Tipping Behavior and Delays: Evidence from Chicago's Transportation Network Providers

Gustavo Pellegrini Dias

October 7th 2022

Abstract

In this paper we consider evidence from 11.4 million rideshare trips that occurred in the city of Chicago between the years of 2018 and 2022 to examine the theory that service quality influences how much a client tips. In an app ride, the passenger's experience depends almost entirely on their interaction with the driver. Therefore, if the tip behavior is determined by the quality of a server's work, this relationship is more likely to appear in rideshare than in most other settings. We run a linear regression between the trips' deviation from average path time and the indicator function of tips controlling for fixed effects of the different route characteristics to estimate the effect of travel time delay on a client's probability of tipping. We also control for observables such as the passenger's estimated income and an indicator function of shared/not shared trip. The results of this paper give us new insight to the debate of whether tipping can be modeled as a rational behavior and if social pressure increases or decreases an individual's probability of tipping.

1 Introduction

1.1 Of The Paper's Structure

This paper is divided in 6 sections. In order of appearance, they are: 'Introduction', 'Theoretical Framework', 'TNP and Descriptive Data', 'Empirical Framework', 'Results', 'Discussion and Conclusion'. The first section, the 'Introduction', concerns itself with a description of the paper's structure and a brief exposure of the main motivation and ideas relevant to the investigation in this paper. Next, 'Theoretical Framework' is divided into 3 subsections: a review of the tipping behavior literature, the construction of the causal theory linking tip probability and trip delays, and a description of the control variables selected for our analysis.

The 'TNP and Descriptive Data section' includes a breakdown of the Transportation Network Providers (TNP) database from which we extract all information associated with the ride-hail trips used in our analysis. The second part of this section provides plots and tables that describe the main variables and controls in this study. This includes frequency tables and scatter plots illustrating the co-relationship between variables such as tipping probability and household income.

The second half of the paper starts with the 'Empirical Framework', where we detail the model used to estimate the causal effect of delays on tipping probability. Following this is a section where we report the results of our model. Next, we consider the impact of these findings in the discussion of whether tipping behavior can be modeled as a rational behavior or not. Finally, we present our concluding remarks, emphasizing the importance of observing more diverse and generalizable tipping behavior for the development of the Behavioral Economics field.

1.2 Of The Paper's Main Motivation and Ideas

From an economic standpoint tipping is a very intriguing behavior as it suggests that consumers are opting to increase their costs for no clear increment in their own utility. Tipping therefore poses a challenge to models of consumer behavior that rely on the Rational Choice Theory (RCT). In the RCT's framework, rational individuals are assumed to always make choices that maximize their net benefits, prompting social scientists to ask questions such as "why pay extra?" and "why do people leave money to strangers when they are not legally obligated to do so?" (Azar, 2005; Lynn 2010).

This paper tries to answer these questions by focusing on evidence of tipping from the ride-hail industry. Historically, studies on tipping behavior have centered around evidence from randomized control trials in restaurants. However, the choice of analyzing data from rideshare service may be more appropriate for testing the relationship between tips and amount of value generated to a client by a server's labor. The experience in an app ride depends almost fully if not entirely on the interaction between a driver and the passenger. Therefore, if the tip decision is determined in any way by the quality of a server's work, this relationship is much more likely to show itself in ride-hail than in settings such as restaurants.

And, assuming time is the determinant factor in a person's decision to order a ride through an app, there exists a clear measure of a driver's service quality in terms of how many minutes they are able to save their passengers in their commute. We can derive this measure by taking the difference between a trip's travel time for a particular route and the average travel time of trips through that same route. The outcome variable we are interested, in this case, is an indicator function of tipping where 0 represents "tip in dollars = 0", and 1 represents "tip in dollars > 0". With these variables at hand we can run a regression with selection on observables to estimate the effect of trip delays, or service quality, in the probability of tipping.

The data which we use to run our analysis comes from the Transportation Network Providers (TNP) database publicly available in the Chicago Data Portal. Particularly, we investigate the subset of ride-hail trips that occurred on weekdays, from 8AM to 10AM. This filtering of our data is important not only because we get to observe trips at similar conditions of traffic, but also because, at this date and time, a trip's pickup census tract is more likely to be the tract where the passenger resides. This allows us to crosswalk the TNP data with the results of the 2020 American Community Survey (ACS) with reasonable accuracy. This adds relevant data for our regression's controls that should raise the internal validity of our study at a very low cost to its external validity.

In the final section of this paper we discuss the main finding of our analysis: an statistically significant evidence of a negative effect on the likelihood of a passenger tipping caused by trip delays, plus some other systematic characteristics of passengers tipping behavior we observe. But for now we will concern ourselves with detailing the historical theories on tipping behavior, their limitations, and the contributions this paper makes to the field. The next section of this paper contains three subsections. First, we give a review of the literature of tipping behavior, a careful construction of the causal theory between tip probability and trip delays, and the reasoning behind each observable selected as controls for our regression.

2 Theoretical Framework

2.1 Background Literature

Traditionally, answers to the question of why people tip have been divided between explanations founded on RCT and explanations founded on the Social Norms Theory (SNT). A counterpart to RCT, the SNT proposes that human behavior is often determined by unwritten rules, that is cultural expectations of how one should behave. These societal expectations overcome individuals' will to maximize their utility, and that would explain the existence of some irrational behaviors such as tipping. In truth, most of the recent literature on tipping behavior concludes that tips are a product of social norms, and that economic models will always fail to answer such questions unless they account for the effect of social pressure and feelings in an agent's economic decision (Azar, 2020). Nevertheless, there are still those that claim tipping could have a rational explanation. A particular rational argument proposed by Bodvarsson and Gibson that remains to be tested is that tipping exists because it is an efficient pricing strategy for businesses in which customers have more information than employers about a worker's contribution to firm's revenues (Bodvarsson and Gibson 1988; Jacob and Page 1980). And from the passenger's perspective, the existence of tips is explained as a rational decision to reward good drivers and punish bad ones, increasing the expected value of their next ride. Tipping, then, would be the result of consumers acting like a firm, monitoring and pricing the marginal product of a worker's labor. Of course, this functional explanation implies that tipping varies systematically according to service quality, server friendliness, and other variables through which a server can add value to a customer's experience. And it is exactly this variation that we intend to test in our analysis.

To understand whether Bodvarsson's and Gibson's theory has any validity, data on the relationship between probability of tipping and labor-generated value to a client must be analyzed. Previous studies have tried to shed a light on why people tip by comparing tip sizes with variables such as service quality (Lynn, 2003), dining-party size (Pearl, 1988), and customer frequency (Lynn and Grassman, 1990). A 2003 paper by Michael Lynn compared randomized service quality surveys in restaurants to tip sizes and found a weak correlation between satisfaction with service and tipping. According to Lynn, evaluations of service quality accounted for less than 2% of the variability in tips expressed as a percentage of the bill. In Thrane et al. 2020, researchers performed a survey experiment with Norwegian students to test peer effects on tipping behavior. They find that students are 12 to 14% more likely to tip when their friends tip. This study also found a significant effect of perceived service quality and tipping, which is inconsistent with the previous findings in Lynn, 2003. A 2021 paper by Brewster et al. tests the effects of wearing a mask on tipping. The authors conclude from their results that wearing a mask is not likely to, on average, have a meaningful effect on how much restaurant customers tip their servers.

But it was bill size, out of all the variables tested by recent research, that showed the strongest correlation with absolute tip. Correlation with bill size was also the most consistent result, showing up as the best predictor of tips in three independent studies (Lynn et al. 1993; Thrane et al. 2020; Bodvarsson et al. 1997). This suggests that a customer's gratuity could be in fact the result of a hidden social agreement in which tips should be a certain percentage of the bill size. That would explain why when the bill size increases, so too does the tip in dollars.

These studies, however, have focused mainly on restaurant settings. There are many issues that arise from this limitation. In a restaurant it is very common for a waiter to serve more than one customer at once, or for a table to be served by more than one waiter during the same meal. This makes it difficult to measure the amount of value a waiter generated to one single customer. Additionally, a client's experience in a restaurant is much more complex than the waiter or waitress service can account for. There are many other variables at play such as the ambiance of the restaurant and the chef's quality. All of these factors could mean that customers are not actually apt to discern how much value a worker adds to their experience in a restaurant as Bodvarsson 1988 and Jacob 1980 suggested. This raises an important question of whether the persistent tip percentage that these studies observed is really a product of social norms or rather a consequence of clients' lack of information. Perhaps, customers resort to a tipping norm that varies little with their actual experience not because of unwritten social rules, but because they don't have enough information to estimate the value of a worker's labor, just like the firms. This creates a necessity for studies that observe tipping in settings where clients have a better understanding and can quantify the impact of a server's labor in their utility more precisely.

And this is where the choice of ride-hail data enters. Ride-hail is one of the few services that sits at the intersection between industries where firms cannot monitor their worker's activity and clients have full information about the value that a server adds to their experience. Specially When it comes to value of time, the ride-hail industry concentrates economic interactions in which the client is fully aware of the time input necessary to perform a service. This is information that consumers don't have access to in most other markets. Consumers do not know how long it takes for a firm to produce a fridge, a television, or a car. But they have a pretty good idea of how long it takes a driver to take them to their home or their workplace. So in the next section we translate the rational choice theories of tipping behavior to the context of ride-hail service, using time as the main measure of service value.

2.2 The Theory of Tips and Delays

The causal theory we want to test in this paper is that delays in rideshare trips cause passengers to not tip their drivers. The main variables that concern our theory are the binary tip and the deviation on average trip duration. The diagram below illustrates the theory of change we propose, where tipping is the dependent variable and deviation on average trip duration is the independent variable.

Figure 1: Theory of Tips and Trip Delays

Ceteris paribus, passengers should derive a greater utility from trips that are faster. Therefore, a good measure of how drivers create value to their passengers' is how much of their passengers' time they can save by covering a trip in less time than expected. Similarly, a good measure of how drivers destroy value to their passengers' experience is how much of their passengers' time they waste by covering a trip in more time than expected.

The causal mechanism that we suggest is the passenger's rational decision to reward good drivers and punish bad ones, increasing the expected value of their next ride. Better drivers are the ones who arrive earlier than expected, thereby generating more value to the passenger, and bad drivers are the ones that arrive later.

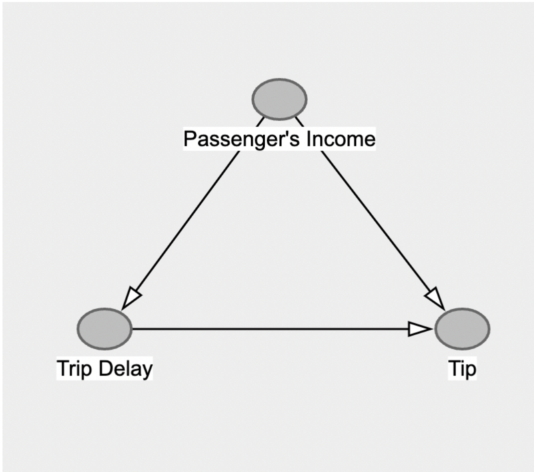
Whether someone gave tip or not to their driver is estimated by an indicator function that assigns 0 to every ride that had a tip equal to $U\$0.00$ and 1 to every ride that had a tip greater than $U\$0.00$. Estimating the deviation on average trip duration associated with each trip requires a bit more work. We start by grouping trips that have a similar path. We create subsets of trips that have a common starting Census tract and a common destination Census tract. Then, we calculate the average trip duration for each of these paths. Finally, we subtract the average trip duration of a path from the total duration of each of the trips that belong to that path subset. This will relate each trip with the difference between their duration and the average duration of a rideshare trip along that path.

This theoretical design needs to account for some important confounding variables. These variables can be divided into two groups: those related to aspects of the trip itself, and those related to passenger's characteristics. In the following section we will describe all observable confounders we wish to control for in order to improve the internal validity of our analysis.

2.3 Controlling for Potential Confounding Variables

Passenger's Income: We began this paper by defining tips as voluntary payments that a client makes to a worker after a service is performed. Therefore, there are good reasons to believe that the decision to tip is determined by a passenger's budget to some degree. There are also reasons for us to believe that a passenger's budget can affect their trip's delay. For instance, it could be argued that people who have lower budgets reside in neighborhoods that have less reliable roads, or that drivers want to take extra caution when they drive by.

Figure 2: DAG with Passenger's Income as a Confounder Variable



Income is a variable that generally does a good job at indicating an individual's budget size. Therefore, our analysis will control for a passenger's income. However, as we will see in the next sections of this paper, the data publicly available on ride-hail trips does not allow us to identify personal characteristics of passengers, including their income. Instead, we will use Census data to generate an adequate estimate of the income of each passenger.

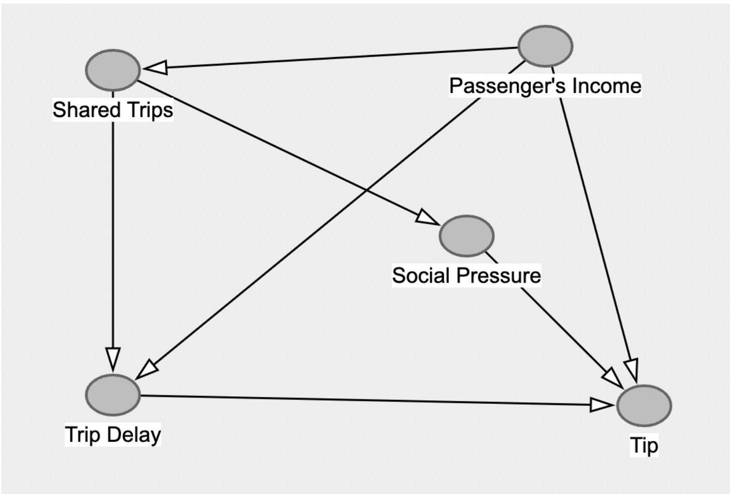
A theoretical argument that permits us to join income date from the Census with the ride-hail database is that we can select rides that occurred only in particular time frames that are characteristic of passenger's who reside in the pickup Census tract. For example, trips that happened between 8AM and 10AM on weekdays are very likely to be ordered by passengers that actually live in their pickup Census tract. Therefore, the average income data of that Census tract is somewhat representative of this passenger's income.

The only problem with this approach is that we will loose a big share of our trip data. Still, because more than 50 million trips occured during 2020, we will have a significant amount of observations from which we can make relevant inferences even after applying this filter. This seems to be a good trade-off as it allows us to add any Census information to our analysis at the low cost of slightly diminishing our test's external validity.

All of these filters reduce the original database from 50 million trips to around 2.2 million trips. We apply these filters, however, so that we can join mean household income information to our analysis

Shared Trips: A good critique to the theory of delay is that there exists a social pressure mechanism in ride-hail trips that is related to trip delays and affects how much tip a passenger awards their drivers. For example, it could be argued that passengers who share trips are exposed to increased levels of social pressure, which causes them to tip higher. At the same time, shared trips are more likely to have more than one stop, which causes delays to be longer. Beyond that, passengers who have lower income are more likely to share trips since it is a more cost-efficient option of transportation than single rides. All of these dynamics are illustrated in the diagram below.

Figure 3: DAG of All Observable Causal Model



There is also a rhetorical value to controlling our results for shared trips. As detailed in the introduction of this paper, one of the reasons why rideshare trips are a better setting than restaurants to test clients' tipping behavior is because it is a less noisy environment and passengers have their experience determined only by the driver's service. By controlling for shared trips we make this assumption stronger since we are essentially controlling for other people in the car. Of course, not every trip that is actually shared is reported as shared on the app, so there is an effect of company in rides that we cannot control for. But controlling for what we observe helps us make this assumption a bit more pertinent.

It is also important to note that there is very little reason to suspect of reverse causality. Ride-hail drivers generally complete a trip with no previous knowledge of the tip size. A passenger could try to influence the duration of their ride while the trip is ongoing by suggesting they will give higher tips if the driver arrives to the final destination earlier. However, this is a rare practice. And even if it does happen, there is no guarantee that the driver will respond to this incentive due to the uncertainty of the passenger honoring this promise.

3 Data and Descriptive Data

3.1 Transportation Network Providers

The data analyzed in this paper comes from the 2018-22 Transportation Network Providers (TNP) database. The TNP is a database made publicly available by the city of Chicago Data Portal. Its unit of analysis are ride-hail trips. In total, there are around 262 million trips recorded in this database. The TNP was constructed in 2018 after updates in Chicago's transportation legislation required ride-hail companies to report trip, driver, vehicle, and compensation data to the city's Commissioner. As stated in the Municipal Code of Chicago, "Each licensee [Uber, Lyft, and Via] shall provide the following data to the Commissioner:

(1) *Trip request data.* A record of each request for a trip made through the licensee's Internet-enabled application or digital platform by a potential passenger;

(2) *Trip data.* A record of each trip which shows where a passenger is picked up and dropped off;

(3) *Driver data.* A record of each of the licensee's drivers who is authorized to pick up passengers using the licensee's Internet-enabled application or digital platform;

(4) *Session data.* A record of each driver session on the licensee's Internet-enabled application or digital platform. A driver's session begins when a licensee's driver activates a mode in the licensee's Internet-enabled application or digital platform, signaling the driver's readiness to receive and respond to trip requests. A driver's session ends when the driver deactivates the mode and is no longer able to receive and respond to trip requests;

(5) *Vehicle data.* A record of each vehicle that is used by each of the licensee's drivers for picking up passengers through the licensee's Internet-enabled application or digital platform;

(6) *Location data.* For every transportation network vehicle and driver combination, location snapshots captured at specified intervals for all times the driver is in session. Each snapshot shall indicate the vehicle's precise location and corresponding date and time;

(7) *Compensation data.* A record of each of the licensee's drivers who is paid an hourly rate, and any other record needed to capture actual driver pay information that is not reflected in licensee's hourly rate compensation records."

The authenticity of this data relies on the honest reporting of the TNP companies Uber, Lyft, and Via, and on the transparency of the Chicago Data Portal. Reports by these institutions tend to be trustworthy since they all come from either public companies or governments. This makes this a very interesting sample to study since it provides us with very accurate information on trip duration, tips, fares, start location, and end location of trips to perform our analysis. It also contains more than 50 million trips recorded, giving us a very representative sample of all ride-hail trips performed in 2020. The database has 21 variables in total. All the variables are listed below. The list items are in the format "Variable Name (Type): Description".

- Trip ID (Plain Text): A unique identifier for the trip;
- Trip Start Timestamp (Date and Time): When the trip started, rounded to the nearest 15 minutes;
- Trip End Timestamp (Date and Time): When the trip ended, rounded to the nearest 15 minutes;
- Trip Seconds (Number): Time of the trip in seconds;
- Trip Miles (Number): Distance of the trip in miles;
- Pickup Census Tract (Plain Text): The Census tract where the trip began;
- Dropoff Census Tract (Plain Text): The Census tract where the trip ended. This column often will be blank for locations outside Chicago.
- Pickup Community Area (Number): The Community Area where the trip began. This column will be blank for locations outside Chicago.
- Dropoff Community Area (Number): The Community Area where the trip ended. This column will be blank for locations outside Chicago.
- Fare (Number): The fare for the trip, rounded to the nearest $2.50.

- Tip (Number): The tip for the trip, rounded to the nearest dollar. Cash tips will not be recorded.
- Additional Charges (Number): The taxes, fees, and any other charges for the trip.
- Trip Total (Number): Total cost of the trip. This is calculated as the total of the previous columns, including rounding.
- Shared trip Authorized (Checkbox): Whether the customer agreed to a shared trip with another customer, regardless of whether the customer was actually matched for a shared trip.
- Trips Pooled (Number): If customers were matched for a shared trip, how many trips, including this one, were pooled. All customer trips from the time the vehicle was empty until it was empty again contribute to this count, even if some customers were never present in the vehicle at the same time. Each trip making up the overall shared trip will have a separate record in this dataset, with the same value in this column.
- Pickup Centroid Latitude (Number): The latitude of the center of the pickup Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.
- Pickup Centroid Longitude (Number): The longitude of the center of the pickup Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.
- Pickup Centroid Location (Point): The location of the center of the pickup Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.
- Dropoff Centroid Latitude (Number): The latitude of the center of the dropoff Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.
- Dropoff Centroid Longitude (Number): The longitude of the center of the dropoff Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.

- Dropoff Centroid Location (Point): The location of the center of the dropoff Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.

3.2 Descriptive Statistics

As stated in the introduction and the theoretical framework, this analysis is to the subset of approximately 11 million ride-hail trips from the TNP database that occurred form 2018 to 2022, on weekdays, from 8AM to 10AM. Having established our observations, we must now detail the statistics of the main variables in our regression analysis. They are 'tip', 'deviation on average trip duration', and the control variables 'shared authorized approved', and 'mean household income'. Each of these variables is listed below with a brief description and its data type.

- Tip: Indicator function for whether the trip had a tip or not. Data type numeric, binary.

- Deviation on Average Trip Duration (or "Delay"): Difference between total trip duration and average trip duration from the ride's pickup Census tract to its Drop Off Census tract rounded to the nearest minute. Data type numeric, integer.

- Shared Trip Authorized: Indicator function for whether the trip was shared with another rider through the app. Data type numeric, binary.

- Census Tract Household Mean Income: Mean income for each pickup Census tract where a trip started. Income values are rounded to the nearest dollar. Data type numeric, integer.

First, we look at trip delays. The minimum delay found among our sample was -240 minutes. This means that the least delayed trip arrived around 4 hours earlier than the expected path time. The maximum delay was 1274 minutes, meaning that the most a trip was delayed in our sample was 21 hours. Delays of this magnitude are very uncommon and unrealistic, so these are likely to be irregular rides, or just very rare phenomena.

Figure 4: Distribution of Travel Time Delays

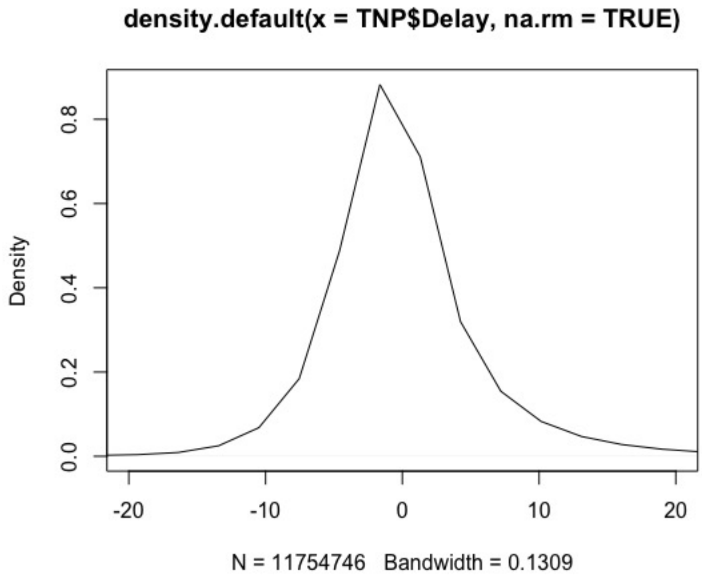


Table 1: Trip Delays (in Minutes)

Min	1^{st} Quartile	Median	Mean	3^{rd} Quartile	Max
-240.116	-2.926	-0.661	0.000	2.002	1266.023

Most trips are concentrated between the intervals of +10 and -10 minutes of delay. This can be seen very clearly in the density graph below. Therefore, a more reliable statistic is the 1^{st}, 2^{nd}, and 3^{rd} quartile, which show that the majority of trips are somewhere 2.5 minutes early or delayed with respect to the average path time, with a very symmetric distribution centered near 0.

Next, we have the variable Tips in percentage. Tips have a very skewed distribution to the left, as we can observe from Figure 5. Most values are concentrated around the minimum of U$0.00. Out of the 11 million trips observed in this study, 81.28% — almost 9 million trips — did not give any tip. So it was expected that the distribution of tips would be skewed to 0. The average tip in percentage was near 3%.

Figure 5: Distribution of Tip in Percentage

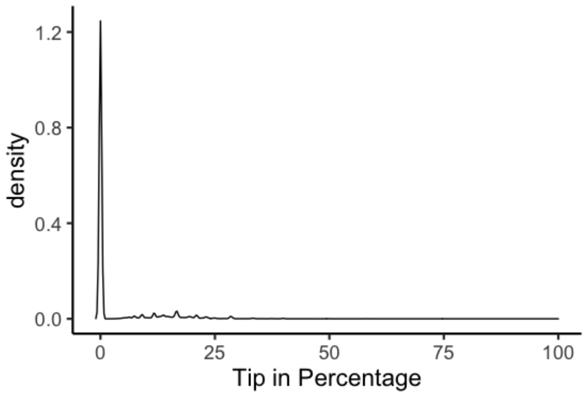
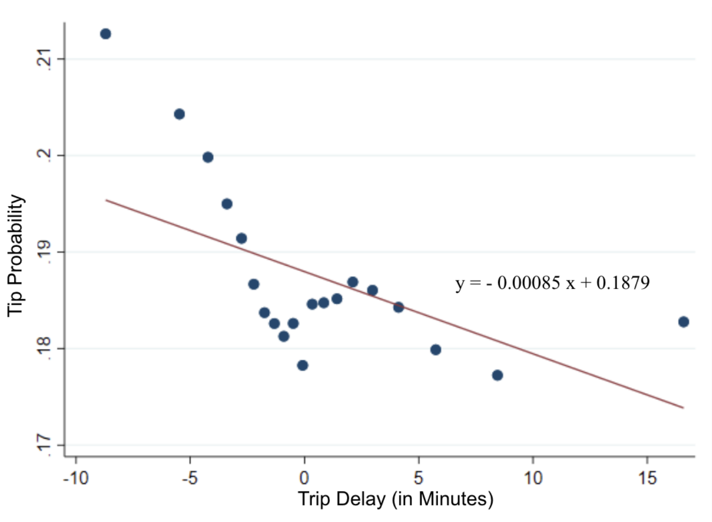


Figure 6 shows us the correlation between our dependent and independent variable: trip probability and trip delay. This scatter plot was created by splitting the 11 million trips into 20 bins with the same number of observations and mapping their average to the plot. The trend line is an OLS regression that uses all the 11 million trips observed. Without any control for confounding variables or fixed effects, the relationship between tipping probability and delay is negative, as expected by the theory of tips and delays. However, it is a very weak relationship if we consider it at the individual trip level. Every 10 minutes a driver is delayed, for instance, is associated with a drop in the tip probability of only 0.85%.

Figure 6: Correlation Between Tipping Probability and Trip Delay



Now, we move to the description of our control variables. Our first control variable, Shared trip Authorized, is a binary variable that can only take up the values of 0 and 1. In the sample we of trips we observed there was a mean of 0.1353 shared trips. This means that around 13.5% of trips were shared between different passengers. This is likely an underestimation of the amount of passengers that actually shared ride-hails due to under reporting from the passenger's end. That is, many passengers share trips but are not willing to report it on the app most likely because they do not want to or cannot share the total cost of the ride with the other passenger.

Our second control variable, passenger's expected household income, is a continuous variable and has its distribution represented in Figure 7 below.

Figure 7: Frequency of Trips by Passenger's Income

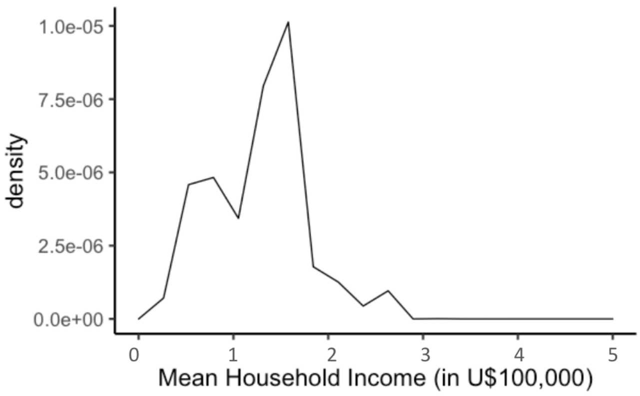
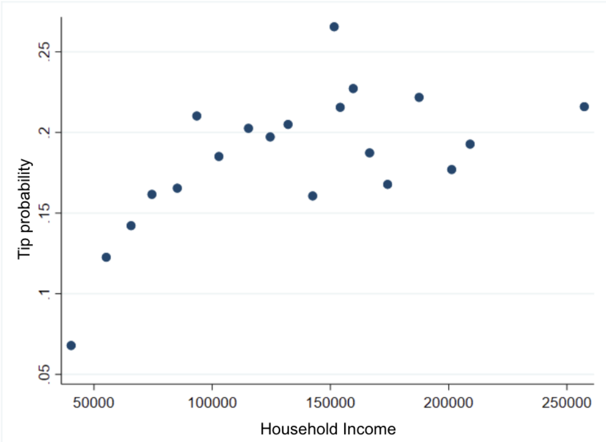


Table 3: Mean Household Income (in Dollars)

Min	1st Quartile	Median	Mean	3rd Quartile	Max
21,704	88,810	133,564	133,321	168,992	448,853

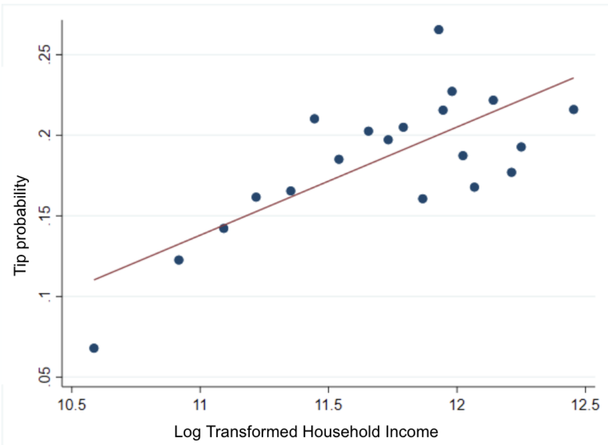
This table shows us that the Census tract that has the minimum mean household income has a mean household income of U$21,704. The maximum has an income of U$448,853. We antecipate there to be significant systematic differences associated with tipping behavior between tracts due to the differences in their mean income. We can find evidence for this suspicion from the next two plots.

Figure 8: Correlation between Tip and Passenger's Expected Income



The scatter plot in Figure 8 suggests that the relationship between tip probability and household income is positive for lower income household, but then reaches a plateau as it reaches the higher values of household income. Figure 9 shows how, by performing a log transformation on household income we arrive at a scatter plot with an an approximately linear relationship between tipping and household income.

Figure 9: Correlation between Tip and Log Transformation of Passenger's Expected Income



Because we suspect income influences tipping probability and that it may also influence trip delays, it is important for the external validity of our study to have an estimate of passenger's income as one of our control variables.

The last table in this section is a compilation of all the correlations between the dependent and independent variables and the control variables.

Table 4: Correlation Between Variables of Interest and Control Variables

	Tip Probability	Tip Probability	Delay	Delay
Shared Trip	-0.1279156***	-	3.892113***	-
	(0.0003356)	-	(0.0047529)	-
Mean Household Income	-	4.86e-07***	-	-3.67e-07 ***
	-	(2.16e-09)	-	(3.16e-08)
Observations	11379883	10745414	11381372	10730671

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

One interesting statistic is the correlation between Delay and Shared Trip. There is an increment in trip delay of 3 minutes associated with sharing a trip instead of not sharing a trip. This matches the expectation of the relationship between these variables that we set in the construction of our theoretical model, namely that sharing trips increases the number of stops between pickup and destination, and therefore increases the deviation from average path time. Of course, this is just a descriptive data. Nevertheless it is useful by reaffirming the necessity of this control variable in our model. All of the betas from these correlations are statistically significant. Note, however, that this is not unexpected since we are studying a relatively big dataset.

4 Empirical Framework

The statistical test used to investigate the causal effect associated with trip delay and tips is a linear regression of deviation from average trip duration on whether a trip had tips or not (TipIndicator), controlling for the confounding variables 'shared trips' and 'passenger's income'. The sample regression equation is written below, with TipIndicator as the dependent variable and deviation from average trip duration as the independent variable. The coefficient we are interested in is $\hat{\beta}_1$. $\hat{\beta}_1$ represents the approximate change in $TipIndicator$ associated with an additional unit of $Delay$ accounting for the controls and route fixed effects.

$$TipIndicator_i = \hat{\beta}_1 Delay_i + \hat{\beta}_2 Shared_i + \hat{\beta}_3 Income_i + RouteFixedEffects + u_i$$

With $i = 1, \ldots, n$.

Our causal identification strategy is achieved through a combination of the linear regression and two key assumptions. The first assumption is that the relationship between the dependent and independent variables is in fact linear. The second is the conditional-independence assumption, that is that the common variables that affect delays and tips be observable. The dependence between delay and whether a passenger tips or not can be removed by conditioning on observable variables passenger's income and shared trips. Having both of these assumptions we can reach a new interpretation of $\hat{\beta}_1$ as the approximate change in $TipIndicator$ (in percentage) caused by an additional unit of $Delay$ (in minutes).

If the causal theory of tips and delays is true, then this test will show us a negative correlation between probability of tipping and trip delay within the entire 95% confidence interval. This result should be both statistically (p-value $<$ 0.5%) and substantively significant (non-trivial change in probability of tipping promoted by change in trip delay). But whether this theory is rejected or not, the findings of this paper will contribute to the understanding of why people tip. If we are able to reject this causation theory, then this paper will serve as evidence to the prevalent social norm interpretation of tipping behavior. If we are not able to reject it, then there will still be reasons to suspect that tips are motivated by rational choice.

5 Results

5.1 Testing the Theory of Tips and Delays

The results of our linear regression controlled by both 'shared trips' and 'passenger's income' as well as the route fixed effects can be seen in the Table 5 below. $\hat{\beta}_1$, in particular, is equal to -0.000145. This tells us that the likelihood of a passenger tipping decreases by approximately 0.015% with each additional minute of delay. This result is significant at the $p < 0.001$ level. Again, not surprising. And the 95% confidence interval is $(-0.000021, -000124)$, which means that the causal relationship between tipping probability and delay is confidently below zero. That is, it would be extremely rare to observe a null or positive effect of delays on tipping behavior.

Table 5: Regression of Tips and Delay with Fixed Effects

	Tip Probability
Delay	-0.000145***
	(0.0000210)
Shared Trip	-0.0948***
	(0.000369)
Household Income	0
	(.)
Constant	0.294***
	(0.000432)
Observations	10652140
FE (Routes)	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6 in the next page contains the same estimated effect plus partial results from OLS regressions with different combinations of controls. The regression controlling only for route fixed effects shows us that for every unitary increment in delay (in minutes) the likelihood of tipping decreases by 0.153%. This was the largest effect of delay on tipping we found. A graph of this linear model can be seen in Figure 10 on page 22.

This stronger negative association, however, seems to dissipate as we control for more variables, and we end up with the weaker negative effect that we originally observed in Table 5. The results of all these regressions is statistically significant (p-value < 0.001). Once more it is important to note that all of these regressions had a unusually big sample size, so other adjustments should be in order in case future work wants to apply this model for other scenarios beyond the specific one that we limit ourselves to in this paper.

An interesting observation is that the effect of delay on tipping behavior drops much more significantly when we introduce control for shared trips than for any other confounding variable or fixed effect. In particular, the effect of delay drops by a factor of 10, approximately. What is more curious is that the effect sharing a trip has on tip probability is negative. This has important consequences for the main arguments that the Behavioral Economics have about the effect of being observed by friends or other individuals during our tipping decision. We shall discuss this topic in more depth in the next section of this paper.

Table 6: Table with All Regressions

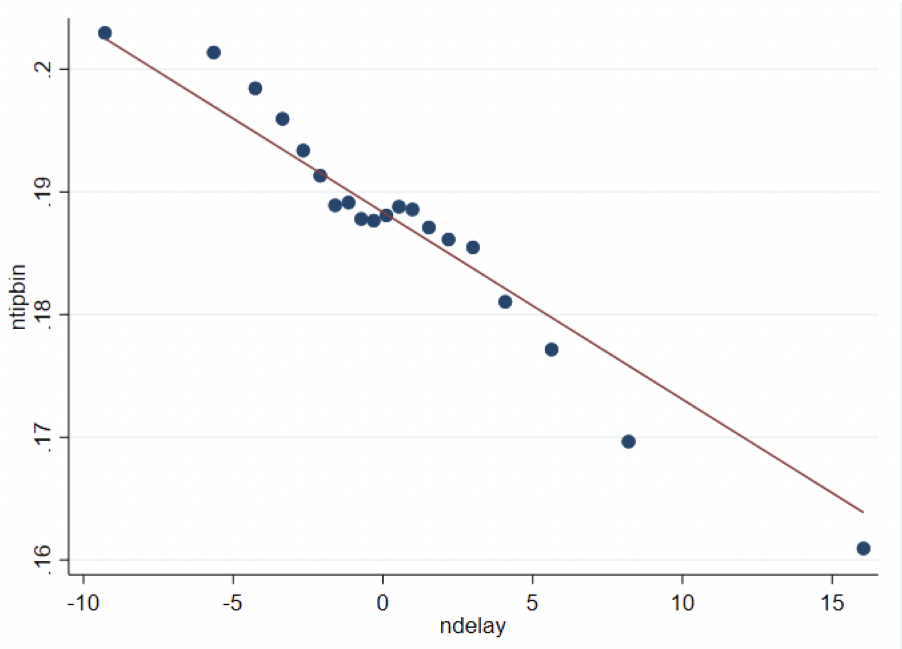
	Tip Probability	Tip Probability	Tip Probability	Tip Probability
Delay	-0.00153***	-0.00155***	-0.000110***	-0.000145***
	(0.0000200)	(0.0000204)	(0.0000206)	(0.0000210)
Income		0		0
		(.)		(.)
SharedTrip			-0.0962***	-0.0948***
			(0.000362)	(0.000369)
Constant 0.188***	0.189***	0.187***	0.297***	0.294***
(0.000116)	(0.000113)	(0.000115)	(0.000423)	(0.000432)
Observations	11301507	10652140	11301507	10652140
FE (Routes)	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As we anticipated, the next plot illustrates the effect of travel time delays on tipping probability while controlling only for the fixed effects of the routes. To create this plot we divided our sample in 20 bins, each with the same number of observations. Then, we took the average tip probability of the observations in each bin and plotted the data points back to the graph. Controlling for fixed effects of the trip routes we get the trend line that is graphed below. This trend line shows that at 0 minutes of delay passengers have nearly 19% of giving a tip. If delay increases by 5 minutes, then the probability of tipping decreases by 0.85%.

Figure 10: Relationship Between Tipping and Delay with Routes FE



Beyond the partial measure of the effect of delays in tipping behavior, this plot also suggests some other characteristics of passengers tipping behavior. First, we can observe how passengers display higher tolerance for delays around 0 to 3 minutes. This is evidenced by the fact that tipping probability remains approximately constant between delays of -3 to 3. So the same way that they do not penalize small delays harshly, they also do not reward small early arrivals generously. This plot also gives support to the claim that passengers are less willing to reward drivers for saving their time then they are willing to punish drivers for wasting their time. We can observe this from the fact that trips that arrive earlier than expected have their tip probability increase at a slower rate than trips that arrive later have. For instance, the average tip likelihood of trips that had around 9 minutes of delay is around 17%, or approximately -1.9% than if they had arrived on time. At the same time, trips that arrived around 9 minutes earlier than expected had a tip probability of 20.1%, or approximately 1.2% higher than if they had arrived at the expected time.

6 Discussion and Conclusion

The 3 main findings from the analysis of the 11 million sample of trips are:

1. Evidence of a statistically significant negative causal effect of trip delays

on the likelihood of a passenger tipping.

2. Evidence that passengers interpret small delays and early arrivals similarly.

3. Evidence of punishing large delays at a harsher rate than they reward large time saving outcomes.

Let us begin by considering the impact of the 1st result in our theory of tipping behavior and delays, as well as its impact in the broader service quality and behavior modelling debate. According to the rational theory of tips and delays described in this paper, after controlling for all observables and fixed effects, we should have observed a negative effect of delays in the likelihood of tips. This is exactly what we see. As travel time delay increases, probability of tipping decreases. If we can trust the statistical significance level and the standard errors of our analysis, then we have enough evidence to reject the theory that travel time delays cause passengers to tip more or have no effect on passengers' tipping behavior. And as we suggested in the beginning of the paper, this is a results that supports the idea of a rational model that can explain the tipping behavior, despite what is mainly believed in the Behavioral field.

It is also true, however, that this is a relatively small effect at the level of individual trips. Arriving at the destination 10 minutes earlier or 10 minutes later represents a change of approximately 0.3% in likelihood of a driver receiving a tip. The substantive significance of this finding, therefore, is not as clear. At the industry level, however, this estimate gains more significance. The yearly average number of trips in the period between 2018 and 2022 was 52.4 million TNP trips in Chicago. Out of these 52 million, only around 9 million give tips assuming there are no delays per ride. An average delay of 2 minutes, however, would cost the industry 1.62 million tipping rides. And this has a somewhat substantive significance for the income of individual drivers.

The 2nd finding gives us hints of what a rational model of tipping would look like in the context of our empirical analysis. By providing evidence that passengers interpret small delays and early arrivals similarly, our 2nd finding suggests that the linear model used in this paper cannot properly address what seems to be a systematic tipping behavior. Future modellings of tip behavior would have to take this into account.

The 3$^{\text{rd}}$ finding also gives us characteristics that a more predictive tipping model would have to take into account. To be more precise, the model would have to consider the fact that great losses are punished more harshly than very early arrivals. But more importantly, our 3$^{\text{rd}}$ finding sheds light on a topic that is very dear to the field of Behavior Economics: loss aversion. It is often claimed by behavioral economists that negative events have a larger weight than positive events. According to these scientists, individuals have a bias to respond to losses more poignantly than they do to gains. The original evidence of this phenomena was reported by Kahneman and Tversky (Econometrica 47:263–291, 1979).

However, recent critics suggest that loss aversion proponents have over interpreted these findings. While commenting on Kahneman and Tversky's work, Echiam 2019 states that "Specifically, the early studies of utility functions have shown that while very large losses are overweighted, smaller losses are often not." (Psychological Research 83, 1327-1339, 2019). Our finding is consistent with this characterization of loss aversion as we also do not find evidence for this bias in small delays, but only in larger delays. And evidence of loss aversion limited only to large delays is not supportive of behavioral explanations to tipping behavior. In fact, it is quite the opposite. Avoiding large losses more than valuing large gains is compatible with rational behavior. Large losses, be them money or time losses, have the potential to destroy your life. It is not a cognitive bias to take extra caution when a decision can put your life at risk, make you lose your house, or harm your family. No matter the upside, it makes sense to avoid extreme and definitive losses.

In conclusion, the analysis of tipping contained in this paper adds to methodological and substantial debates within the Tipping Behavior literature and the broader Behavioral Economics field. Our study contributes to the field by finding a new environment to conduct our analysis, moving away from the restaurant market. It innovates in analytical method by conducting a study on data with large datasets, which are not common for the field. And all these innovations come with improvements to the quality and confidence of our results.

There are, of course, limitations to our analysis. The lack of personal identifiable data on both passengers and drivers associated with the trips, for example, prevents us from controlling for other relevant variables such as driver's app rating, passenger's age, or driver ethnicity. And our main findings, though consistent with previous literature, are also limited and must be submitted to further inquiries in order to improve our understanding of tipping and other behaviors claimed to be irrational.

References

Conlisk, Sarah. "Tipping in crises: Evidence from Chicago taxi passengers during COVID-19." Journal Of Economic Psychology (2022): 102475.

Azar, Ofer H. "Why pay extra? Tipping and the importance of social norms and feelings in economic theory." The Journal of Socio-Economics 36, no. 2 (2007): 250-265.

Azar, Ofer H. "The social norm of tipping: A review 1." Journal of Applied Social Psychology 37, no. 2 (2007): 380-402.

Azar, Ofer H. "What sustains social norms and how they evolve?: The case of tipping." Journal of Economic Behavior Organization 54, no. 1 (2004): 49-64.

Lynn, Michael. "Restaurant tips and service quality: A weak relationship or just weak measurement." International Journal of Hospitality Management 22, no. 3 (2003): 321-325.

Lynn, Michael, George M. Zinkhan, and Judy Harris. "Consumer tipping: A cross-country study." Journal of Consumer Research 20, no. 3 (1993): 478-488.

Saunders, Stephen G., and Michael Lynn. "Why tip? An empirical test of motivations for tipping car guards." Journal of Economic Psychology 31, no. 1 (2010): 106-113..

Chapman, Gretchen B., and Jennifer R. Winquist." The magnitude effect: Temporal discount rates and restaurant tips." Psychonomic Bulletin Review 5, no. 1 (1998): 119-123.

Azar, Ofer H. "The economics of tipping." Journal of Economic Perspectives 34, no. 2 (2020): 215-36.

Thrane, Christer, and Erik Haugom. "Peer effects on restaurant tipping in Norway: an experimental approach." Journal of Economic Behavior Organization 176 (2020): 244-252.

Brewster, Zachary W., and Kenneth Gourlay. "The “Mask Effect” on the tips that customers leave restaurant servers." International Journal of Hospitality Management 99 (2021): 103068.

Bodvarsson, Örn B., and William A. Gibson. "Economics and restaurant gratuities: Determining tip rates." American Journal of Economics and Sociology 56, no. 2 (1997): 187-203.

CHICAGO, IL., MUNICIPAL CODE OF CHICAGO ch. 9, § 115 (2022).

City of Chicago, 2022. Transportation Network Providers – Trips [data file]. Chicago Data Portal. Retrieved from: https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p.

Theory References

Kahneman, Daniel, and Amos Tversky. “Prospect Theory: An Analysis of Decision under Risk.” Econometrica, vol. 47, no. 2, 1979, pp. 263–91. JSTOR, https://doi.org/10.2307/1914185. Accessed 29 Sep. 2022.

Echiam, E. Acceptable losses: the debatable origins of loss aversion. Psychological Research 83, 1327–1339 (2019). https://doi.org/10.1007/s00426-018-1013-8

Data Set References

Hegde S, Abkarian H, Mahmassani HS. Characterizing Ride-Hailing Driver Attrition and Supply in the City of Chicago Through the COVID-19 Pandemic. Transp Res Rec. 2022 Aug 24:03611981221117542. doi: 10.1177/03611981221117542. PMCID: PMC9412150.

Credit, Kevin, Gustavo Dias, and Brenda Li. "Exploring neighbourhood-level mobility inequity in Chicago using dynamic transportation mode choice profiles." Transportation research interdisciplinary perspectives 12 (2021): 100489.

Methodological References

Correia, Sergio. 2017. "Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator" Working Paper. http://scorreia.com/research/hdfe.pdf

Noah Constantine, Sergio Correia, 2021. reghdfe: Stata module for linear and instrumental-variable/GMM regression absorbing multiple levels of fixed effects. https://ideas.repec.org/c/boc/bocode/s457874.html