



Estimating US Retail Spending by Income Quartile

Project Lead: Aleksandr Popov

Analysts: Thomas Colagiuri, Aidan O'Neil, Iñigo Sanchez-Asiain, Diego Williams*

Spring 2024

*All of the listed authors study at the University of Chicago. Special thanks to Rebecca Hutchinson at the U.S. Census Bureau for support and guidance throughout our research. We would also like to thank Shivant Krishnan at the University of Chicago for help in this project.

Contents

1	Team and Project Overview	3
1.1	Charles River Economics Labs	3
1.2	Partner Organization	3
1.3	Project Overview	3
2	Introduction and Motivation	4
2.1	Motivation	4
2.2	Summary of Approach	4
2.3	Retail Spending	5
3	Literature Review	5
3.1	Monthly Retail State Sales	5
3.2	Opportunity Insights	6
3.3	JPMorgan Chase, BofA	8
4	Data & Methodology	9
4.1	Data - Affinity	9
4.2	Data - Diary of Consumer Payment Choice	10
4.3	Accounting for Changes in Coverage	11
4.4	Bias Correction and the Raking Algorithm	11
4.5	Bootstrapping	13
4.6	Testing Assumptions in our Method and Existing Literature	14
5	Results	15
5.1	Initial Consideration of Affinity Data	15
5.2	Changes in Retail Spending by Income Level	15
5.3	Consumer Habits and the Impact of Credit Cards	20
6	Conclusion and Next Steps	23
6.1	Conclusion	23
6.2	Next Steps	24
7	Appendix	27
7.1	Raking Algorithm Visualization	27
7.2	Figures: Bootstrapped Standard Errors for Cumulative Changes in Spending	27
7.3	Figures: Cardholders vs. Non-Cardholders	29
7.4	Data Sources	35

1 Team and Project Overview

1.1 Charles River Economics Labs

Charles River Economics Labs is a group of undergraduate researchers at the University of Chicago, Harvard, and Stanford that completes pro-bono economics research for governments and other social enterprises. The organization is advised by graduate researchers and faculty members at Harvard and the University of Chicago. Its past and current research partners include the United Nations Development Programme, the Department of Transportation, World Bank, the US Department of State and the Census Bureau, among others.

1.2 Partner Organization

The US Census Bureau’s Associate Director for Economic Programs (ADEP) team operates under the leadership of Nick Orsini, Associate Director. The directorate conducts more than 60 monthly, quarterly, and annual surveys covering various sectors of the economy, including 12 economic indicators and an economic census and census of governments in years ending in two and seven. Much of their current work focuses on using statistical methods to create new, synthetic data from existing data sources, departing from the money- and time-consuming nature of surveys. This research project is a first step towards creating a data product that would estimate spending in the retail sector by income quartile, using only existing data sources.

1.3 Project Overview

Our project demonstrates the ability to use existing, incomplete credit card data in order to create unbiased estimates for changes in retail spending by income levels. We also test the intrinsic differences between card holders and non-card holders, which we cannot account for with demographic-based bias correction. Our results demonstrate a need to incorporate either novel statistical methods or data sources containing cash spending in order to get a fully unbiased estimate. Finally, this project lists further steps to creating a full data product for the Census , including some analysis of other data sources that can be incorporated and areas for improved statistical methods and assumptions.

2 Introduction and Motivation

2.1 Motivation

With credit card debts having hit historic highs, and savings rates reaching all-time lows, the wake of the COVID-19 pandemic has caused consumer spending habits to undergo drastic changes. Many question whether these changes in preferences will prove to be permanent.

The pandemic left its mark on different demographics in different ways. While some managed to maintain similar spending habits before and after the pandemic, the chaos of the pandemic provoked immense change in others. It is of particular importance to gauge the differences in spending habits of individuals belonging to different income levels. What income levels experienced the most change in monthly spending and to what extent did spending change? This project aims to construct a method to answer this question and track the changes in retail spending across different income levels. More specifically, our objective is to use existing demographic, retail, and other data sources to determine the changes in spending patterns among different income groups. We additionally look to explore novel statistical methods and incorporate them into already existing metrics to bolster our estimates.

2.2 Summary of Approach

Our approach draws heavily from the methods used by Opportunity Insights in their “Track the Recovery” economic tracker. Using credit card transaction data provided by Affinity Solutions, the tracker measures changes in weekly spending across several sectors (not just retail) for consumers of different income levels. Affinity does not explicitly give information on the income of card holders. Instead, it assigns an income level to a transaction based on the median income of the zip-code in which the transaction took place. In simplest terms, Opportunity Insights looks at the amount spent in a zip-code as a proxy for the amount spent by the median income of that zip-code. This is tracked across all the zip-codes the data provides, and changes are observed from there. This strategy is the backbone of our project, and we have found and implemented several improvements. This paper discusses both the improvements that we have implemented, and that which we have not quite implemented, but have still considered extensively. Of note, we implement a “raking algorithm” to provide weights to the transactions Affinity tracks to ensure fair representation.

These weights will account for other demographic factors that can impact spending within each quartile. Finally, we gauge the impacts of only accounting for the transactions of card holders on our figures. How representative are the spending habits of card holders of the spending patterns of non-card holders? Is there a way to account for non-card transactions?

2.3 Retail Spending

By retail spending we are referring to the types of transactions whose North American Industry Classification System (NAICS) code is 44 or 45. Labeled as “retail trade,” transactions under this category can be divided into store and non-store retailers. Store retailers are stores that are operated at fixed point-of-sale locations, which are intended to draw in several walk-in customers. They generally sell goods to the public for personal or household use, though some also cater to businesses and institutions. Beyond selling products, certain store retailers offer additional services like repairs and installations. On the other hand, non-store retailers encompass various methods such as kiosks, carts, vending machines, direct sales, telemarketing, direct marketing, and online retailing (1). Importantly, these transactions are included in the Census’ Monthly Retail Trade Survey (MRTS). This survey is sent to about 13,000 different retailers and produces detailed figures on retail purchases in different industries across all states. Using the MRTS as well as other third party data sources, the Census can provide estimates of total retail spending across different industries and states.

3 Literature Review

3.1 Monthly Retail State Sales

The Census’ methods for providing these estimates use both a “top down” and “bottom up” approach of analysis upon the MRTS and third party data to create an estimate for Monthly State Retail Sales (MSRS) data [2]. These methods were greatly influential in our initial approach to determining changes in retail spending, and ultimately steered towards our final methodology. The top down approach combines MRTS data with payroll data to produce an estimate for retail sales within a state, in a specific tabulation industry (3-digit NAICS code) and period. This is accomplished by summing over the MRTS data for retail sales of each estimation industry (3-,

4-, or 5-digit NAICS code) within the specified tabulation industry and weighing them by the proportion of the gross annual payroll of the specified state against the gross annual payroll of the nation in the estimation industry indexed. Effectively, the payroll weight allows a proportion of the MRTS data for sales in an industry to be attributed to a specific state. Succinctly, this is given by the following sum:

Spending in state g , and in tabulation industry i , Y_{ig} is

$$Y_{ig} = \sum_{j \in i} \left(\frac{X_{jg}}{X_j} \right) Y_j$$

X_{jg} is the payroll in state g and estimation industry j

X_j is the payroll of the nation in estimation industry j

Y_j is the total national spending in estimation industry j , from the MRTS data.

The bottom up approach takes a different route, instead looking to directly sum the MRTS and third party data for a specific tabulation industry, state, and period, and then applies additional terms to adjust for potential error from the raw data. This draws from Circana data, which tracks most of the major retail providers in the U.S.

As will be discussed later, the approach we take is somewhat similar to the Census' bottom up approach. Much like the census approximates the amount spent using the transaction data in their Circana sample, we use the transactions we track in zip codes of different income levels in order to create our approximations. Even though our approximations and methodology ended up being of a very different, the MSRS, was very informative early on in the project. It was important to study ways in which one could estimate how spending could be estimated across different groups. With some further data sources, it might be possible to combine a top-down approach with the bottom-up approach in a similar way to improve the estimates.

3.2 Opportunity Insights

We draw the most inspiration from the Opportunity Insights Economic Tracker, "Track the Recovery", run by a team of researchers based at Harvard University and Brown University ([3]). The tracker measures changes in weekly spending across sectors (retail, transportation, hospitality, healthcare, grocery, entertainment and recreation). It also has a separate metric for low, middle,

and high-income consumers, as well as each U.S. state since January 2020, at the outset of the COVID-19 pandemic. As described in its accompanying paper "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data" (4), the input dataset is provided by Affinity Solutions, which accounts for about 10% of credit card transactions and is our primary dataset as well. The data is made up of observations at the intersection of industry, county, date, and income quartile, reporting spending and card count for each cell. Income quartile for any given transaction is determined by the residential zip code of where the transaction took place.

Since raw changes in spending might reflect changes in Affinity's data suppliers or changes in the data obtained by those suppliers rather than actual changes in consumer spending, Opportunity Insights works to identify and control for coverage changes. In particular, they investigate sudden discrete changes in spending from particular providers that might indicate changes in coverage, then, in accordance with their findings, adjust the data by imposing continuity. In addition, the research recognizes that much of the decrease in spending during the early days of COVID-19 is due to a decrease in card usage, where spending per card did not see a decline. Therefore, indexing at card level is insufficient to account for this. Therefore, the method in Figure 1 is used to adjust for coverage for the entire dataset:

$$\text{adjusted_spending}_{l,i,t} = \begin{cases} \frac{\text{spending}_{l,i,t}}{\text{cards}_{l,i,t}} \times \overline{\text{cards}_{l,i,t' \in \text{Jan } 2020}} & \text{if } t \leq \text{January } 2020 \\ \frac{\text{spending}_{l,i,t}}{\text{cards}_{l,i,t}} \times \overline{\text{cards}_{l,i,t' \in \text{Jan } 2020}} \times \frac{\overline{\text{cards}_{l,i,t' \geq t | t' \in \text{Feb } 2020}}}{\overline{\text{cards}_{l,i,t' \in \text{Feb } 2020}}} & \text{if } t \in \text{February } 2020 \\ \frac{\text{spending}_{l,i,t}}{\text{cards}_{l,i,t}} \times \overline{\text{cards}_{l,i,t' \in \text{Jan } 2020}} \times \frac{\overline{\text{cards}_{l,i,t}}}{\overline{\text{cards}_{l,i,t' \in \text{Feb } 2020}}} & \text{if } t \in [\text{March } 2020, \text{July } 2020] \\ \frac{\text{spending}_{l,i,t}}{\text{cards}_{l,i,t}} \times \overline{\text{cards}_{l,i,t' \in \text{Jan } 2020}} \times \frac{\overline{\text{cards}_{l,i,t' \leq t | t' \in \text{Aug } 2020}}}{\overline{\text{cards}_{l,i,t' \in \text{Feb } 2020}}} & \text{if } t \in \text{August } 2020 \\ \frac{\text{spending}_{l,i,t}}{\text{cards}_{l,i,t}} \times \overline{\text{cards}_{l,i,t' \in \text{Jan } 2020}} \times \frac{\overline{\text{cards}_{l,i,t' \in \text{Aug } 2020}}}{\overline{\text{cards}_{l,i,t' \in \text{Feb } 2020}}} & \text{if } t \geq \text{September } 2020 \end{cases}$$

Figure 1: Method to Adjust for Changes in Credit Cards

We aimed to mimic this method. However, given our constraints, we choose to account for changes in coverage by normalizing over card count, indexed by the original number of cards tracked in January 2019 (see the Data & Methodology section).

Also note that, to account for seasonal changes in spending, Opportunity Insights controls for seasonal differences in consumer behavior by indexing spending for any given week relative to the same week in 2019. By its construction, the Affinity dataset, from which Opportunity Insights findings are derived, only contains data on holders of credit cards. This introduces bias due to the possibility that cardholders exhibit different spending habits than their non-cardholding counterparts within their income bracket. To explore the impact of this, Opportunity Insights gathered information from CoinOut, a service that tracks grocery receipts. It can therefore give us information on cash spending. They claim that, in practice, "card and cash spending track each other closely". This provides evidence that card spending can be used as a proxy for overall card spending. Nevertheless, when looking at spending at the level of income, we might see slightly different trends and Affinity data might not exactly reflect the true change in spending. Later, we compare spending patterns between consumers with credit cards and those without. We combine this with our analysis of data derived from cardholder purchases to gain a more accurate understanding of true changes in retail spending over time.

3.3 JPMorgan Chase, BofA

Banking institutions have studied the behavior of their own cardholders to examine trends in spending over time, particularly in the wake of the pandemic. JPMorgan Chase publishes a "Household Pulse" product [5], which tracks inflation-adjusted changes in monthly cash balances, both by dollar amount and by percent change, based on data from 8.6 million of its customers since January 2020, reporting averages by race and income. Similarly, Bank of America produces a "Consumer Checkpoint" product [6], which uses the bank's proprietary data to track changes in real spending and income by income group and generation over time. Definitionally, these sources only track credit card spending and only do so from one provider, which introduces bias that composite sources seek to eliminate. At the same time, these institutions do not suffer from problems that arise from changes in data providers, since they collect their own information, but observed changes in spending by a bank might reflect fluctuations in credit card preferences among consumers rather

than true changes in population retail spending.

A separate report conducted by JP Morgan on consumer spending in the early weeks of the COVID-19 Pandemic gave additional insights into tracking the spending of card holders [7]. Along with some other goals, the report looked to determine how much individual spending had fallen, and whether this drop varied across households. This was of interest to us as it aligned closely with our own objective. To determine these changes, JP Morgan first broke down the spending drop into non-essential and essential spending. The decision to do this was motivated by the closure of many non-essential businesses during the pandemic. They also studied spending that had occurred before the pandemic. Lastly, they tested for the spending of individuals employed in different sectors to determine the impact that pandemic-induced job losses had on spending. For our purposes, it was important to note this way in which JP Morgan cut up their data into different time periods, industry, and employment so as to tailor the data to the question that they were trying to answer. Later, the findings served as a means of comparison, allowing us to gauge our own results against those of a reliable source with a similar objective.

4 Data & Methodology

4.1 Data - Affinity

Affinity Solutions provides the card spending data that we use to determine our figures. Affinity is the data source that we spent the most time with, and is the only data set we ended up using in our final product to construct our estimates. While we explored other data sets, due to either challenges acquiring them or challenges working with them, we settled on only using Affinity. We will explore some of these other data sets later in this paper. Affinity accounts for about 10% of credit card transactions. The data is from tracked credit cards, but is aggregated at the county level. For each transaction that Affinity reports, it provides the NAICS code of the industry in which the transaction occurred and the income quartile of the zip-code where it occurred. The quartile is internally determined by Affinity solutions and the data does not actually report the zip code. The fact that Affinity provides the industry code is of particular importance as well as it allows us to filter only for retail spending, which has NAICS code beginning with 44 or 45. Finally, Affinity gives the number of cards that it tracks from each county within each industry. This fact

is crucial in accounting for possible changes in coverage of the Affinity data: It is the means by which Opportunity insights accounted for such changes, and the means by which we will as well.

It is worth noting that we did not directly have access to the Affinity data due to privacy concerns. Instead, we created a synthetic dataset, with 1 million rows, to mimic the format of the actual dataset in order to test our code. We would then send our code to the Census to run on the raw data on their internal servers. We would then receive the aggregated results on which we base our analysis.

4.2 Data - Diary of Consumer Payment Choice

The Diary of Consumer Payment Choice (DCPC) is a survey conducted annually during the month of October by the federal reserve banks of Atlanta, Boston, and San Francisco, [8]. Participants are recruited from a panel maintained by the University of Southern California’s Understanding America Study (UAS), and are asked to complete a short survey and then keep a three-day diary recording their payment activities. They receive financial incentives for completing the study. The results of this study offer us a snapshot into consumer payments in the United States, since the DCPC records the type of payment instruments and transaction type, the value and frequency of transactions, and their preferences and attitudes towards the different payments. This means that through this data source we have access to data on just below 5,000 people and 18,000 transactions in October of each year.

This data is very useful since the provision of a survey and a diary gives us the ability to merge on id, thus providing a deeper insight into consumer behavior. Furthermore, since we know the store location of all transactions, it allows us to filter by retail spending. This data source was therefore critical in helping us approach our concerns about cash spending, which is a significant portion of retail spending. We also used it to analyze our methodology with indexing at card count as a proxy for indexing on an individual level. We first used the DCPC data to check whether or not changes in number of credit cards per person can be observed. Then, we ran regressions to observe whether owning a credit card affected total retail spending. These, checks were vital in order to make our assumptions hold when using the Affinity dataset. A slight concern with the DCPC is its low sample size, and its significant attrition over the days of the survey.

4.3 Accounting for Changes in Coverage

A key feature of almost all existing card data, and in particular our final choice of Affinity, is that it is incomplete, and coverage is always changing as card providers are added or removed. However, we want to track real changes in consumer habits, and as such to remove changes in coverage specific to the data providers. These changes do not represent any broader trends among U.S. consumers, and we need a way to adjust for coverage. In this project, we index all spending at the number of cards tracked in each income quartile as a proxy for the number of consumers tracked. This essentially means that we track spending per card in each income bracket,. We test the robustness of using this methodology in Section 4.5. In addition, other literature using similar methodologies has a different way of adjusting for coverage in the early days of COVID-19. It appears that per-card spending did not significantly change during this time- instead, people used fewer cards and likely had fewer transactions. The impact of this fact will be discussed in our results.

4.4 Bias Correction and the Raking Algorithm

Our main way of improving the raw data is through bias correction. We derive our methodology from the way that the Atlanta Fed diary accounts for potential demographic biases in the data- through use of the Raking algorithm.

We know that our data source has several angles of bias- for example several geographical regions are over-represented. Therefore looking at simple changes in spending, even if adjusted for coverage through indexing, leaves us exposed to the extremely strong assumption that the only factor determining consumer spending in retail is income. While no literature exists to test this fact in American retail spending, previous research suggests that demographics have an impact on financial decisions. [9] Therefore, in order to avoid making such assumptions about consumer behavior, we aim to make our sample representative of the entire population.

We do this through a Raking algorithm, which is in essence a weighting algorithm. We "rake" over 3 different variables- age, ethnicity and geography (we use the state as the variable). The most granular information we have about the location of the transaction is the county but other data sources are more granular- giving the zip code or even the exact geolocation of the transaction.

Therefore, we make a similar assumption to that which we use to estimate the income of the consumer- that in each county we track a demographically representative consumer for that county. In order to arrive at the target proportions in each group, we use the Census County Population by Characteristics: 2020-2022 data. [10] This dataset provides a population count in each county, age group and ethnicity. We aggregate these across the counties in different states in order to arrive at the target values in each group. In the process, we match the county FIPS to the state and name of the county using the Harvard Dataverse County FIPS Matching Tool. [11] In this way, we arrive at the "target" proportions for each group for our Raking algorithm.

In preparation of the credit card data, we aggregated the card count across all retail sectors in each county. We split the cards in a representative fashion across the various demographics of each county. For example, if 50% of the population in a county is of a certain demographic group, we assume that 50% of the cards we track are used by people from that demographic group. We then aggregated these cards at the state level. We arrive at the "actual" demographics of the cards tracked in our dataset. We want to weight these to make them representative of the population of the United States.

All of these variables are interconnected- for example, the age profile of a region is not disconnected from its location in the United States. Therefore, we cannot use a simple weighting method for each of our variables of interest. In fact, simply multiplying the weights in each of these categories yields a weighted spend around 8x larger than unweighted spend in our sample. It is worth mentioning that, as we are keeping the weighted card counts constant in our sample instead of the spend, we also leave ourselves exposed to potential fluctuations of the weighted total spend. We analyze whether we get this side effect, by taking a ratio of weighted and unweighted spend in our full data.

Instead of simple weighting, we use the weightipy library in Python. [12] This library provides an iterative implementation of the raking algorithm (RIM). A visualization of the way that this algorithm works can be seen in Figure 16. In each iteration, we weight our results on some axis (one of variables) in order to fit to our target results for that axis. However, in doing so, we have changed the other variables in a way that is not representative of the population, due to the interplay between the variables. For example, cannot weight the age variable without also changing the proportions of different ethnicities. We therefore use the results from the first weighting iteration, where the

spending is weighted to be representative on a certain axis, to weight on a different axis, making that representative. Again, this makes the first axis unrepresentative, so we have to weight it again, given the new weighted spends. We repeat this process, switching between the 3 variables, until we converge to a set of weights for each group that would make our sample representative of the state, age and ethnic composition of the US. We ensured that this was followed and that we converged to a solution that would make our card counts representative. While not a closed-form solution, it achieves the expected representation in each of our target ethnicities, regions and age groups.

This process gives us a weight for each ethnicity and age group combination in each state. We go back to the Census County data, and assign a weight to each county based on the demographics of that county, and what proportion they have in a certain group. In essence, we use the following formula:

$$\text{weight}_{\text{county}} = \sum_i p_i \cdot w_i$$

Where the sum is taken over all demographic groups i , p_i is the proportion of the population of the county in group i , and w_i is the weight attributed to group i by our Raking algorithm. In calculating our final changes in spending, we multiplied the total retail spending in each income bracket in each county by the county's weight before summing them together across each income quartile. As a result of this weighting, the weights for counties range between around 0.1 and 5- a reasonable range in our framework.

As mentioned above, we also tested this methodology to ensure that this raking does not skew the overall spending.

4.5 Bootstrapping

Given the complex methodology above, with many weightings and adjustments, it is extremely difficult to estimate the intrinsic error in our methodology. Instead, we test the robustness of our methodology using standard errors derived from bootstrapping. In our implementation, we take the entire size of the sample in terms of the number of observations in our dataset. We randomly sample our observations, such that some are counted several times, while others are left out entirely from our sample. We then calculate the indexed spending per card, before and after weighting using the raking algorithm. We conduct this $n=1000$ times, and take the standard deviation of our

calculations as the error. In essence, this methodology allows us to test how robust our methodology is to the particular data captured in our sample. If our estimates vary wildly depending on the particular data captured by our provider, it cannot be seen as a reliable estimate.

It is worth noting that, due to the size of our dataset and the computing power demands of our methodology, we did not recalculate the weights in each iteration. Instead, we used the weights calculated on the entire dataset. This likely led to an overestimation of our standard error as we do not keep the weighted card count constant, creating further variance in our estimates according to the weights assigned to the particular sample. If we were to re-rake in every iteration of the bootstrapping, we would keep card counts representative and remove one source of variance in our bootstrapping.

4.6 Testing Assumptions in our Method and Existing Literature

In the above analysis we make a key assumption in the aggregation procedure, indexing on credit card count. This amounts to essentially calculating the total spent and accounting for changes in total card count. This means that if a new card is added to the dataset it is assumed that this card belongs to a new individual and not an additional tracker of the same individuals. If for example the affinity data started tracking twice as many cards but each of these were held by the same set of people before the change in number of cards our estimate would be very inaccurate. As a result we must confirm that number of cards held by individuals is held roughly constant over time. If linear trends in time can be detected then comparing affinity amounts year over year would need to be adjusted for this reality. We turn to two different sources to test this question. The results can be observed in [28](#) [29](#) and [30](#). In these the graphs show small variations over the years in the sample but when difference of means fails to reject the null of sharing a common level over the years. The first two graphs use data from the Consumer Diary of Choice from The third image shows a comparison to another source, Statista which also does not find a common linear trend with cards held. These results strengthen the use of indexing on card count in our data.

5 Results

5.1 Initial Consideration of Affinity Data

We first looked at the overall composition of the Affinity data. In the first graph below, we fit the density of transaction tracked by Affinity by the median income level of the county in which they took place to a lognormal distribution. This had a log mean of 11.18 and a standard deviation of 0.22. On the right, we have the population distribution of the United States as a whole based on the median income of the county of residence, taken from the American Community Survey 2022 data for each county. This log-normal fit has a log mean of 11.17 and standard deviation of 0.27.

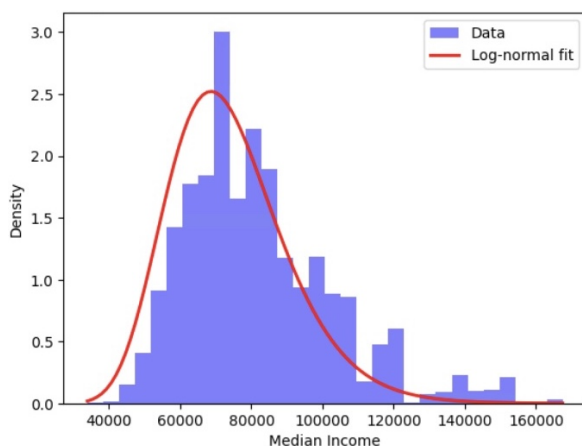


Figure 2: Density of Transactions Tracked by Affinity by Median Income Level of the County

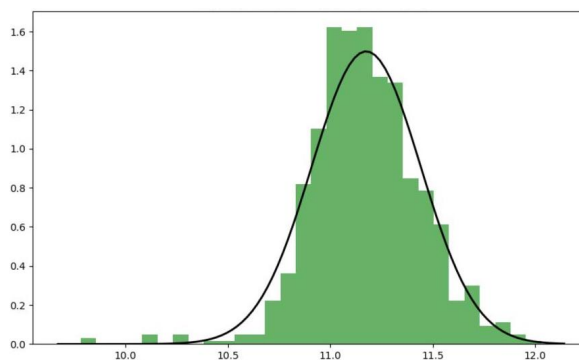


Figure 3: Density of U.S. Population Based on the Median Income of County of Residence

From these charts, we can see that Affinity’s coverage is on average very close to being representative of the incomes of the U.S. population, albeit slightly higher. However, there are fewer deviations from the mean, as indicated by the far lower standard deviation. This may mean that we have fewer readings on the higher and lower income groups- something that we see affecting the standard error of our estimates.

5.2 Changes in Retail Spending by Income Level

In this section, we discuss our examination of changes in retail spending by income level, considering our results from both regular indexing by card count and our implementation of the raking algorithm in order to see how much bias impacts our estimates. For each income quartile and each methodology, we looked at three metrics between the beginning of January 2019 and the

end of March 2024:

1. Week-over-week changes in retail spending, i.e., percent change from the previous week.
2. Year-over-year changes in retail spending, i.e., for each week after the first year of our coverage, the percent difference between that week and the same week of the previous year.
3. Cumulative changes in weekly spending, i.e., the percent difference between each week and the week for which our coverage begins.

First, we notice that week-over-week spending is highly volatile, likely due to seasonal changes in consumer behavior. Most changes are under 5 percent, but certain weeks see changes over 10 percent, with, for example, spending decreasing by almost 15 percent for all income quartiles between December 24 and December 31, 2023 (this post-Christmas holiday drop being a perfect example of predictable seasonal changes in consumer habits). For week-over-week spending, it is difficult to see differences between income groups, and results gleaned from the raking algorithm are similar to those obtained from indexing by card count. The corresponding figures are below, where Quartile 1 is the bottom quartile of earners, and Quartile 4 is the top quartile:

WoW Changes for Quartile 1

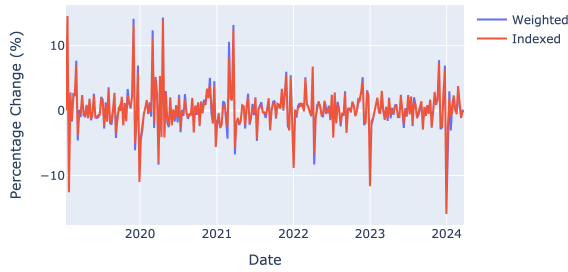


Figure 4: Week-Over-Week Changes in Spending, Q1

WoW Changes for Quartile 2



Figure 5: Week-Over-Week Changes in Spending, Q2

WoW Changes for Quartile 3

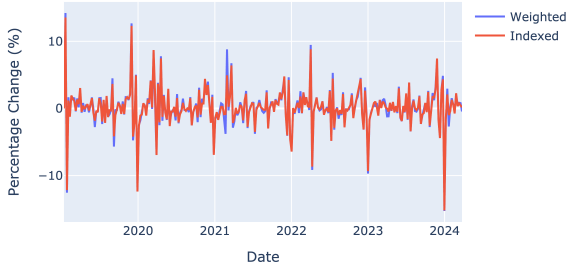


Figure 6: Week-Over-Week Changes in Spending, Q3

WoW Changes for Quartile 4



Figure 7: Week-Over-Week Changes in Spending, Q4

Second, we look at year-over-year changes in spending. For all income levels, year-over-year changes are negative in early 2020, positive and decreasing for the rest of 2020 and early 2021, dropping close to zero in 2021 and then increasing until early 2022, then decreasing once again and ultimately dropping below zero in early 2023, and subsequently recovering through 2023 and into early 2024. Year-over-year spending volatility appears to be somewhat greater for lower income quartiles, which aligns with our expectation that shocks would impact the behavior of low-income individuals more than that of high-income people. The corresponding year-over-year figures are below:

YoY Changes for Quartile 1

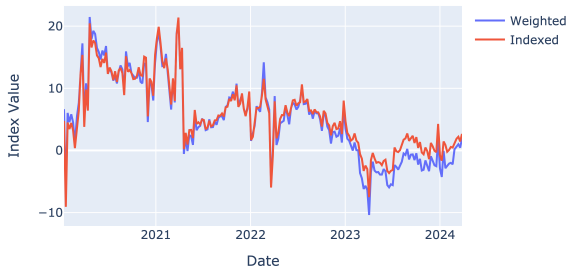


Figure 8: Year-Over-Year Changes in Spending, Q1

YoY Changes for Quartile 2

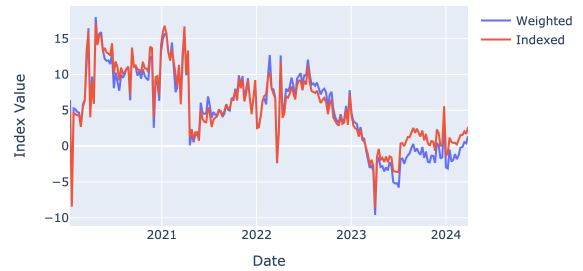


Figure 9: Year-Over-Year Changes in Spending, Q2

YoY Changes for Quartile 3

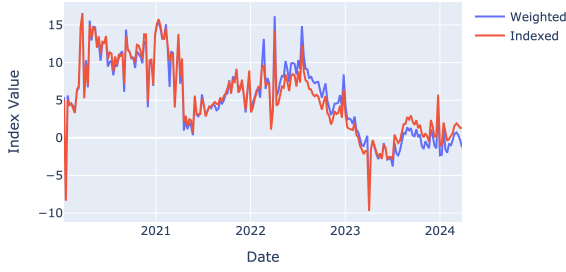


Figure 10: Year-Over-Year Changes in Spending, Q3

YoY Changes for Quartile 4

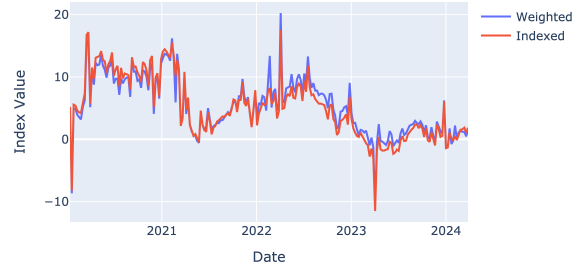


Figure 11: Year-Over-Year Changes in Spending, Q4

Third, we consider cumulative changes in spending. For all income levels, cumulative spending follows a general increasing trend along regular large dips and peaks, which we believe are in large part due to the aforementioned regular seasonal changes in consumer activity. When we look at total changes in spending, we get differing results for each methodology. Using the regular card-count indexing methodology, we observe a 37.75% total positive change in spending for the lowest income quartile, a 35.85% change for the second quartile, a 36.66% change for the third quartile, and a 31.69% difference for the top quartile - in other words, we observe that weekly spending has increased more since January 2019 for lower income quartiles. The total increases are smaller and closer to one another when we implement the raking algorithm: a 34.96% change for the lowest quartile, a 34.76% change for the second quartile, a 35.42% change for the third quartile, and a 33.97% change for the top quartile. It is worth noting that, relative to regular indexing, the weighting generally estimates greater changes in spending for the third and fourth income quartiles after 2022. However, the changes in spending before then are smaller, while indexing gives us smaller changes in spending for the first and second quartiles beginning in 2023 and greater changes before that point. The corresponding figures depicting cumulative changes in spending are as follows. After accounting for bias, we see that overall the income quartiles have had similar cumulative changes. However, we witness larger volatility in the lower income quartiles. This phenomenon was touched on by Opportunity Insights, where stimulus packages in the early days of the COVID-19 pandemic meant that recover was much faster in the lower income groups. We see that Quartile 1 has YoY increases of almost 20% throughout 2020- levels not seen in any of the other income brackets:

Cumulative Changes for Quartile 1

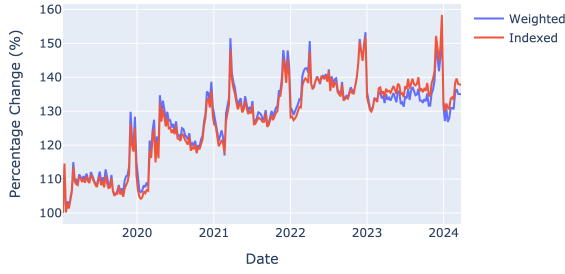


Figure 12: Cumulative Changes in Spending, Q1

Cumulative Changes for Quartile 2

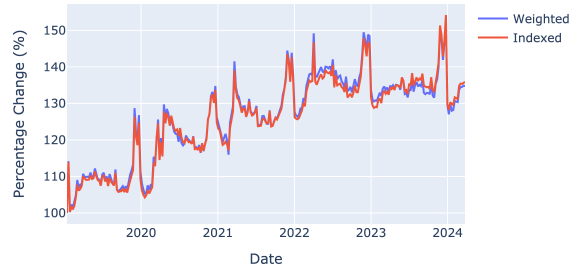


Figure 13: Cumulative Changes in Spending, Q2

Cumulative Changes for Quartile 3

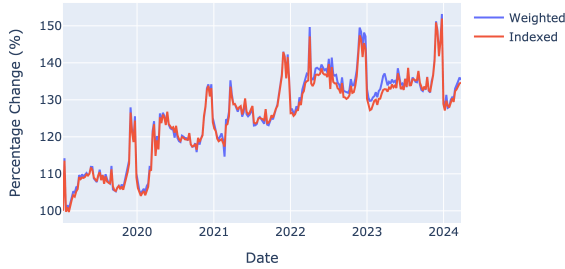


Figure 14: Cumulative Changes in Spending, Q3

Cumulative Changes for Quartile 4

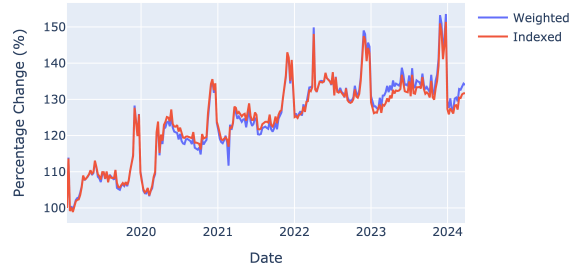


Figure 15: Cumulative Changes in Spending, Q4

It is important to note that we observe that we do not observe the expected large decrease in retail spending associated with the outset of COVID—19 March 2020. This is because much of the decrease in spending in March can be attributed to a decrease in the total number of credit cards in use at that time. Between the weeks of March 15 and March 29, 2020, the number of credit cards tracked by Affinity in the first quartile decreases by 21.2%. Therefore, when we index by card count to account for changes in coverage, we nullify this effect. In their "Track the Recovery" product, Opportunity Insights adjusts for this problem after they index by card count, as noted in the literature review. Implementing a similar procedure is an essential next step to improve our analysis of spending patterns at various stages during and after the COVID—19 pandemic.

Alongside the above results, we obtained bootstrapped standard errors for cumulative spending for each income quartile, though above we have omitted standard error bars from our visualization for clarity. Instead, we plot them on separate graphs for each income quartile, which are located

in Figures 17, 18, 19 and 20 of the Appendix. We can see that the standard errors are generally higher for the lower income quartiles- likely an indication that we have fewer data points for these. In fact, at some points our standard errors were over 4% for Quartile 1. This is consistent with our suspicions that Affinity is skewed towards higher-income individuals in its sampling. Bootstrapped standard errors are larger for raking algorithm estimates than they are for estimates derived from regular indexing, a difference that generally increases over time. This is intuitive, since, in order to improve the accuracy of our estimates, raking gives higher weights to characteristics for which we have less information. When we weight observations in this way, we make an inherent bias-variance tradeoff. However, as mentioned in the methodology, we are likely overestimating our standard errors as we are not recomputing the weights with each iteration. Doing this would give us a fairer representation of the standard errors in our results. Computing bootstrapped standard errors for year-over-year changes is another important next step in validating our results.

5.3 Consumer Habits and the Impact of Credit Cards

In this section we explore potential differences between card holders and non-card holders. A key aspect of our analysis is that the data we use tracks only credit card/debit card spending in the US economy. The potential problems with this can be observed in Figure 21. In this data it is clear that credit card value only accounts for 60.4% of all transactions. This presents some potential problems with the interpretation of our results as measures of overall changes in retail spending. If credit cards do not act as a valid proxy for all spending then using this 60% of transactions would account for a biased sample as a result of this selection. Given stronger data there are many ways to account for this (notably retail scanner data that collects other spending types), however with public data analyzing presents some challenges.

We decide to leverage the Federal Reserve’s Survey and Diary of Consumer Payment Choice between 2016 to 2020 to explore this topic. This data source has two key components, the survey and the diary with about $n=4721$ respondents each year and $n=22388$ transactions. The survey asks numerous questions about the demographics of the household, such as household income, race, state, and most importantly number of credit cards held. The diary portion has respondents record purchases made including key details such as location of purchase, amount, type of payment, type of payments allowed, etc. These two sections allow us to analyze the spending patterns of individuals

by demographic groupings. A couple of issues exist with this dataset. Primarily the sample size is greatly reduced when compared with Affinity and some of the other existing datasets. Additionally, the survey is only run once a year in the months at the end of the year over the course of three days per respondent. This rules it out of being used to estimate changes in spending over the course of the entire year and reduces external validity of estimates produced from it. Finally, there exists attrition in the data which complicates the interpretation of some results since it is hard to verify if this is random attrition. Nonetheless we proceed with this dataset to explore some of the questions that arise on this topic.

A few topics stand out as key areas to analyze. First we explore whether individuals who own credit cards are observably different than those who do not own credit cards. Along key demographic angles (age, race, state, and income) we graph the differences between individuals in sample who own and do not own credit cards. These can be seen in figure 22, 23, 24, and 25. The important takeaway from these graphs is that there are large differences between cardholders on the basis of race and income but not necessarily age or state. This implies that the analysis of differences should control for these in order to understand the unobservable differences. It is important to note that in the following analysis we only compare credit card holders to non-credit card holders instead of card holders (debit or credit) to non card holders.

There are some problems that arise from needing to control from income and race in this sample. The main cause of this problem is the lack of sample size. Adding these controls requires increased data to fit the model. Since the data does not have a large sample size we must impose additional parametric assumptions on the models we employ to estimate differences. The first parametric difference is to assume that differences are linear in income bucket. Second we assume no interaction between the different demographic controls. This is the extent to which extra parametric assumptions are added to the models. The first set of models estimate the difference in overall spending between credit card holders and non-credit card holders.

Figure 26 shows the result of three different OLS models and one IV model. We begin by analyzing the three OLS models. This shows that by adding additional controls we can explain the difference observed in column 1. The key variable to analyze in each of these columns is `cc_adapt` coefficient which represents the controlled difference between spending in credit card holders and non-credit card holders. Coefficients here will be interpreted as semi-elasticities for convenience,

but the sign and significance is the important aspect. It is clear in these models that observable differences account for at least 75% of the original 0.4 log spending differences. However, some of these observable differences are of some potential concern. Even within groups of people with the same income, same number of trips, same age, etc we still observe a 10% difference in spending. After controlling for race fixed effects in a model not shown here that number stays essentially the same. This demonstrates that these effects are already accounted for in our other variables. Assuming state or other demographic controls do not influence spending significantly after these controls it is obvious there is an observable difference between the spending level of these two groups. However, this does not pin down the causal link between these variables. In column four we implement an IV regression as an example of a causal estimate. This model does not satisfy exclusion so we leave it as an example of a potential method moving forward to estimate a causal link with different data/instruments. In this IV we control for credit card history and instrument for owning a credit card with reported convenience of using a credit card. These models suggest unobservable spending differences between card holders and non-card holders. This does not imply our results from earlier are biased since we are focusing on changes in spending as opposed to their absolute value. However, it does suggest that these groups may behave differently in ways that demographic controls (implemented in our case through the raking algorithm) do not fix. This implies further research needs to be done or more comprehensive data needs to be explored to confirm that the volatility of non card spending matches that of card based spending.

Building on this we explore how card holders spend cash. For robustness of our estimates, ideally card holders do not spend cash in a systematically different way than with cards. In order to evaluate this concern we implement OLS models to estimate correlations. In figure 27 we show various estimates to analyze this question. Across these models we find that adoption of credit cards is correlated with less spending of cash even when controlling for other potential covariates. This implies that individuals who own credit cards spend less cash per trip where they pay with cash than those who do not. There are numerous possible reasons for this difference, which may lead to lower year over year differences. As a result this table strongly implies that even among the individuals who do own credit cards there may be uses for cash that are less volatile. It is still important to track these.

In summary, we observe strong reasons to suspect that the use of cash in the economy cannot

just be corrected for by scaling for the proportionate value of transactions. It is clear that cash transactions play a different role among people who own cards and individuals who do not own credit cards spend less as a whole. This is true even when we account for demographic characteristics that could influence spending habits. These differences are indicative of the potential that our main results may be biased by leaving out this critical element of the retail economy. As a result, we will recommend in the future steps of this project to explore data opportunities which include native cash spending, as is common in retail scanner based datasets.

6 Conclusion and Next Steps

6.1 Conclusion

Our original motivation in pursuing this project was to observe the ways in which the retail spending habits of different income groups have changed over the last few years. We have constructed a tool that gets us closer to answering this question. Our work built on the raw Affinity data and took inspiration from the work Opportunity Insights. Most notably, we used a raking algorithm. This helped us account for biases in the demographics of age, state, and ethnicity that were present in the set of counties tracked by the Affinity data. Doing so allowed us to produce figures that were more representative of the spending of the U.S. We also only considered retail spending (NAICS 44 and 45), and added features specific to income spending, including adjusting for the coverage of each income bracket.

The second part of our project is our consideration of non-card holders and cash transactions, as well as validating the assumptions made in our methodology. The importance here is two-fold. First is the fact that our card-indexing method of accounting for coverage implicitly assumes that the number of cards a cardholder uses remains constant from period to period. Our preliminary analysis shows that this method is robust. The second is the fact that there are many who do not own cards or choose to spend cash even while owning a card. This research explored the impact of this on our estimates, and the way that this group can be included in future estimates.

As discussed in the introduction, this project serves as a first step to creating a data product to estimate retail spending by income level. There are several areas of limitations for this project, including the heavy aggregation of our main data source of Affinity making it difficult to make

substantial improvements on the data. What follows is a brief overview of next steps we identified over the course of our project to improve the estimate further, along with future plans for this work.

6.2 Next Steps

Looking forward, the U.S. Census Bureau is continuing this project in-house, where future areas of focus are: improving assumptions; including new data sources; and adjusting methodology. The best way to improve our assumptions is to adjust the raking algorithm to include more dimensions, such that it rakes over more demographics thus creating a more representative sample. Furthermore, it might be useful to change the code to allow for the growth of the population, which is not considered when looking at per-card spending. Using more contemporary demographic statistics instead of only the one from 2022 would give a more accurate picture of the biases in demographic compositions.

Secondly, a large improvement would be to synergize all the data sources available to the U.S. Census Bureau which are summarized in [31](#) and include: Affinity Solutions; SafeGraph; Circana; Booth-NielsenIQ; Diary of Consumer Payment Choice; and MRTS. This would lead to an improvement in coverage and a reduction in variance.

Finally, our method could be adjusted with the other data sources to more closely align with their modus operandi by presenting data in quintiles instead of quartiles. A last area for improvement would be to refine the age group classifications to better reflect credit card ownership patterns, specifically by creating a new group for ages 0-15 to account for the lack of credit card usage in this demographic. This adjustment would enhance the accuracy and representativeness of locations with a significant young population.

References

- [1] NAICS Association. Naics code description. <https://www.naics.com/naics-code-description/?code=44-45#:~:text=44%2D45%20%2D%20Retail%20Trade&text=The%20Retail%20Trade%20sector%20comprises,to%20the%20sale%20of%20merchandise>, 2023.
- [2] U.S. Census Bureau. Monthly state retail sales technical documentation. https://www.census.gov/retail/mrts/www/statedata/msrs_technical_documentation.pdf, 2023.
- [3] Opportunity Insights. Economic tracker. <https://tracktherecovery.org/>, 2023.
- [4] Raj Chetty, John N. Friedman, Michael Stepner, and the Opportunity Insights Team. The economic impacts of covid-19. 2023.
- [5] JPMorgan Chase & Co. Household pulse: Balances through october 2023. <https://www.jpmorganchase.com/institute/research/household-income-spending/household-pulse-cash-balances-through-october-2023>, 2023.
- [6] Bank of America. Consumer checkpoint. <https://institute.bankofamerica.com/content/dam/bank-of-america-institute/economic-insights/consumer-checkpoint-january-2024.pdf>, 2024.
- [7] JP Morgan Chase. The initial household spending response to covid-19. <https://www.jpmorganchase.com/institute/research/household-income-spending/initial-household-spending-response-to-covid-19#:~:text=Average%20weekly%20credit%20card%20spending%20per%20household%20>, 2020.
- [8] Federal Reserve Bank of Atlanta. 2022 survey and diary of consumer payment choice: Summary results. https://www.atlantafed.org/-/media/documents/banking/consumer-payments/survey-diary-consumer-payment-choice/2022/sdcpc_2022_report.pdf, 2023.
- [9] Nisha Prakash, Subburaj Alagarsamy, and Aparna Hawaldar. Demographic characteristics influencing financial wellbeing: a multigroup analysis. 2022.
- [10] U.S. Census Bureau. County population totals: 2020-2022. <https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-detail.html>, 2022.
- [11] Harvard Dataverse. County fips matching tool. <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/OSLU4G>, 2021.

- [12] Remi Sebastian Kits. weightipy: A python package for weighted analysis. <https://pypi.org/project/weightipy/>, 2023.
- [13] W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table: An iterative method for the estimation of probabilities. *The Annals of Mathematical Statistics*, 11(4):427–444, 1940.

7 Appendix

7.1 Raking Algorithm Visualization

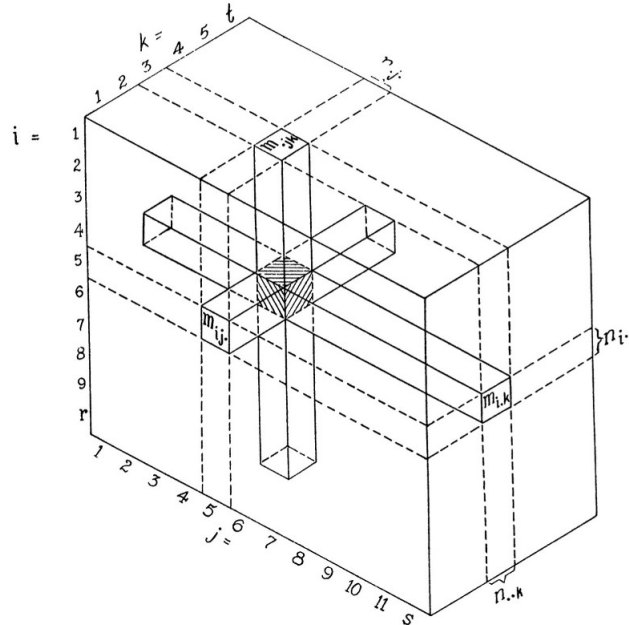


FIG. 2. SHOWING THE SYSTEM OF NOTATION FOR THE CELL FREQUENCIES AND MARGINAL TOTALS IN THE THREE DIMENSIONAL SAMPLE

Figure 16: Raking Algorithm over 3 Dimensions, where the totals in each row are known [13]

7.2 Figures: Bootstrapped Standard Errors for Cumulative Changes in Spending

Bootstrapped SEs for Quartile 1

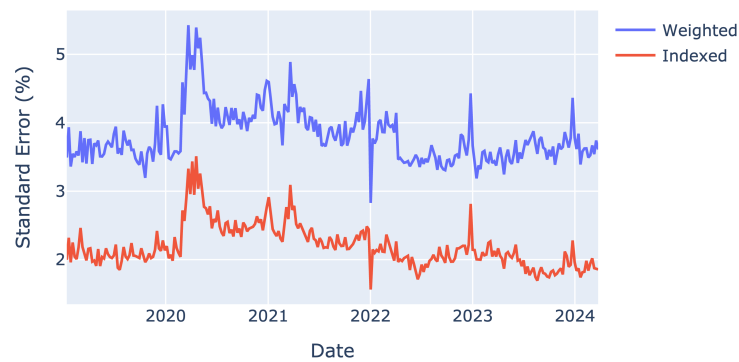


Figure 17: Bootstrapped Standard Errors, Quartile 1

Bootstrapped SEs for Quartile 2

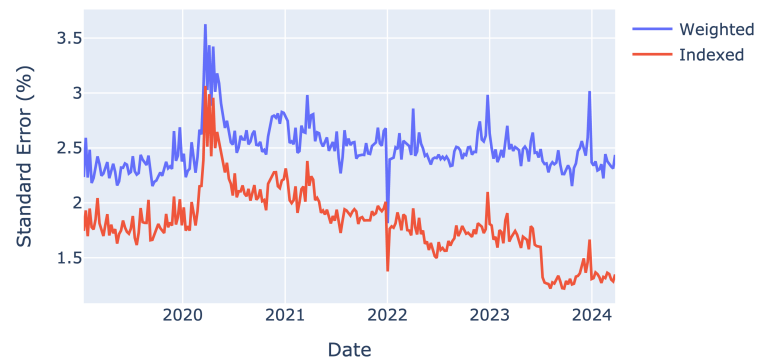


Figure 18: Bootstrapped Standard Errors, Quartile 2

Bootstrapped SEs for Quartile 3

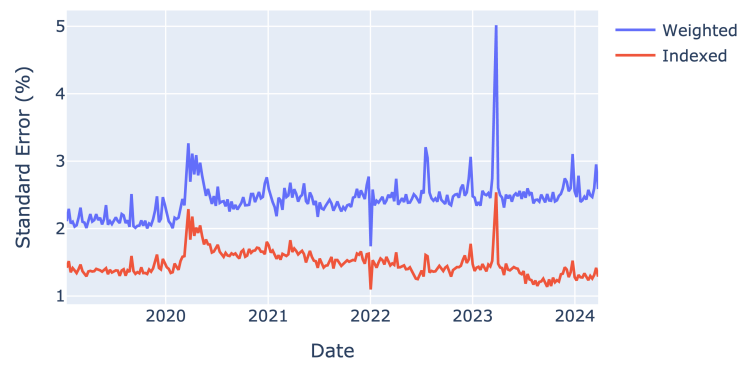


Figure 19: Bootstrapped Standard Errors, Quartile 3

Bootstrapped SEs for Quartile 4

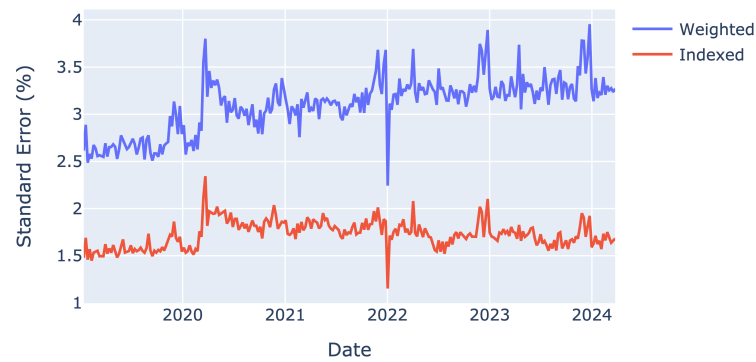


Figure 20: Bootstrapped Standard Errors, Quartile 4

7.3 Figures: Cardholders vs. Non-Cardholders

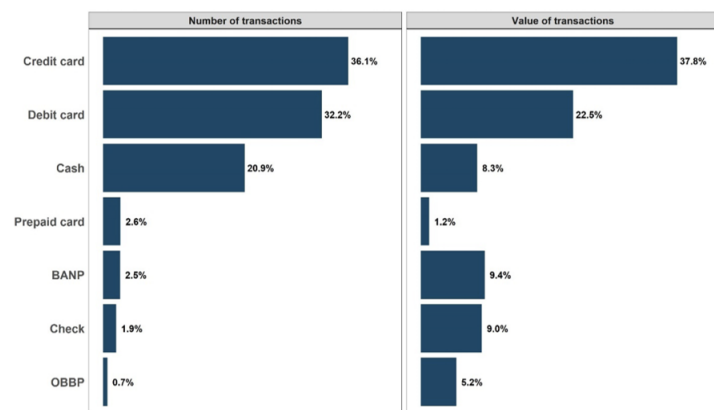


Figure 21: Spending Volumes 2022

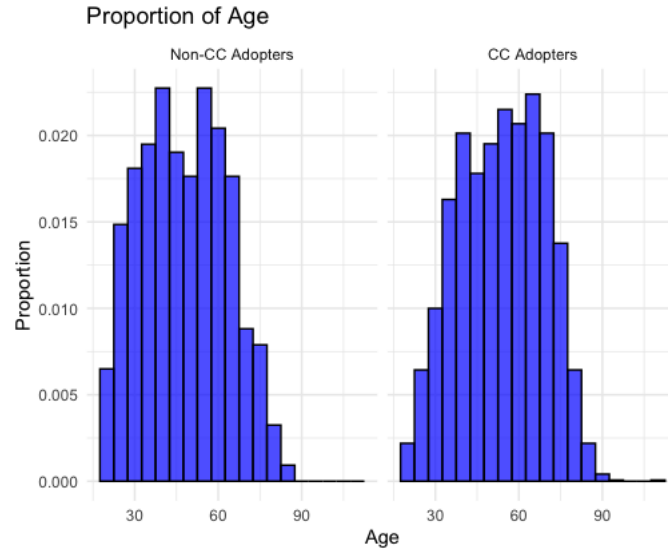


Figure 22: Age Distribution by Card Holding

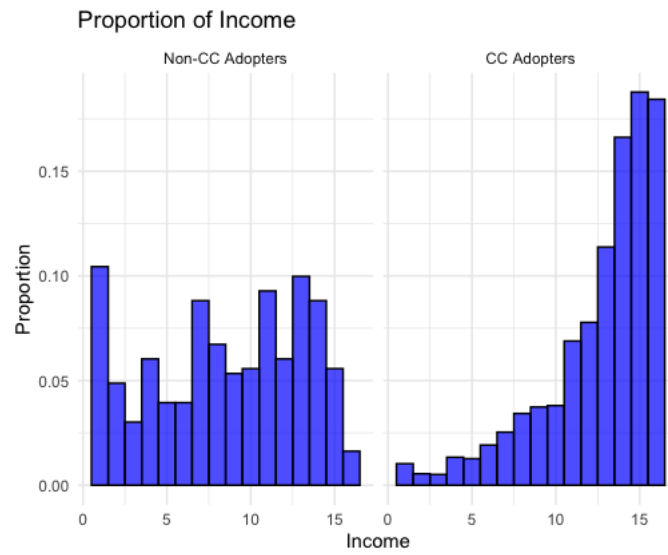


Figure 23: Income Distribution by Card Holding

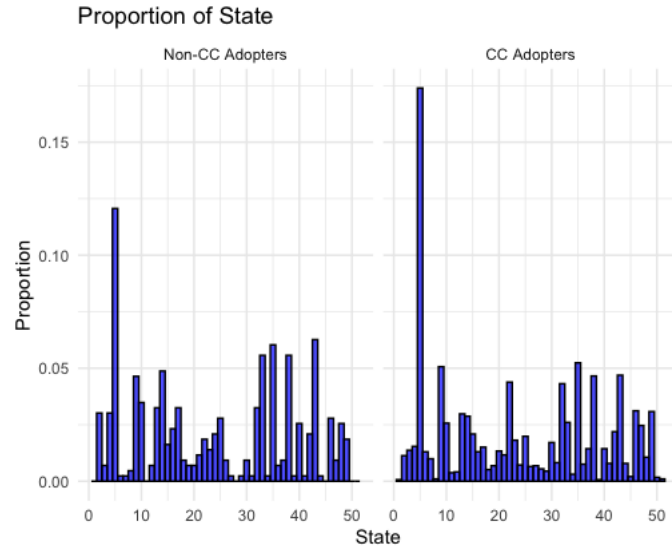


Figure 24: State Distribution by Card Holding

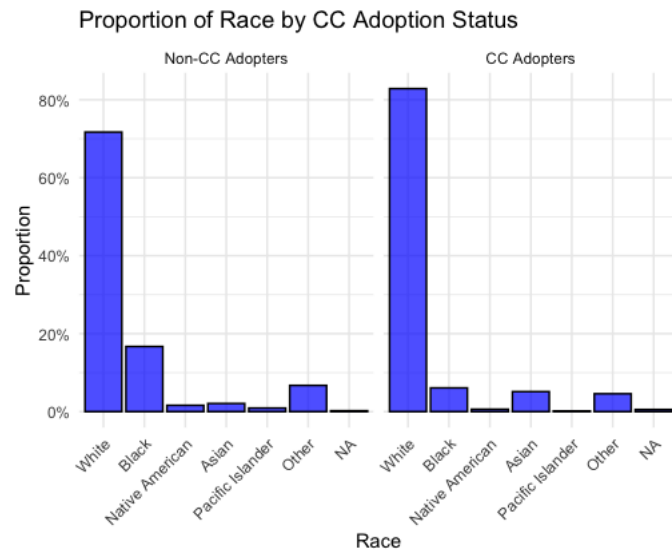


Figure 25: Race Distribution by Card Holding

Table 1:

	<i>Dependent variable:</i>			
	log(amnt)			
		<i>OLS</i>		<i>IV</i>
	(1)	(2)	(3)	(4)
hhincome		0.059*** (0.005)	0.045*** (0.004)	0.040*** (0.010)
as.factor(Year)2022		0.069** (0.031)	0.075*** (0.026)	0.076** (0.034)
as.factor(as003_d6)2				0.306 (0.502)
as.factor(as003_d6)3				-0.001 (0.469)
as.factor(as003_d6)4				-0.253 (0.476)
as.factor(as003_d6)5				-0.183 (0.481)
cc_adopt	0.443*** (0.042)	0.204*** (0.046)	0.109*** (0.039)	0.203 (0.300)
age		0.002** (0.001)	0.0002 (0.001)	0.001 (0.001)
num			0.391*** (0.008)	0.406*** (0.010)
Constant	3.852*** (0.039)	3.176*** (0.075)	2.586*** (0.064)	2.656*** (0.465)
Observations	6,222	6,213	6,213	3,567
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01			

Figure 26: Spending Models

Table 1:

	<i>Dependent variable:</i>			
	cash_proportion	cash.num	card_per_trip - cash_per_trip	log(cash.amnt)
	(1)	(2)	(3)	(4)
num		0.238*** (0.010)		
hhincome	-0.014*** (0.003)	-0.020*** (0.005)	5.386 (6.683)	-0.010 (0.007)
as.factor(Year)2022	0.004 (0.019)	-0.009 (0.038)	49.936 (44.607)	0.076 (0.049)
cc_adopt	-0.215*** (0.025)	-0.196*** (0.051)		-0.162*** (0.062)
cash.amnt.no.card	0.001*** (0.0001)	0.002*** (0.0002)		0.006*** (0.0003)
cash.num	0.021* (0.011)			0.424*** (0.028)
age		0.003** (0.001)		
Constant	0.869*** (0.035)	1.008*** (0.082)	-21.975 (89.375)	2.241*** (0.088)
Observations	1,580	1,580	521	1,580
<i>Note:</i>			*p<0.1; **p<0.05; ***p<0.01	

Figure 27: Card Holder Models

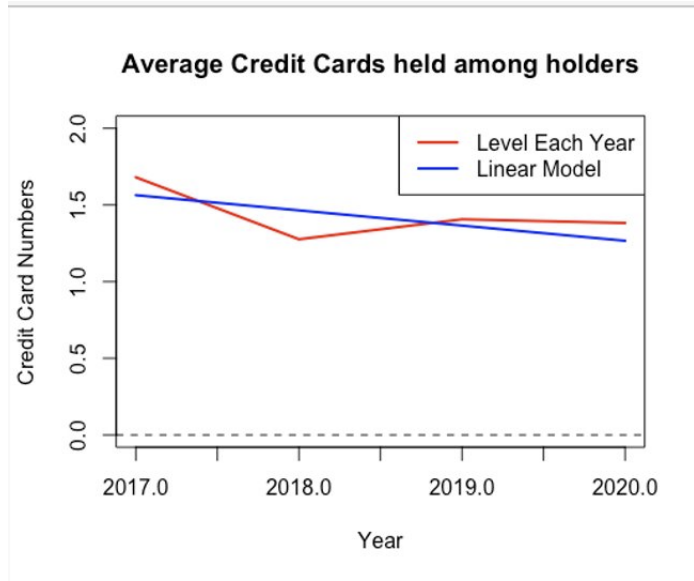


Figure 28: Trends of Card Holders 1

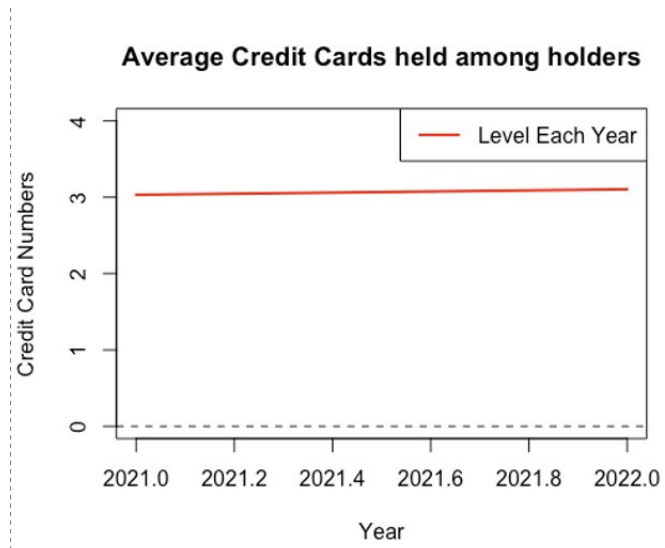


Figure 29: Trends of Card Holders 2

Characteristic	2011	2014	2017	2021	2024*
Ownership	61.9%	60.1%	65.6%	66.7%	68.5%
Used a credit card	-	57.1%	61.9%	62.5%	65.3%
<i>Gender:</i>	-	-	-	-	-
Male	67.8%	58.7%	68%	65.6%	65.8%
Female	56.3%	61.5%	63.3%	67.7%	71.2%
<i>Age group:</i>	-	-	-	-	-
15-24 years	38.7%	24.3%	39.5%	37.3%	37.8%
25 years and older	66.8%	67.1%	70.2%	71.8%	73.5%
<i>Income level:</i>	-	-	-	-	-
Poorest 40%	43.3%	40.2%	43.6%	46%	46.3%
Richest 60%	74.3%	73%	80.2%	80.4%	83.3%

Figure 30: Statista Survey Results

7.4 Data Sources

Source	Key Features
Affinity	- Tracks ~10% of all credit card transactions - Aggregated on county level, NAICS code, and income quartile of the zip code where the transaction took place
SafeGraph	1.5% of transactions in the U.S. since 2019. - Per location offers store name, geolocation, # of transactions, # of customers by income - Difficult to account for changes in coverage since we can't identify unique consumers
Circana	Measurement of more than 4 trillion in global consumer spend covering over 30 million products and representing 75% of the world's GDP
Booth – NielsenIQ	Retail Scanner Data - Weekly product data for 2.6-4.5 million UPCs
Atlanta/SF/Boston Fed	- Diary of Consumer Payment Choice tracking just below 5,000 people and 18,000 transactions in October each year - Records retail spending data from all sources, including cash
MRTS	- Mail-out/mail-back survey of about 13,000 retail and food services businesses. - Provides data on monthly retail sales

Figure 31: Key Features of the Main Data Sources