior using below-target and above-target observations. That is, we compare the effect of an unanticipated increase in taxi demand on the probability that drivers end their shift when drivers' cumulative earnings in the day are below or above their income target. We replicate this analysis using 100 different samples, where each sample g is created by randomly selecting one-fifth of the drivers observed in the full dataset. By comparing the two set of estimates, it will be possible to assess the extent to which behavioral biases play a role in drivers' labor supply decisions. Moreover, we focus only on those observations that are nearby the income target. That is, we compare drivers' responses to unexpected subway disruptions when they are just below or above their income target.³¹

Corresponding results are displayed in Fig. 3. The first column in panel (A) reports the effect of subway disruptions on drivers' stopping behavior obtained using all observations where earnings of the day are in the range $[0.75 \text{ Income target}_{is}; 1.25 \text{ Income target}_{is}]_g$ (i.e., where the absolute distance between cumulative earnings and the target is lower than 25%), irrespectively of being executed while earnings of the day are above or below drivers' target.³² The gray bar represents the average effect obtained using 100 different samples while the horizontal bars indicate the 99.9% confidence intervals.³³

³⁰ Online Appendix C provides further details on the construction of income targets.

³¹ Results obtained using the full set of trips are confined to the Online Appendix V.

³² As explained in the previous section, the magnitude of subway disruptions on drivers' stopping behavior is computed using the ratio between the coefficient of the variable of interest and the average probability of stopping observed in the corresponding sample (see discussion in Online Appendix U).

³³ All estimates were obtained using the model presented in column (6) of Table 2.

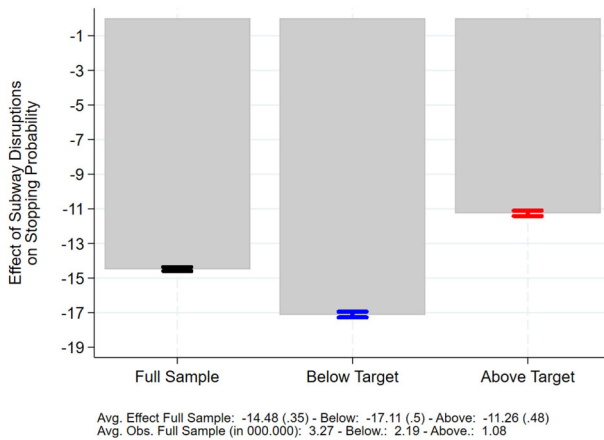


Fig. 3 Subway disruptions, drivers’ responses and targeting behavior. NOTE: The figures display the estimated effect for subway disruptions on drivers’ stopping probability using 100 different samples, where each sample g is created by randomly selecting one-fifth of the drivers observed in the full dataset. The gray bars display the average effect observed while the horizontal bars indicate the 99.9% confidence intervals. The service disruption data was constructed using the official Twitter account of the MTA-NYC Transit Subway Service (@NYCTSubway), following the approach described in Sect. 3.2. Results were obtained using the model presented in column (6) of Table 2. All estimates were obtained using the full set of driver fixed effects, hour fixed effects, calendar day fixed effects, census-track fixed effects and an hourly indicator for observed rainfall observed in Central Park plus total earnings of the driver in the day ($Cum. earnings_{i,STX}$) and cumulative shift hours ($Cum. hours worked_{i,STX}$) as additional explanatory variables. Estimates presented in the first panel are obtained using all observations where earnings of the day are in the range $[0.75 Income target_{iS}; 1.25 Income target_{iS}]_g$. The second (third) column reports estimates derived with below-target observations within the window $[0.75 Income target_{iS}; Income target_{iS}]_g$ ($[Income target_{iS}; 1.25 Income target_{iS}]_g$). As explained in the main text, $Income target_{iS}$ represents the sample average income prior to the current shift s for the driver i

The second and third columns in panel (A) of Fig. 3 display the estimates obtained using observations below and above drivers’ income target, respectively. Two clear results emerge.

On the one hand, our results show that taxi workers adjust their labor supply decisions as a consequence of a surge in taxi demand (caused by subway disruptions) both when they are above and when they are below their income target. In both columns, the estimated effect of subway disruptions is always negative and economically relevant: drivers’ appears to respond to labor demand shocks irrespective of the fact of being above or below their income target.

On the other hand, drivers’ responses vary widely. Estimates imply that the probability of ending their shift during subway delays is around 20% lower when drivers are below their income target. The effect of demand shocks obtained using above-target observations is still negative, but the underlying magnitude is more than halved: the probability of ending their shift during subway breakdowns is (only) 10% lower when drivers have reached their income target. This pattern provides clear evidence of the relevance of behavioral biases in labor supply decisions of taxi drivers: when earning opportunities are temporarily higher, the fact of surpassing the target significantly reduces drivers’ labor supply responses. The difference between the two set of esti-

mates is large and different in a statistically significant way: the average difference between the effect of subway disruptions between trips performed when drivers are below or above their income target is of around ten percentage point (t-statistic = 150).

The results presented in the first panel of Fig. 3 are in sharp contrast with the prediction of the standard model of labor supply but are consistent with the prediction of the model where both the standard component and the behavioral component coexist.³⁴ While drivers' behavior seems largely consistent with the prediction of a standard model of labor supply, the large difference between below-target and above-target responses suggests that targeting behavior plays a (relevant) role in drivers' decisions.

Corresponding estimates are reported in the second panel of Fig. 3. Results presented in the first column are obtained using all trips that are *close* to the drivers' target. The second column reports estimates derived with below-target observations within the window $[0.75 \text{ Income target}_{is}; \text{Income target}_{is}]_g$. Finally, the last column in panel (B) displays the effect of subway disruptions obtained using observations where the drivers have reached their income target and that are in the range $[\text{Income target}_{is}; 1.25 \text{ Income target}_{is}]_g$.

Overall, the adoption of alternative samples confirms previous findings and results presented in panel (B) are in line with the first set of estimates. Drivers broadly behave like neoclassical workers, and they are always less likely to stop when earnings opportunities are higher; however, behavioral biases play a non-secondary role in their labor supply decisions and, after reaching their target, drivers are less likely to respond to an increase in taxi demand.

Due to the properties of the gain-loss utility function (Kőszegi and Rabin 2006), drivers should become more responsive to an increase in taxi demand as they approach their daily target and be less prone to adjust their labor supply once their daily cumulative earnings reach and surpass their rational expectations regarding income.

We further investigate drivers' responses using observations further away from drivers' income target. In doing so, we consider two additional sub-samples. The first one contains below-target trips performed when drivers' cumulative earnings are within the window $[0.50 \text{ Income target}_{is}; 0.75 \text{ Income target}_{is}]_g$. The other sub-sample is made by all observations where earnings in the day are above drivers' target and fall in the window $[1.25 \text{ Income target}_{is}; 1.5 \text{ Income target}_{is}]_g$. Similarly to the exercise presented above, we estimate the effect of subway disruptions on drivers' stopping behavior for each distance bracket using 100 different samples, where each sample is created by randomly selecting one-fifth of the drivers observed in the full dataset.

Corresponding results are reported in Fig. 4. Estimates are ranked on the horizontal axis based on the percentage distance from the reference point, and the gray dashed line separates estimates obtained using observations below or above the income target (to the left and the right, respectively). A clear pattern emerges from a visual inspection of the figure. Overall, the effect of subway disruption on drivers' stopping behavior is always economically relevant, and drivers are generally less likely to end their shift when subway disruptions are observed. However, the underlying magnitude of demand shocks follows a clear non-linear pattern. When drivers are below their income target,

³⁴ See Online Appendix H.

the closer they get to their daily target the stronger is the impact of labor demand shocks. This pattern is fully reversed once drivers reach their daily goal. Estimates presented to the right of the dashed line suggest that the effect of demand shocks decreases the farther drivers are from their target. In other words: when drivers are below (above) their target, the lower is the distance between daily income and their goal, the more (less) prone they are to adjust their labor supply to the unexpected surge in taxi demand caused by subway breakdowns.

To back up this finding, we use effects displayed in Fig. 4 to estimate the following model:

$$|Effect\ of\ Subway\ Disr.|_{gb} = \beta_0 + \beta_1\ Above\ Target_{gb} + \beta_2|Distance\ from\ Target|_{gb} + \epsilon_{sb} \tag{2}$$

where the dependent variable is the estimated magnitude (in absolute value) of the effect of subway disruptions on stopping behavior obtained using the distance bracket b in the sample g , where g is one of the 100 different samples created by randomly selecting one-fifth of the drivers observed in the full dataset. The variable $|Distance\ from\ Target|_{gb}$ represents the average absolute distance from the target observed in the sample g for the distance bracket b and $Above\ Target$ is a dummy that takes value 1 for the two above-target brackets [$Income\ target_{is} : 1.25\ Income\ target_{is}$] $_g$ and [$1.25\ Income\ target_{is} : 1.50\ Income\ target_{is}$] $_g$ (and value 0 for the two below income target brackets [$0.50\ Income\ target_{is} : 0.75\ Income\ target_{is}$] $_g$ and [$0.75\ Income\ target_{is} : Income\ target_{is}$] $_g$).

Table 4 below displays the results. The estimated coefficient of the variable $Above\ Target$ is negative and significant at the 1% level and demonstrates that drivers are less influenced by demand shocks once they reach their target. Coefficient estimates also indicate a negative and statistically significant relationship between the magnitude of drivers' response and absolute distance from their daily income target: the higher is the distance from the target, the lower will be the effect of subway disruptions on drivers' stopping probability. Overall, we take this new set of estimates to provide further support for the results presented in Fig. 4. The effect of subway disruptions on drivers' stopping behavior is always negative but appears to be stronger (i.e., more negative) when they are closer to the target. As soon as drivers move away from the target the effect of an increase in taxi demand on stopping probability, while being always negative, decreases in the magnitude.

It is also important to note that while providing further evidence of the role of behavioral biases in the labor supply behavior of taxi drivers, the results obtained with this second set of estimates also allows us to rule out *fatigue* as an alternative explanation for our findings. In a model of labor supply with fatigue, drivers' response to demand shocks should be negatively related to the effort exerted during the day. If drivers' earnings and drivers' effort are related, the higher the earnings in the day, the lower should be the relevance of subway disruptions on drivers' stopping behavior (i.e., in a model with fatigue the relevance of demand shocks should always be not-decreasing). However, our results show that when drivers are below the target, the lower the level of earnings, the lower is drivers' response. This evidence is in sharp

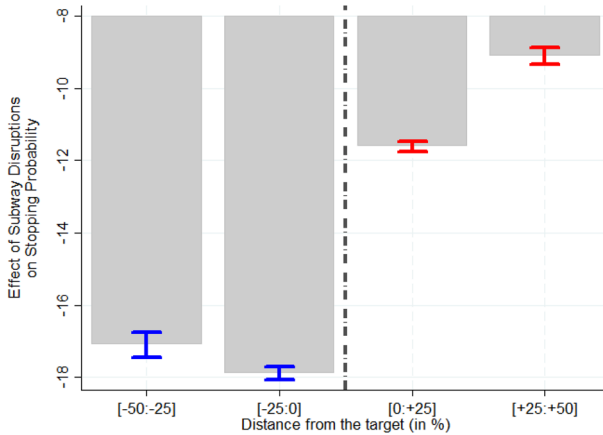


Fig. 4 Subway disruptions, drivers' responses and distance from the target. NOTE: The figure displays the estimated effect for subway disruptions on drivers' stopping probability using 100 different samples, where each sample g is created by randomly selecting one-fifth of the drivers observed in the full dataset. The gray bars display the average effect observed while the horizontal bars indicate the 99.9% confidence intervals. The service disruption data was constructed using the official Twitter account of the MTA-NYC Transit Subway Service (@NYCTSubway), following the approach described in Sect. 3.2. Results were obtained using the model presented in column (6) of Table 2. All estimates were obtained using the full set of driver fixed effects, hour fixed effects, calendar day fixed effects, census-track fixed effects and an hourly indicator for observed rainfall observed in Central Park plus total earnings of the driver in the day ($Cum. earnings_{iSTX}$) and cumulative shift hours ($Cum. hours worked_{iSTX}$) as additional explanatory variables. Estimates are ranked on the horizontal axis based on the percentage distance from the drivers' income target. The gray dashed line separates estimates obtained using observations below or above the income target (to the left and to the right, respectively). As explained in the main text, $Target_{is}$ represents the sample average income prior to the current shift s for the driver i

contrast with the prediction of a standard model with fatigue but it is consistent with alternative theories that explain drivers' labor supply decisions with the presence of a behavioral component.

Overall, the evidence presented in this section is in line with the prediction of a model of labor supply where both the standard component and the behavioral component coexist. On the one hand, drivers' behavior seems to be broadly in line with the prediction of a standard model of labor supply and their response to positive changes in labor demand is always substantial and economically relevant (both when they are below and above their income goal). On the other hand, we observe a significant difference between below-target and above-target responses: the underlying magnitude of the effect of demand shocks on drivers' stopping behavior is still negative but is around 40% lower when drivers have already reached their income goal.

Robustness and placebo checks Results presented in the previous paragraphs were further validated by a comprehensive battery of robustness and sensitivity checks.

Firstly, we investigate whether previous results hold when different approaches to identify subway disruptions are adopted. To this end, we replicate the analysis presented in the previous section when (i) we consider only non-technical disruptions events (i.e., all disruptions caused by medical, police and fire-related emergencies) and (ii) we combine the two different sources used to identify disruption episodes.

Table 4 Subway disruptions, drivers' responses and distance from the target: regression results

Dep. variable: Effect of subway disruptions _{gb}	(1)	(2)	(3)
Above income target _{gb}	-7.148*** (0.118)		-7.345*** (0.0851)
Absolute dist. from target _{gb}		-0.0326** (0.0152)	-0.0678*** (0.00345)
Observations	400	400	400
R-squared	0.901	0.011	0.950

The dependent variable is the estimated magnitude (in absolute value) of the effect of subway disruptions on stopping behavior obtained using the distance bracket *b* in the sample *g*. Estimates are obtained using 100 different samples and each sample is created by randomly selecting one-fifth of the drivers observed in the full dataset (additional details are provided in the main text). The service disruption data was constructed using the official Twitter account of the MTA-NYC Transit Subway Service (@NYCTSubway), following the approach described in Sect. 3.2. The variable *Absolute distance from target_{gb}* represent the average absolute distance from the target observed in the sample *g* for the distance bracket *b* and *Above income target_{gb}* is a dummy that takes value 1 for distances above income target brackets [*Income target_{1s}*; 1.25 *Income target_{1s}*]_g and [1.25 *Income target_{1s}*]_g (and value 0 for the two below income target brackets [0.50 *Income target_{1s}*; 0.75 *Income target_{1s}*]_g and [0.75 *Income target_{1s}*; *Income target_{1s}*]_g. Standard errors are reported in parentheses. Statistical significance is represented by * *p* < 0.10, ** *p* < 0.05, *** *p* < 0.01

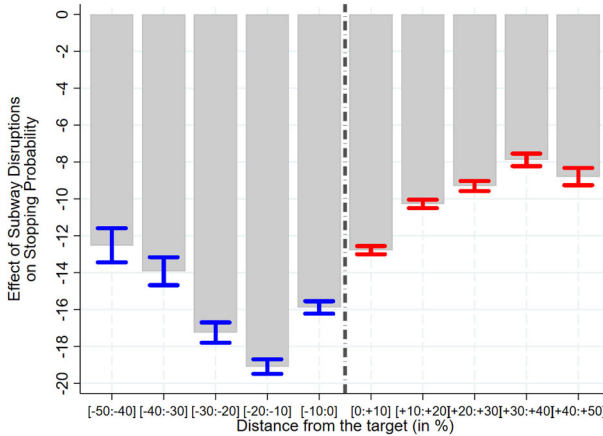


Fig. 5 Robustness targeting behavior: Alternative income-brackets. NOTE—The figure displays the estimated effect for subway disruptions on drivers’ stopping probability using 100 different samples, where each sample g is created by randomly selecting one-fifth of the drivers observed in the full dataset. The gray bars display the average effect observed while the horizontal bars indicate the 99.9% confidence intervals. Results were obtained using the model presented in column (6) of Table 2. All estimates were obtained using the full set of driver fixed effects, hour fixed effects, calendar day fixed effects, census-track fixed effects and an hourly indicator for observed rainfall observed in Central Park plus total earnings of the driver in the day ($Cum. earnings_{ISTX}$) and cumulative shift hours ($Cum. hours worked_{ISTX}$) as additional explanatory variables

Corresponding results are reported in Online Appendix W. Overall, results obtained using alternative measures of disruptions are in line with previous findings and confirm the importance of behavioral biases in the labor supply decisions of drivers.

Secondly, we perform an additional sensitivity test geared toward assessing whether previous results hold when alternative income-brackets are adopted. To this end, rather than grouping observation in 4 different groups,³⁵ we split the sample in 10 equally spaced groups. The 10 different groups range from $[0.50 Income target_{is} : 0.40 Income target_{is}]_g$ to $[1.40 Income target_{is} : 1.50 Income target_{is}]_g$. Figure 5 below displays the results (additional results are provided in Online Appendix X). Reassuringly, our main findings remain very similar. The only significant difference lies in the fact that the *turning point* (i.e., when drivers’ responses to demand shock start shrinking) appears to be slightly before the gray dashed line that separates estimates obtained using observations below or above the income target. However, this finding needs to be interpreted with caution, given that the average difference in terms of cumulative earnings between the groups $[0.20 Income target_{is} : 0.10 Income target_{is}]$ and $[0.10 Income target_{is} : Income target_{is}]$ (i.e., the two groups to the left of the target) is less than 30 \$.³⁶

³⁵ The 4 different groups are the ones displayed in Fig. 4: two below target groups $[0.50 Income target_{is} : 0.75 Income target_{is}]_g$ and $[0.75 Income target_{is} : Income target_{is}]_g$ and two above income target groups $[Income target_{is} : 1.25 Income target_{is}]_g$ and $[1.25 Income target_{is} : 1.50 Income target_{is}]_g$.

³⁶ To provide some context, it is worth highlighting that the average cumulative earnings at the end of the shift observed in the sample is over 300 \$.

Additional robustness exercises We replicate the results presented in Fig. 4 by using weighted estimates, where weights are equal to the number of observations used to compute the income target. We compute sample proxies for drivers' rational expectations using driver-specific sample average income before the current shift. Since sampling errors tend to be more substantial in the first part of the sample period, weighted estimates reduce the bias due to sampling variation. Results from weighted regression, where sampling variation is considered, are reported in Online Appendix Y. Estimates obtained with unweighted and weighted regression are essentially the same.

Finally, we explore whether alternative approaches to identify drivers' target affect our results.

Other papers have also considered the existence of multiple targets. For example, Crawford and Meng (2011) estimate a model of labor' supply with two targets, defined as specific sample average income and hours prior to the current day. In the previous analysis, we have only considered one target (i.e., income target) for two main reasons. Firstly, most of the literature has focused on the presence of an income target. Secondly, and more importantly, we observe a high correlation between income and hour targets. While most of the trips happen when drivers are contemporaneously below income and hour targets (84% of the trips in the sample), the majority of the remaining observations is recorded when drivers are contemporaneously above both targets (8.5% of total trips). Observations recorded when drivers are only above their hour target constitute only a small proportion of total trips (2.3%), while the number of trips observed when drivers have only reached their income target is almost double (4.2%). For these reasons, our main analysis was conducted using the driver's rational expectations concerning daily earnings as the primary target. However, we explored whether our results hold when alternative approaches are considered. To this end, we replicate the analysis presented in this section using (i) a model where drivers set their target based on the number of hours worked in the day and (ii) a model where both income target and hours target are jointly considered. Corresponding results are confined to the Online Appendix Z and are in line with the main estimates presented in the previous paragraphs.³⁷

Overall, results displayed in this section (and in the Online Appendix) provide further support for the fact that, while drivers positively respond to demand shocks, behavioral biases still play a relevant role in the daily labor-supply decisions of taxi drivers.

7 Conclusions

There is a growing list of influential papers documenting a series of behavioral anomalies that question the validity of standard economic models. The implications of behavioral biases may be particularly important within the context of labor economics. For example, the current design of income taxes or unemployment insurance policies

³⁷ Results are also very similar when alternative approaches to identify drivers' target are adopted. Online Appendix Z reports results where targets are constructed as using drivers day-of-the-week specific sample average earnings in the day, prior to current shift (i.e., the target for driver i for day-of-the-week d is the average income of driver i in the previous days-of-the-week d up to but not including the day in question).

is entirely based on the prediction of the standard model of labor supply: if behavioral factors play an important role in labor supply decisions, then policies that neglect such anomalies are going to be sub-optimal. An assessment of the extent to which behavioral anomalies affect workers behavior is of crucial importance to academics and policymakers alike.

In this paper, we investigate how New York City taxi drivers respond to positive changes in labor demand caused by subway service disruptions and whether targeting behavior affect drivers' response. To the best of our knowledge, this is the first paper to use an identifiable, exogenous and transitory demand shocks to quantify the role of behavioral biases in the labor supply decisions of taxi drivers.

We show that drivers' broadly behave like neoclassical workers and they work more when earnings opportunities are more significant. However, we provide evidence of the fact that targeting behavior does play an essential role in determining the labor supply of taxi drivers. Results show that once they reach their income target, drivers are less likely to respond to an increase in taxi demand.

Further research is needed to investigate labor supply responses using truly exogenous demand shifters to understand and to quantify the importance of potential complementarities between neoclassical and behavioral components better.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40797-021-00162-3>.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abeler J, Falk A, Goette L, Huffman D (2011) Reference points and effort provision. *Am Econ Rev* 101(2):470–92
- Andersen S, Brandon A, Gneezy U, List JA (2018) Toward an understanding of reference-dependent labor supply: theory and evidence from a field experiment. Working Paper
- Angrist JD, Caldwell S, Hall JV (2017) Uber vs. taxi: a driver's eye view. Working Paper
- Ashenfelter O, Doran K, Schaller B (2010) A shred of credible evidence on the long-run elasticity of labour supply. *Economica* 77(308):637–650
- Camerer C, Babcock L, Loewenstein G, Thaler R (1997) Labor supply of New York City cabdrivers: one day at a time. *Q J Econ* 112(2):407–441
- Chang T, Gross T (2014) How many pears would a pear packer pack if a pear packer could pack pears at quasi-exogenously varying piece rates? *J Econ Behav Organ* 99:1–17
- Chen MK, Sheldon M (2015) Dynamic pricing in a labor market: surge pricing and flexible work on the uber platform

- Crawford VP, Meng J (2011) New York City cab drivers' labor supply revisited: reference-dependent preferences with rational-expectations targets for hours and income. *Am Econ Rev* 101(5):1912–32
- Doran K (2014) Are long-term wage elasticities of labor supply more negative than short-term ones? *Econ Lett* 122(2):208–210
- Dupas P, Robinson J, Saavedra S (2018) The daily grind: cash needs, labor supply and self-control. Working Paper
- Farber HS (2005) Is tomorrow another day? The labor supply of New York City cabdrivers. *J Polit Econ* 113(1):46–82
- Farber HS (2008) Reference-dependent preferences and labor supply: the case of New York City taxi drivers. *Am Econ Rev* 98(3):1069–82
- Farber HS (2015) Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *Q J Econ* 130(4):1975–2026
- Fehr E, Goette L (2007) Do workers work more if wages are high? Evidence from a randomized field experiment. *Am Econ Rev* 97(1):298–317
- Giné X, Martínez-Bravo M, Vidal-Fernández M (2017) Are labor supply decisions consistent with neoclassical preferences? Evidence from Indian boat owners. *J Econ Behav Organ* 142:331–347
- Haggag K, Paci G (2014) Default tips. *Am Econ J Appl Econ* 6(3):1–19
- Haggag K, McManus B, Paci G (2017) Learning by driving: productivity improvements by New York City taxi drivers. *Am Econ J Appl Econ* 9(1):70–95
- Hai Long D, Chu J, Yao D (2019) Unanticipated income shocks and daily labor supply: taxi drivers' response to cancellations and no-shows. Working Paper
- Kőszegi B, Rabin M (2006) A model of reference-dependent preferences. *Q J Econ* 121(4):1133–1165
- Martin V (2017) When to quit: narrow bracketing and reference dependence in taxi drivers. *J Econ Behav Organ* 144:166–187
- Oettinger GS (1999) An empirical analysis of the daily labor supply of stadium vendors. *J Polit Econ* 107(2):360–392
- Schmidt M-A (2018) The daily labor supply response to worker-specific earnings shocks. Working Paper
- Stafford TM (2015) What do fishermen tell us that taxi drivers do not? An empirical investigation of labor supply. *J Labor Econ* 33(3):683–710
- Thakral N, Tô LT (2021) Daily labor supply and adaptive reference points. Working Paper

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.