Ridesharing, Spatial Frictions and Urban Consumption Patterns

Zhe Zhang*, Beibei Li†

November 2018

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 platforms can influence individuals' movement frictions, 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 consumption 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 credit card and debit card transactions from January 2012 through May 2016 from a major 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 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.

*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 or ridesharing 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 impacts, and policy efforts made to block their entry into metropolitan areas. Today, several years after their entry into major cities though, 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. With the large popularity of these services, have they just replaced previous transportation modes, or does their popularity signal a change in individual behaviors?

In contrast to efforts to curb ridesharing, other recent urban policies promote the use of ridesharing services, and promote it as a potentially new form of public transit. In 2017, Innisfil, Canada chose to subsidize Uber for its citizens as a form of public transit, instead of adding an additional bus line. Other cities have subsidized Uber or Lyft for specific purposes, such as Pinellas, Florida subsidizing Uber for low-income workers who work late-night shifts; Evesham, New Jersey subsidizing rides at night originating from select bars or restaurants in town; 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?

With the fast-growing popularity and observed paid usage of ridesharing services, we theoretically frame this mechanism as a negative shock to the spatial frictions of a subset of consumers. This provides a case study for the research not only on the impact of ridesharing specifically, but also on urban spatial frictions and transportation infrastructure. The goal of this paper is to study the first-order effect of ridesharing services, the actual usage of these services, on the behavior of those affected consumers. Looking forward too, there will likely be a substantial change in urban transportation via the entry of autonomous, self-driving vehicles. Cities are considering how the availability of these vehicles, which may provide very cheap and spatially-unrestricted transportation, could affect urban life and behavior.

In this paper, we 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 a panel of credit card accounts and their individual transactions from 2012 - 2016, including over 10K 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.

The major confounding concern is that there are unobserved, individual-specific time trends that affect both ridesharing usage and consumption behavior. 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.

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 couple areas of literature, that studying the impact of ridesharing (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.

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, Y. Zhu, and Gurbaxani, 2016), and public transit usage (J. D. 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 (J. V. Hall and Krueger, 2018; Cook et al., 2018; Angrist, Caldwell, and J. V. 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 F. 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 our data and measuring the outcome and treatment variables. In Section 3 and 4, we discuss the main methodology and results of studying our first on consumption frequency, as well as heterogeneity in the effects. In Section 5, we use the same methodology to study consumption spatial diversity. In Section 6, we consider a variety of robustness checks and alternative specifications on our main findings. We conclude in Section 6 and discuss managerial and policy implications.

1.2 Theoretical Framework & Mechanisms

We describe a simple spatial framework of consumption that motivates our hypotheses. This is inspired by theories of urban spatial interaction as a set of origins and destinations, with costs of movement, as surveyed in Batty (2013). In short, we assume for each set of consumers, there is a fixed set of origins $o \in \mathcal{O}$ and destinations $d \in \mathcal{D}$. If in a time period t, a consume chooses to from a location o to a destination d, they incur an individual-specific movement cost, or spatial friction, C^*_{itod}; and in the next time period, $t+1$, d is the next origin. Lastly, consumers also have an incentive to move in each time period, V_{itod}, which is individual-specific, and is normalized to the outside option of not moving, which is normalized to 0. We assume that V_{itod} stays fixed in our analysis for all consumers and time periods, so the key mechanism we focus on is C^*_{itod} and $(V_{itod} - C^*_{itod})$. Across a set of time periods $[t_1, T]$, the consumer chooses their set of origins and destinations.

$$\max_{\{o,d\}\in O\times D} U_{i,t_1,T} = \sum_{t_1}^{T} (V_{itod} - C^*_{itod})$$

C^*_{itod} consists of being the lowest-cost transport mode among several possible transport options: such as bus, personal car, walk, ridesharing services, or asking a friend. Each mode has its own associated costs. We assume the consumer chooses the lowest cost mode at time t, $C^*_{itod} = \min\{C^{bus}_{itod}, C^{car}_{itod}, C^{walk}_{itod}, C^{ridesharing}_{itod}, \ldots\}$.

The first takeaway is that, in our data, we observe when a trip is taken and $C^*_{itod} = C^{ridesharing}_{itod}$. We are interested in the direct effects of ridesharing, i.e. the actual usage of it, so we measure how often $C^*_{itod} = C^{ridesharing}_{itod}$. In other words, we measure ridesharing usage as the number of ridesharing trips taken. The effect of ridesharing on consumers spatial frictions can vary widely, some may be affected a lot while others who may already own a personal car may not see much change. For examples, for consumers j, we observe no trips, and then we conservatively assume that ridesharing has not had any marginal effect on their transport costs. Alternatively, for the consumers i, we observe many trips, then we assume that their spatial frictions have been substantially impacted. In other words, $C^{*,postadoption}_{itod} < C^{*,preadoption}_{itod}$, while $C^{*,postadoption}_{jtod} = C^{*,preadoption}_{jtod}$.

A second takeaway is that for consumers i who have $C^*_{itod} = C^{ridesharing}_{itod}$, we assume this to leads to an increase in $(V_{itod} - C^*_{itod})$ for those consumers; while for other consumers, e.g. j, who have no or few observed trips, we assume that $\left(V_{jtod} - C^*_{jtod}\right)$ has not changed. This naturally raises our research questions: if spatial frictions are being reduced for those consumers i who are affected, does this simply replace their existing transportation options, or does this reduction in costs increase the number of trips with positive utility and hence lead to more trips taken? This shift in consumption frequency is illustrated, as model-free evidence, in Figure 8.

Third, whether or not the number of trips changes, for those affected by ridesharing, does the shift in spatial frictions affect the set of destinations d, e.g. spatially or in diversity, or do the chosen destinations remain the same?

2 Panel Data on Detailed Credit Card Transactions

From a large U.S. financial institution, we have access to data from a random set of credit card accounts, including some over-sampling in two particular metropolitan areas of interest to the institution. For each sampled account, we observe all customer-level credit card transactions over 4+ years (53 months), from January 2012 through May 2016; not all accounts are active

throughout this time range though. This date range is appropriate because while both UberX and Lyft (popular 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. Among the metropolitan regions where we observe the most transactions, ridesharing services were not available until late 2013 or throughout 2014. In the appendix, we provide a list of major metropolitan areas and when Uber or Lyft entered that area.

For each sampled credit card account, we observe all transactions, and for each transaction, we observe the account ID, transaction date, amount spent, the merchant's standardized 4-digit Standard Industrial Categorization (SIC) code, city location, and the merchant name[1]. The SIC code is each merchant's self-identified categorization of their goods or services. Additionally, we also observe each account's monthly metadata, including credit limit, city of that month's mailing address. Our two main datasets are thus: a panel of raw transactions and a monthly of account metadata.

There are some data quality issues that we make adjustments for though. There are two batches of raw transaction data: [Jan 2012 - Dec 2014] and [August 2014 - May 2016]. We do not know the reason for the split, but we do know that the account IDs are consistent between the two. The newer batch of transactions is our reference, so from the older batch, we keep only account IDs that we observe in the newer batch. We combine these two batches by using the older data for the entire years of 2012, 2013, 2014; and use the newer data for 2015 - 2016. There are a few months of overlap between these two batches, and they match well, with all the accounts observed in the new batch also being observed in the old batch.

We remove duplicate transactions and credit card bill payment transactions. Duplicate transactions are defined as having the exact same merchant name, transaction amount, and date of transaction; if this occurs, we keep only one of the transaction. Credit card payments are one of the most popular transactions, with a merchant name like "PAYMENT*THANK YOU" or "ONLINE CREDIT CARD PMT"[2]. Since these have large transaction amounts and occur frequently in our data, we remove them from our sample. The data is also missing almost all transactions in the month of January, thus, we remove all January transactions to ensure fair comparisons within and between individuals.

A summary of the data is provided in the form of visualizing the various distributions of outcome variables and treatment variables, in order to be most transparent. This is shown

[1]Any stores with fewer than 10 transactions will be truncated to ensure it does not have personally identifiable information. For example, in some credit card transactions, the transaction number will be included in the merchant's 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.

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

in Figures 4, 5, and 7.

2.1 Ridesharing Usage Data

2.1.1 Defining Ridesharing Adoption Date and Usage

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 merchants. Uber has several false positives, so we only use a whitelist of known Uber merchant names and mark those as being Uber transactions. The specific whitelist (and examples of popular false positives) are in the appendix. These transactions include 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. However, there may be unobserved usage of Uber/Lyft. If ridesharing indeed has a positive impact on consumption and this unobserved usage occurs before our observed adoption date, this makes our results conservative by raising the consumption levels before adoption. If this occurs after our observed adoption date, and if this unobserved usage is more likely to be among consumers with low observed usage, this will likely also make our results conservative. Adopters with little usage are one of our control groups. Or if consumers were previously paying for ridesharing using a different method, this would affect our pre-adoption period and similarly make our result a conservative estimate.

2.1.2 Pre- and Post-Adoption Time Periods

Our central econometric identification is a two-period analysis, comparing individuals within themselves, between the immediate post-adoption period, and a corresponding pre-adoption time period. This is motivated by the discussion of serial correlation in difference-in-difference models in Bertrand, Duflo, and Mullainathan (2004), so we simplify our identification into this simple comparison of pre-adoption and post-adoption periods.

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 but a year before adoption. As we discuss in our econometric model, this removes individual-specific day-specific effects—i.e., if a post-adoption period includes

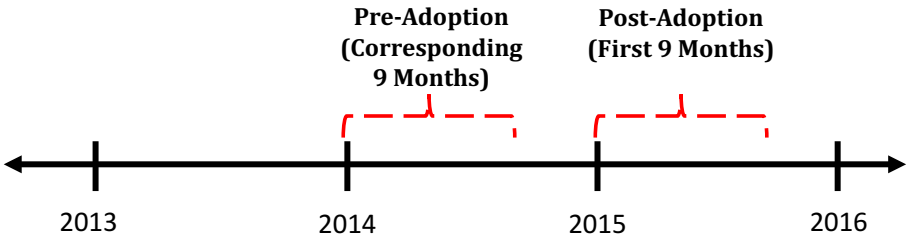


Figure 1: A hypothetical consumer who adopts ridesharing in Jan 2015. Jan - Sep 2015 is considered their post-adoption period, and as an apt comparison, we identify Jan 2014 - Sep 2014 as their pre-adoption baseline comparison period.

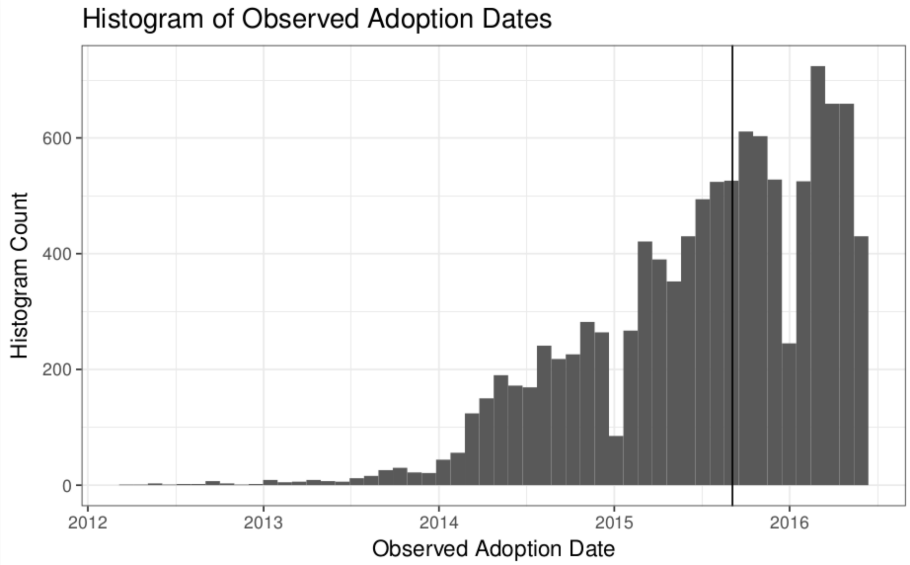


Figure 2: The distribution of observed adoption dates (i.e. first observed ridesharing transaction), among all the observed adopters in our data. The vertical black line represents our cutoff date for the sample of adopters we consider (all to the left of the black line), to ensure we observe at least 270 days after adoption.

a birthday or a child's specific events, the pre-adoption period will also include those. To be specific, if a consumer is observed to adopt 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 periods will be determined by their adoption date. Figure 1 shows an example of the adoption 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 include them in two robustness checks: smaller time windows, like $T = 180$, and in a fixed-effects model.

2.1.3 Minimum Activity Levels

To ensure that we are using fair comparisons, and to control for possible entry/exit of credit card consumers, we make two adjustments. First, we require that each consumer have a minimum amount of monthly activity to be considered in our sample. For the 9-month time window (270 days), since we remove January, we consider 8 months as the max amount of possible activity. From this, we allow for some vacation and periods of non-use, and choose 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 to the number of months observed in each period by dividing the total number of transactions by the number of unique months observed.

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 a fixed effects model in our robustness checks.

2.1.4 Discussion

Potential Use of Multiple Credit Cards. The requirement of a minimum activity level reduces some of the concern that consumers may have multiple or alternative credit cards. The comparison of individuals within themselves also seeks to address this. Further, for robustness, we consider a smaller dataset with the high and low outliers removed, to attempt to remove those who may only infrequently use this credit card, or those who use the card for business purposes. Additionally, to the best of our knowledge, the financial institution did not run any incentive campaigns to encourage spending on their credit card(s) during the data's time period, nor any ridesharing-specific campaigns.

Comparing the credit card users within themselves and requiring minimum account activity helps control for consumers potentially having multiple credit cards. It ensures that the credit card account was sufficiently active before adoption occurs, and that ridesharing usage is not being simply driven by new accounts. It is a possible confounder though that a consumer shifts additional consumption towards this credit card account, resulting in both additional spending and ridesharing usage. To adjust for this, we test for signs of this potential confounder in some of our robustness checks, such as considering the average amount spent or comparing the effect of ridesharing across different categories of merchants.

Further, it is anecdotal, but other major, national credit cards offer more rewards on travel spending, such as Uber or Lyft, than we understand that these credit card accounts

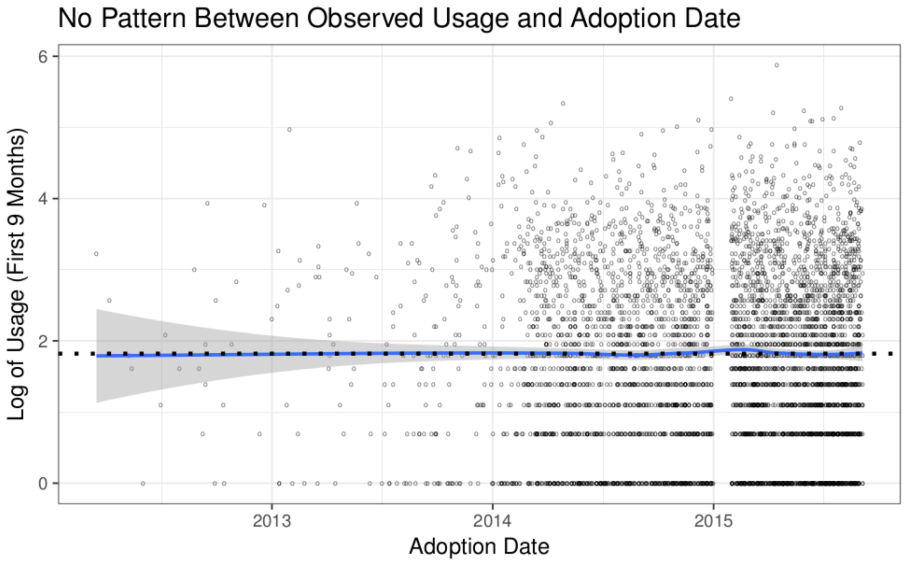


Figure 3: The mean of ridesharing usage is plotted as black dotted lines (on average, 11 transactions in the 270 days post-adoption). A flexible, local-linear line is then fit to the data. It shows there is no relationship between the future usage of adopters and the time of adoption. This reduces the 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.

do. Thus, we assume that if a consumer is choosing to pay for Uber and Lyft using this credit card, it is likely to be their "main" or only credit card they use. As discussed above, if consumers or adopters with low observed usage are using having unobserved ridesharing using another payment source or with friends, this makes our result a conservative estimate. Or if consumers were previously paying for ridesharing using a different method, this would affect our pre-adoption period and similarly make our result a conservative estimate.

Date of Adoption & Ridesharing Usage. There exists an overall increasing, over time, trend in total spending in the data. It is a thus a concern that those who adopt ridesharing later in time also use ridesharing more. We find that this is not the case. Among our adopters with minimum activity levels, we find no relationship between adoption date and ridesharing usage in the window of $T = 270$. Further, in that time window, the usage ridesharing appears to be roughly constant throughout that time period, for example, all the usage is not all concentrated in the early time periods. Figure 3 illustrates this.

Future Ridesharing Usage & Pre-Adoption Variables. Among the sample of adopters, we do not observe any relationship between an adopter's main pre-adoption consumption

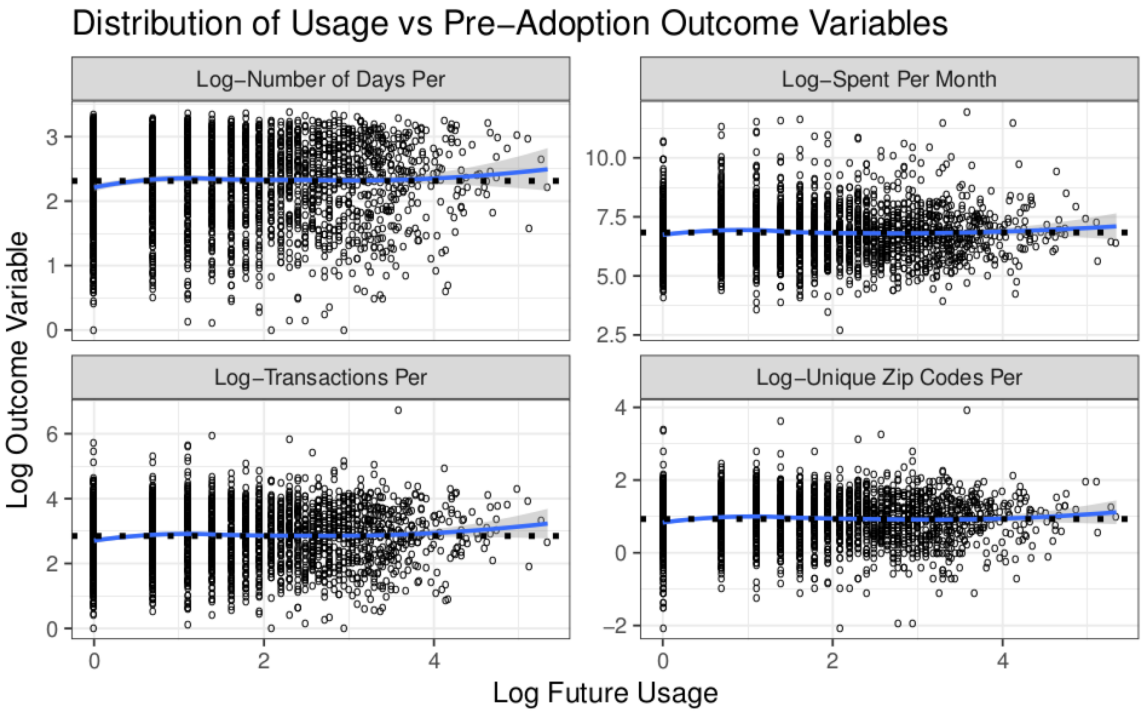


Figure 4: Each of the four plots shows a scatter plot of a pre-adoption consumption metric versus each consumer's future, post-adoption ridesharing usage. The black dotted line is the mean of each consumption metric, and the blue line is a flexible, local-linear fit to the scatterplot. As seen, the confidence 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. This again reduces concern that there may be individual-specific time trends that are associated with pre-adoption consumption metrics. For example, we do not see that higher spending adopters or adopters with more frequent transactions pre-adoption also use ridesharing more.

metrics and their resulting choice of ridesharing usage. This reduces the concern about potential individual trends that may be driving some consumers to both transact more and use ridesharing more in their post-adoption period. We demonstrate this in Figure 4.

As additional concern about individual time trends, we also consider a falsification test on a simulated adoption date that occurs a year before the actual adoption date.

Potential Unobserved Confounders. As we discuss in the econometric identification section, the central concern of the 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, such as rising income over time, an income boost, or perhaps a lifestyle change that coincides with ridesharing adoption.

Or, perhaps a consumer is choosing to spend more using this credit account than before, which results in a shift in the number of overall transactions observed and ridesharing transactions.

As a falsification test on these possibilities, we also consider the average amount spent per transaction as an outcome variable of interest. If consumers are using ridesharing as a result of an income boost in the post-adoption period, we hypothesize that there would also be a corresponding change in taste, as reflected by more expensive goods. On the other hand, no change in the average amount spent reduces the concern of a time-varying income boost.

2.2 Consumption Frequency Outcome Variables

The key outcome variable of interest is transaction frequency. This follows from the model of spatial frictions: if a consumer's spatial frictions are lowered by ridesharing, we expect their willingness to consume locally to increase. The first outcome of interest are transactions normalized per unique month observed, though the results are similar if we normalize or not.

Importantly, we remove online and ridesharing transactions from being measured in our outcome variables. We remove ridesharing transactions of course because they are the treatment variable of interest. We also remove online transactions because, in our model of spatial frictions, we model online transactions as having zero spatial frictions, so should not be affected directly by ridesharing services. Thus, our outcome variables seek to measure only plausibly 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.

We also consider a few other outcome variables as robustness checks: the number of unique days with a transaction, and amount spent per month observed. We expect to find similar results for these two outcome variables. The number of unique days with a transaction is a useful robustness check; since it is not affected by possible individual-specific time trends in consumption that increase the number of credit card transactions per day. Rather, the number of days outcome variable measures if ridesharing simply increases the number of days with transactions.

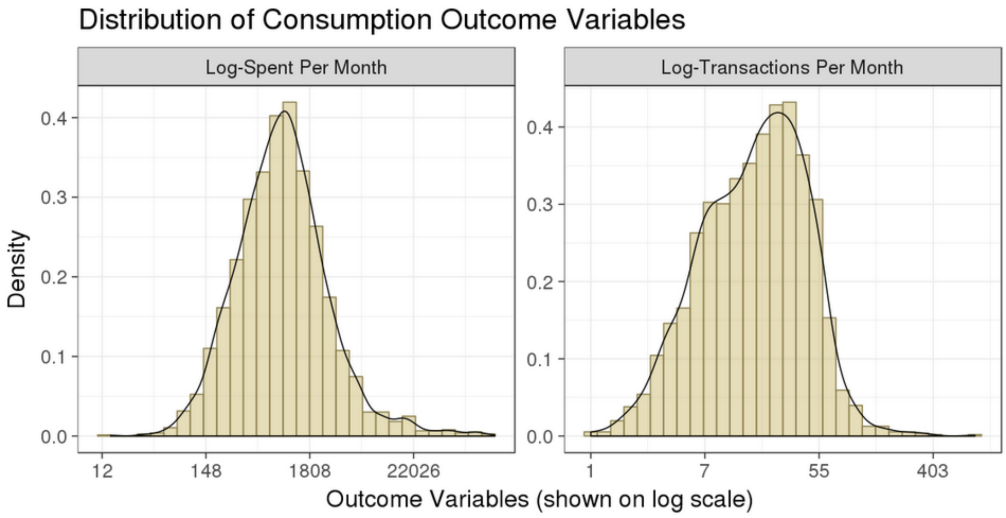


Figure 5: Each of the four plots shows a histogram of the distribution a consumption metric across consumers, from the pre-adoption periods. This provides one summary our data, in addition to Figure 7.

2.3 Geographic Outcome Variables

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 diversity, we consider two similar outcomes: the number of unique merchant zip codes visited in the post-period (or the version normalized by the number of observed months). We discuss these results overall in Section 5.

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

Using the overall count unique zip codes as an outcome metric could be confounded consumers who happen to travel more in their post-adoption period than in the pre-adoption period. Thus, we consider robustness measure where we only consider consumption in a geographic boundary of City A, the city with the most transactions in our data. City A's geographic boundary removes the possibility of this result being influenced by travel spending.

As an additional measure of the diversity of consumption, we consider 3 additional measures on the "web" or spread of a consumer's travel. In our outcome variables, we only

consider transactions that occur in these zip codes; this prevents the concern about inter-city travel spending affecting our geographic outcome variables, since its limited to spending in just this one city. Further, to avoid including those who travel to City A, we consider only those accounts with at least 10 restaurant, bar, or grocery transactions in City A in the time period before adoption.

For each 5-digit zip code, we assign a latitude and longitude from the 'zipcode' package from R based on the zip code's centroid. Each transaction is then treated as point-data instead.

We evaluate a couple simple distance-based metrics on :

- distance from consumer centroid. For each individual, based on their transactions with zip codes, we identify their transaction-weighted consumption centroid in both the pre and post periods (by taking the average latitude and average longitude). Then, we calculate the average distance from that centroid of transactions in that period. This is a rough measure of if a consumer is traveling further away from their estimated "center".

- pairwise distance of transactions. Among all transactions at merchants in City A, we calculate the average and median pairwise distance between all transactions. This is a loose approximation at the "web" of travel for each consumer. This attempts to consider the spread of a consumers area covered inside of City A. This is a useful metric to approximate the entropy or "web" of consumption. Even if a consumer is not traveling further from their "center", this metric would increase if a consumer, post-adoption, is consuming in a more diverse set of zip codes.

3 Econometric Methodology

In this section, we introduce our two econometric identification and estimation models. The first is a simple, parsimonious model that uses only the subsample of consumers who are adopters, and compares pre- and post-adoption periods adopters within themselves. It estimates both an adoption effect and an effect from ridesharing usage, though we hypothesize there to only be an effect from usage. The second model is our main model, which adds an additional control group of matched consumers and online spending as controls for potential unobserved, individual-specific time trends. In the robustness section, we introduce other potential controls and various falsification tests.

3.1 Simple Model: Ridesharing Usage is Cause of Change

In our conceptual model, we focus on highlighting the role of the usage of Uber/Lyft on changes in consumer local consumption. Based on the discussion of serial correlation in Bertrand, Duflo, and Mullainathan (2004), we simplify our identification into a simple comparison of two wide time-periods: pre-adoption and post-adoption periods. In this simple model, we only use the subsample of consumers who we observe to have adopted, i.e., they have at least one ridesharing transaction.

The identification of this conceptual model is a first-differences model, for an outcome variable between the pre-adoption and post-adoption time periods. With this variation in these consumption metrics, we then seek to explain this variation using just two variables: an adoption effect and an individual's usage of ridesharing. This pre-adoption vs post-adoption first-differences period is equivalent to a model with individual fixed effects, with two time periods and using relative time (to adoption). As an alternative, we also use a fixed-effects model, without relative time, in Section 6.

Since we are looking at spending and transaction behavior, which are skewed, we use log-transformed outcomes. This first-differences model removes individual, time-invariant variation in transaction/spending behavior. Figure 7 shows the distribution of the first-differences for four different outcome variables: transactions per month, spent per month, number of transaction days per month, and the number of unique zip codes in the period (normalized by months observed).

We start with a parsimonious model of consumption behavior on individual fixed effects, ridesharing usage, and individual-day time dummies. For a time period t, we define the time period as including days $d \in t$. This gives us a model for each time period:

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

With two time periods, a pre-adoption period and a post-adoption period, which have the same day dummies, we then calculate the first difference estimating equation. This equation differences out both the individual fixed effect and the individual day dummies:

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

We conservatively assume there is no ridesharing in the pre-adoption period before we observe usage, so we set $\ln Usage_{i0} = 0$.

In this estimating equation, the τ_a constant estimates the average adoption effect dummy, representing the conditional effect of being in the post-adoption period, relative to the same

days the year before[3]. This term also captures the average individual-specific time trend, if any. Even though consumers adopt at a wide range of different times, if there is any individual-specific time trend, this term will reflect that average, since it reflects the average difference between each consumer's pre- and post-adoption periods, after conditioning on $Usage_{i1}$. We theorize this to have no first-order effect and for $\tau_a = 0$.

Rather, we hypothesize that only those adopters who choose to utilize ridesharing—thus, those whose spatial frictions are affected by ridesharing—will show evidence of changing consumption. In other words, actual usage is revealed preferences that reveals the individuals whose frictions are affected more, ceteris paribus, more; and vice versa, those whose frictions are not. Further, we hypothesize that this is generally increasing, but diminishing in ridesharing usage: the more a consumer utilizes ridesharing, the more we hypothesize their spatial frictions are reduced, and the larger the effect would be. Thus, we estimate β, a coefficient on conditional linear relationship between $\Delta \ln Y_i$ and $\ln Usage_{i1}$, and hypothesize $\beta > 0$.

The null hypothesis here, as a comparison, is that ridesharing services actually has no effect on offline consumer consumption behavior, but instead, simply replaces consumers' existing transportation options. In such case, we wouldn't expect ridesharing usage to explain any variation in the consumers' change in spending between pre- and post-adoption periods.

Adoption of ridesharing is self-selected. It may be related to individual time-trends. We control for this first by only considering the subsample of adopters, and second, by estimating the average adoption effect, if any.

We don't explicitly control for time period fixed effects in our simple model. This is because of two empirical factors: the adoption event occurs in a wide range of times, and the time of adoption in empirically uncorrelated with a consumer's future choice of ridesharing usage. This is illustrated this in Figure 3 and is robust to different time windows and removal of outliers.

3.2 Main Model: Matched Control Group and Online Spending Controls

This simple model on usage controls for individual fixed effects (time-invariant consumption behavior) and time of year (due to how time periods are defined). However, both the adoption and the usage of ridesharing are self-selected.

[3]For 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].

If there are unobservable, individual time trends that affect both consumption and adoption, this would affect our estimated coefficient on the adoption effect. Since we both hypothesize this to be zero and observe it to be zero, we do not focus on this possible confounding.

More importantly, the amount of ridesharing usage may be affected by unobservable, individual time trends that also affect a consumer's post-adoption consumption. We discuss examples of potential confounders in Section 2.1.4. If present, these may bias our estimate of the coefficient on ridesharing usage on explaining post-adoption variation in spending. To first address this, in our main model, we introduce a few controls for individual, unobservable time trends. Later in the paper, we address this by testing for signs of potential confounding and testing our findings using falsification tests.

We expand our sample of consumers to include a select, matched set of non-adopting consumers. We select this sample using estimated propensity to adopt, as described in the subsection below. This control group has no usage nor an adoption-date though. To assign each matched control consumer a synthetic adoption date and ridesharing usage, we use nearest-neighbor, 1:1 matching and 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 reflect some of the unobservable, individual time trends that the adopters may have—except in a counterfactual world where they do not utilize ridesharing.

In other words, if the matched control group has similar unobservable individual time trends, they will provide an additional better baseline comparison than just the adopters' pre-adoption periods. Those matched control users who are assigned higher levels of synthetic usage have similar estimated propensity to adopt as the actual adopters with higher levels of usage; and we hypothesize that they will also have similar individual time trends.

Lastly, we remind the reader of the discussion in Section 2.1.4, where we highlight the empirical observation that a consumer's future choice of usage is uncorrelated with several pre-adoption consumption metrics like spending per month or transactions per month.

A second control we introduce in our main model for individual time trends is a consumer's online spending. 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—affecting both ridesharing usage and changes in consumption—we expect this to also affect their online spending. We measure both changes in their online spending per month ($\Delta Spent_{online}$) and in their online transactions per month ($\Delta Trans_{online}$) to capture any potential unobservable individual time trends. Further, we add a flexible quadratic linear model, $f()$, consisting of two coefficients, on each of these to better fit the model. We do not use a log-transformation on these in order to avoid removing consumers from the sample from may have no change, or

a change of 1 (e.g. using a "log-plus-one" transformation). In the robustness section though, we do check for robustness using a log-transformation on changes in online spending.

The resulting main model is:

$$\Delta Y_i = \gamma + \tau_a Adopt_i + \beta_1 \ln Usage_{i1} + \beta_2 \ln Usage_{i1} \cdot Adopt_i + f(\Delta Spent_{online}) + f(\Delta Trans_{online}) + \Delta\epsilon_i$$

The γ coefficient reflects a very 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 unobservable individual time trends of the control group—and β_2 is our actual coefficient of interest, reflecting the additional explanatory power of the adopters' actual ridesharing usage. As stated earlier, $f(\Delta Spent_{online})$ and $f(\Delta Trans_{online})$ reflect flexible quadratic models on online spending metrics to better act as proxies for potential confounders.

3.2.1 Matched Control Group

We identify a set of similar non-adopting consumers as a control group. For this set of control consumers, we match each one to an adopter with the most similar estimated propensity to adopt, and synthetically assign them that adopter's adoption date and subsequent post-adoption usage. We consider that this similar control group may also have similar, unobserved individual-specific time trends to their matched adopters, and help control for them if they exist.

We match consumers on their propensity to adopt ridesharing, using a propensity score model derived from a lasso logistic regression on many possible predictors, based on various overall consumption metrics, category-specific consumption metrics, and credit limit. We use a 1:1 nearest neighbor match, and also remove a small portion of adopters with very high or low propensity who do not have a close control group match. The use of a propensity score matching greatly improves balance on covariates between adopters and non-adopters, though there are still appear to be some slight differences. In an extension, we seek to improve this using exact matching. Figures 6 shows a comparison of propensity score balance and pre-adoption consumption.

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-

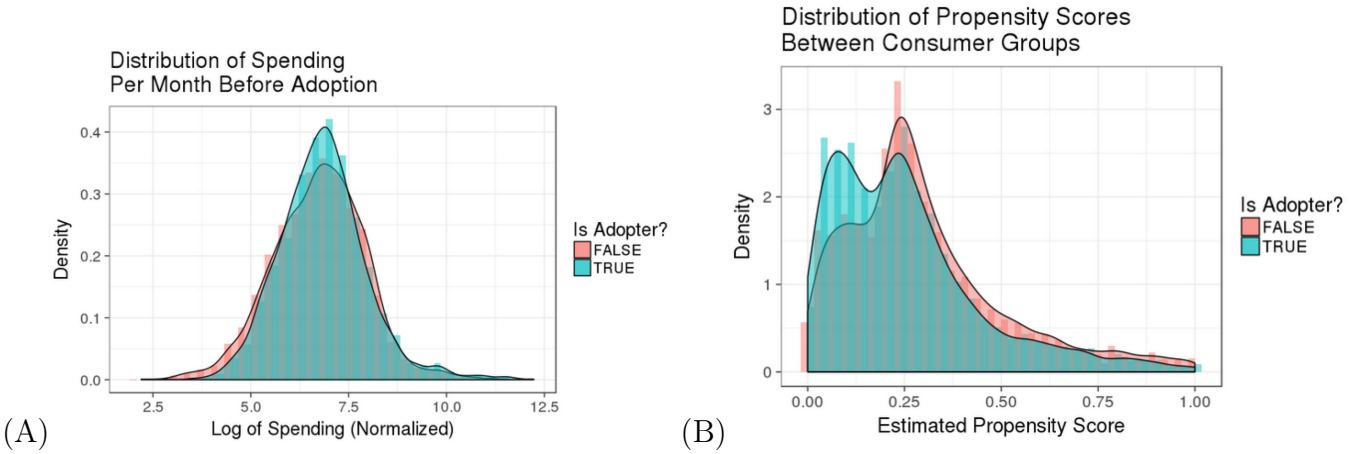


Figure 6: A comparison of the matched consumer group with the adopter group. In red are the matched consumer group, and in blue is the adopter group. On the left, in part (A), the two groups show a very similar distribution in spending per month, before adoption. On the right, in part (B), the adopters have a slightly higher concentration of lower propensity consumers.

consistent standard errors.

3.3 Data Summaries

Figure 7 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 the period (normalized by months observed). In Figure 5, we see that these outcome variables are fairly smoothly and normally distributed. Figure 8 shows the model-free evidence of both our simple conceptual

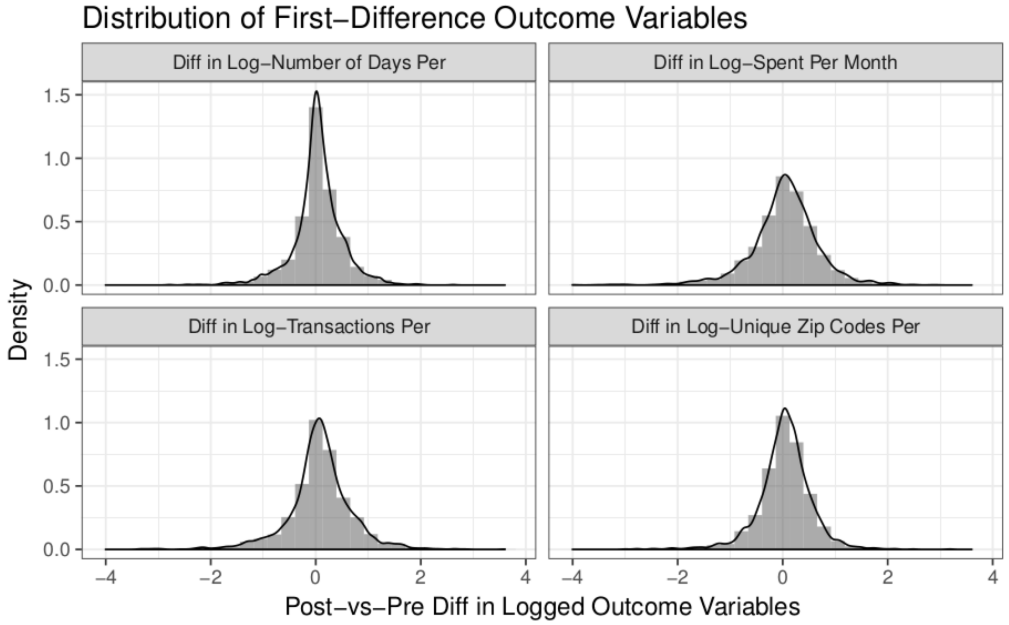


Figure 7: This plots the summary of the actual outcome variable that we're analyzing. Among all the adopters in the 270-day window, these are histograms of the comparison between the post-adoption and corresponding pre-adoption periods of transactions, spending, number of zip codes, and number of days (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, $\ln(Y_{i1}/Y_{i0})$.

model and model with matched consumers. For each scatter plot, on the y-axis, it plots the change in transactions per month for the sample of consumers (adopters or adopters + matched consumers), and on the x-axis, it plots the ridesharing usage of each consumer post-adoption. On top of each scatterplot, flexible local-linear lines are plotted for each consumer group. The adopters show a significant and substantial relationship between their usage and their post-adoption change in transactions per month.

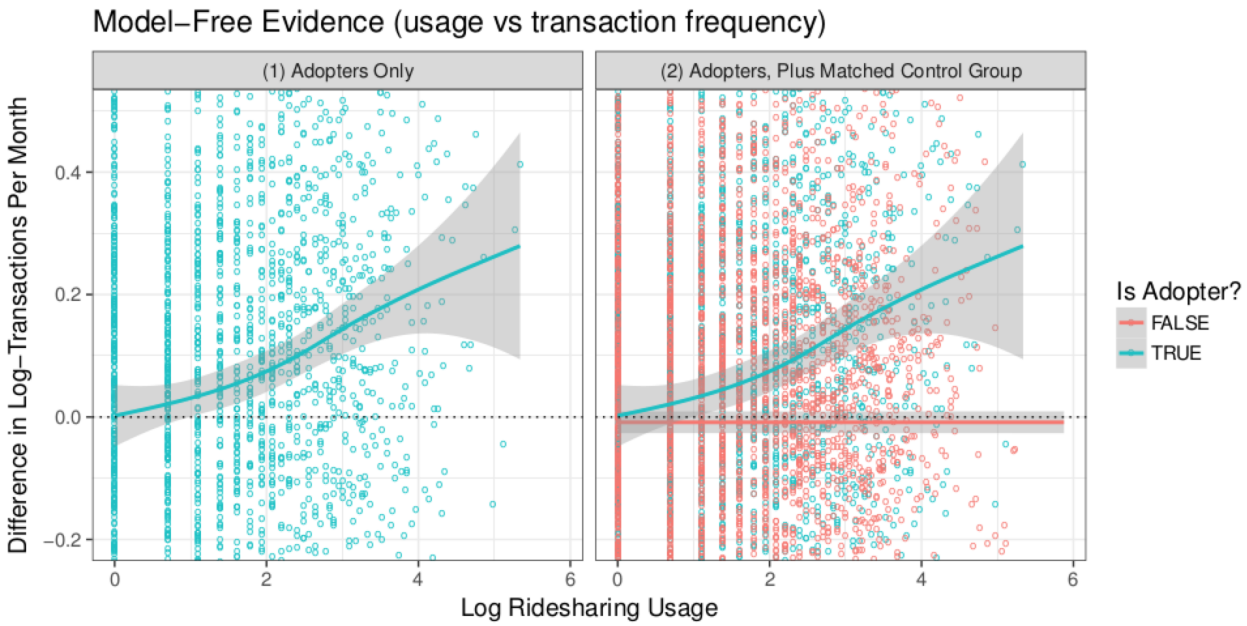


Figure 8: This scatterplot illustrates model-free evidence for the relationship between post-adoption ridesharing usage (x-axis) and between-period changes in transaction frequency (y-axis). The left plot shows the scatterplot of just the sample of adopters. The right plot adds an additional matched 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 trend 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 Consumption Frequency Results

Table 1 shows the results of our simple and main models on consumer transaction frequency. Column (1) only estimates the adoption effect, without controlling for each consumer's actual ridesharing usage, and shows a positive increase in the post-adoption period for adopters. However, in Column (2), once we introduce the actual ridesharing usage of adopters, this captures all the significant effect of ridesharing. There is no significant adoption effect in Column (2). This is consistent with our hypothesis and model; those who utilize ridesharing are those, by revealed preferences, whose spatial frictions are substantially affected, and as a consequence, increase their consumption frequency post-adoption. This goes against the null hypothesis, that those utilize ridesharing are simply replacing their existing transportation options and not changing their consumption choices.

Columns (3) and (4) of Table 1 introduce our main model, which adds a set of matched

control consumers for comparison and then adds controls for individual changes in income or propensity to consume based on observed changes in online consumption. The coefficient on being in the adopters group is insignificant in Columns (3) and (4), suggesting that there is no significant mean-difference between adopters and the matched group. Further, in both columns, the matched group's assigned usage also shows no relationship. If the potential individual-specific time trends—that affect both usage and spending—are related to an individual's propensity to adopt, the matched consumers, with the assigned synthetic usage may also reflect these unobservable individual-specific time trends; however, we do not see any significant post-adoption relationship between the matched group's usage and their consumption changes. Column (4) adds controls for differences in online spending and transactions between periods. This shows some predictive value, but does not appear to substantially affect the coefficient estimate on the effect of ridesharing usage.

The key coefficient of interest is that of how actual ridesharing usage for the adopters group explains post-adoption changes in transaction frequency. This is in the last row of Columns (3) and (4), and is significantly positive in all columns. Along with an insignificant adoption effect, this suggests that only those whose spatial frictions are affected by ridesharing services, as demonstrated by actually utilizing the service, show signs of increasing their offline transaction frequency.

All else equal, at least 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. In our robustness section, we consider other point estimates.

4.1 Main Robustness & Falsification Checks

Since the usage of ridesharing is self-selected, the key concern is that there may be an unobserved individual time trends (e.g. higher spending habits or rising incomes in post-adoption period) affecting both our treatment variable (observed usage) and outcome variables (consumption metrics). To test for this, we highlight a few robustness and falsification checks. We expand upon these checks in Section 6 too.

First, as a falsification test, we consider the average amount spent per transaction in 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, this confounder would show a positive coefficient between usage and changes in the average amount spent per transaction. However, we do not hypothesize that

	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***	−0.018	−0.016	−0.010
	(0.013)	(0.021)	(0.016)	(0.016)
Log-Usage for Matched		0.053***	0.004	0.003
		(0.011)	(0.007)	(0.007)
Actual Adopter Effect			−0.002	−0.007
			(0.027)	(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***	0.046***
			(0.013)	(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 $\log\left(y_{\text{after}}/y_{\text{before}}\right)$ 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.

ridesharing or lowering spatial frictions would have a first-order effect on the price-level of merchants that adopters patronize. Thus, as a falsification test, in Column (4) of Table 2, we use our main model specification with the first-difference in logged average amount spent per transaction as the outcome variable. Column (4) no longer shows a significant relationship between ridesharing usage and any changes in the average amount spent. This reduces the concern of there being unobserved confounders that would affect a consumer's average amount spent.

Second, as additional robustness, we demonstrate a 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}. Columns (2) and (3) of Table 2 use total transactions and spent per month as additional robustness checks. Since we don't observe average amount spent changing with usage, but do observe an increase in transaction frequency, it makes sense that there is an increase in the amount spent per month too.

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 to see if, among adopters, there are individual-specific time trends occurring before adoption. If we are concerned of individual time trends affecting our results between post- and pre-adoption periods, we posit that there may be evidence of these same time trends may be occurring in time periods before adoption too. To check for these, we perform a falsification test by shifting back our pre/post periods by exactly 1 year, i.e. synthetically assigning an adoption date 1 year before the actual adoption date[4]. We also synthetically assigning each adopter their future usage to the fake "post" period. Each consumer's future usage represents to proxy to test for any individual time trends that are present before adoption that may be related to both future usage and consumption. Figure 9 shows an example of how synthetic post and pre periods

[4]For 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].

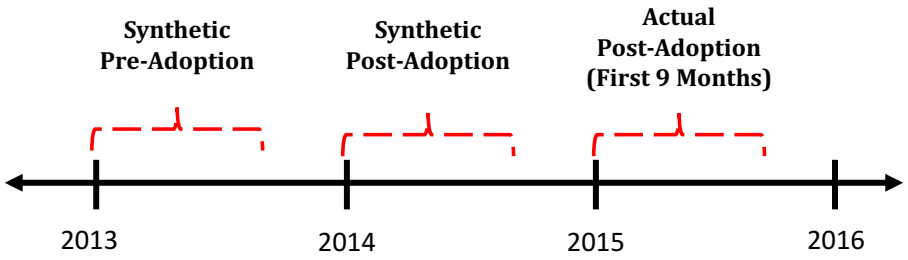


Figure 9: A hypothetical consumer who adopts ridesharing in Jan 2015. Jan - Sep 2015 is their *actual* post-adoption period. Thus, Jan 2014 - Sep 2014, instead of being the actual pre-adoption period, is now treated as the *synthetic* post-adoption period, and Jan 2013 - Sep 2013 is their *synthetic* pre-adoption period. We then compare the synthetic periods.

are collected, for a hypothetical consumer who adopted ridesharing in Jan 2015. If these individual time trends—rising incomes or increasing consumption preferences over time—are occurring before adoption, then we will observe similar positive coefficients associated with the synthetic "usage". As a precaution, we also do the same "yearback" shift for the matched consumers too.

The results of these falsification regressions 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* pre- and post-adoption period consumption metrics (rather than synthetic).

4.2 Heterogeneous/Detailed Effects

In our model, we hypothesize that leisure consumption may be more elastic with respect to spatial frictions. Other consumption that may be considered more necessary to an individual may already be occurring in the presence of higher spatial costs, but leisure consumption may be more marginal. This elasticity could either be due to relatively large transportation costs for leisure (or unscheduled) consumption—for example, for spontaneously going to a restaurant, 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 a consumer is drinking alcohol.

In the terms of our theoretical framework, this means that ridesharing has a larger reduction in C^*_{itod} when d is a leisure consumption destination like a restaurant or bar. An alternative explanation is that leisure consumption is more marginal for consumers. $|V_{itod} - C^*_{itod}|$ may be smaller for consumers when consider leisure consumption and thus more sensitive to changes in spatial frictions. A consumer may have a very high value in

	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	0.001	−0.062***	−0.052***
	(0.013)	(0.017)	(0.019)	(0.014)
Log-Usage for Matched	0.002	0.003	0.011	0.007
	(0.006)	(0.008)	(0.009)	(0.006)
Actual Adopter Effect	−0.009	−0.012	0.010	0.017
	(0.021)	(0.028)	(0.030)	(0.020)
Δ Online Spending	3.612**	4.231	6.460***	3.623***
	(1.809)	(2.646)	(2.400)	(1.180)
Δ Online Spending, Squared	1.208**	0.729	0.262	0.019
	(0.478)	(0.786)	(0.678)	(0.374)
Δ Online Tran	0.629	1.341	0.007	−2.218***
	(1.178)	(1.769)	(1.485)	(0.690)
Δ Online Tran, Squared	−5.960***	−7.938***	−8.859***	−2.232**
	(1.567)	(2.313)	(2.155)	(1.116)
Log-Usage for Adopters	0.032***	0.056***	0.042***	−0.004
	(0.010)	(0.014)	(0.014)	(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	−0.064***	0.030	−0.285***
	(0.020)	(0.016)	(0.019)	(0.036)
Log-Usage for Matched	0.011	0.012	0.020**	0.007
	(0.009)	(0.008)	(0.008)	(0.017)
Actual Adopter Effect	0.170***	0.032	0.139***	−0.008
	(0.032)	(0.025)	(0.030)	(0.054)
Log-Usage for Adopters	−0.015	−0.018	−0.021	0.101***
	(0.016)	(0.012)	(0.015)	(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.

traveling to work or to pick up children, and thus while ridesharing may lower their costs, these travel choices would occur with or without ridesharing.

We consider restaurant and bar transactions[5] as leisure consumption, and model the difference in just restaurant and bar transactions between periods. In Column (1) of Table 4, we see 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 stores[6] and shows a similar result to our main result. While grocery shopping may be a routine for some people, some grocery trips may be more marginal for others. Columns (3), measures transactions at merchants selling tradable and durable goods[7] 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 different times. In Column (4), we do not see a relationship between changes in gas station spending and consumers who use ridesharing. In Column (5), while there is a positive 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 it reduces 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); e.g., rising incomes may affect the number of transactions at restaurants and goods merchants. While restaurant and bar spending is more elastic to ridesharing's shock to spatial frictions, it is solely driving our findings; if we remove these transactions are from the data, we continue to find that ridesharing usage explains increases in consumer transactions in the post-adoption period.

Consumer Demographics. To assess potential heterogeneity in the effect of ridesharing, we consider a few available demographic aspects. 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 usage in pre-adoption spending,

[5]SIC codes 5812, 5813, 5499
[6]SIC codes 5411, 5921
[7]SIC 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***	0.037	−0.069**	−0.026	0.046
	(0.028)	(0.028)	(0.031)	(0.029)	(0.092)
Log-Usage for Matched	−0.019	−0.011	0.024*	0.005	−0.038
	(0.013)	(0.013)	(0.014)	(0.014)	(0.039)
Actual Adopter Effect	−0.043	−0.061	0.028	0.032	−0.065
	(0.044)	(0.045)	(0.049)	(0.045)	(0.119)
Δ Online Spending	4.100	3.285	−2.602	2.282	−3.574
	(2.967)	(2.729)	(5.242)	(1.422)	(4.428)
Δ Online Spending, Squared	0.138	0.467	−1.082	0.850	−1.657
	(0.971)	(0.843)	(1.292)	(1.256)	(1.047)
Δ Online Tran	0.116	0.850	4.358	4.619***	2.472
	(2.072)	(1.877)	(3.832)	(1.085)	(3.651)
Δ Online Tran, Squared	−8.362***	−6.914***	−1.588	−4.664**	−0.328
	(2.601)	(2.395)	(4.123)	(1.829)	(2.648)
Log-Usage for Adopters	0.111***	0.055**	0.001	0.001	0.073
	(0.021)	(0.021)	(0.023)	(0.021)	(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}.

but no heterogeneity based on the age of the account.

In Columns (1) and (2) of Table 5, we see that among those in the bottom 50% of pre-adoption spending (Column 1), those who utilize ridesharing also see a marked increase in their consumption. Those in the top 50% of pre-adoption spending though, shown in Column (2), do not see a significant relationship between their ridesharing usage and changes in their consumption. This suggests that our main results showing around the impact of ridesharing usage is being driven by those consumers who are in bottom 50-percentile of pre-adoption spending. This raises a potential concern of pre-trends driving our post-adoption observations. However, two key empirical observations reduce this concern: one, there is no pre-trend relationship between pre-adoption spending and post-adoption usage, and two, amongst the bottom 50-percentile of spenders, ridesharing usage is still the driving factor and there is no general adoption effect.

Also, in Columns (3) and (4) of Table 5, we see a similar notable difference between the accounts in bottom 50% of account age versus those in the top 50%. Column (3), with the subsample of relatively younger accounts shows the significant relationship 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 power. For example, if younger and lower purchasing power people have 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 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.

In the robustness section, we also look into the time of adoption and do not find a notable difference based comparing early vs later adopters.

Driven by Young, Low-Income Consumers? Out of 5471 consumers (adopters and matched), 4250 (78%) we have information on their account age. Among these consumers,

	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^{***}	-0.102^{***}	0.036	-0.049^{**}
	(0.023)	(0.021)	(0.030)	(0.021)
Log-Usage for Matched	−0.014	0.026^{***}	−0.0004	0.016
	(0.011)	(0.009)	(0.013)	(0.010)
Actual Adopter Effect	−0.026	0.033	-0.091^{*}	0.089^{**}
	(0.040)	(0.034)	(0.047)	(0.036)
Δ Online Spending	1.919^{***}	−0.003	0.332	0.470
	(0.715)	(1.757)	(2.131)	(0.399)
Δ Online Spending, Squared	−0.508	−0.354	−0.048	0.071
	(0.519)	(0.528)	(0.414)	(0.352)
Δ Online Tran	2.096^{***}	3.369^{***}	1.056	2.556^{***}
	(0.767)	(1.105)	(1.605)	(0.534)
Δ Online Tran, Squared	-1.630^{***}	-4.053^{**}	-3.416^{**}	−0.777
	(0.474)	(1.590)	(1.596)	(0.559)
Log-Usage for Adopters	0.087^{***}	−0.007	0.074^{***}	0.0001
	(0.019)	(0.015)	(0.022)	(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.

there is little correlation between those with low pre-adoption spending and those with young accounts. The correlation between the logged-pre-adoption-spending and logged-age-of-account is 0.06. This suggests that it may not just be young, low-income consumers who are driving the consumption changes that we see.

Persistent Effect? In our main analysis, we consider the aggregate 270 days after an individual's adoption.

5 Spatial Outcome Results

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 the overall frequency though, we are interested in how the effects of ridesharing manifests geographically. We hypothesize that, for those affected by ridesharing services, the city is, in effect, has become smaller to them. 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.

To consider this, we introduce a few additional outcome variables not on consumption frequency, but instead on geographic aspects of consumption. We are limited by our data in not clearly observing home locations nor the specific locations of individual transactions. Instead, we utilize 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.

As our central metric of interest, we compare each consumer's unique number of merchant zip codes in 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 how many more

	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***	−0.019	−0.032**	−0.028*
	(0.010)	(0.018)	(0.015)	(0.015)
Log-Usage for Matched		0.036***	0.004	0.003
		(0.009)	(0.007)	(0.007)
Actual Adopter Effect			0.013	0.009
			(0.024)	(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***	0.030***
			(0.011)	(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)
Note:				*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.

merchant zip codes they visited. This coefficient is similar across all four columns. The null hypothesis here is that consumers, even if they increase their transaction frequency due to ridesharing, still go to merchants in the same set of neighborhoods that they did pre-adoption, just more often. We find evidence against this null hypothesis.

These results are tangential to our findings on increasing consumption frequency. If a consumer increases their consumption frequency, or propensity to consume, it does not necessarily mean that they will visit a new set of merchants. It could be that they only visit the same set of merchants and zip codes, but more often. Instead, in this result and section, we find evidence that consumers who are affected by ridesharing do visit a wider range of zip codes. It also offers additional robustness on our findings of consumption frequency and our mechanism on spatial frictions: if an unobserved confounder, such as higher incomes, is causing a consumer to increase both their consumption and use of ridesharing, it does not necessarily increase the number of zip codes visited.

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 7 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. In the robustness section, we investigate additional aspects and details of consumer travel.

5.1 Zipcode Centric Perspective

We are interested in how ridesharing may 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. This has consequences on the local economy however. In this section, we introduce an exploratory analysis on which zipcodes may be more affected by ridesharing. Unfortunately, in the data, our zip code information is limited to mostly Pittsburgh-metropolitan area zipcodes.

Our main finding is that that we identify some zipcodes that, among those who use ridesharing significantly, are seem to be affected substantially both *positively* and *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 though, these local-geographic-economy impacts may not be generalizable to beyond Pittsburgh; they may be due to specific ways that Pittsburgh populations and consumption amenities are distributed.

The analysis is conducted in a simple, naive manner. We use the data from the post-adoption and pre-adoption 270-day periods. We identify the top-35 zipcodes with the most observations in the post-adoption period. We then run 35 simple diff-in-diff analyses, among

	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	0.014	0.028	0.034
	(0.025)	(0.025)	(0.025)	(0.037)
Log-Usage for Matched	−0.012	−0.002	−0.007	0.007
	(0.011)	(0.011)	(0.011)	(0.018)
Actual Adopter Effect	−0.035	−0.002	−0.010	−0.038
	(0.036)	(0.042)	(0.042)	(0.059)
Δ Online Spending	0.013	−3.161	−3.086	2.305
	(1.557)	(3.923)	(3.768)	(1.923)
Δ Online Spending, Squared	−1.281***	−0.242	−0.257	−0.273
	(0.237)	(0.318)	(0.302)	(0.375)
Δ Online Tran	−0.400	3.311	3.040	−1.893
	(1.222)	(3.177)	(3.056)	(1.417)
Δ Online Tran, Squared	−0.503	1.465	1.402	−2.087
	(1.034)	(2.215)	(2.126)	(1.498)
Log-Usage for Adopters	0.049***	−0.002	0.010	0.006
	(0.017)	(0.021)	(0.020)	(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 7: 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.

adopters only, grouping the adopters into two coarse groups of low-usage and high-usage adopters, based on the median post-adoption usage of 5 trips. 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 yet adjust for multiple testing however. 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 any time trends or group-differences. The coefficient I_3 is the coefficient of interest, which provides an average estimate on the high-usage consumer's average consumption frequency in that zipcode changes in the post-adoption period.

Figure 10 shows the estimated post-adoption consumption frequency coefficients for the top 35 zipcodes in our data. Most of these are in the Pittsburgh metropolitan area. 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. To further demonstrate the geographic aspects of this potential

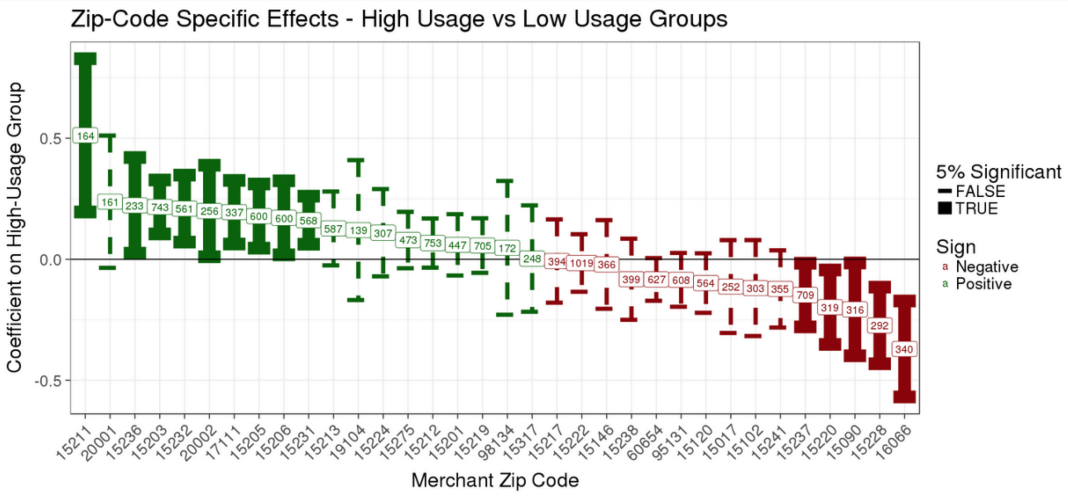


Figure 10: The estimated post-adoption consumption frequency coefficients for the top 35 zipcodes in our data. Most of these are in the Pittsburgh metropolitan area. The fact that some zipcodes suggest significant increases in activity, and some suggest significant decreases in activity suggest that 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.

shift in consumption, in Figure 11 we show a map of the Pittsburgh metropolitan area and its zipcodes among these top 35. In this figure, we are identifying the differences in activity between the high-usage group and the low-usage group. It's notable that most of the green crosses (positive, but not significantly different from zero), and the green dots (positive and significant) are concentrated in the city-center; while the negative estimated coefficients tend to lie in the suburbs surrounding the city. A possible mechanism explanation for 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 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 subsamples.

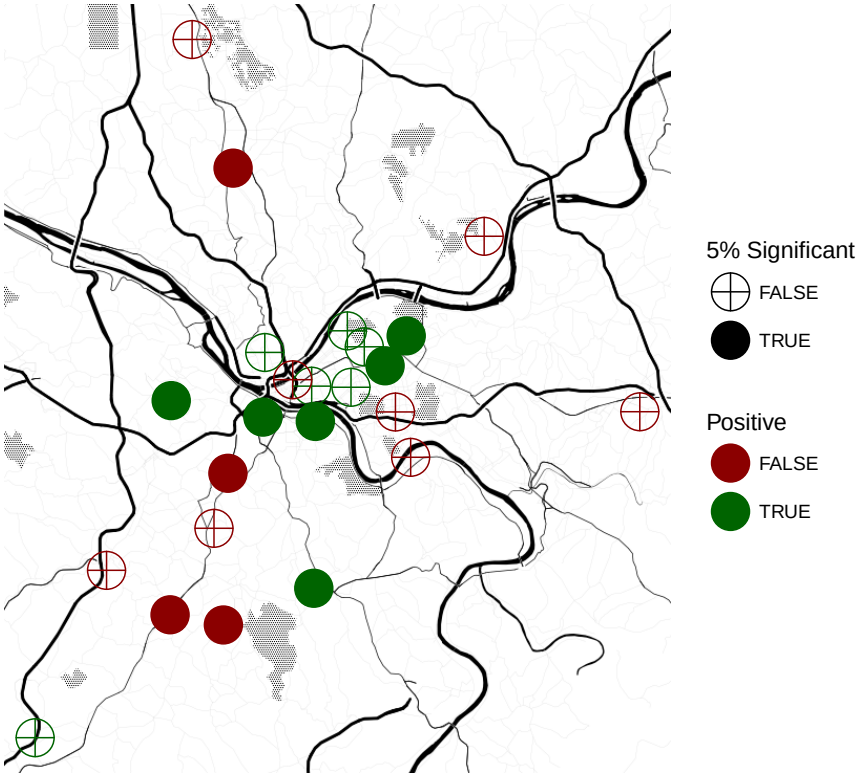


Figure 11: A map of the estimated high-usage group coefficients in the Pittsburgh metropolitan area. Most of the zipcodes with suggested increases in activity for the high-usage group (relative to the low-usage group) lie near the city center, while those with relative decreases in activity lie outside the city.

	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**	0.042	−0.028	−0.322***
	(0.055)	(0.037)	(0.018)	(0.088)
Log-Usage for Matched	0.006	0.010	0.012	0.051
	(0.026)	(0.017)	(0.009)	(0.041)
Actual Adopter Effect	0.050	0.011	0.052*	0.088
	(0.088)	(0.059)	(0.030)	(0.133)
Δ Online Spending	8.172	8.388*	2.218	35.745***
	(6.144)	(4.739)	(3.121)	(8.043)
Δ Online Spending, Squared	−2.471	1.900	−0.375	3.550
	(1.694)	(1.369)	(1.299)	(2.962)
Δ Online Tran	12.180***	3.913	2.431	−22.151***
	(4.250)	(3.268)	(2.315)	(5.200)
Δ Online Tran, Squared	−22.561***	−16.549***	−6.094**	−23.211***
	(5.156)	(4.072)	(2.697)	(7.256)
Log-Usage for Adopters	0.116***	0.069**	0.029**	−0.016
	(0.043)	(0.028)	(0.015)	(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
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 8: 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.

6.1 Robustness in Model Specifications

Logged-Difference Outcome Variables In our main model, we consider a first differences model in 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.

Column (1) of Table 11 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 8 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:

$$\ln Y_{it} = \delta_t + \alpha_i + \tau_a Adopt_{it} + \beta \ln(Usage_{it} + 1) \cdot Adopt_{it} + \epsilon_{it}$$

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, $(\ln Usage_{it} + 1)$. 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 9 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

	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	0.016	0.006	−0.015	−0.013*	−0.016**
	(0.013)	(0.011)	(0.010)	(0.010)	(0.008)	(0.007)
Log-usage (before adoption)						
Log-usage (after adoption)	0.140***	0.109***	0.062***	0.153***	0.120***	0.085***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(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 9: 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)).

the stochasticity of consumption, however, the differences are minimal.

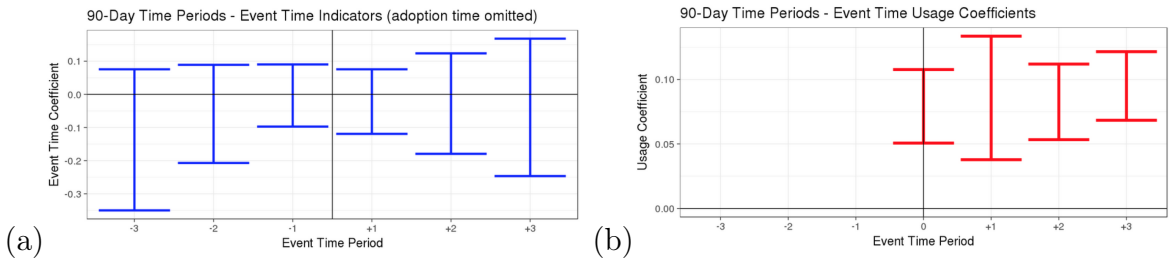
Event Time Model In our main model, we only consider two periods, which are defined around the adoption event, and have the same corresponding calendar days. To get a sense of the potential dynamics of the effect of ridesharing, we use the calendar time model, with event-time indicators. We find the 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 12, on the left, we show that there is no statistically significant event-time pattern. This is an important robustness consideration, reflecting similar null results to the yearback falsification test. The null coefficients for the post-adoption periods also suggest robustness to



(a) (b)

Figure 12: 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.

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 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 10 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;

	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	−0.014	−0.028*	−0.049***
	(0.016)	(0.013)	(0.016)	(0.014)
Log-Usage for Matched	0.005	0.004	0.004	0.005
	(0.008)	(0.006)	(0.007)	(0.007)
Actual Adopter Effect	−0.004	−0.006	0.006	0.012
	(0.027)	(0.021)	(0.024)	(0.021)
Δ Online Spending	2.751	3.531**	2.076	3.570***
	(2.207)	(1.792)	(1.503)	(1.178)
Δ Online Spending, Squared	0.251	1.203**	0.410	0.015
	(0.700)	(0.473)	(0.548)	(0.375)
Δ Online Tran	2.261	0.656	1.651*	−2.214***
	(1.461)	(1.169)	(0.948)	(0.691)
Δ Online Tran, Squared	−6.571***	−5.894***	−4.876***	−2.168*
	(1.963)	(1.550)	(1.395)	(1.114)
Log-Usage for Adopters	0.044***	0.029***	0.033***	−0.0002
	(0.014)	(0.011)	(0.012)	(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 10: 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.

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.

6.2.2 Robustness in Non-Movers

Among the 5471 consumers in the main sample, we know the city of the account's mailing address for each month (~84%). 4500 consumers did not change cities in the time period that we consider.

6.2.3 Robustness in Early Adopters

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 weather[8] 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 11 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 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

[8]High 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	−0.010	−0.065	−0.055
	(0.036)	(0.041)	(0.054)	(0.037)
Log-Usage for Matched	0.0001	−0.001	−0.004	−0.003
	(0.015)	(0.018)	(0.024)	(0.017)
Actual Adopter Effect	−0.048	−0.108	−0.191**	−0.083
	(0.057)	(0.066)	(0.093)	(0.062)
Δ Online Spending	0.437	4.798	11.118***	6.320
	(3.137)	(2.971)	(3.852)	(4.595)
Δ Online Spending, Squared	−0.199	1.096***	0.756	−0.340
	(0.324)	(0.318)	(0.491)	(0.396)
Δ Online Tran	−0.025	−1.294	−6.526**	−5.232
	(2.640)	(2.575)	(3.145)	(3.791)
Δ Online Tran, Squared	−2.790	−6.937***	−9.495***	−2.557
	(1.801)	(1.589)	(2.422)	(2.648)
Log-Usage for Adopters	0.056**	0.076**	0.099**	0.022
	(0.027)	(0.032)	(0.043)	(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 11: 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.

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 Local Economic Implications

In our analysis from a consumer perspective, we identify that some subset of consumers—those who choose to use Uber and Lyft because it lowers their transportation costs—are affected by ridesharing and this subsequently positively impacts their local economic activity.

A subsequent interesting question is, in which ways are economy activity affected? In looking at consumer patterns, we do see that some types of merchants are disproportionately affected by ridesharing; e.g., consumers who are affected show significant increases in their restaurant and bar activity, but not in durable goods or gasoline.

Geographically, we also want to be able to characterize the impact of ridesharing. To do so, we adjust our analysis to consider a zip-code level perspective.

In the year before adoption (among the adopters in the 270-day main sample, with a minimum of 6 months of activity), we observe 529K transactions. Of those, 223K have merchant zip code information (42%). The range of total transactions among the top-100 zip codes is [297, 13144] and [188, 13144] among the top-150 zip codes.

8 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 consumption 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 credit card and debit card transactions from January 2012 through May 2016 from a major 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, Ziru, 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.

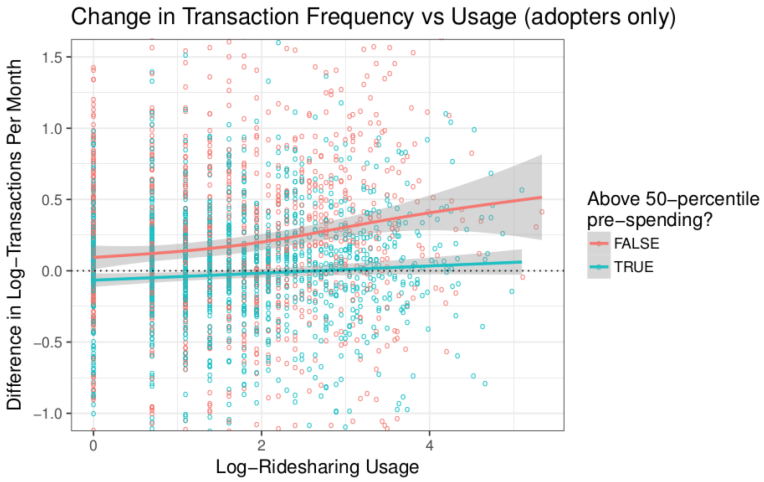


Figure 13: The scatterplot compares the change in transaction frequency in the post-adoption period (y-axis) versus the ridesharing usage in the post-adoption period (x-axis), among adopters only. The adopters are split into two groups based on their pre-adoption spending, above or below the 50th-percentile. 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 4, remember that there is no relationship between usage and spending in pre-adoption.

9 Appendix

9.1 Additional Figures

1. Figure 13 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.

9.2 Data Appendix

9.2.1 Uber and Lyft entry dates

- 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

- 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/

9.2.2 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"}

9.2.3 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.

9.2.4 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)}.