

Ridesharing, Spatial Frictions, and Urban Consumption Patterns

Zhe Zhang*, Beibei Li†

April 2019 (*DRAFT - Private*)

Abstract

With the pervasiveness of mobile technology and location-based computing, new forms of smart urban transportation, such as Uber and Lyft, have become increasingly popular. These new ridesharing (or e-hailing) platforms can influence individuals' movement frictions, in turn influencing local consumption patterns and the economic performance of local businesses. To gain insights about the future impact of urban transportation changes, in this paper, we analyze individuals' urban consumption patterns before and after their adoption and usage of the ridesharing services. Our study is validated using a novel, anonymized panel dataset of fine-grained individual credit card and debit card transactions from January 2012 through May 2016 from a large U.S. bank, including from over 10K observed adopters of ridesharing with 7M transactions. Based on revealed preferences, we hypothesize that those who choose to use ridesharing more are those whose mobility frictions are affected more by the availability of ridesharing, and vice versa. Our findings demonstrate a significant positive impact from the usage of ridesharing services on individuals' consumption frequency, as well as the spatial diversity of their spending. Our results also indicate strong heterogeneity in such effect. The effect becomes significantly stronger in increasing restaurant and bar transactions, and also stronger among younger customers and those who spend less before adopting the ridesharing services. Lastly, we see evidence of the heterogeneous impact of spatial frictions between different neighborhoods of a major city.

*Heinz College, Carnegie Mellon University, Pittsburgh, PA - 15213; zhezhang@cmu.edu

†Heinz College, Carnegie Mellon University, Pittsburgh, PA - 15213; beibeili@andrew.cmu.edu

1 Introduction

Peer-to-peer car transportation (i.e., ridesharing or e-hailing) services have become increasingly popular globally. In the United States, Uber and Lyft are the two most popular businesses providing peer-to-peer platforms for these new forms of transportation. During the initial growth and rise in popularity of these platforms, there was much discussion on their urban impacts, and policy efforts made to block their entry into metropolitan areas. Today though, even several years after their entry into major U.S. cities, there continue to be urban policy efforts surrounding these services. Some of this effort seeks to regulate or curtail the use of ridesharing services, as in New York City’s 2018 proposed cap on the number of available Uber providing vehicles. To help evaluate potential urban management and regulation of these services, it is important to know: with the large popularity of these services, is ridesharing simply replacing previously existing transportation modes, or does their popularity signal that ridesharing has also changed urban behaviors?

On the other hand, in contrast to efforts to curb ridesharing, other recent urban policies are promoting the use of ridesharing services as a new form of public transit. In 2017, Innisfil, Canada chose to subsidize Uber for its citizens as a form of public transit, in lieu of adding an additional bus line. Other cities have subsidized Uber or Lyft for specific purposes or demographic groups. This includes Pinellas, Florida subsidizing Uber for low-income workers who work late-night shifts; Evesham, New Jersey subsidizing rides at night originating from a select area of town with bars and restaurants; or Boston, Massachusetts subsidizing Uber for paratransit users. If peer-to-peer ridesharing services are taking on a role as an additional form of public transit or urban infrastructure, what impacts does this new infrastructure have? Further, does the impact of ridesharing differ meaningfully along demographic characteristics?

Based on the above motivations, in this paper, we use a novel dataset to study these questions. Looking forward too, cities, urban planners, and businesses are interested in understanding how the entry of autonomous, self-driving vehicles that provide cheap and spatially-unrestricted transportation might affect urban life and activity too. Both ridesharing and autonomous vehicles can be framed as a reduction in spatial frictions (i.e., a generalized measure of transportation costs). The goal of this paper is to use the entry of ridesharing services in the U.S. as a case study on a shock

to spatial frictions. In this paper, we observe the fast-growing popularity of ridesharing and we observe individual ridesharing transactions, and can frame this new service as a shock that reduces the spatial frictions of a subset of consumers, that are revealed by their choice to pay for these services. We study the first-order effect of ridesharing services on these affected consumers. The insights of this paper are useful for understanding this new form of urban transportation globally, but also as a case study for future changes in urban transportation and spatial frictions.

In this paper, we specifically focus on two empirical questions on the first-order impact of ridesharing services. First, among those who utilize ridesharing services, has this affected their offline consumption behavior? It is not clear, even if spatial frictions are in fact lower, that offline consumption choices will be affected; ridesharing may just replace existing transportation. Second, among that same group, has ridesharing affected their spatial and geographic patterns of their offline consumption? From an urban studies perspective, there is interest in the demographic segregation of urban consumption (Davis et al., 2017). We are interested to see how a shock to spatial frictions affects consumers' geographic consumption patterns. Additionally, in our results, we consider potential heterogeneity in the effects of ridesharing, including on the demographics of consumers or the categories of merchants.

We focus on those who utilize ridesharing services because they are the consumers who demonstrate, via revealed preferences, that their spatial frictions, and thus, choice in transportation modes, have been affected by the availability of ridesharing. I.e., we assume that those who choose to pay for these ridesharing services are those who are affected by it, specifically their spatial frictions are reduced; while for others who do not use it, we make a simplifying and conservative assumption that they are not affected. Subsequently, the key driver or treatment variable of interest is a consumer's observed usage of ridesharing. Since both the adoption and usage of ridesharing is self-selected, we seek to identify establish a causal effect of ridesharing usage by carefully testing against and controlling for potential unobserved individual trends that may affect both consumption and ridesharing usage. If a consumer does not utilize ridesharing's new services, we assume that their spatial frictions have not been affected; further, we assume that consumers who utilize ridesharing more are also those whose spatial frictions have been affected more.

The key data we utilize are an anonymized panel of anonymized individual credit card transactions from 2012 - 2016, including from over 10K anonymous accounts which we identify as being users of the Uber or Lyft ridesharing platforms. Our central identification focuses on studying the relationship between an account’s usage of ridesharing and changes in their post-ridesharing-adoption period compared with their consumption in a corresponding pre-adoption period. We compare ridesharing users against both low-usage adopters and a matched control group, and also seek to control for potential individual time trends or unobserved confounders. We supplement the analysis with various falsification and robustness tests. Further, there are no pre-trends between consumption patterns and an adopter’s future usage of ridesharing; usage is uncorrelated with pre-adoption spending frequency, spending amount, or spatial patterns. In the data, we are also able to remove online and ridesharing transactions, thus considering ostensibly offline transactions.

In our main results, we find that, consistent with our model of revealed preferences and reduced spatial frictions, there appears to be no impact from adoption itself, but instead only in those who utilize ridesharing more. Post-adoption, relative to their earlier, pre-adoption consumption in the same time-of-year period, those who utilize ridesharing seem to increase both their consumption frequency and their number of unique zip codes visited. In other words, a consumer’s actual usage of ridesharing services significantly explains variations in consumer spending post-adoption, relative to pre-adoption spending in that same time-of-year period. We also find substantial heterogeneity in this relationship of usage and post-adoption changes, along the dimensions of merchant category, account age, and account pre-spending.

While we control for individual behavior before ridesharing availability, the major confounding concern is that there are unobserved, individual-specific time trends that affect both ridesharing usage, and consumption behavior. Ridesharing adoption may also be related to individual time trends; however, this paper does not focus on nor identify any adoption effects. Rather, its focus is on those consumers who choose to use ridesharing frequently. We test against several possible types of confounders that may show up in other signals, and find that, instead, our main results are robust to robustness tests and do not show signs of the presence of confounders. These tests include a propensity-score matched control group, controls for online spending, average spending metrics,

simulated earlier adoption dates to test for pre-trends, alternative model specifications, comparisons across merchant categories, and focusing our sample on extreme weather days.

To motivate the rest of the paper, we show a figure that captures in visual, model-free evidence the argument of the analysis. Figure 1 shows data from the subset of ridesharing users that we observe in our data. On the y-axis, it plots the change in consumption between post-adoption and pre-adoption consumption frequency; on the x-axis, it plots how much each user chose to use ridesharing in the first 9 months post-adoption. A nonparametric flexible line plotted on this scatterplot visualizes the main identification and theme of this paper: those low-usage adopters have no change on average in their consumption behavior, but those higher-usage adopters, who reveal themselves to be affected by ridesharing, show significant positive increases in local urban consumption activity.

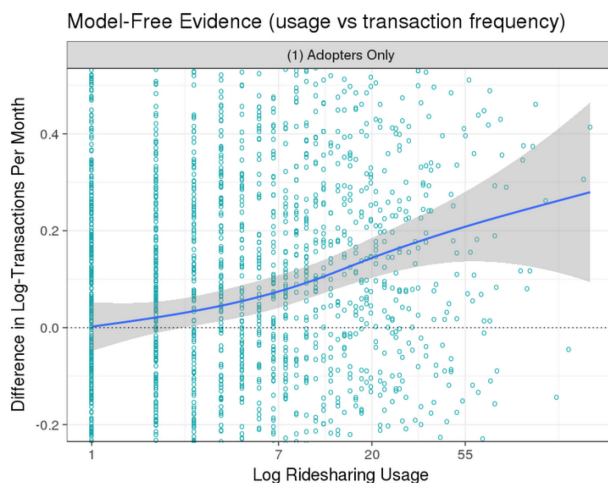


Figure 1: Model-free evidence that previews a central argument of the paper, that the higher-usage adopters, who reveal that their spatial frictions are affected by ridesharing, are the only consumers who show systematic changes after ridesharing availability.

Overall, this research contributes to a better understanding on the impacts of transportation infrastructure and of spatial frictions. Managerially, looking forward, this helps urban managers consider how self-driving vehicles may influence consumer choices, as well as, adding contemporary and recent evidence on the potential pros and cons of the recent growth of not only Uber & Lyft, but ridesharing services in general. While Uber & Lyft may be stable services in U.S. cities, today, there is still a variety of political discussion on the regulation or promotion of such ridesharing

services.

1.1 Related Literature

Our work contributes specifically to a few areas of literature. It contributes to study of the impact of the large ridesharing phenomenon (e.g. Uber) on cities and individuals, and also an area of urban economics studying the urban spatial frictions, city consumption amenities, and public transit infrastructure. It also contributes to the research area on the societal and broader impacts of information technology.

On the recent research that has studied specifically the recent phenomenon of the growth of ridesharing services and its societal impacts, it has been at the macro-level rather than the individual consumer level. Studies on the entry of Uber on impacting urban or consumer aspects, using differences-in-differences analyses and regional variations in ridesharing entry times, have been studied on the regional effects of reducing drunk-driving incidents (Greenwood and Wattal, 2017), traffic congestion (Li, Hong, and Zhang, 2016), entrepreneurial activity measured using Kickstarter campaigns (Burtch, Carnahan, and Greenwood, 2018), individual bankruptcy rates (Nian, Zhu, and Gurbaxani, 2016), and public transit usage (Hall, Palsson, and Price, 2018; Babar and Burtch, 2017). Our study differs in focusing on consumption activity, at the individual-level, and highlighting new aspects such as spatial diversity and individual heterogeneity in the impacts of ridesharing. An individual-based analysis of Uber on consumers is (Cohen et al., 2016), which uses individual-level observations and price-demand variation to estimate the consumer surplus generated by the Uber ridesharing platform. In our paper, we make the assumption, by revealed preferences, that those who choose to utilize ridesharing are gaining surplus and lower spatial frictions, but we are interested to see if these affect their subsequent offline choices. Other studies with more detailed data on the Uber platform have focused on the platform and its market aspects. For example, data using individual Uber-related data to study the impact of ridesharing has been used to study consumer surplus from Uber Cohen et al. (2016) and labor supply insights on the Uber driver-side behaviors and motivations (Hall and Krueger, 2018; Cook et al., 2018; Angrist, Caldwell, and Hall, 2017).

In urban economics, there is interest in the role that space and spatial frictions have on consumer

lives and behaviors. Davis et al. (2017) use a discrete-choice model to highlight the importance of segregation in urban consumption, and how both spatial and social frictions influence this segregation in New York City restaurant visits among Yelp users. They find a “first-order role for spatial frictions in determining the geography of consumption [...], halving the minutes of travel time to a venue implies that a user would be two to nearly four times more likely to visit the venue from that origin”. Our paper complements this work by considering the geographic spread of an individual’s consumption and suggesting an increase in the diversity of neighborhood consumption among those affected by ridesharing. We also provide empirical evidence of a shock to spatial frictions and how this shock affects consumers’ subsequent consumption choices. Black, Kolesnikova, and Taylor (2014) empirically study the role of commuting times, a specific form of spatial frictions, specifically on labor supply of married women and find a particularly strong effect of commute times on women. Glaeser, Kahn, and Rappaport (2008) argue that low-income people live in cities because their preference for access to public transportation and the high costs of automobile ownership outweigh the higher land costs; they also point out that low-income people are more reliant on being close to transit hubs. Our paper relates to these by considering potential heterogeneity via income in the consumption-ridesharing elasticity. Glaeser, Kolko, and Saiz (2001), Couture and Handbury (2017), and Couture (2013) highlight the importance of urban consumption amenities, the latter two using consumption benefits to explain city-center revival and show that benefits urban density come in non-tradable variety-seeking. Both of these findings are consistent with our paper, which shows some non-tradable services (e.g. restaurants) having the strongest elasticity with respect to the entry of Uber availability, and also shows an increase in neighborhood variety-seeking due to ridesharing lowering spatial frictions. Eizenberg, Lach, and Yiftach (2016) study the role of spatial frictions in explaining variation in grocery prices in Jerusalem.

More broadly, our work relates to studies on the growth of the broader sharing economy, and on the societal impacts of changes in information/digital technology. In this vein, papers have looked at the entry of Craigslist had impacts beyond on the classified ads sector, such as on local newspapers (Seamans and Zhu, 2013), or on HIV transmission and prostitution trends Chan and Ghose (2014) and Mojumder, Chan, and Ghose (2016). Forman, Goldfarb, and Greenstein (2012)

study the impact of Internet investment growth on local wages.

Overall, relative to these bodies of literature, we contribute an analysis of offline and local consumption, using a panel detailed individual consumption data. Using this data, we also contribute an analysis of the spatial nature of consumption, getting a sense of individual travel patterns and interactions with the city. As noted in the above literature review, in addition to the new questions addressed here, our findings also complement and expand in detail upon the insights of the earlier literature. Further, we are able to note differences in the impact of ridesharing on those who use it a little versus those who use it substantially, and how ridesharing may influence the consumption choices differently between different types of consumers.

In the rest of this section, we present a theoretical framework that motivates our hypotheses. In Section 2, we describe the data and measurement of the outcome and treatment variables. In Section 3 and 4, we discuss the main methodology and the results on consumption frequency, as well as heterogeneity in the effects. In Section 6, we use the same methodology to study the geographic impacts of ridesharing, including on the spatial diversity of individual consumption. In Section 6, we consider a variety of robustness checks and alternative specifications on our main findings. We conclude in Section 7 and discuss managerial and policy implications.

1.2 Theoretical Framework & Motivation

To motivate some of our hypotheses, we describe a simple revealed-preferences framework of spatial frictions (i.e. transportation costs). This is not a model, but a motivating framework. This framework centers around the cost of traveling at any given moment t , from origin o to destination d —where “cost” represents spatial frictions beyond just monetary price, such as time, effort, or discomfort. For conciseness, we refer to this set of costs as C_{itod}^* . This is similar to theories of urban spatial interaction modeled as pairs of origins and destinations, each with costs of movement, as surveyed in Batty (2013). Importantly, we conceive C_{itod}^* that represents the lowest-cost mode of transport for each $\{t, o, d\}$ tuple. This includes various possible options such as bus, personal car, walk, ridesharing services, or asking a friend. Each mode has its own associated costs, and we assume C_{itod}^* is equal to the lowest cost mode at time t , i.e., $C_{itod}^* = \min\{C_{itod}^{bus}, C_{itod}^{car}, C_{itod}^{walk}, C_{itod}^{ridesharing}, \dots\}$.

The first takeaway from this framework is that observed ridesharing transactions reveal that

ridesharing was the lowest-cost mode of transport for that given trip. In our data, we do observe usage of specific transport modes, importantly including individual ridesharing transactions. (We also observe some bus, taxi, or gasoline transactions, though these transactions may not reflect individual trips.) We assume that, if a ridesharing transaction is observed, it was the lowest-cost mode of transport for that individual to take that trip. This implies that the spatial frictions of consumers who take ridesharing rides have been lowered after adoption, with the first-order assumption that the costs of other modes do not change. In other words, if individual i is observed to use ridesharing a substantial amount after adoption, then we know that they have lower spatial frictions, i.e., $C_{itod}^{*,postadoption} < C_{itod}^{*,preadoption}$.

Second, for each individual possible trip $\{t, o, d\}$ there is also a corresponding utility value from taking that trip, V_{itod} . We assume the value is not first-order affected by the availability of ridesharing. Thus, we care about the following utility change from ridesharing:

$$\Delta_{itod}^{adoption} = (V_{itod} - C_{itod}^*)^{preadoption} - (V_{itod} - C_{itod}^*)^{postadoption}.$$

Among trips $\{t, o, d\}$ where $\Delta_{itod}^{adoption}$ is larger, this means that the adoption of ridesharing has had a larger impact on the utility of taking that trip—in other words, ridesharing has made this trip more valuable to individual i . This term $\Delta_{itod}^{adoption}$ will motivate several research questions and hypotheses in our analysis.

1. Consider 3 individuals: i who uses ridesharing substantially (“adopter”), j who does not use ridesharing at all (a “non-adopter”), and k who only uses ridesharing a few times (“low-usage adopter”). Since ridesharing is not observed to have affected j ’s behavior at all, we assume that $\Delta_{jtod}^{availability} = 0$ (using availability here instead of adoption since they do not adopt). Individual j may have a personal car and only be interested in trips where using the car is of low-cost, such as daily commuting and trips to nearby grocery stores. Next, we generalize and argue that on average, $\Delta_{itod}^{adoption} > \Delta_{ktod}^{adoption}$. While this may not hold true for all cases, this is a general argument that those consumers who use ridesharing more are those whose utility of spatial movement have been affected more.

Therefore, *if ridesharing does affect consumption behavior, high-usage individuals should be*

affected more than low-usage or no-usage individuals. This is one of our key identification strategies in our paper.

- (a) To reiterate, the impact of ridesharing on consumers' spatial frictions can vary widely. Some consumers (e.g. individual i) may be affected a lot while others who may already have lots of low-friction modes of transport available (e.g. individuals j and k) may not be affected much. In our identification, we argue that in general, those individuals who reveal to use ridesharing more are those whose trip utilities ($\Delta_{itod}^{adoption}$) have been affected more.
2. If $\Delta_{itod}^{adoption} > 0$, which we assume it is for adopters, does this lead to an increase in the number of trips taken and subsequently, the number of local transactions? This is our first major research question. While we assume, via revealed preferences as argued above, that ridesharing reduces spatial frictions, does this simply replace their existing transportation options and have no change in trips taken, or does this reduction in costs ($\Delta_{itod}^{adoption}$) increase the number of trips?
 - (a) For example, if we consider the origin o of trips as individual i 's home, are there more times (e.g. days) with positive value trips after ridesharing adoption? Does $\sum_t 1 \{(V_{itod} - C_{itod}^*)^{postadoption} > 0\} > \sum_t 1 \{(V_{itod} - C_{itod}^*)^{preadoption} > 0\}$? If so, the increase in time periods (days) with positive value trips implies increased local activity.
 3. If $\Delta_{itod}^{adoption} > 0$, does the composition of $\{o, d\}$ pairs change? This is our second major research question. Does the reduction in spatial frictions affect the set of destinations that are visited? Or, do the visited destinations remain the same? Even if the ridesharing increases the frequency of local activity, it is not clear that the spatial makeup of activity will change.

The ideas from this presented framework motivate the research hypotheses and analyses of our paper.

2 Panel Data on Detailed Credit Card Transactions

2.1 Credit Card Transactions and Account Metadata

From a large U.S. financial institution, we have access to an anonymized set of credit card transactions from a random set of accounts. These accounts were uniformly randomly sampled from the institution’s accounts across the U.S. except with over-sampling of accounts in two metropolitan areas of interest to the institution. The analysis of this paper does reflect the geography of this institution, so it may not reflect the impact of ridesharing in areas where the institution has relatively less presence. Each sampled credit card account is assigned a randomly generated anonymous, unique account identifier. We only observe this anonymous account identifier; when we refer to an individual account or account identifier, we are referring to this anonymous identifier, not the actual account ID number.

For each sampled account, we observe all of its anonymized credit card transactions over 4+ years (53 months), from January 2012 through May 2016; though not all accounts are active throughout this time range, which we discuss below. This date range is appropriate because while both UberX and Lyft (the most popular U.S. peer-to-peer ridesharing transportation platforms) were both introduced in San Francisco in mid-2012, their national expansion and growth mostly occurred in 2013 and 2014. As we show in a histogram of observed adoption dates in Figure 3, we have sufficient data both before and after the expansion and adoption of ridesharing services. For most of the sampled accounts with available mailing addresses, ridesharing services were not available in those areas until late 2013 or throughout 2014. In the appendix, we provide a table of some major metropolitan areas and when Uber or Lyft entered those areas.

For each credit card transaction, we observe: transaction date, transaction amount spent, the merchant’s standardized 4-digit Standard Industrial Categorization (SIC) code, merchant city location, and the merchant name¹. SIC codes are a standardized set of merchant and service categories

¹Any merchant names with fewer than 10 transactions will be truncated to ensure it does not have personally identifiable information. This includes, for example, some credit card transactions that include a transaction number in the merchant name. These unique numbers are scrubbed from our dataset to ensure privacy, but the more general store information, if no longer personally identifiable, is kept. This ensures that store names are consistent even if they are scrubbed. It also helps identify transactions as belonging to the same merchant, even if the transactions’ merchant name may have slight differences in date, order number, or time.

that each merchant is associated with, such as SIC code 5812 is a frequent one that represents “Retail-Eating Places”. In addition to an account’s transactions, we also observe each account’s monthly metadata, including credit limit, and for a subset of accounts, the city of the mailing address that month. Our two main datasets are thus: panels of anonymized transactions from a random set of accounts and panels of monthly account metadata.

Removed Transactions Before any analysis, we remove a few types of transactions. First, we remove all transactions that appear to be duplicate transactions. This is a small fraction of the data, but there are some that have the exact same merchant name, transaction amount, and date of transaction; if this occurs, we choose to keep only one of the transactions. Second, we remove the very frequently occurring credit card bill payment transactions, with merchant names such as “PAYMENT*THANK YOU” or “ONLINE CREDIT CARD PMT”². Lastly, due to issues in the data, the data is missing almost all transactions from the month of January. Thus, we choose to remove all January transactions to ensure equal comparisons within and between individuals, and January is considered not a month in our data. Related, to reiterate, when we consider measures of consumption behavior, we distinguish between online and local transactions, and try to consider only local transactions. This is discussed in Section 2.3.

2.2 Account and Time Period Selection

Identifying Ridesharing Adopters and Adoption Date In our data, since we observe the merchant name, we can identify whether each transaction is an Uber or Lyft transaction. Simply, we do so by identifying merchants that have the word “UBER” or “LYFT” in their name. Lyft transactions have no false positive merchant names. Uber has several false positive merchants, so we only consider a whitelist of known Uber merchant names and mark those as being Uber transactions. The specific whitelist (and examples of popular false positives) is in the appendix. This whitelist includes purchases made using proxies, such as Square, PayPal, or Google Pay.

Each consumer’s first observed Uber or Lyft transaction is deemed their adoption date. This adoption date becomes a demarcator in event time, which we use to define pre and post-adoption

²We list all the specific text strings in the appendix.

time periods. Figure 3 shows a histogram of the adoption dates in our data. We discuss the potential issue of unobserved ridesharing transactions (before or after the observed adoption date) in Section 2.5 and why potential unobserved usage strengthens our findings.

Pre and Post-Adoption Time Periods Our central econometric identification is an event-time difference-in-difference analysis, first comparing individuals within themselves, between the immediate period post-adoption, and a corresponding pre-adoption time period, and second comparing this difference across individuals (e.g. high-usage versus low-usage adopters versus matched non-adopters). We focus the main specification on just two periods, as motivated by the discussion of serial correlation in difference-in-difference models in Bertrand, Duflo, and Mullainathan (2004). In the robustness checks of Section 6, we do show two additional analyses showing the dynamics of this analysis using multiple event-time time periods, and also using a calendar-time fixed effects model.

The post-adoption period is defined as the first T days after adoption, inclusive, where $T \in \{180, 225, 270, 315, 365\}$. In the main text, we choose to focus on 270 days (9 months) as a window that allows for a larger sample size and more statistical power, but is also sufficiently lengthy to observe consumer behavior. As we show in the appendix, our results remain robust to the choice of time window.

The pre-adoption period is defined as 365 days before the post-adoption period, i.e., the same calendar days as post-adoption period but in the year beforehand. Emphasizing the same calendar days in both periods removes individual-specific day effects—if the post-adoption period includes a birthday or an annual event important to that individual, the pre-adoption period will also include them. To be specific, if a consumer adopts Uber/Lyft on 2015-05-01, with a window of $T = 270$ days, the post-adoption period is [2015-05-01, 2016-01-26]. That consumer’s pre-adoption period is thus [2014-05-01, 2015-01-26]. Each consumer’s unique pre and post-adoption (event-time) periods is determined by their adoption date. Figure 2 shows an example of the time periods for a hypothetical consumer that adopts ridesharing at the beginning of 2015.

Since the last day of our data is 2016-05-31, a time window of $T = 270$ requires that we only consider those who adopt ridesharing at least before 2015-09-04. This removes a substantial portion of our observed ridesharing adopters. We are able to include these consumers in two robustness

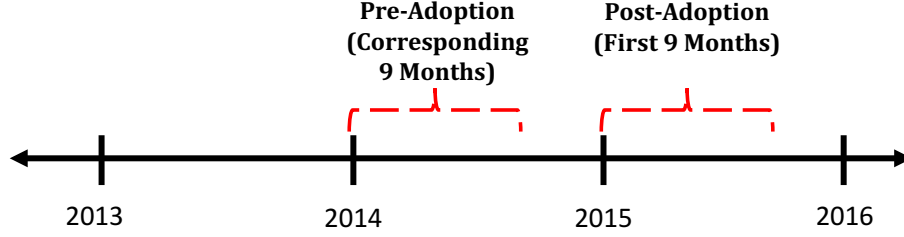


Figure 2: A hypothetical consumer who adopts ridesharing in Jan 2015. Jan - Sep 2015 is considered their post-adoption period, and as an apt comparison, we identify Jan 2014 - Sep 2014 as their pre-adoption baseline comparison period.

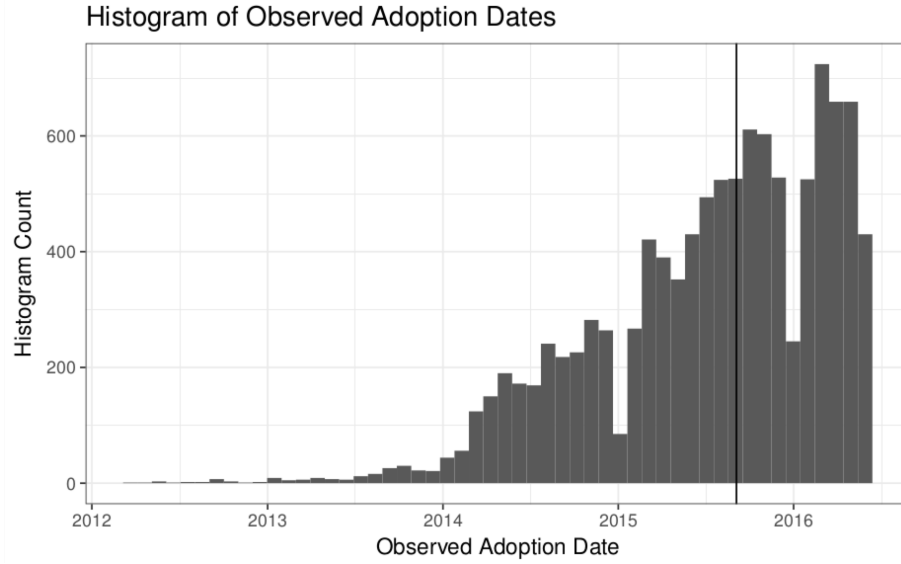


Figure 3: The distribution of observed adoption dates (i.e. first observed ridesharing transaction), among all the observed adopters in our data. The vertical black line represents our cutoff date for the sample of adopters we consider (all to the left of the black line) when we set $T = 270$ and ensure we observe at least 270 days after adoption.

check models: one with smaller time windows, like $T = 180$, and in a calendar-time fixed-effects model.

Minimum Activity Levels Another aspect of specifying the pre-adoption period in this manner is that it ensures that consumers were active on this credit card account at least 1 year beforehand. We are interested in analyzing customers who appear to be regular users of this credit card account, not those whose credit card usage is sporadic. Even though comparing customers within themselves already controls for individual static differences consumption patterns, we want to add some additional robustness by considering regular card users. To do so, we make two adjustments.

First, we require that each consumer has a minimum amount of monthly activity to be considered in our sample. For the 9-month time window ($T = 270$), since we remove January, we consider 8 months as the max amount of possible activity. From this, we allow for some vacation or periods of non-use, and require that accounts have a minimum of 6 unique observed months in both the pre *and* post-adoption periods. This ensures that the set of consumers we consider are not entering/exiting the credit card right before/after ridesharing adoption. Second, we normalize our outcome variables by the number of months observed in each period by dividing by the number of unique months observed for each account-period.

This results in a substantially smaller sample of 2,392 adopters who adopt at least 270 days before June 2016. While this limits our set, this sample results in the cleanest identification, and also focuses on the subset of adopters who adopt relatively earlier. We consider the entire set of observed adopters in the calendar-time fixed-effects model in our robustness check models.

2.3 Consumption Behavior Metrics

Above, we discuss the selection of accounts and event-time periods. For each period, we measure several measures on consumption frequency and spatial consumption. The goal of the paper is to analyze if these measures increase due to ridesharing. These measures are also referred to as the outcome variables of the model. We describe these measures in this section.

Importantly, we remove online and ridesharing transactions from counting in these consumption measures. We remove ridesharing transactions of course because they are the treatment variable of interest. We also remove online transactions because we consider online transactions as having zero spatial frictions, so they should not be first-order affected by ridesharing services. Thus, our consumption measures seek to measure only plausibly local, offline consumption. Online transactions are deemed to be those with 4-digit SIC merchant categories of {5942, 5735, 5968, 7399, 8999, 7311, 4816}; these are detailed in the appendix. Both in their SIC definition, and empirically in our data, these SIC codes are highly correlated with transactions with online companies, such as iTunes or Amazon.com or Netflix.

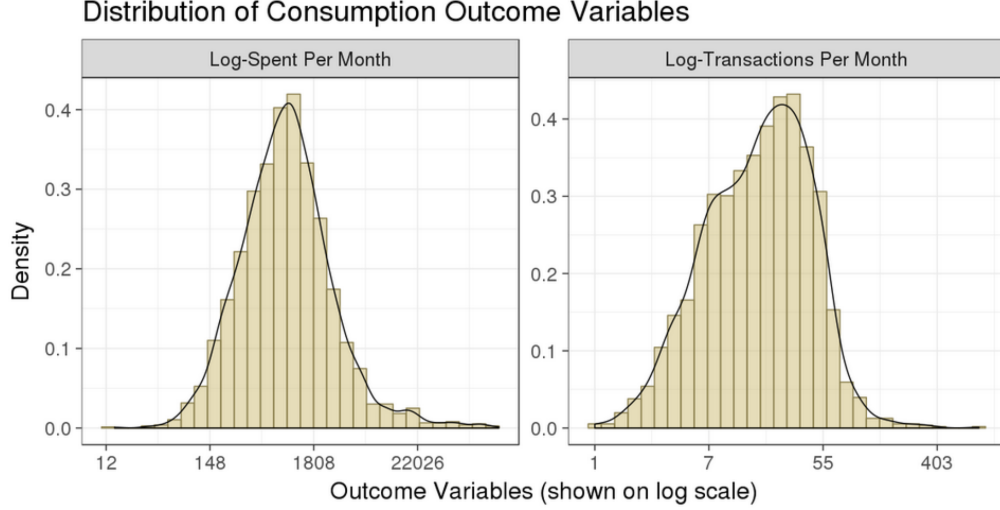


Figure 4: Two histograms of the distribution of the transactions per month and amount spent per month, from the pre-adoption period. This provides one summary our data, in addition to Figure 8.

Consumption Frequency The main outcome variable of interest is transaction frequency. This follows from the framework of spatial frictions: if a consumer’s spatial frictions are lowered by ridesharing, we hypothesize their willingness to consume locally to increase. For each consumer, we count the total number of local transactions in each period. Then we get the transactions per month by normalizing the total number by the number of unique months observed for each individual—though the results are similar if we normalize or not.

We also consider other consumption frequency measures as robustness checks: the (normalized) number of unique days with a transaction, and the (normalized) amount spent per month. We expect to find similar results for these two outcome variables. The number of unique days with a transaction is a useful robustness check because it is not affected by individual time trends that may increase the number of credit card transactions per day. Rather, the number of days outcome variable reflects if ridesharing reduces spatial frictions such that a consumer has more days with local activity.

Spatial Consumption Behavior We are interested not only in consumption frequency, but also how that consumption may be occurring spatially and geographically within a city. Our primary source of the spatial aspect of consumption are the 5-digit zip codes of merchants. Thus, as a

measure of spatial diversity, we consider the number of unique merchant zip codes visited in each period. For example, in the $T = 270$ days window, we consider all the unique zip codes among all the 9 months of transactions (not the number of unique zip codes per individual month). However, we refer to the normalized unique number of zip codes visited by again normalizing by the number of observed individual months.

The availability of merchant zip code data is not pervasive. There is zip code information for about 42% of transactions amongst adopters and 16% of unique merchant names among adopters. We draw this information from anonymized debit card transactions, as well as an external database on chain store zip codes. However, if a merchant name has a zip code, it has that zip code throughout the time of the data; thus the availability of merchant zip codes is not affected by time.

Using the number of unique zip codes as an outcome metric could be confounded by consumers who happen to travel more in their post-adoption period than in the pre-adoption period. High-usage adopters may be traveling more than low-usage adopters. Thus, as a robustness measure, we only consider unique zip codes visited within a geographic boundary of City A. City A is the city with the most transactions in our data and its geographic boundary removes the possibility of this result being influenced by travel spending.

2.4 Ridesharing Usage Data

Ridesharing usage is the key explanatory variable in our econometric models. Ridesharing usage is measured as the total number of Uber or Lyft transactions observed in the post-adoption period (e.g. first 9 months after the adoption date). We use ridesharing usage to identify those consumers who are affected by ridesharing, via revealed preferences. We show that the impact of ridesharing is increasing on those who use ridesharing more, whether we consider it as a continuous variable, or use it to separate consumers into high-usage and low-usage groups. The histogram of adoption dates is shown in Figure 3, but in this section, we further discuss descriptive statistics on the ridesharing usage we observe.

Date of Adoption & Ridesharing Usage. There exists an overall, increasing-over-time trend in total spending in the data. It is thus a concern that if those who adopt ridesharing later in time are

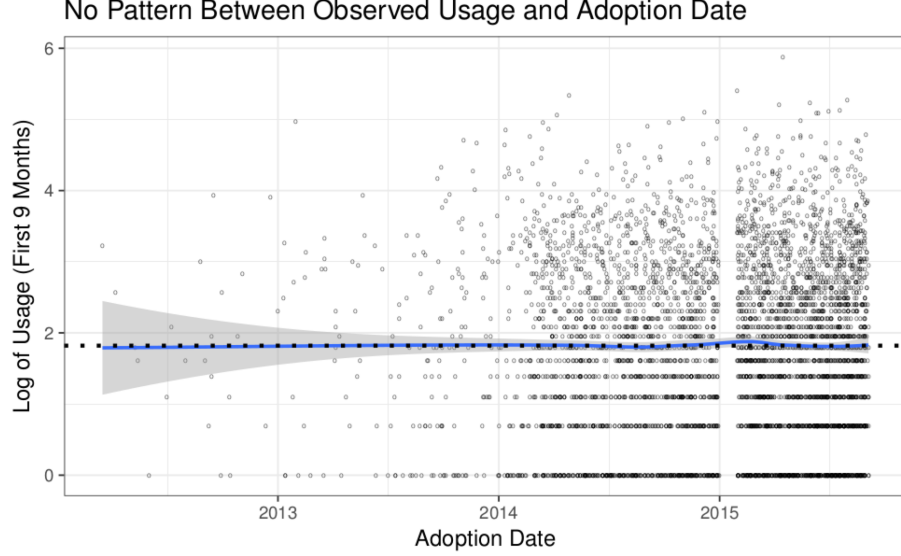


Figure 5: The mean of ridesharing usage is plotted as black dotted lines (on average, 11 transactions in the 270 days post-adoption). A flexible, local-linear line is then fit to the data. It shows there is no relationship between the future usage of adopters and the time of adoption. This reduces the concern about overall time trends affecting our analysis of comparing pre- versus post-adoption periods. If, for example, later adopters also used ridesharing more, this would introduce a concern that any relationship with usage could also be driven by time trends later in the our sample.

also those who use ridesharing more, this overall time trend may confound our analysis. However, this is not the case. Among our set of adopters, there is no relationship between adoption date and ridesharing usage. Figure 5 illustrates this using a post-adoption window of $T = 270$. Further, within these first 9 months, ridesharing usage appears to be approximately constant throughout, i.e., all the usage is not all concentrated in the first couple months but is rather spread out throughout the 9 months.

Post-Adoption Ridesharing Usage & Pre-Adoption Consumption One concern that we will test for throughout this paper is that of potential individual trends that may drive some consumers to both transact more and use ridesharing more in their post-adoption period. This hidden consumption trend would be confounded with the high-usage consumers we focus on.

As one assessment of this, we look at pre-trends between pre-adoption consumption behavior and post-adoption ridesharing behavior. Among the adopters, there is no observed relationship between their pre-adoption consumption metrics and the post-adoption choice of ridesharing usage. We demonstrate this visually in Figure 6 using flexible non-parametric models for four different

Distribution of Usage vs Pre-Adoption Outcome Variables

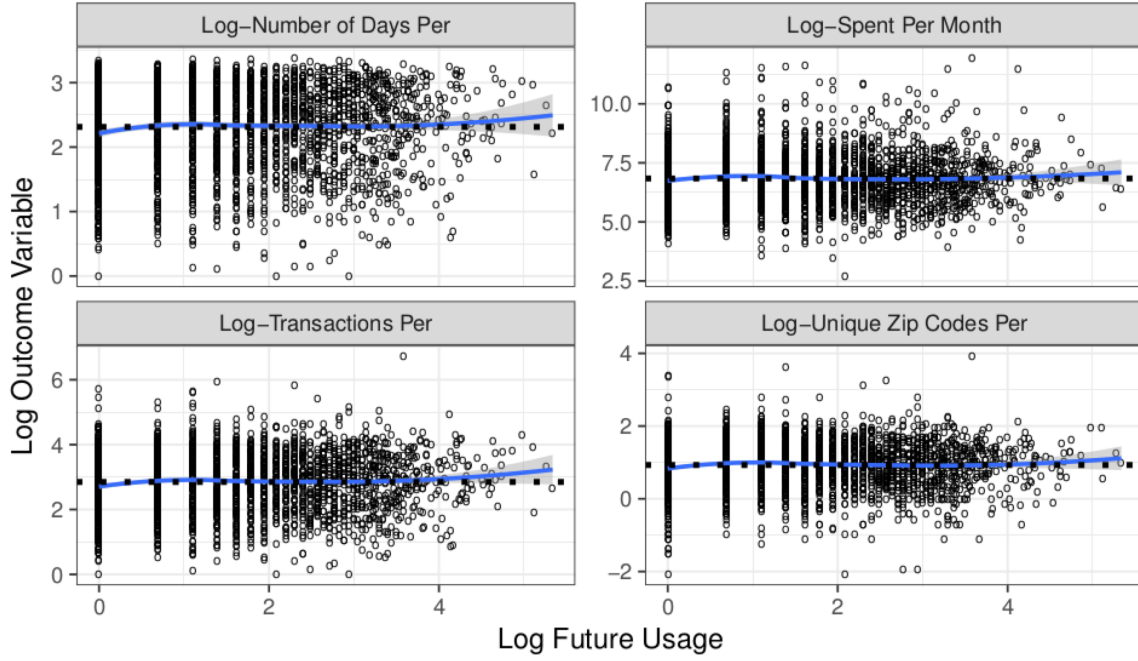


Figure 6: Each of the four plots shows a scatter plot of a pre-adoption consumption metric versus each consumer’s future, post-adoption ridesharing usage. The black dotted line is the mean of each consumption metric, and the blue line is a flexible, local-linear fit to the scatterplot. As seen, the confidence interval of the blue line mostly overlaps the black line.

There is no observed relationship between the future ridesharing usage of adopters and consumption metrics pre-adoption. Higher spending adopters or adopters with more frequent transactions pre-adoption are not associated with using ridesharing more post-adoption.

consumption metrics. This reduces the concern about individual trends, but we assess this further in the event study model of Section 6. Another way we address the concern about individual time trends, is to test for individual trends that may be existing before adoption—we do so in Section 4 by performing a falsification test by simulating an adoption date a year before the actual adoption date and running our same models.

Unobserved Ridesharing Usage There may be unobserved usage of Uber/Lyft. If this unobserved usage occurs before our observed adoption date (e.g. paying for ridesharing using a different payment method), and ridesharing indeed has a positive impact on consumption activity, this makes our results conservative by raising the consumption levels of the before adoption comparison period. If this unobserved usage occurs after our observed adoption date, we should consider whether high-

usage or low-usage consumers are more likely to have unobserved ridesharing usage. If low-usage consumers are, this makes our results conservative, since they are a control group; if high-usage consumers are, this would suggest that the relationship between usage and increases in consumption activity may be steeper, but the underlying mechanism is still valid. This is because ridesharing usage is proxying for individual-specific reductions in spatial frictions that affect behavior.

2.5 Discussion

There are two particular data concerns that we discuss. One, there may be issues with only observing one credit card rather than an individual’s whole set of consumption, and two, there may be unobserved confounders that dynamically affect both ridesharing usage and increases in consumption activity. In this subsection, we point out these concerns and mention the various tests we will use to show evidence against them.

Potential Use of Multiple Credit Cards There is a concern that individuals may have multiple credit cards or unobserved spending. Since we are able to observe individual spending over time, the concern is more so limited to dynamic changes in an individual’s usage of the observed credit card. If an individual uses approximately half of their total spending on this credit card, but this is static throughout the pre and post-adoption periods, this does not influence identifying the influence of ridesharing availability. However, if some individuals have a positive shift in their usage of this observed credit card account and they are associated with more ridesharing usage, this would confound our identification. We check for this in several ways, including, but not only: using average amount spent as a falsification test, limiting our sample to only those who adopt soon after availability, using a subsample of extreme weather days as a check on the direction of the estimated effects, and looking at the effects of ridesharing on different merchant categories.

There may be other quirks in the use of credit card accounts as a source of data. To reduce the concern of having consumers who only use their account sporadically—even if this is static—we use the requirement of a minimum amount of monthly activity in both pre and post-adoption periods. This also prevents the analysis from being influenced by new accounts that were created shortly before ridesharing availability, accounts must have been active at least 1 year before adoption. In

the robustness checks of Section 6, we also consider a smaller sample of consumers with the low and high outliers in consumption and ridesharing usage removed, to address potential outsized statistical influence of outliers or high spenders.

An additional concern lies in the choice to pay for ridesharing on this credit card. To the best of our knowledge, among the financial institution’s cards, during the study’s time period, they did not run any travel-specific or Uber/Lyft-specific marketing campaigns that encouraged ridesharing transactions. Even if they had, to confound our analysis of consumption patterns, this marketing would need to target only a subset of consumers and also concurrently coincide with marketing to get them to spend more on non-ridesharing transactions too.

Related to this, it is a relevant point to consider if low-usage adopters differ in their credit card usage than the high-usage adopters. If so, this would raise some doubt in our comparison of these two groups. While it is a difference-in-differences analysis and their static differences would be removed, it could still add doubts that the high-usage consumers behave differently over time. However, we show no evidence of this, as shown in Section 2.4 and Figure 6. There, there is no relationship between pre-adoption consumption measures and how much a consumer chooses to use ridesharing in the post-adoption period. (If this were not the case, for example, high-spending consumers also use ridesharing more, it may suggest autocorrelation or heteroskedasticity issues.) Also, there are statistical checks on the differences between high and low-usage groups: using a simulated, earlier adoption date; coarsening into just two groups rather than treating usage as a continuous value; and checking for changes between these groups in different merchant categories.

Potential Unobserved Confounders Beyond potential concerns of the credit card data, there is also a concern that individual trends may confound our paper’s identification. Individual consumers are being compared to their past pre-adoption behavior, so a confounder would need to be dynamic, influencing individual consumption dynamically and also influencing the consumer’s choice of ridesharing usage. So, the central concern of identification is that there are unobserved, individual-specific, time-varying variables that increase both the consumer’s usage of ridesharing and their consumption. We refer to these in short as individual-specific time trends. Since we focus the analysis on the event-time of ridesharing adoption, examples of such confounders that

may coincide closely with ridesharing adoption include rising income over time, an income shock, or a lifestyle change like moving locations.

Unobserved confounders that simultaneously affect usage and consumption are what especially concern us, not especially the confounders that affect adoption and consumption, since the core of our identification is comparing high-usage versus low-usage consumers.

We already control for some individual consumption stochasticity by using the same calendar days in pre and post-adoption periods. Then, in the main models, we use changes in online spending to control for potential income changes. To further test for evidence of potential income or propensity-to-spend changes around the adoption date, we: use average amount spent as a falsification test model, use a multiple-period event study to look for pre-trends in consumption before adoption, compare spending changes across merchant categories, and use a limited subsample of adopters who adopt more soon after ridesharing is available. Another important test is using a subsample of consumption only on extreme weather days, since we expect increased propensity-to-spend to be dampened on extreme weather days (since spatial frictions are higher) but the ridesharing’s effect to be larger (since ridesharing has a higher effect on frictions in extreme weather days).

3 Econometric Methodology

In this section, we our econometric identification and estimation models. The first is a simple model that compares adopters-only to estimate both an adoption effect and an effect on those consumers who use ridesharing more—we hypothesize that we should only see an effect on the consumption of those consumers who have higher usage, since this is the mechanism of interest. The second model is our main difference-in-differences model comparing high-usage adopters versus the control groups of a matched group and of the low-usage adopters, and also adds online spending controls.

As a clarification, in this paper, we are interested in essentially the treatment effect on the treated, rather than considering ridesharing as a policy where we would be interested in the overall urban treatment effect. This is because the important takeaways from this analysis are the heterogeneity in this effect demographically and spatially.

3.1 Adopters-Only Model:

Ridesharing Usage, Not Adoption, Is Related to Consumption Changes

This simple model is important to reflect the conceptual argument of our paper. By looking only among the sample of adopters, we avoid the concerns of adoption being self-selected. This model focuses on the distinction between the effect of ridesharing adoption and the effect of reduced spatial frictions from ridesharing. Low-usage consumers have the former, but not the latter because, by revealed preferences, their behavior has not really been changed from ridesharing. In contrast, the higher-usage consumers reveal that their transportation choices have been affected by ridesharing’s availability. So, in this model, all the consumers have at least one ridesharing transaction.

We simplify the model into a simple comparison of two long time-periods: pre-adoption and post-adoption periods. Based on the discussion of Bertrand, Duflo, and Mullainathan (2004), doing so also reduces concerns of serial correlation across periods. We use more periods in the robustness checks. Thus, the identification of this conceptual model is a first-differences model, which is equivalent to a two-period model with individual fixed effects. We use several outcome variables Y_{it} : transactions per month, spent per month, average amount spent per transaction, and the normalized number of unique zip codes per period. The first-difference outcome variable is thus ΔY_i , for each of these four outcomes. Figure 8 shows the distribution of these first-differences.

We seek to explain the variation in ΔY_i using just two variables: an adoption indicator and each individual’s observed ridesharing usage ($Usage_i$). Since we are looking at spending and transaction behavior, which are skewed, we use natural-log-transformed outcomes. This first-differences model removes time-invariant variation across individuals in transaction/spending behavior. It also removes individual-specific calendar-day influences on consumption behavior.

To illustrate how these time-invariant influences are differenced out, consider the parsimonious linear model of consumption behavior below for a consumption measure Y_{it} , for time periods $t = 0$ and $t = 1$. Consumption is modeled as made up of an individual fixed effect α_i , calendar day dummy-variables Day_{id} for all the calendar days $d \in t$, the influence from ridesharing usage $Usage_{it}$, and an error term ϵ_{it} .

$$\ln Y_{it} = \alpha_i + \beta \ln Usage_{it} + \sum_{d \in t} \delta_{id} Day_{id} + \epsilon_{it}$$

Taking the first difference, the identifying estimation equation removes both the individual fixed effect and the individual day dummies:

$$\Delta \ln Y_i = \ln Y_{i1} - \ln Y_{i0} = \tau_a + \beta \ln Usage_{i1} + \Delta \epsilon_{i,1,0} \quad (1)$$

We conservatively assume there is no ridesharing in the pre-adoption period before we observe usage, so we set $\ln Usage_{i0} = 0$. The τ_a constant estimates the average adoption effect, representing the conditional effect of being in the post-adoption period, relative to the same days the year before³. Also, if there are unobserved individual time-trends that are related to adoption, this term also captures the average individual-specific time trend, if any. It reflects the average difference between each consumer’s pre and post-adoption periods, after conditioning on $Usage_{i1}$. We theorize adoption to have no first-order effect and for $\tau_a = 0$.

Instead, we hypothesize that only those adopters who choose to utilize ridesharing—those whose spatial frictions have been revealed as affected by ridesharing—will show evidence of changing consumption behavior. Ridesharing transactions reveal individuals whose spatial frictions are affected more, *ceteris paribus*, and vice versa. Additionally, as discussed in Section 1.2, we hypothesize that generally the reduction in spatial frictions is increasing, but diminishing in observed ridesharing usage. Therefore, we estimate β , the coefficient on the conditional linear relationship between $\Delta \ln Y_i$ and $\ln Usage_{i1}$, and hypothesize $\beta > 0$ to reflect the influence of ridesharing availability.

The null hypothesis, if ridesharing availability actually has no effect on local consumption behavior, is that ridesharing simply replaces consumers’ existing transportation options. In this case, we would not expect observed usage to explain any variation in the consumers’ change in spending measures between pre and post-adoption periods.

We don’t explicitly control for time period fixed effects in our simple model. This is because of two empirical factors. One, the adoption event occurs in a wide range of times, so including calendar-

³For example, if a consumer is determined to have adopted Uber/Lyft on 2015-05-01, with a window of $N = 270$ days, then the post-adoption period is [2015-05-01, 2016-01-26]. That consumer’s specific pre-adoption period is then [2014-05-01, 2015-01-26].

time dummies would reduce the sample size of adopters. Two, there is no empirical relationship between the time of adoption and post-adoption usage, as illustrated in Figure 5. This lack of relationship is robust to different time windows and removal of outliers, so it reduces the concern that high-usage adopters are systematically earlier/later adopters. In the robustness section, we do include calendar-time dummies however.

3.2 Main Model: Matched Control Group and Online Spending Controls

In the main model specification, we introduce two additional controls for the higher-ridesharing-usage consumers. The simple model already controls for a potential constant self-selection adoption effect by comparing against the post-adoption periods of low-usage consumers. In this model, we introduce each individual’s change in online spending behavior as a control for income changes, and introduce a matched control group of non-adopters as a another control for potential individual time trends.

We hypothesize that online spending, which has no spatial frictions, is unaffected by changes to a consumer’s spatial frictions. If a consumer has higher purchasing power or higher propensity-to-consume in their post-adoption period—which increases both ridesharing usage and consumption—we expect this to also affect their online spending. We identify the online transactions that we removed from the consumption measures, as discussed in Section 2.3. With these, we measure both changes in their online spending per month ($\Delta Spent_{online}$) and in online transactions per month ($\Delta Trans_{online}$). To add further flexibility to the model, we do not simply model a linear relationship from these measures, but a flexible quadratic model, $f(\cdot) = \zeta_1 \cdot + \zeta_2 \cdot^2$, to better capture potential spending-propensity changes. We do not use a log-transformation here to avoid removing consumers from the sample from may have no or negative changes in online spending. Observing each individual’s online spending behavior is a helpful signal of spending shocks and helps control for some potential confounders in ridesharing usage. As discussed in Section 2.5, in addition to modeling controls, we also use various subsamples and falsification checks to also look for signals of potential confounders.

In this model, we expand our sample of consumers to include an additional comparison group for baseline time trends. Using estimated propensity to adopt, as described in the subsection below, we

identify a 1:1 nearest-neighbor matched set of consumers who do not have any observed ridesharing transactions. It is possible that these consumers have easy personal transportation access, prefer to consume in convenient areas, or are using ridesharing on a different credit card. As we show in the subsection below, these matched consumers do show quite a similar distribution of pre-adoption consumption measures. To assign each matched control consumer a synthetic adoption date and ridesharing usage, we assign each matched control consumer the adoption-date and usage of their matched adopter. Using this synthetic assignment, we expect the matched control group to generally reflect some of the unobservable, individual time trends that the adopters may have in a counterfactual world where they did not actually utilize ridesharing. The matched non-adopters provide an additional comparison group alongside the low-usage adopters.

The resulting main model is:

$$\begin{aligned} \Delta Y_i = & \gamma + \tau_a \text{Adopt}_i + \beta_1 \ln \text{Usage}_{i1} + \beta_2 \ln \text{Usage}_{i1} \cdot \text{Adopt}_i \\ & + f(\Delta \text{Spent}_{\text{online}}) + f(\Delta \text{Trans}_{\text{online}}) + \Delta \epsilon_i \end{aligned} \quad (2)$$

The γ coefficient reflects the general time trend, the conditional average difference between pre and post-adoption periods, among both adopters and control consumers.

The τ_a coefficient then reflects any actual adoption effect, which is the conditional average difference-in-differences of the outcome variable between the adopter group and the control group, conditional on their ridesharing usage.

β_1 is the estimated coefficient on the (synthetic) usage of the control consumers, reflecting the individual time trends of the synthetic high-usage consumers of the control group.

β_2 is our actual coefficient of interest. It reflects the additional explanatory power of the adopters' actual ridesharing usage.

As stated earlier, $f(\Delta \text{Spent}_{\text{online}})$ and $f(\Delta \text{Trans}_{\text{online}})$ reflect flexible quadratic-linear models on changes in online spending to better act as proxies for potential income confounders.

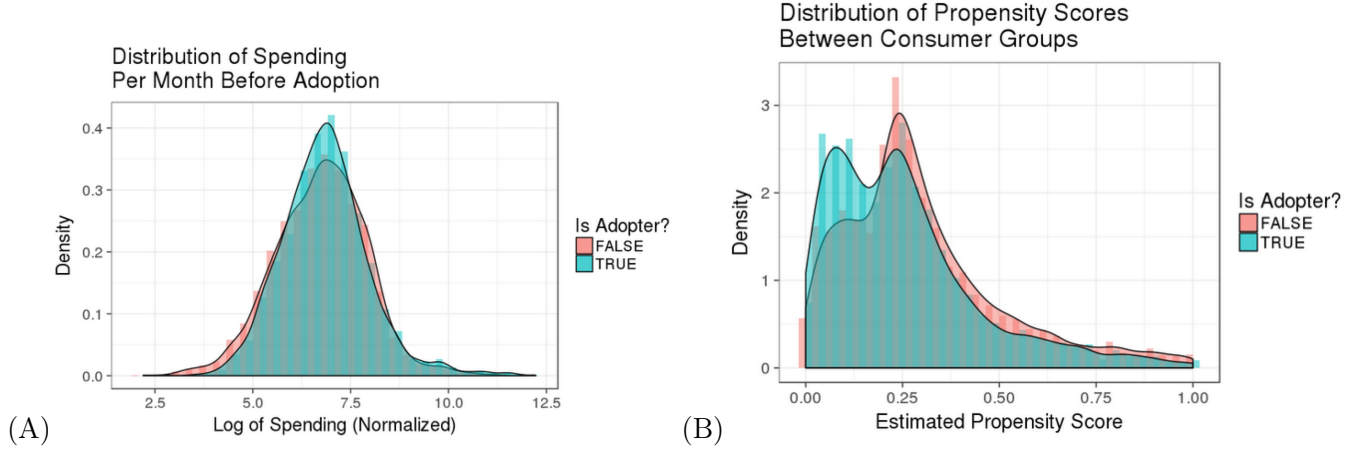


Figure 7: A comparison of the matched consumer group with the adopter group. In red are the matched consumer group, and in blue is the adopter group. On the left, in part (A), the two groups show a very similar distribution in spending per month, before adoption. On the right, in part (B), the adopters have a slightly higher concentration of lower propensity consumers.

Heteroskedasticity Since our main treatment variable of interest, observed ridesharing usage, is a continuous variable, all of the results in the paper are shown with heteroskedasticity-consistent standard errors.

3.2.1 Matched Control Group

We identify a set of similar non-adopting consumers as a control group. We match consumers on their propensity to adopt ridesharing, using a propensity score model derived from a lasso logistic regression (i.e. penalized logistic regression) on many possible predictors. This model is estimated using predictors that are all from the adopters' year before adoption, including various overall consumption metrics, merchant-category-specific consumption metrics, and the average credit limit in the pre-adoption data. We use a 1:1 nearest neighbor match. In the matching process, we also remove a small number of adopters who do not have a close propensity match, which reduces the number of adopters in the sample slightly. Compared with the general sample of non-adopters, using the matched non-adopters has a very large improvement on the balance of covariates. This is likely because the non-adopters include a large proportion of infrequent card users. Figures 7 shows the balance of some of these covariates, including balance on propensity scores and balance on pre-adoption spending per month.

3.3 Consumption Measure Summaries and Descriptives

Figure 8 shows four histograms of the outcome variables of interest, the first difference between pre and post-adoption periods of: transactions per month, spent per month, number of transaction days per month, and the number of unique zip codes in each period (normalized by months observed). In Figure 8, we see that these outcome variables are mound-shaped and approximately symmetric, which helps reduce concerns about statistical outliers affecting estimation. Figure 9 shows model-

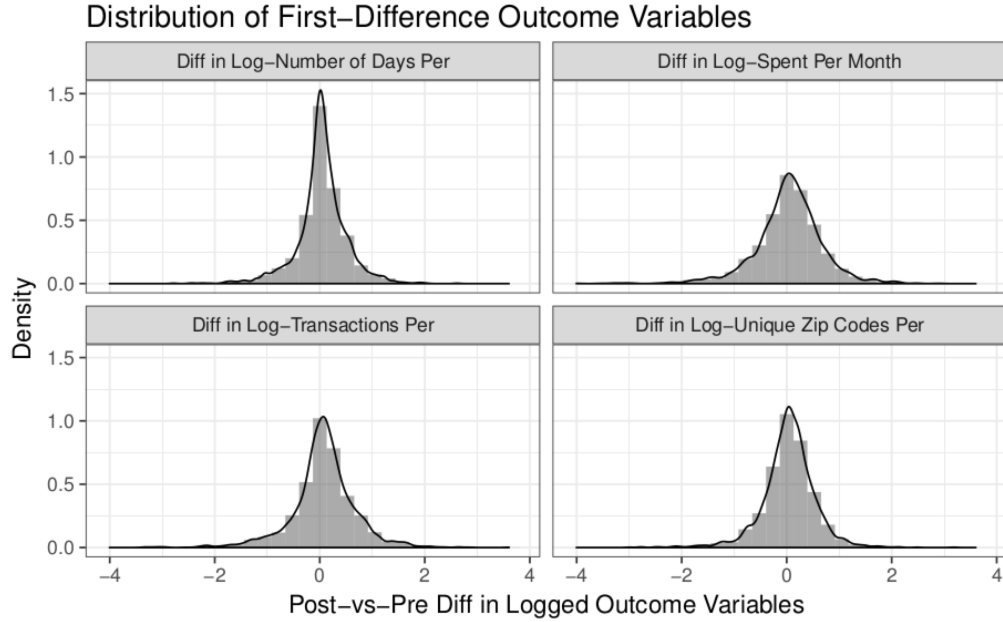


Figure 8: This plots the summary of the actual outcome variable that we’re analyzing. Among all the adopters in the 270-day window, these are histograms of the comparison between the post-adoption and corresponding pre-adoption periods of transactions, spending, number of zip codes, and number of days (normalized by number of observed months). Specifically, for each of the Y_i mentioned, we calculate $\Delta Y_i = \ln Y_{i1} - \ln Y_{i0}$ (or in other words, $\ln(Y_{i1}/Y_{i0})$).

free evidence of both our simple conceptual model and model with matched consumers. For each scatter plot, the y-axis plots the post-adoption change in transactions per month for the sample of consumers (adopters; or adopters and matched consumers), and the x-axis plots the ridesharing usage of each consumer post-adoption. Within each scatterplot, flexible local-linear nonparametric models are plotted for each consumer group. The adopters show a significant and substantial relationship between their observed usage and their post-adoption change in transactions per month.

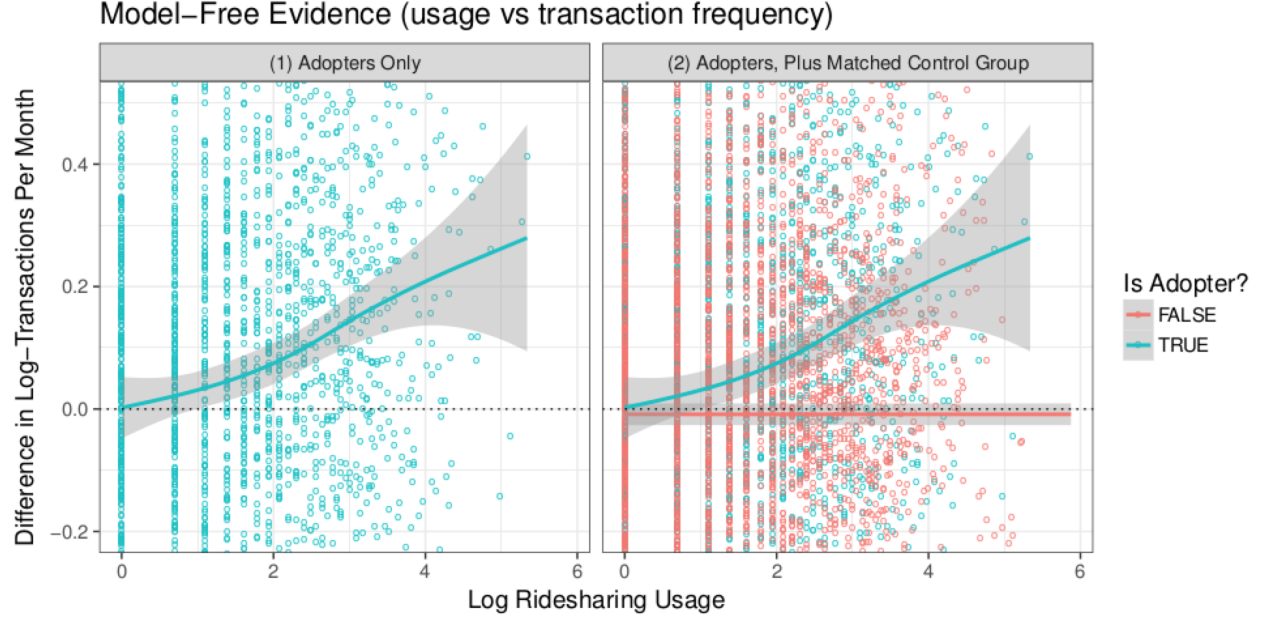


Figure 9: This scatterplot illustrates model-free evidence for the relationship between post-adoption ridesharing usage (x-axis) and between-period changes in transaction frequency (y-axis). The left plot shows the scatterplot of just the sample of adopters. The right plot adds an additional matched control sample, with synthetically assigned usage (shown as red points).

On top of the scatterplots, there are two flexible, local-linear lines plotted for the groups (adopters or matched control) that show the flexible nonparametric model between usage and between-period changes. As seen, those adopters who do not utilize ridesharing show little consistent difference post-adoption, but those who do use ridesharing show a positive change post-adoption, which increases in usage.

4 Results: Consumption Frequency

Table 1 shows the results of our adopter-only and main models on the transaction frequency outcome. Column (1) estimates only an adoption effect and shows an overall positive increase in the post-adoption period for adopters. In Column (2), once we introduce each adopter’s actual observed ridesharing usage, this captures all the significant explanatory effect of ridesharing. There is no longer a significant adoption effect. This is consistent with our hypothesis; those who utilize ridesharing are those whose spatial frictions are meaningfully affected, by revealed preferences, and as a consequence, are those who increase their consumption frequency post-adoption. This is evidence against the null hypothesis that those utilize ridesharing are just simply replacing their existing transportation choices.

Columns (3) and (4) of Table 1 introduce the additional controls of our main model, a set of matched non-adopter consumers and measures of changes in online consumption. There is no significant explanatory power of being an adopter versus a non-adopter in Columns (3) and (4), suggesting that, conditional on ridesharing usage, there is no significant difference between adopters and the matched group. The matched group’s synthetic usage also shows no relationship. If the potential individual-specific time trends are related to an individual’s propensity to adopt, the matched consumers should also reflect some of these time trends in their consumption changes. However, there is no significant post-adoption relationship. between the matched group’s usage and their consumption changes. Column (4) adds controls for changes in online spending and transactions between periods. This shows some predictive value, but does not substantially affect the coefficient estimate on ridesharing usage.

The key coefficient of interest is that of observed ridesharing usage. This has significant explanatory power for the adopters group in explaining post-adoption changes in transaction frequency. This is in the last row of Columns (3) and (4), and is significantly positive in both. Combined with an insignificant adoption coefficient, this is consistent with our hypothesis and suggests that only those whose spatial frictions are affected by ridesharing show signs of increasing their offline transaction frequency.

All else equal, in this model with logged outcomes and logged usage, comparing someone who uses twice as much utilization of ridesharing services, e.g. from 5 to 10 times in the period, it predicts the higher-usage person will see a 4.6% increase in the number of offline transactions per month relative to the lower-usage consumer. Importantly, we interpret observed ridesharing usage as a signal of reduced frictions, not as each transaction as being the sole direct cause of increased transactions. In addition to direct transportation services, ridesharing also reduces spatial frictions it its existence value by being a transportation safety-net option, or also in its peer-effects of encouraging activity among friends or shared rides that may be unobserved.

4.1 Other Consumption Outcomes & Falsification Tests

We consider alternative outcome measures of consumption frequency, which support the robustness of the findings. In Section 6, there are additional robustness checks too.

	9-month time window			
	log (ratio of transactions per month) (OLS)			
	Adopters-Only		w/Matched Group	w/Online Controls
	(1)	(2)	(3)	(4)
Constant Δ	0.072*** (0.013)	-0.018 (0.021)	-0.016 (0.016)	-0.010 (0.016)
Log-Usage for Matched		0.053*** (0.011)	0.004 (0.007)	0.003 (0.007)
Actual Adopter Effect			-0.002 (0.027)	-0.007 (0.027)
Δ Online Spending				2.838 (2.224)
Δ Online Spending, Squared				0.243 (0.705)
Δ Online Tran				2.226 (1.469)
Δ Online Tran, Squared				-6.627*** (1.980)
Log-Usage for Adopters			0.049*** (0.013)	0.046*** (0.013)
Observations	2,396	2,396	5,471	5,471
Adjusted R ²	0.000	0.010	0.010	0.031
Residual Std. Error	0.616 (df = 2395)	0.613 (df = 2394)	0.553 (df = 5467)	0.547 (df = 5463)

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 1: With the $\log(y_{\text{after}}/y_{\text{before}})$ outcome on total transactions, we see a consistent significant positive coefficient between the usage of Uber/Lyft and the change in the # of transactions after adoption.

First, as a falsification test, we consider the changes in average amount spent per transaction between the pre and post-adoption periods. If individuals have an unobserved income shock or increase in propensity-to-consume, they may also have a higher willingness-to-pay and spend more per transaction. If so, we could see evidence of this confounder in a positive relationship between ridesharing usage and increases in the average amount spent per transaction. However, Column (4) of Table 2 shows no evidence of this. It uses the same main model specification and consumers as Column (4) of Table 1, except with the different outcome variable, logged average amount spent per transaction. This reduces the concern of there being unobserved confounders that would affect a consumer’s average amount spent. This result is consistent with our framework since we do not hypothesize that ridesharing or lowering spatial frictions would have a first-order effect on the price-level of merchants that adopters patronize.

Second, as additional robustness using the same main model, we find evidence of the significant relationship of ridesharing usage on other offline consumption metrics: {the number of unique days with a transaction, total unnormalized transactions, and spending per month}. These are in Columns (1), (2), and (3) respectively of Table 2. The number of unique days with a transaction is a particularly useful robustness check. It is not affected by multiple credit card transactions per day. If our main results were being driven only by more transactions per day (rather than more days with transactions), this would raise concerns on our hypothesis of lower spatial frictions. However, Column (1) of Table 2 shows a significant and similarly positive coefficient between adopters ridesharing and changes in the number of days with transactions. This suggests that consumers who are affected by the availability of ridesharing not only show an increase in transactions, but also an increase in the number of days with transactions. This is consistent with a hypothesis that consumers whose spatial frictions are lowered are more willing to leave their home for local transactions.

Third, we use a second falsification test and check for earlier pre-adoption evidence of individual-specific time trends among adopters. If there are individual time trends that affect our findings between pre and post-adoption periods, we posit that there may be some evidence of these same trends in time periods before adoption too. To check, we perform a falsification test by shifting

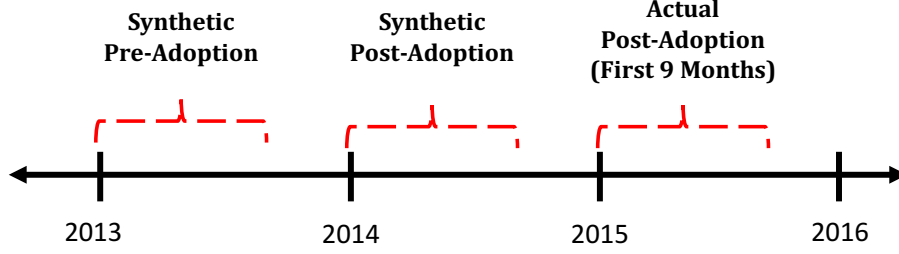


Figure 10: A hypothetical consumer who adopts ridesharing in Jan 2015. Jan - Sep 2015 is their *actual* post-adoption period. Thus, Jan 2014 - Sep 2014, instead of being the actual pre-adoption period, is now treated as the *synthetic* post-adoption period, and Jan 2013 - Sep 2013 is their *synthetic* pre-adoption period. We then compare the synthetic periods.

back our pre/post periods by exactly 1 year, i.e., synthetically shifting back the adoption date by 1 year before the actual adoption date⁴. Figure 10 shows an example of how synthetic post and pre periods are identified, for a hypothetical consumer. In doing so, we also synthetically assigning each adopter their future usage to the fake “post” period. This future usage is a proxy to test for individual time trends that are present before adoption that may be related to both future usage and consumption. We refer to this as a “yearback” test. As a precaution and additional control, we also do the same “yearback” shift for the matched non-adopters too. We also perform another test for pre-trends among adopters in the event study of Section 6.

The results of this yearback falsification test are shown in Columns (1) - (3) Table 3. These hold if we exclude or include the matched consumers too. Of note, the number of accounts in these regressions are smaller because for some consumers, we do not observe enough data for them 2 years before their adoption date. Column (4) confirms that, with this smaller sample, our main results continue to hold. It replicates the main model using the *actual* (not synthetic) pre and post-adoption period consumption metrics, except with the smaller sample.

4.2 Heterogeneous Effects

Merchant Categories This paper proposes that the mechanism influencing changes in consumption is that ridesharing availability reduces spatial frictions for a subset of consumers (as revealed by usage). As a test of this mechanism, this reduction in frictions is should generally be more

⁴For example, for a consumer who adopts on 2015-05-01, we synthetically simulate that they adopted on 2014-05-01, and assign the [2014-05-01, 2015-01-26] as the fake “post” period and compare it with the fake “pre” period of [2013-05-01, 2014-01-24].

	9-month time window (log-ratio OLS models)			
	Days w/Transaction	Total Transactions	Spent per Month	Average Amount Spent
	(1)	(2)	(3)	(4)
Constant Δ	-0.011 (0.013)	0.001 (0.017)	-0.062*** (0.019)	-0.052*** (0.014)
Log-Usage for Matched	0.002 (0.006)	0.003 (0.008)	0.011 (0.009)	0.007 (0.006)
Actual Adopter Effect	-0.009 (0.021)	-0.012 (0.028)	0.010 (0.030)	0.017 (0.020)
Δ Online Spending	3.612** (1.809)	4.231 (2.646)	6.460*** (2.400)	3.623*** (1.180)
Δ Online Spending, Squared	1.208** (0.478)	0.729 (0.786)	0.262 (0.678)	0.019 (0.374)
Δ Online Tran	0.629 (1.178)	1.341 (1.769)	0.007 (1.485)	-2.218*** (0.690)
Δ Online Tran, Squared	-5.960*** (1.567)	-7.938*** (2.313)	-8.859*** (2.155)	-2.232** (1.116)
Log-Usage for Adopters	0.032*** (0.010)	0.056*** (0.014)	0.042*** (0.014)	-0.004 (0.010)
Observations	5,471	5,471	5,471	5,471
Adjusted R ²	0.024	0.031	0.028	0.001
Residual Std. Error (df = 5463)	0.430	0.592	0.602	0.424

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 2: Replacing the outcome variable with average amount spent, we don't see any relationship with either the Uber/Lyft adoption nor usage.

	9-month time window (log-ratio OLS models)			
	Transactions Per	Avg Spent Per	Zips per Month	Future Transactions Per
	(1)	(2)	(3)	(4)
Constant Δ	-0.006 (0.020)	-0.064*** (0.016)	0.030 (0.019)	-0.285*** (0.036)
Log-Usage for Matched	0.011 (0.009)	0.012 (0.008)	0.020** (0.008)	0.007 (0.017)
Actual Adopter Effect	0.170*** (0.032)	0.032 (0.025)	0.139*** (0.030)	-0.008 (0.054)
Log-Usage for Adopters	-0.015 (0.016)	-0.018 (0.012)	-0.021 (0.015)	0.101*** (0.026)
Observations	3,384	3,384	3,384	3,384
Adjusted R ²	0.018	-0.0001	0.012	0.015
Residual Std. Error (df = 3380)	0.525	0.434	0.487	0.880

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 3: To test for potential confounding individual time-trends, we create simulated adoption a year before the actual adoption, creating fake pre/post periods, but synthetically inserting the actual usage. In this falsified data, we don't actually see any of the relationship between those consumers who *will* use ridesharing in the future.

impactful on consumption that is more elastic with respect to spatial frictions; for example leisure, weekend consumption may be more elastic than consumption related to a daily work schedule. This elasticity could be due to either relatively large transportation costs for leisure (or non-routine) consumption. If spontaneously going to a restaurant in a different neighborhood, bus lines may not run as frequently at night, there may be more discomfort in walking at night, or driving may be removed as an option if parking is limited or the consumer is drinking alcohol. As an alternative, leisure consumption may also be considered more marginal consumption, and so be more sensitive to changes in spatial frictions. In contrast, to test this mechanism, we do not expect consumption of durable and household goods to be as sensitive to changes in spatial frictions. These purchases may be less frequent and may be more easily scheduled by the consumer in low spatial frictions moments.

Therefore, we expect the effect of ridesharing to differ between different categories of merchants, and to differ in ways that are consistent with the hypothesis. In Table 4, we estimate the main model across several merchant categories. Column (1) looks at changes in spending at restaurant and bar transactions⁵. It shows that the coefficient of ridesharing usage on restaurant transactions is more than twice as large as the coefficient in our main results. The direction of this result is consistent with our hypothesis. In the pre-adoption period, among both adopters and matched consumers, restaurant and bar transactions are 17.6% of the total offline transactions.

The other columns of Table 4 also are consistent with our hypothesis and provide additional robustness for our findings. Column (2) measures only transactions with merchants categorized as grocery stores⁶ and shows a similar-sized coefficient to our main result. While grocery shopping may be a routine for some people, some grocery trips may be more marginal for others. Column (3) measures transactions at merchants selling tradable and durable goods⁷ and the coefficient is smaller and statistically insignificant. We expect this relationship to be weaker, if any, because while durable or tradable goods are a broad category, they generally are less time-sensitive and can be purchased at more convenient times. Column (4) shows no relationship between changes in gas station spending and consumers who use ridesharing. In Column (5), while there is a positive

⁵SIC codes 5812, 5813, 5499

⁶SIC codes 5411, 5921

⁷SIC codes 5999, 5200, 5310, 5651, 5311, 5732

	9-month time window (log-ratio OLS models on total transactions)				
	Restaurant/Bar	Grocery	Gas Stations	Goods	Transit
	(1)	(2)	(3)	(4)	(5)
Constant Δ	0.078*** (0.028)	0.037 (0.028)	-0.069** (0.031)	-0.026 (0.029)	0.046 (0.092)
Log-Usage for Matched	-0.019 (0.013)	-0.011 (0.013)	0.024* (0.014)	0.005 (0.014)	-0.038 (0.039)
Actual Adopter Effect	-0.043 (0.044)	-0.061 (0.045)	0.028 (0.049)	0.032 (0.045)	-0.065 (0.119)
Δ Online Spending	4.100 (2.967)	3.285 (2.729)	-2.602 (5.242)	2.282 (1.422)	-3.574 (4.428)
Δ Online Spending, Squared	0.138 (0.971)	0.467 (0.843)	-1.082 (1.292)	0.850 (1.256)	-1.657 (1.047)
Δ Online Tran	0.116 (2.072)	0.850 (1.877)	4.358 (3.832)	4.619*** (1.085)	2.472 (3.651)
Δ Online Tran, Squared	-8.362*** (2.601)	-6.914*** (2.395)	-1.588 (4.123)	-4.664** (1.829)	-0.328 (2.648)
Log-Usage for Adopters	0.111*** (0.021)	0.055** (0.021)	0.001 (0.023)	0.001 (0.021)	0.073 (0.050)
Observations	4,957	4,719	4,389	4,669	1,188
Adjusted R ²	0.023	0.010	0.007	0.012	0.005
Residual Std. Error	0.899 (df = 4949)	0.846 (df = 4711)	0.878 (df = 4381)	0.862 (df = 4661)	1.042 (df = 1180)

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4: Analyzing categories of transactions individually, the clearest and strongest signal we see is in restaurant and bar spending. The columns, respectively, look at subsamples of {restaurants and bars, grocery merchants, gas stations, tradable/durable goods, public transit services}.

association between transit spending and ridesharing usage, it is statistically insignificant, partly because there are also fewer consumers that have transit spending in both periods and the estimate has reduced power. Columns (3) - (5) provide additional falsification checks because they reduce the concern of the coefficient on ridesharing usage being driven by unobserved confounders (such as rising incomes or increased propensity to spend in the post-adoption period). For example, if rising incomes were driving our findings, we may expect it to affect transactions at both restaurants and goods merchants. As a related note, increases in restaurant and bar transactions is more elastic to ridesharing availability, it is not solely driving our findings; if we remove these transactions from the data, we continue to find that ridesharing usage explains increases in consumer transaction frequency in the post-adoption period.

Consumer Demographics Of particular interest to urban managers and policy makers who are considering policies to subsidize or regulate ridesharing services, we consider how ridesharing may

have affected consumers differently along demographic lines. We consider (A) the spending amount of each adopter pre-adoption, as a proxy for income or purchasing power, and (B) the age of the credit card account at the time of adoption, as a rough proxy for the age of the consumer. We find significant heterogeneity in the effect of ridesharing along a consumer’s pre-adoption spending amounts—note however, in Section 2.4, that there is no relationship between a consumer’s pre-adoption spending and their choice of how much to use ridesharing.

Columns (1) and (2) of Table 5 compares the main model estimates between consumers in the bottom 50% of pre-adoption spending (Column 1) versus the top 50% (Column 2). The estimated coefficient on observed usage is significantly positive in Column (1), but insignificant in Column (2). This implies that our main consumer frequency results are being driven by those consumers in bottom 50-percentile of pre-adoption spending. By itself, this would raise a concern about pre-trends driving our findings. However, one, there is no pre-trend relationship between pre-adoption spending and post-adoption usage; and two, even amongst the bottom 50-percentile of spenders, consumption changes are still explained entirely by ridesharing usage not a general adoption effect. In Section 6’s event study, there is also no evidence of pre-trends.

In Columns (3) and (4) of Table 5, we see a similar notable difference between the estimated main model on accounts in bottom 50% of account age versus those in the top 50%. Column (3), with the subsample of younger accounts shows the significant coefficient on ridesharing usage that we observe in our main results, but Column (4), with the subsample of relatively older accounts, no longer shows a relationship between usage and post-adoption changes in consumption.

For the main sample of the 270-day window, the 50% percentile of pre-adoption spending per month, among adopters, is \$934, and the 50% of account age at time of adoption is 3.56 years.

These results around the heterogeneity of our estimated effects suggest a couple hypotheses. Assuming that account age and pre-adoption spending are strong proxies for demographic age and purchasing power, this suggests that ridesharing has a larger impact on the spatial frictions for those younger and with lower purchasing power. For example, if younger and lower purchasing-power people are those with less access to personal vehicles, ridesharing may in fact have a larger impact on their spatial frictions; while those who have higher purchasing power may

	9-month time window (log-ratio OLS models)			
	<50-percentile Spenders	>50-percentile Spenders	<50-percentile Account Age	>50-percentile Account Age
	(1)	(2)	(3)	(4)
Constant Δ	0.070*** (0.023)	-0.102*** (0.021)	0.036 (0.030)	-0.049** (0.021)
Log-Usage for Matched	-0.014 (0.011)	0.026*** (0.009)	-0.0004 (0.013)	0.016 (0.010)
Actual Adopter Effect	-0.026 (0.040)	0.033 (0.034)	-0.091* (0.047)	0.089** (0.036)
Δ Online Spending	1.919*** (0.715)	-0.003 (1.757)	0.332 (2.131)	0.470 (0.399)
Δ Online Spending, Squared	-0.508 (0.519)	-0.354 (0.528)	-0.048 (0.414)	0.071 (0.352)
Δ Online Tran	2.096*** (0.767)	3.369*** (1.105)	1.056 (1.605)	2.556*** (0.534)
Δ Online Tran, Squared	-1.630*** (0.474)	-4.053** (1.590)	-3.416** (1.596)	-0.777 (0.559)
Log-Usage for Adopters	0.087*** (0.019)	-0.007 (0.015)	0.074*** (0.022)	0.0001 (0.018)
Observations	2,756	2,715	1,918	2,332
Adjusted R ²	0.035	0.049	0.029	0.022
Residual Std. Error	0.602 (df = 2748)	0.469 (df = 2707)	0.570 (df = 1910)	0.490 (df = 2324)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: Among subsamples of the consumers, comparing the relationship between ridesharing usage and changes in consumption. Columns (1) and (2) compare low versus high spenders pre-adoption, and Columns (3) and (4) compare those with shorter versus longer account age. It appears that among those whose spatial frictions are affected, those who have younger account ages and those who had lower pre-adoption spending demonstrate all of the consumption elasticity, whilst the older accounts and those that spent more, appear to show no significant relationship between usage and consumption changes.

simply use ridesharing to replace their existing transportation options. Alternatively, this lower-spending demographic may also be more sensitive to changes in spatial frictions; they may have a higher elasticity of offline consumption with respect to spatial frictions.

5 Results: Spatial Consumption Behavior

Our second main question of interest is how a shock to spatial frictions affects the geographic nature of individual's consumption. Ridesharing services lower spatial frictions to neighborhoods that are further away, or ones that are less accessible — for example, not near home or work, or not near convenient public transit. In our analysis of the first question, we find evidence that those consumers who are affected by Uber/Lyft do appear to increase the frequency of their offline consumption.

Beyond increases in frequency though, we are interested in how the effects of ridesharing mani-

festations geographically. Do these increases happen simply in the same neighborhoods that consumers were already going to, the null hypothesis? Or, does ridesharing’s availability effectively “shrink” the city to some consumers? We hypothesize the latter and that among those consumers who are affected by ridesharing, there should be an increase in the diversity of spatial consumption. As a consequence, farther away or less frequently traveled neighborhoods will be more accessible. For example, they may be able to travel downtown more easily or frequently than before, or they may be able to visit a more diverse set of neighborhoods. This is in line with the arguments of Couture (2013) that consumers are variety-seeking in cities.

To study this, we focus on the available data on merchant locations. We observe the 5-digit zip code of merchant locations that consumers transact at. As detailed in the data section, the availability of merchant zip code data is not complete though. Unfortunately, in the data, we do not have complete data on individual consumer home locations though.

We use the main model and compare each consumer’s (normalized) unique number of merchant zip codes between their pre and post-adoption periods. Again, we have attempted to best remove online transactions and ridesharing transactions, so this is attempting to reflect transactions in unique offline zip codes. We count all the unique zip codes of merchants in the 9 month period, and normalize this by the number of months observed in the period.

Table 6 reflects our general findings, using the same simple and main identification models that we use in evaluating consumption frequency. Column (4) shows a significantly positive relationship between how much a consumer used ridesharing services and an increase in unique merchant zip codes visited. This coefficient is similar across all four columns.

These results are tangential to our findings on increasing consumption frequency. It suggests that while some consumers may be increasing their consumption activity due to ridesharing, it also means that they may be patronizing a wider array of merchants and neighborhoods. As we get into in our zip-code-centric analysis, this can have meaningful implications for urban management. Further, this finding again adds additional robustness to our hypothesis of the mechanism of reduced spatial frictions. If an unobserved confounders such as income were increasing consumption frequency, it

	9-month time window			
	log (ratio of number of (normalized) unique zips post vs pre) (OLS)			
	Adopters-Only	w/Matched Group		w/Online Controls
	(1)	(2)	(3)	(4)
Constant Δ	0.041*** (0.010)	-0.019 (0.018)	-0.032** (0.015)	-0.028* (0.015)
Log-Usage for Matched		0.036*** (0.009)	0.004 (0.007)	0.003 (0.007)
Actual Adopter Effect			0.013 (0.024)	0.009 (0.024)
Δ Online Spending				2.073 (1.512)
Δ Online Spending, Squared				0.417 (0.558)
Δ Online Tran				1.689* (0.954)
Δ Online Tran, Squared				-4.889*** (1.407)
Log-Usage for Adopters			0.032*** (0.011)	0.030*** (0.011)
Observations	2,396	2,396	5,471	5,471
Adjusted R ²	0.000	0.007	0.007	0.022
Residual Std. Error	0.507 (df = 2395)	0.505 (df = 2394)	0.485 (df = 5467)	0.481 (df = 5463)
<i>Note:</i>			*p<0.1; **p<0.05; ***p<0.01	

Table 6: Consumers who use ridesharing service more also seem to increase the number of unique zip codes they visit in the post-adoption period.

perhaps would not be affecting spatial consumption behavior.

It is possible that the positive estimated coefficient between ridesharing use and an increase in unique zip codes visited could be confounded by consumers who travel frequently in the post-adoption period. To test against this confounder, in the appendix, we consider only zip codes within a specific city’s geographic boundary and continue to show similar results and a similarly positive coefficient.

5.1 Zipcode/Neighborhood Perspective

Ridesharing does not just affect consumer choices; as a consequence, ridesharing may also be affecting the local economy. We have so far focused our analysis on those consumers whose local spatial frictions appear to be lowered by ridesharing. In this section, we consider which zipcodes and neighborhoods may be more strongly affected by ridesharing. We focus on the geography of a large US city, City A, with a lot of zip code availability. We identify all the observed zip codes that fall within City A’s geographic boundaries and analyze if there are changes in consumption patterns

in these zip codes due to ridesharing.

Our main finding is that, among high-usage adopters, there is meaningful variation in the consumption pattern changes among the zip codes of City A. There does not to be a uniform change in the various zip codes, but rather, some seem to be affected positively, and others negatively. Furthermore, the zip codes that appear to be affected positively seem to be geographically centered in the city-center, while those that are affected negatively appear to be geographically in the city-suburbs. Of note, these local-geographic-economy impacts may not be generalizable to beyond this large US city; they may be due to specific ways that the city’s population and consumption amenities are distributed. This evidence of variation in how consumption in zip codes change post-adoption is further supportive evidence of our hypothesis.

The first analysis is on all the top 35 zipcodes with the most observations in the post-adoption period (not restricted to City A). The analysis is conducted in a simple difference-in-differences setup, between the high-usage and low-usage groups. We use the same post-adoption and pre-adoption 270-day periods. We then estimate 35 simple diff-in-diff analyses, among adopters only, by grouping the adopters into two coarse groups of low-usage and high-usage adopters, based on the median post-adoption usage of 5 transactions. Finally, we then run 35 independent regressions, one for each zip code, keeping only consumers who have at least one transaction in that zipcode either pre-adoption or post-adoption. This analysis does not adjust significance estimates for multiple testing, since it is instead interested how the estimates vary across zipcodes. Each diff-in-diff regression is thus: $\ln Y_{ip} = I_1 Period_p + I_2 High_i + I_3 High_i Period_p$. The coefficients I_1 and I_2 control for a general event-time trend and static high-low group differences respectively. The coefficient I_3 is the coefficient of interest, which provides an average estimate on the high-usage consumers’ change in average consumption frequency in that zipcode, in the post-adoption period, relative to the change of the low-usage consumers.

Figure 11 shows these estimated I_3 post-adoption coefficients for the top 35 zipcodes in our data, irrespective of city location. The zipcodes are hidden for anonymization. The fact that some zipcodes show significant relative increases in activity, and some show significant relative decreases in activity suggest that ridesharing has indeed affected behavior, not just overall, but also shifting

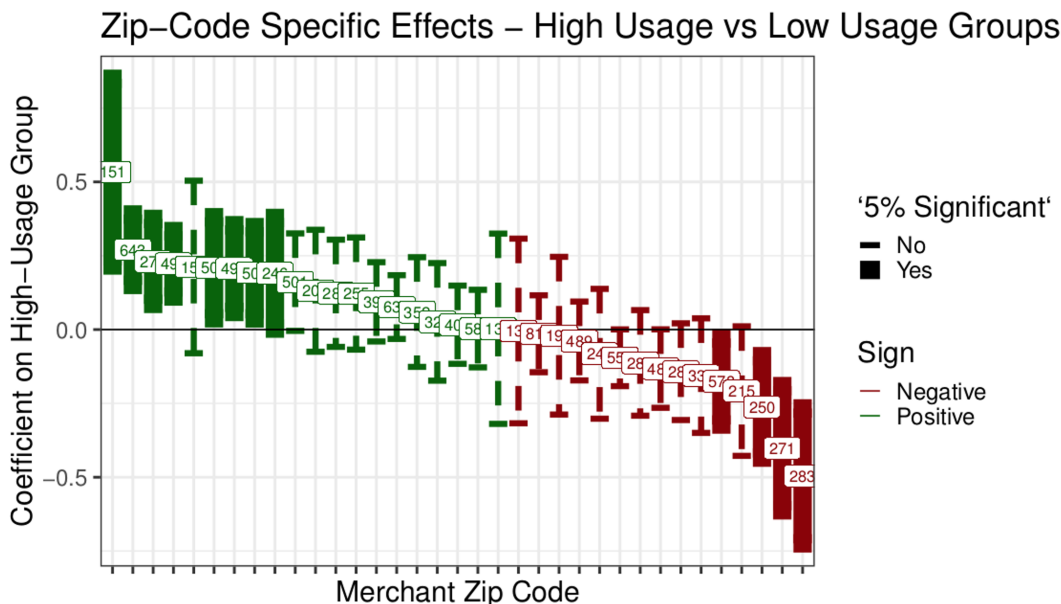


Figure 11: The estimated post-adoption consumption frequency coefficients for the top 35 zipcodes in our data, determined by pre-adoption consumption activity. The fact that some zipcodes suggest significant increases in activity, and some suggest significant decreases in activity suggest that ridesharing has indeed affected behavior, shifting it geographically. It also suggests that our results are not just confounded by an overall increase in consumption frequency.

it geographically. It also suggests that our results are not just confounded by an overall propensity-to-spend increase.

In the second analysis, to further evaluate the spatial aspects of this shift in consumption behavior, we focus on the zip codes in the geographic boundaries of City A. Figure 12 shows a map of this metropolitan area and its top zipcodes. In this figure, the green points (solid or crossed) represent a positive estimated coefficient (ignoring statistical significance and multiple testing), where the high-usage consumers spend relatively more frequently in those neighborhoods post-adoption, relative to what the low-usage consumers do. Conversely, the red points represent areas where the high-usage consumers have fewer transactions there in the post-adoption period, relative to the low-usage consumers. It's notable that most of the green points are concentrated in the city-center, while the red points tend to lie in the suburbs surrounding the city. A possible mechanism explaining this, consistent with anecdotes and prior literature on drunk-driving incidents, is that this may be driven by those who live outside the city and before ridesharing, chose to consume locally rather than in the city. This could be motivated by parking, traffic, car-access, or drunk-driving

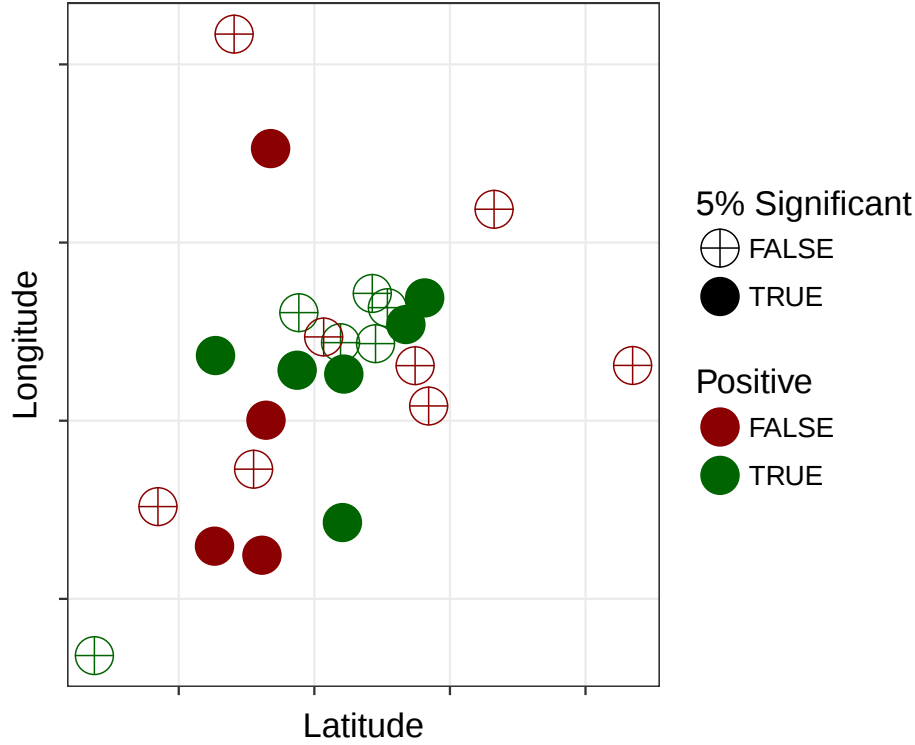


Figure 12: A map of the estimated high-usage group coefficients in a large US metropolitan area. Most of the zipcodes with suggested increases in activity for the high-usage group (relative to the low-usage group) lie near the city center, while those with relative decreases in activity lie outside the city.

considerations.

6 Robustness Checks

To check the robustness of our main findings on consumer consumption patterns, we utilize a variety of alternative model specifications, falsification tests, and consumer subsamples.

6.1 Robustness in Model Specifications

Logged-Difference Outcome Variables In our main model, we consider a first differences model using the logged outcome variables (e.g. $\Delta Y_i = \ln Y_{i1} - \ln Y_{i0}$). However, to ensure that our specification is not driving the results, we can also consider model with the logged difference as the outcome variable (e.g. $\Delta Y'_i = \ln(Y_{i1} - Y_{i0})$). This does not have as clear an underlying behavior model per-period, since the log-transformation is applied after the first difference.

	9-month time window			
	log (first-diff of transactions per month) (OLS)			
	Transactions Per	Number of Days	Zips Per	Avg Amt Per
	(1)	(2)	(3)	(4)
Constant Δ	0.115** (0.055)	0.042 (0.037)	-0.028 (0.018)	-0.322*** (0.088)
Log-Usage for Matched	0.006 (0.026)	0.010 (0.017)	0.012 (0.009)	0.051 (0.041)
Actual Adopter Effect	0.050 (0.088)	0.011 (0.059)	0.052* (0.030)	0.088 (0.133)
Δ Online Spending	8.172 (6.144)	8.388* (4.739)	2.218 (3.121)	35.745*** (8.043)
Δ Online Spending, Squared	-2.471 (1.694)	1.900 (1.369)	-0.375 (1.299)	3.550 (2.962)
Δ Online Tran	12.180*** (4.250)	3.913 (3.268)	2.431 (2.315)	-22.151*** (5.200)
Δ Online Tran, Squared	-22.561*** (5.156)	-16.549*** (4.072)	-6.094** (2.697)	-23.211*** (7.256)
Log-Usage for Adopters	0.116*** (0.043)	0.069** (0.028)	0.029** (0.015)	-0.016 (0.063)
Observations	5,471	5,471	5,471	5,471
Adjusted R ²	0.040	0.027	0.026	0.002
Residual Std. Error (df = 5463)	1.859	1.236	0.634	2.768
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01		

Table 7: Replacing the outcome variable with a different transformation, the log-transformation on the difference between periods (rather than the difference between logged outcomes), the results are robust and directionally the same.

Column (1) of Table 10 shows the results for transactions per month. The coefficient on ridesharing usage is on a different scale, but it remains significantly positive in the same manner as our main model. Similarly, there also continue to be no significant adoption effects nor differences between the adopters and control group. Columns (2) and (3) of Table 7 show the same style of findings, except with number of days per month and unique number of zip codes per month as the logged-difference outcome variables.

Calendar Time Models Instead of considering an event-time model centered around an individual’s adoption date, we can instead use specific calendar-time periods. This model then has fixed effects for individuals and specific time periods. This does not have pre or post-adoption periods, but rather the calendar year is split up into a block of months, such as {2, 3, 4, 6} month-size blocks. For example, if we use 3-month blocks, the time periods are quarters of the year: Jan-Mar,

Apr-Jun, Jul-Sep, Oct-Dec, for each year.

We consider these larger month blocks because we want to reduce confounding from individual unobservables in particular time-periods. For example, if we used week blocks and a consumer's birthday falls in that week, both their propensity to consume and propensity to use ridesharing may increase, but we do not observe that it was their birthday that week. Our main model, of using pre- and post-adoption periods, controls for this by using the same calendar days in both periods, but just separated by a year. In this model, without relative time, we instead seek to control for this using large blocks of time that smooth out these individual consumption stochasticities. The results remain robust using even half-year (6 month) blocks.

The simple model, when comparing just adopters amongst themselves is:

$$\ln Y_{it} = \delta_t + \alpha_i + \tau_a \text{Adopt}_{it} + \beta \ln(\text{Usage}_{it} + 1) \cdot \text{Adopt}_{it} + \epsilon_{it}$$

This controls for time-period and individual fixed effects (δ_t and α_i), and uses a differences-in-differences model moderated by a continuous treatment variable, $(\ln \text{Usage}_{it} + 1)$. In time periods after an individual adopts, an individual may not be observed using ridesharing, thus we add 1 to the usage in each period, for all individuals. We use a difference-in-differences setup because there is always zero usage time periods before an individual adopts. We are interested in β , the effect of ridesharing usage in periods including and after when an individual adopts. τ_a estimates the adoption effect. β estimates the continuous treatment impact of ridesharing usage after adoption.

Table 8 shows the results of this fixed effects model for just the sample of adopters. Within this subsample of self-selected adopters, and controlling for specific time-period and account fixed effects, the results are consistent with our main findings and central hypotheses. Generally, we do not see any effects of adoption itself on consumer spending. However, if a consumer, via revealed preferences in their actual ridesharing usage, shows reduced spatial frictions in the time-periods, only then do they show signs of both increased consumption frequency and neighborhood diversity. Columns (1) - (3) show results for 6-month time periods, and Columns (4) - (6) show results for 3-month time periods. We expect the 3-month time period estimated coefficients to be slightly higher because of potential confounding in the stochasticity of consumption, however, the differences are

	Various logged-outcomes (month-window size in parentheses)					
	Transactions Per (6)	Days Per (6)	Zips Per (6)	Transactions Per (3)	Days Per (3)	Zips Per (3)
	(1)	(2)	(3)	(4)	(5)	(6)
After-Adoption Indicator	0.019 (0.013)	0.016 (0.011)	0.006 (0.010)	-0.015 (0.010)	-0.013* (0.008)	-0.016** (0.007)
Log-usage (before adoption)						
Log-usage (after adoption)	0.140*** (0.005)	0.109*** (0.004)	0.062*** (0.004)	0.153*** (0.005)	0.120*** (0.004)	0.085*** (0.003)
Time-window Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Acct Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49,608	49,608	49,608	92,228	92,228	92,228
Adjusted R ²	0.683	0.643	0.632	0.664	0.621	0.620

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 8: Results from the alternative specification fixed effects model, which does not use time relative to adoption, but instead considers the influence of the usage of ridesharing in wide time windows. Within the sample of adopters, the results are consistent with main results, where low-usage adopters show no significant pattern whether we use 6-month time windows (Columns (1) - (3)) or 3-month time windows (Columns (4) - (6)).

minimal.

Event Time Model In our main model, we only consider two periods, which are defined around the adoption event, and have the same corresponding calendar days. To get a sense of the potential dynamics of the effect of ridesharing, we use the calendar time model, with event-time indicators. We find the influence of ridesharing usage is fairly constant over time.

We consider a 3-month time window. The calendar time window that contains the adoption date is event time period 0. We then consider the three periods before, and the three time periods after the adoption period. We also ensure a balanced panel, by ensuring that all consumers have all 7 event time periods. Lastly, we only consider adopters in this analysis. Thus, for each consumer, we observe, at a minimum, both 270 days after adoption and 270 days before adoption.

The estimating model accounts for both calendar time fixed effects, individual fixed effects, and event time fixed effects. The coefficient of interest is the interaction of ridesharing usage with event time indicators:

$$Y_{ite} = \alpha_i + \delta_t + E_e + E_e \cdot \ln Usage_{it} + \epsilon_{it}$$

In our discussion of these results, we omit the individual and calendar-time fixed effects. In Figure 13, on the left, we show that there is no statistically significant event-time pattern. This is an

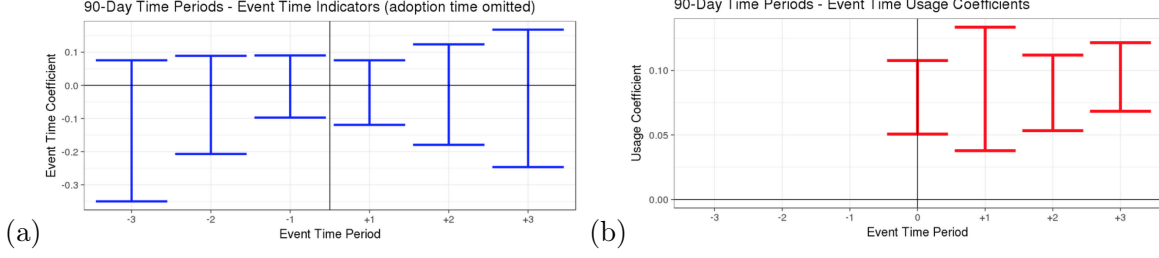


Figure 13: On the left, we show that there appear to be no significant pre-adoption trends for adopters, nor a post-adoption effect for the low-usage users. This is consistent with our earlier evidence that suggests no pre-trends and no adoption effect.

On the right, the usage coefficients remain statistically constant over event-time after adoption. This is consistent with our hypothesis of lower frictions having a constant impact, rather than a confounding, adoption-time-specific effect.

important robustness consideration, reflecting similar null results to the yearback falsification test. The null coefficients for the post-adoption periods also suggest robustness to our result that low-usage adopters have no change and there is no significant adoption effect. On the right, we consider if the coefficient on usage is changing over time (e.g. does it drop off substantially over time). We find no significant evidence that the usage coefficient changes over event time, consistent with the hypothesis that the lower frictions from ridesharing may have a fairly constant impact on local economic consumption.

Furthermore, this result is robust if we remove the usage coefficient, and instead, coarsen our analysis by splitting the consumers into two groups of high versus low-usage consumers.

6.2 Robustness in Account Selection

6.2.1 Robustness in Removing Extreme Outliers

Lastly, even though our outcome variables are log-transformed, we may be concerned that our results could be driven by extreme outliers in the data. As we see in the data summary, though we consider log-transformed ridesharing usage, there are still a few accounts with a particularly high number of ridesharing transactions in the post-adoption period. For example, in the $N = 270$ time window, the 98th, 99th, and 100th (i.e. max) percentile of ridesharing transactions are 67, 87, and 356 respectively. Log-transformed, these are not as extreme (4.2, 4.5, and 5.9), however, we still want to ensure that our results are robust to these extreme outliers. For each time window, we

	9-month time window (top 2% ridesharing users removed)			
	Transactions per month	Days per	Unique zip per	Avg amount spent
	(1)	(2)	(3)	(4)
Constant Δ	−0.013 (0.016)	−0.014 (0.013)	−0.028* (0.016)	−0.049*** (0.014)
Log-Usage for Matched	0.005 (0.008)	0.004 (0.006)	0.004 (0.007)	0.005 (0.007)
Actual Adopter Effect	−0.004 (0.027)	−0.006 (0.021)	0.006 (0.024)	0.012 (0.021)
Δ Online Spending	2.751 (2.207)	3.531** (1.792)	2.076 (1.503)	3.570*** (1.178)
Δ Online Spending, Squared	0.251 (0.700)	1.203** (0.473)	0.410 (0.548)	0.015 (0.375)
Δ Online Tran	2.261 (1.461)	0.656 (1.169)	1.651* (0.948)	−2.214*** (0.691)
Δ Online Tran, Squared	−6.571*** (1.963)	−5.894*** (1.550)	−4.876*** (1.395)	−2.168* (1.114)
Log-Usage for Adopters	0.044*** (0.014)	0.029*** (0.011)	0.033*** (0.012)	−0.0002 (0.010)
Observations	5,364	5,364	5,364	5,364
Adjusted R ²	0.031	0.023	0.022	0.0005
Residual Std. Error (df = 5356)	0.548	0.430	0.482	0.424

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 9: As a robustness check against extreme ridesharing users, which is the treatment of interest, we remove the top 2% of ridesharing users. This moves the 98th-100th percentiles from 67, 87, and 356 to 47, 57, and 67. Columns (1) - (3) confirm that our main results hold around consumption frequency, and Column (4) confirms that a falsification test also continues to hold.

remove the top 2% of accounts in ridesharing usage and verify our main findings. In the $N = 270$ time window, with a 6-month minimum in each time period, the resulting new 98th, 99th, and 100th percentiles are 47, 57, and 67 respectively.

Table 9 shows the results of this robustness check. Columns (1) and (2) shows the main results around consumption frequency, with transactions per month and days per month; Column (3) shows the main result around the number of unique zip codes; and Column (4) shows a falsification test on the average amount spent per transaction. In all columns, the results are similar in magnitude and significance to the main results.

6.3 Robustness in Outcome Variables

6.3.1 Extreme weather.

Another falsification test we consider is considering only consumption on extreme weather days. We hypothesize that days with particularly cold, rainy, or snowy weather⁸ will have higher spatial frictions in general, and if present, some potential confounders would have less influence on these days. For example, if a post-adoption income boost is occurring, we hypothesize that this confounder will have a more limited effect in extreme weather days (due those days having inherently higher spatial frictions). In the opposite direction though, if a consumer is actually being affected by ridesharing, we hypothesize that they will have a larger ridesharing shock on spatial frictions on extreme weather days. We hypothesize that this larger shock of spatial frictions on extreme weather days should be in the opposite direction of potential confounders.

Thus, in this robustness check, we only consider transactions that occur on these extreme weather days, and compare extreme weather consumption activity pre-adoption vs post-adoption. We focus on the weather patterns only of City A and thus use a smaller subset of consumers whose mailing addresses are in City A or City A's surrounding suburban area, and have at least one extreme weather transaction pre-adoption and post-adoption. While due to variation in weather, the number of extreme weather days can differ between an individual's pre- and post-adoption periods, overall, there is no relationship between the difference in extreme weather days and adoption dates, nor any relationship between the difference in days and post-adoption usage.

To confirm our main results for this smaller sample, Column (1) of Table 10 replicates the main results for all days of the pre- and post-periods. The outcome here is total transactions, unnormalized by the number of months to compare with Column (2), where we also consider the total transactions also, but only on extreme weather days. The point estimate for the coefficient on ridesharing usage is higher in Column (2) than in Column (1), which is directionally consistent with the hypothesis that ridesharing lowers frictions more on extreme weather days. Column (3) demonstrates the robustness of this estimate by showing a similar relationship for ridesharing usage and post-adoption spending on extreme weather days. Additionally, of note, the point estimates are

⁸High temperature < 18 degrees F or precipitation > 0.206 inches or snow > 0.38 inches. These numbers come from the median values in City A for precipitation and snow, and from temperatures on winter days.

	9-month time window (log-ratio OLS models)			
	Tot Tran	Weather Tran	Weather Spent	Weather Avgamt
	(1)	(2)	(3)	(4)
Constant Δ	0.035 (0.036)	-0.010 (0.041)	-0.065 (0.054)	-0.055 (0.037)
Log-Usage for Matched	0.0001 (0.015)	-0.001 (0.018)	-0.004 (0.024)	-0.003 (0.017)
Actual Adopter Effect	-0.048 (0.057)	-0.108 (0.066)	-0.191** (0.093)	-0.083 (0.062)
Δ Online Spending	0.437 (3.137)	4.798 (2.971)	11.118*** (3.852)	6.320 (4.595)
Δ Online Spending, Squared	-0.199 (0.324)	1.096*** (0.318)	0.756 (0.491)	-0.340 (0.396)
Δ Online Tran	-0.025 (2.640)	-1.294 (2.575)	-6.526** (3.145)	-5.232 (3.791)
Δ Online Tran, Squared	-2.790 (1.801)	-6.937*** (1.589)	-9.495*** (2.422)	-2.557 (2.648)
Log-Usage for Adopters	0.056** (0.027)	0.076** (0.032)	0.099** (0.043)	0.022 (0.030)
Observations	1,387	1,387	1,387	1,387
Adjusted R ²	0.018	0.043	0.024	0.001
Residual Std. Error (df = 1379)	0.588	0.671	0.901	0.598

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 10: Measuring only days of the pre/post period with extreme weather, our results suggest that spatial frictions continue to be lower in extreme weather days.

more negative for the adopter matched group in Columns (2) and (3), suggesting that adopters, pre-adoption, transact and spend particularly less on extreme weather days than the matched control group. Lastly, Column (4) again uses the average amount spent per transaction on extreme weather days as a falsification test, and as in our main results, there is no significant coefficient on ridesharing usage in this case.

7 Conclusion

With the pervasiveness of mobile technology and location-based computing, new forms of smart urban transportation, such as Uber and Lyft, have become increasingly popular. These new ridesharing platforms can influence individuals' movement frictions and patterns, in turn influencing local consumption patterns and the economic performance of local businesses. To gain insights about future impact of urban transportation changes, in this paper, we analyze individuals' urban con-

sumption patterns before and after their adoption and usage of the ridesharing services. Our study is validated using a novel panel dataset of fine-grained individual anonymized credit card and debit card transactions from January 2012 through May 2016 from a large U.S. bank, including from over 10K observed adopters of ridesharing. Based on revealed preferences, we hypothesize that those who choose to use ridesharing more are those whose mobility preferences are affected more by the availability of ridesharing, and vice versa. Our findings demonstrate a significant positive impact from the usage of ridesharing services on individuals' consumption frequency, as well as the spatial diversity of their spending. Our results also indicate strong heterogeneity in such effect. The effect becomes significantly stronger in increasing restaurant and bar transactions, and also stronger among younger customers and those who spend less before adopting the ridesharing services.

Overall, this research contributes to a better understanding on the impacts of transportation infrastructure and of spatial frictions. Managerially, looking forward, this helps urban managers consider how self-driving vehicles may influence consumer choices, as well as, adding contemporary and recent evidence on the potential pros and cons of the recent growth of not only Uber & Lyft, but ridesharing services in general. While Uber & Lyft may be stable services in U.S. cities, today, there is still a variety of political discussion on the regulation or promotion of such ridesharing services.

References

- Angrist, Joshua D, Sydnee Caldwell, and Jonathan V Hall (2017). *Uber vs. Taxi: A Driver's Eye View*. Working Paper 23891. National Bureau of Economic Research. DOI: 10.3386/w23891. URL: <http://www.nber.org/papers/w23891>.
- Babar, Yash and Gordon Burtch (2017). "Examining the Impact of Ridehailing Services on Public Transit Use". In: *Social Science Research Network*.
- Batty, Michael (2013). *The new science of cities*. Mit Press.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How much should we trust differences-in-differences estimates?" In: *The Quarterly journal of economics* 119.1, pp. 249–275.

- Black, Dan A, Natalia Kolesnikova, and Lowell J Taylor (2014). “Why do so few women work in New York (and so many in Minneapolis)? Labor supply of married women across US cities”. In: *Journal of Urban Economics* 79, pp. 59–71.
- Burtch, Gordon, Seth Carnahan, and Brad N Greenwood (2018). “Can you gig it? An empirical examination of the gig economy and entrepreneurial activity”. In: *Management Science*.
- Chan, Jason and Anindya Ghose (2014). “Internet’s Dirty Secret: Assessing the Impact of Online Intermediaries on HIV Transmission”. In: *MIS Quarterly* 38.4, 955–A3.
- Cohen, Peter et al. (2016). *Using big data to estimate consumer surplus: The case of uber*. Tech. rep. National Bureau of Economic Research.
- Cook, Cody et al. (2018). “The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers”. In: *Working paper*.
- Couture, Victor (2013). “Valuing the consumption benefits of urban density”. In: *Job market paper*.
- Couture, Victor and Jessie Handbury (2017). *Urban revival in America, 2000 to 2010*. Tech. rep. National Bureau of Economic Research.
- Davis, Donald R et al. (2017). *How Segregated is Urban Consumption?* Tech. rep. National Bureau of Economic Research.
- Eizenberg, Alon, Saul Lach, and Merav Yiftach (2016). “Retail prices in a city”. In:
- Forman, Chris, Avi Goldfarb, and Shane Greenstein (2012). “The Internet and local wages: A puzzle”. In: *The American Economic Review* 102.1, pp. 556–575.
- Glaeser, Edward L, Matthew E Kahn, and Jordan Rappaport (2008). “Why do the poor live in cities? The role of public transportation”. In: *Journal of urban Economics* 63.1, pp. 1–24.
- Glaeser, Edward L, Jed Kolko, and Albert Saiz (2001). “Consumer city”. In: *Journal of economic geography* 1.1, pp. 27–50.
- Greenwood, Brad N and Sunil Wattal (2017). “Show Me the Way to Go Home: An Empirical Investigation of Ride-Sharing and Alcohol Related Motor Vehicle Fatalities”. In: *MIS Quarterly* 41.1, pp. 163–187.
- Hall, Jonathan D, Craig Palsson, and Joseph Price (2018). “Is Uber a substitute or complement for public transit?” In: *Journal of Urban Economics*.

- Hall, Jonathan V and Alan B Krueger (2018). “An analysis of the labor market for Uber’s driver-partners in the United States”. In: *ILR Review* 71.3, pp. 705–732.
- Li, Zirui, Yili Hong, and Zhongju Zhang (2016). “Do ride-sharing services affect traffic congestion? An empirical study of uber entry”. In: *Social Science Research Network* 2002, pp. 1–29.
- Mojumder, Probal, Jason Chan, and Anindya Ghose (2016). “The Digital Sin City: An Empirical Study of Craigslist’s Impact on Prostitution Trends”. In: *Proceedings of International Conference on Information Systems 2016*.
- Nian, Tingting, Yuyuan Zhu, and Vijay Gurbaxani (2016). “Uber on: Until you are free of debt”. In: *Workshop on Information Systems and Economics 2016*.
- Seamans, Robert and Feng Zhu (2013). “Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers”. In: *Management Science* 60.2, pp. 476–493.

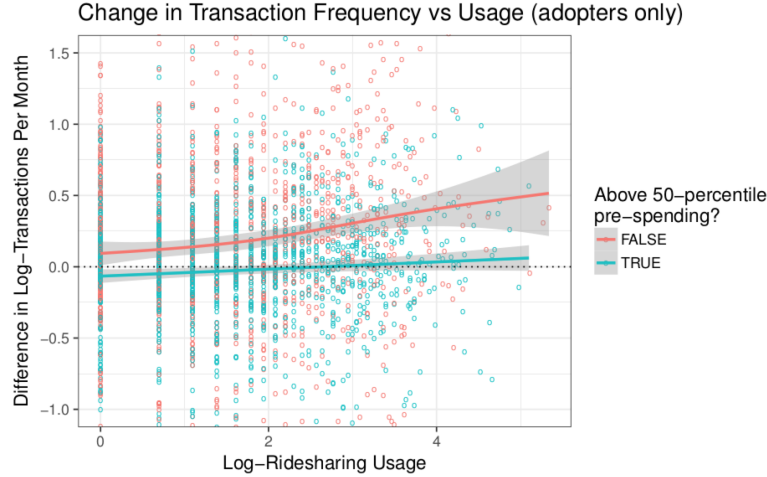


Figure 14: The scatterplot compares the change in transaction frequency in the post-adoption period (y-axis) versus the ridesharing usage in the post-adoption period (x-axis), among adopters only. The adopters are split into two groups based on their pre-adoption spending, above or below the 50th-percentile. Those in the bottom 50% continue to show a similar increasing relationship between usage, and those low-spending adopters who do not use continue to show no change. However, those in the top 50% of pre-adoption spending do not show a notable relationship in their between-period changes with their ridesharing usage.

To reiterate, referring back to Figure 6, remember that there is no relationship between usage and spending in pre-adoption.

8 Appendix

8.1 Additional Figures

1. Figure 14 shows model-free evidence of demographic differences (and non-differences) on spending. In it, on the left, it shows a very notable change in the average change for low-spending adopters vs high-spending adopters. However, on the right plot, it still highlights the importance of ridesharing usage in explaining this change; low-usage, low-spending adopters do not show any patterns in their post-adoption change.

8.2 Additional Analyses

8.2.1 City A Robustness

City A Robustness Our overall metric on the unique number of zip codes is susceptible to a potential confounding from those who travel a lot more in the post-adoption period than in their pre-adoption period. This could cause an increase in both Uber usage and consumption in

a larger number of unique zip codes. Thus, to test the robustness of our findings, we consider only transactions that occur at merchants inside our largest city (hereafter named City A). This geographic limitation provides a clearer estimate and robustness test on our hypothesis of spatial frictions and diversity of consumption.

We consider consumers that have at least 10 restaurant, bar, or grocery transactions in City A in the time period before adoption. We use this to identify consumers who reside in or nearby City A.

Column (1) of Table 11 checks the robustness of our zip codes findings. Considering only transactions in City A, there does appear to be an increase in the number of zip codes with transactions among those who utilize ridesharing more, consistent with our overall main findings on unique zip codes. Column (2), (3), and (4) consider the approximate estimates of distance traveled for consumers, as detailed in Section 2.3. Partly due to the reduced power of this smaller sample and potentially because a sizable fraction of merchants do not have known zip codes, we do find significant results around distance traveled, either in the distance from a consumer’s centroid, or in the pairwise distances between the merchants they transact at.

8.3 Data Appendix

8.3.1 Uber and Lyft entry dates

Below, we have a list of various metropolitan cities in the US and their respective launches of Uber and Lyft transportation services.

- Atlanta, GA: UberBlack (Aug 2012), UberX (June 26 2013), Lyft (Aug 30 2013)
<https://www.cnet.com/news/lyft-launches-in-indianapolis-st-paul-atlanta/>
<https://newsroom.uber.com/us-georgia/introducing-uberx-better-than-a-taxi-for-the-same-price/>
- Baltimore MD: UberX (Oct 17 2013), Lyft (Oct 14, 2013)
<http://technical.ly/baltimore/2013/10/10/lyft-ridesharing-service-to-launch-oct-17/>
- Cincinnati, OH: UberBlack & UberX (March 11 2014), Lyft (March 28 2014)
<http://www.bizjournals.com/cincinnati/blog/2014/03/don-t-be-alarmed-by-the-pink-mustaches-lyft.html>
<https://newsroom.uber.com/us-ohio/uber-cincinnati-launches-with-free-rides-for-2-week>

	9-month time window (logratioOLS models)			
	Zips per month	Avg dist from centroid	Avg pairwise dist	Median pairwise dist
	(1)	(2)	(3)	(4)
Constant Δ	-0.008 (0.025)	0.014 (0.025)	0.028 (0.025)	0.034 (0.037)
Log-Usage for Matched	-0.012 (0.011)	-0.002 (0.011)	-0.007 (0.011)	0.007 (0.018)
Actual Adopter Effect	-0.035 (0.036)	-0.002 (0.042)	-0.010 (0.042)	-0.038 (0.059)
Δ Online Spending	0.013 (1.557)	-3.161 (3.923)	-3.086 (3.768)	2.305 (1.923)
Δ Online Spending, Squared	-1.281*** (0.237)	-0.242 (0.318)	-0.257 (0.302)	-0.273 (0.375)
Δ Online Tran	-0.400 (1.222)	3.311 (3.177)	3.040 (3.056)	-1.893 (1.417)
Δ Online Tran, Squared	-0.503 (1.034)	1.465 (2.215)	1.402 (2.126)	-2.087 (1.498)
Log-Usage for Adopters	0.049*** (0.017)	-0.002 (0.021)	0.010 (0.020)	0.006 (0.029)
Observations	1,733	1,733	1,733	1,733
Adjusted R ²	0.015	-0.001	-0.002	-0.002
Residual Std. Error (df = 1725)	0.391	0.481	0.469	0.681

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 11: As a robustness check, we restrict the geographic boundary to only those transactions with zip codes in City A. The results on the number of unique zip codes is robust to this restriction, showing similar results among consumers who appear to transact in City A. However, there is no significant pattern with respect to other distance traveled metrics, such as the average distance from centroid or the average pairwise distance between transactions.

- Chicago, IL: UberBlack (Sep 2011), UberX (April 22 2013), Lyft (May 9 2013)
<https://newsroom.uber.com/us-illinois/chicagouberx/>
<http://www.latimes.com/business/technology/la-fi-tn-lyft-chicago-name-change-20130509-story.html>
 - Cleveland, OH: UberBlack (April 2014), UberX (April 9 2014), Lyft (April 11 2014)
<http://www.clevescene.com/scene-and-heard/archives/2014/04/09/uber-launches-in-cleveland-today>
http://www.cleveland.com/business/index.ssf/2014/04/-pink-mustached_lyft_launches_its_own_car_service_in_cleveland_today.html
 - Columbus, OH: UberBlack (Oct 2013), UberX (Feb 22 2014), Lyft (Feb 21 2014)
<http://www.columbusunderground.com/lyft-launching-in-columbus-bw1>
<https://newsroom.uber.com/us-ohio/uber-launches-in-columbus/>
- (Stopped in Jan 2015) <http://www.dispatch.com/content/stories/local/2015/01/06/Lyft-pauses-operations-in-Columbus.html>

- (Lyft Relaunched in 2016) <http://www.dispatch.com/content/stories/business/2016/03/24/1-lyft-relaunching-in-columbus.html>
- Indianapolis IN: UberX (Sep 5, 2013), Lyft (Aug 29, 2013)
 - <https://newsroom.uber.com/us-indiana/indy-launches-free-uberx-month-w-rider-zero-paul-george/>
 - <https://www.eventbrite.com/e/lyft-indianapolis-launch-party-tickets-7872582105#>
- Louisville KY: UberX (Apr 24, 2014), Lyft (Apr 24, 2014)
 - <https://newsroom.uber.com/us-kentucky/louisville-your-uber-is-arriving-now-2/>
- Nashville, TN: UberX (Dec 10 2013), Lyft (Dec 6 2013)
 - <http://blog.lyft.com/posts/lyft-rolls-into-nashville>
 - <https://newsroom.uber.com/us-tennessee/nashville-uberx-better-cheaper-faster-than-a-taxi/>
- New York, NY: UberX (Sep 02 2012), Lyft (Jul 24, 2014)
 - <https://techcrunch.com/2012/09/12/uber-ride-sharing/>
 - <http://allthingsd.com/20120702/a-status-symbol-moves-down-market-whats-behind-the-uberx-launch/>
 - <http://www.businessinsider.com/uber-is-finally-going-to-offer-cheaper-cars-2012-7>
 - <https://www.uber.com/blog/san-francisco/sf-vehicle-choice/>
 - <https://www.theverge.com/2012/9/7/3300244/uber-taxi-new-york-travis-kalanick-rogue>
- Raleigh NC: UberX (Apr 24 2014), Lyft (Apr 24, 2014)
- Philadelphia, PA: UberBlack (Oct 2014), UberX (Oct 24 2014), Lyft (Jan 30 2015)
 - <http://blog.lyft.com/posts/philadelphia>
 - <http://technical.ly/philly/2014/10/24/uberx-philadelphia-free/>
- Pittsburgh, PA: UberX (Feb 11 2014), Lyft (Feb 7, 2014)
 - <https://newsroom.uber.com/us-pennsylvania/pittsburgh-your-uber-is-arriving-now/>
 - <http://iheartpgh.com/2014/02/07/lyft-pittsburgh-pink-mustache-cabs/>
- Los Angeles: UberX (June 13, 2013), Lyft (Jan 31, 2013)
 - <http://romsdeals.com/2013/06/18/uber-in-downtown-los-angeles-my-first-time-using-uber/>
 - <https://thehub.lyft.com/blog-c/2013/02/04/lyft-la-celebrates-one-year>
- San Francisco, CA: UberX (July 24, 2013), Lyft (Aug 25, 2012)
- Washington DC: UberBlack (Dec 2011), UberX (Aug 8 2013), Lyft (Aug 2013)
 - <https://newsroom.uber.com/us-dc/dc-uberx-is-arriving-now/>

8.3.2 Uber and Lyft merchant names

We detail specifically how we determine if a transaction is a ridesharing transaction on the Uber or Lyft ridesharing platforms.

Lyft: Any merchant with the word “LYFT” in the merchant name is a Lyft transaction. We checked all matches, and they are all Lyft transactions.

Uber: The word “UBER” in merchant names results in various false positives, such as merchants from cities named “HUBER” or other merchants with the word “Uber” in their name. We thus identify all merchants with the word “UBER” in the name, and arrange in order of their observed total number of transactions. Using this, we create a whitelist of Uber transactions. We may miss a few transactions with odd spelling that we did not notice, this keeps our estimate of ridesharing usage as conservative. The whitelist of Uber merchant names are:

{“uber technologies inc”, “uber technologies inc.”, “uber”, “uber bv”, “paypal *ubertechnol”, “google *uber technologi”, “google *uber”, “google * uber technologie”, “uber technologies”, “uber technologies”, “uber*~~~~” (anonymized), “paypal *uber~~~~”, “paypal *uber bv ~~~~ B”, “ady*uber technologies”, “google *ady*uber techno”, “paypal *ubertechnol ube”, “paypal *uber llc”, “sq *uber ~~~~~~ I”, “sq *uber technologies, in”}

8.3.3 Removing non-merchant transactions

We remove all transactions with the following merchant names:

{“online credit card pmt”, “new account check”, “balance transfer*bc”, “promo convenience check”, “convenience check”, “payment*thank you”, “line of credit advance”, “online banking payment”, “credit card cash advance”}.

As an additional precaution, we also remove transactions to merchants with the SIC code 6010, which are categorized as payments to credit card companies.

8.3.4 Categorizing online transactions

In our transactions, to the best of our knowledge, we remove online transactions. To do so, we use a merchant’s SIC code, because merchant names can be noisy due to misspelling or minor variations

in formatting. We analyzed the SIC codes in the data and identified the following as being closely corresponding to online transactions.

5942 (bookstores, with the large majority being Amazon.com),

5735 (online music and media merchants, such as iTunes, Google Play, or Amazon Video)

5968 (online subscription services, such as Netflix, Spotify, or Dollar Shave Club)

7399 (online payments, such as Venmo, Amazon Web Services)

8999 (online payments, such as Paypal, Stubhub)

7311 (payments to online platforms, such as Google, Facebook, or Microsoft)

4816 (payments for online services, such as Amazon, Go Daddy, or Microsoft)