Estimating the Impact of Airport Pickups on Driver Pay

Andrew McCall, Arwin Rahmatpanah, Dylan Fletcher, and Jordan Lewis

Contents

1   Introduction

In the third quarter of 2022, Uber drivers completed 1.95 billion trips[1]. A concern for drivers and rideshare companies alike is how to maximize the amount of revenue generated from each of these trips. There have been studies showing that drivers in Seattle made over $23 per hour after expenses while other areas had put net earnings well below $20 per hour. Another new study put the figure at less than half that[2].

With so many conflicting reports, it can be a challenge for drivers and rideshare companies to understand what the normal rate of compensation is and how they can increase their take home pay. Without a clear understanding of what types of trips lead to increased pay, many drivers commit to completing trips that lock them into a lower rate that might be earned elsewhere.

Our study aims to understand whether certain trips, specifically those originating at airports, lead to increased compensation for drivers. We utilize trip record data provided by the New York City Taxi and Limousine Commission (TLC). The data show trip timing (pickup, dropoff, duration, etc.), location (e.g. borough, airport, etc.), rideshare company (e.g. Uber, Lyft, etc.) and pay (base driver pay, tips, etc.). Using a regression model we estimate the relationship between driver pay and location of pickup (airport vs non-airport).

2   Data and Conceptualization

The TLC data is collected using technology installed in taxis and from high-volume for hire vehicle (HVFHV) providers, such as Lyft and Uber. While the TLC recognizes that taxi data is reliant on the accuracy of the systems used and completeness of the provided data from dispatchers, we can assume that the data from the HVFHV is much more complete as the collected data is central to their operation. This study focuses on data collected in June 2022. It is the most recent data before a change was made that allows drivers to reject trips based on the destination and could skew results[3].

In total there are over 17 million observations. Due to system constraints the analysis focuses on a random sample of 1 million observations of HVFHV trips. The dataset provides a pick-up and drop-off location variable which allows us to focus on the La-Guardia and JFK Airports to compare against all other trip locations. For drivers to conduct an airport pickup, they must enter a special airport queue, meanwhile for non-airport trips, drivers are told to go elsewhere for pickups. We created a flag for whether a trip is an airport pickup or not. Additionally, we are not using total driver pay in insolation, rather we are including tips in the total compensation. Tips have a significant impact on compensation, and can be dependent on length of trip. Additional variables that will be examined are trip duration (in minutes), wait duration before pick up (in minutes), and time of day for the trip.

Table 1: Airport Pickup Trips

Statistic	N	Mean	St. Dev.	Min	Max
Driver Pay ($)	38,721	47.1	30.0	1.6	960.6
Trip Duration (m)	38,721	39.2	21.5	1.3	297.6
Pay per Minute ($/m)	38,721	1.3	0.5	0.4	11.2
Wait Duration (m)	38,721	7.1	4.7	0.2	105.4

The tables above summarize the numerical variables between airport pickup trips and non-airport pickup trips. Airport pickup trips occur much less frequently, around 4% of total trips. However, airport pickups are generally longer in trip duration, wait duration, and have higher driver pay. Importantly, the driver pay per minute is also higher for airport pickup trips, $1.30 vs. $1.10 This difference is statistically significant based on an unpaired two-sample t-test and provides initial evidence that airport pickup trips pay drivers more.

We operationalize additional variables for rideshare company and trip time of day which both appear relevant to driver compensation. The graph in the appendix shows how the average driver pay for a trip predictably increases

---

[1]Huang, Kalley. "Uber's Revenue Is Up 72 Percent From Last Year, but It Posts a Loss." The New York Times, 1 Nov. 2022, https://www.nytimes.com/2022/11/01/technology/uber-revenue.html?searchResultPosition=1

[2]Scheiber, Noam. "When Scholars Collaborate With Tech Companies, How Reliable Are the Findings?" The New York Times, 12 July 2020, https://www.nytimes.com/2020/07/12/business/economy/uber-lyft-drivers-wages.html

[3]Capoot, Ashley. "Uber Unveils New Features, Including One That Lets Drivers Choose the Trips They Want." CNBC, CNBC, 29 July 2022, https://www.cnbc.com/2022/07/29/uber-will-let-drivers-choose-the-trips-they-want-to-take.html

Table 2: Non-Airport Pickup Trips

Statistic	N	Mean	St. Dev.	Min	Max
Driver Pay ($)	961,279	20.0	16.2	0.1	633.4
Trip Duration (m)	961,279	19.5	14.0	1.0	582.2
Pay per Minute ($/m)	961,279	1.1	0.5	0.02	26.7
Wait Duration (m)	961,279	5.5	3.4	0.02	67.4

and decreases throughout the day. We created a flag to indicate where each ride landed on a spectrum of selections including: early morning, morning, midday, afternoon, evening and late evening. In addition, Uber drivers seem to make consistently more than Lyft drivers.

3 Model and Modeling Decisions

We have specified the model below to estimate the premium that drivers can receive from picking up passengers at an airport as opposed to regular trips. We hypothesize that driver compensation will be higher on airport pickups because travelers have more inelastic demand and are more likely to tip their drivers.

$$ln(\widehat{driver\ pay}) = \beta_0 + \beta_1 \cdot airport\ pickup + \beta_2 \cdot ln(trip\ duration) + \beta_3 \cdot ln(wait\ duration) + \beta_4 \cdot congestion\ zone + \beta_5 \cdot Uber + \mathbf{T}\gamma$$

In our model, *airport pickup* is a boolean variable and the coefficient β_1 represents the additional premium that a driver receives for making an airport pickup. β_2 and β_3 are both expected to be positive because the more time a trip lasts or time a driver spends picking up a passenger, the more the driver will be compensated. The *congestion zone* is a boolean variable denoting if the trip occurred in the city's congestion zone in lower Manhattan. *Uber* is also a boolean variable denoting if the trip was for Uber or Lyft, the only two operators. Finally, $\mathbf{T}$ is a vector of indicator variables for the different periods of day and γ is a column vector of coefficients. We decided to log the driver pay, trip time, and wait time for two reasons. The first reason is that all three fields have large outlier values that may generate large residuals. By logging the fields, we can de-emphasize the importance of outliers and maintain a more constant error variance. The second reason is that coefficients from a log-log model will be more generalizable to drivers. The effect of increasing trip time or wait time to increase pay is relative. Increasing trip duration from 10 minutes to 20 minutes is very different from increasing trip duration from 1 hour 50 minutes to 2 hours. We think our results will be more useful to drivers expressed as percentages.

4 Results

Table 3, below, shows the results for three different regressions run on the data. All variables were highly statistically significant across each of the models. To describe the various indicator variables in our model, we use the Halvorsen-Palmquist correction for dummy variables in a logarithmic equation. This equation states that the percent change in the outcome is described by the following:

$$\triangle y = 100 \cdot (e^{\beta_i} - 1)$$

When examining airport pick-ups exclusively in the first model, the regression reveals that an airport pickup results in a 151% increase in driver pay. However, the R^2 value indicates that this variable alone only accounts for 7% of the variance in driver pay. The third model includes all the variables examined and accounts for 80% of the variance in driver pay. Here we see a decrease in the effect airport pick-ups have when accounting for additional variables, but that it still represents a 31% increase in driver pay. In addition, we estimate that a 1% increase in trip duration results in a 0.85% increase in driver pay. With the airports located approximately 30 minutes outside of the city, we can see how this is also an important variable to describe the impact of airport pick-up trips. Additionally, this data also confirms results that Uber drivers make more than Lyft drivers with a 9.4% increase in driver pay for Uber drivers over Lyft and can be seen in the plot in the Appendix.

Table 3: Estimated Regressions

	Output Variable: ln(Driver Pay)		
	(1)	(2)	(3)
Airport Pickup	0.92***	0.26***	0.27***
	(0.003)	(0.002)	(0.002)
ln(Trip Duration)		0.85***	0.85***
		(0.001)	(0.001)
ln(Wait Duration)		0.04***	0.04***
		(0.001)	(0.001)
Congestion Zone			0.13***
			(0.001)
Uber			0.09***
			(0.001)
Constant	2.77***	0.38***	0.36***
	(0.001)	(0.002)	(0.002)
Time of Day			✓
Observations	1,000,000	1,000,000	1,000,000
R^2	0.07	0.78	0.80
Residual Std. Error	0.64 (df = 999998)	0.31 (df = 999996)	0.29 (df = 999989)

Note:	HC_1 robust standard errors in parentheses. Time of day is broken down into the following groups based on perceived high and low demand periods: early morning, morning, midday, afternoon, evening and late evening

5 Discussion

5.1 Statistical Limitations

- **Heteroskedasticity:** Given that our model has a high degree of heteroskedasticity, we are using Heteroskedasticity-Consistent standard errors as a way of correcting this. The heteroskedasticity is likely being caused by the NYC-imposed minimum for hourly driver pay. This minimum is mostly enforced on short rides, making the driver pay more formulaic and linear. This means our model very accurately estimates the driver pay when driver pay is low for a trip but the estimate becomes less accurate as driver pay increases. See the scale-location plot of residuals in the appendix.

- **Multicollinearity:** In our original model, trip distance and trip time were highly correlated, resulting in multicollinearity. We dropped distance from our model - time is much more likely to be important to drivers. See the appendix with a correlation matrix of all numeric variables.

- **Independent Identically Distributed (IID):** Our data is only sampled from New York City, so it is not necessarily reflective of all rideshare drivers. In addition, our sample suffers from geographical clustering because drivers are operating in the same areas as each other. This causes some concern with our IID assumption as drivers' actions are not totally independent of one another. A driver accepting a trip will affect the availability of trips for other drivers. And in the aggregate, drivers may affect trip pricing and therefore drivers' pay.

- **Normally Distributed Errors:** Originally, the distribution of residuals exhibited a Leptokurtic distribution with a long left tail. Upon review, we found that there were a small number (~50) trips with impossibly low driver pay given the trip time. For instance, the data reported making less than 20 cents on a trip over 30 minutes in duration. We removed the handful of erroneous trips and the error distribution became more normal. We believe the remaining slight positive skew to the residuals is a product of the NYC driver pay minimums imposed on some trips. See the appendix for a histogram of our scaled residuals.

5.2 Structural Limitations

There are two variables omitted from the dataset that could have a material impact on the results of our analysis. First, we do not have any insight into costs incurred by the driver when completing trips (e.g. fuel, insurance, etc.). Given that a driver may earn more revenue from driving more, it is reasonable to assume that their costs increase as well. A more robust analysis would likely include some sort of cost metric to see how net and gross revenues vary by type of trip.

Secondly, we do not have any detail on the type of vehicle used for each trip. Many rideshare services offer different tiers of ride options (e.g. Uber pool, Uber X, Uber XL, etc.) which charge different premiums for the exact same trip. Without this detail we could have two drivers with pickups and dropoffs from the same locations and times of day but receive different revenues.

6 Conclusion

This study examined the impact airport pick-ups had on driver pay for Uber and Lyft, in New York City. Our results indicate that all else held constant, airport trips result in a 31% increase in driver pay. Additionally, we found that increasing trip duration by 1% increases driver pay by 0.85%. We also noticed similar positive effects for whether the trip passed through a designated congestion zone, had an increased wait time, or occurred during a particular time of day. Finally, Uber drivers had a 9.4% increase in driver pay compared to Lyft drivers.

For further studies, we suggest also gathering data to account for additional costs the drivers face for different trips and the level of service provided by the rideshare company. These would help provide a more accurate picture of the revenue earned by drivers for different trips.
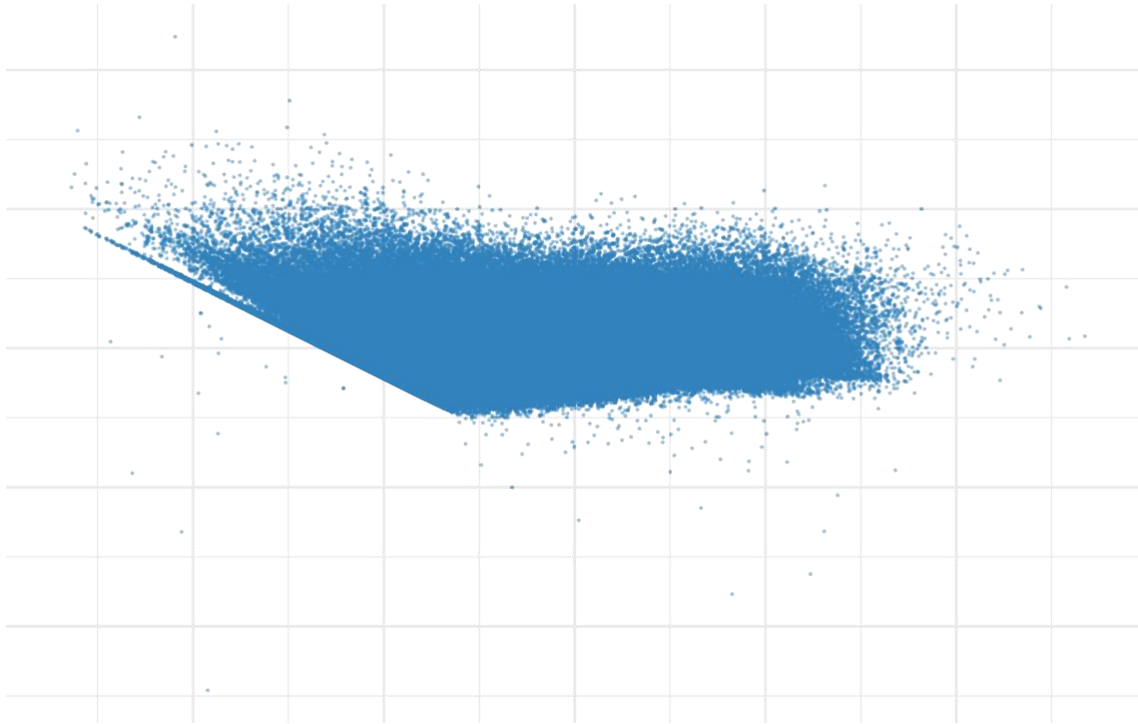
7   Appendix

Average Driver Pay for High−Volume FHV Rides in June 2022
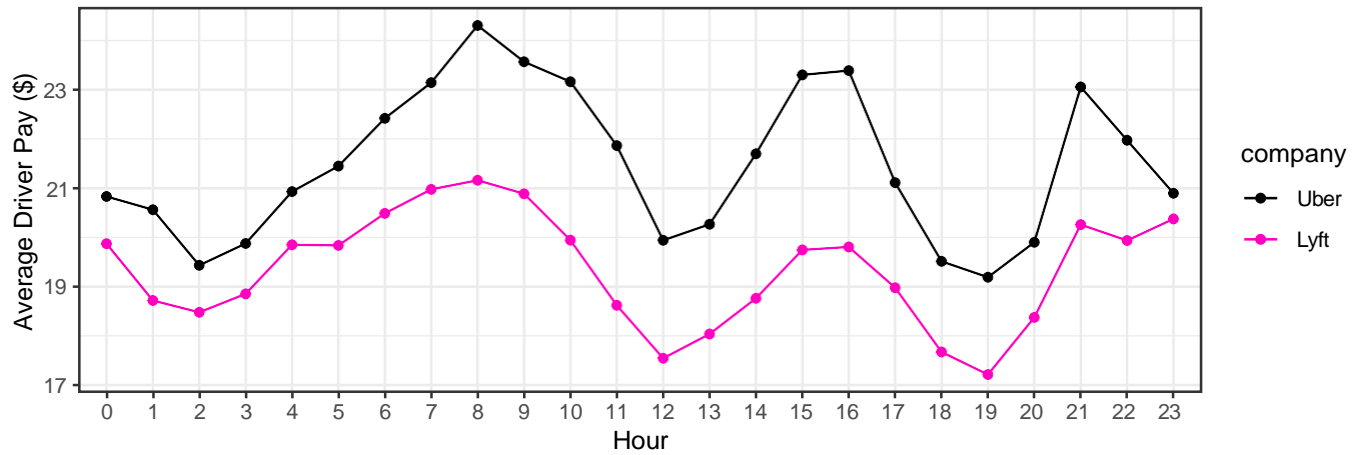
Correlation Between For−Hire Vehicle Fields
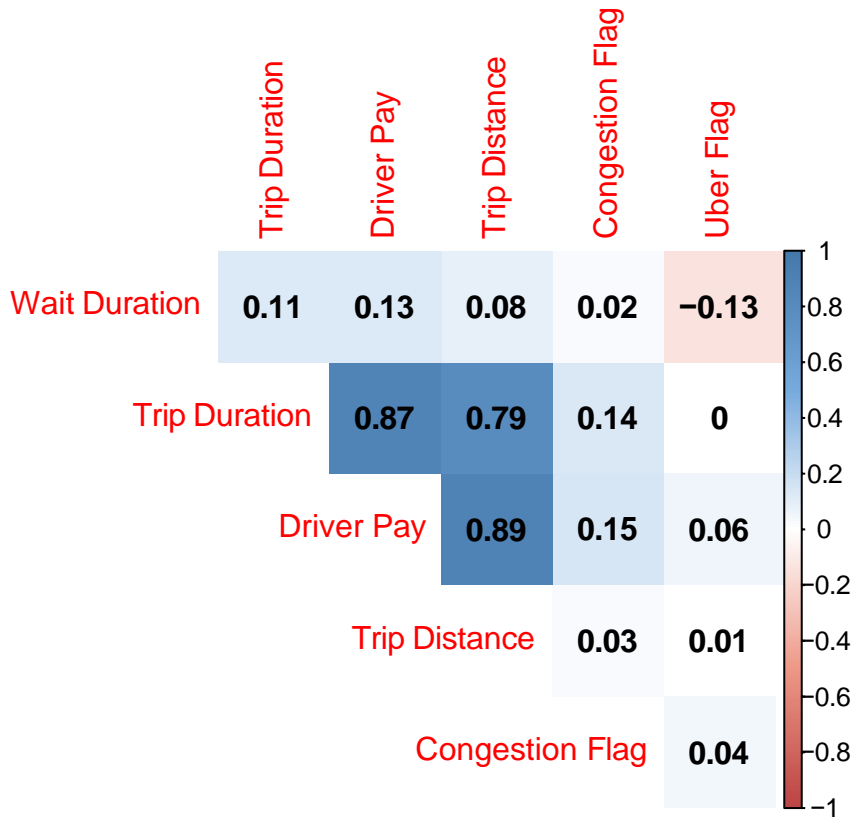
Histogram of Scaled Residuals

Scaled Residuals