Concentrating on Customers: Spending Across Firms and Space [*]

Christina Patterson
Chicago Booth & NBER

Joseph Vavra
Chicago Booth & NBER

March 2024

Abstract

Typical firm-level data can be used to measure the distribution of total spending at different firms but contains no information on their individual customers. Using data from trillions of dollars of card-level transactions for a large credit card company, we document new facts about the distribution of customer spending across firms and space. We propose a new measure of market concentration that is based on the location of the customer, rather than the firm, and use this measure to describe the potential options available to customers across the country in an array of retail and service industries. We find that due to customer travel and online shopping, customers face less concentrated markets that are more similar across space than suggested by firm-level data. Individual customers in markets with more potential options also distribute their spending more evenly across firms, suggesting that access to more potential options leads to an increase in store switching and demand elasticity. In ongoing work, we bolster this conclusion by studying how customer shopping patterns change after firm entry and exit. Finally, we show that within a market, large firms systematically attract customers who shop around less than those at smaller firms.

[*]We thank Alex Blumenfeld, Luisa Pomarici, and Daniela Villacreces for excellent research assistance, and seminar participants and UIC, the Empirical Macro Workshop-Phoenix and the Federal Reserve Bank of Philidalphia for helpful comments.

1 Introduction

How do the number of potential shopping options and levels of competition vary across space? How big are markets, and which firms compete with each other for customers? Answering these questions requires knowing how far customers are willing to travel for purchases and on how segmented customers are across firms within particular locations. In this paper, we use data covering trillions of dollars of spending from the universe of transactions of a major U.S. credit/debit card processor to document new facts about the distribution of firm spending across individual customers.

Little systematic evidence exists on the distribution of spending by individual customers across different firms and how these customer-specific spending patterns vary across space and products. Scanner data sets can be used to measure detailed spending patterns for individual customers, but the number of customers sampled is too small to measure anything about how total spending for individual firms is allocated across different customers.[1]

The Census measures sales by individual firms, and a large literature uses this data to measure firm concentration as a proxy for market dominance. However, this data contains no information on the identity of which customers shop at which firms, and so this analysis must make strong assumptions about the geographic scope of markets and how far customers are willing to travel. This shopping geography likely varies by product (e.g. dry cleaners vs. appliances) and by location (e.g. urban vs. rural). In addition, the growing prominence of online retail makes it hard to measure potential shopping options and market scope using data based on the location of firms since the physical location of online shopping options is typically not well-measured and also may not have any relevance to customers.

Even more fundamentally, two firms operating in the same category located in the same place could potentially draw customers from different distances and thus compete with a different set of firms. Indeed, we begin our empirical analysis by showing there is substantial cross-firm heterogeneity in the geographic scope of customer bases even with the same location and product category. For example, "Restaurant A" (located in the zipcode 60614 in Chicago) makes only 8% of its sales to customers from 60614 while "Restaurant B" (down the block) makes 48% of its sales to customers from 60614.[2] This

[1] For example, the NielsenIQ panel samples around 60k households per year for the entire U.S. meaning that many stores that actually have sizable spending have no customers in the panel, and even those with positive spending by Nielsen households typically only have a handful of such customers.

[2] Data use agreements prohibit particular firm identities from being disclosed.

in turn implies that there is actually no "right" geography that can be applied to correctly group all of the firms in a particular location and category into the same market: any fixed boundary based on firm locations that includes all relevant competitors for one firm will typically include irrelevant competitors for some other firm.[3]

To overcome challenges arising from measuring the distribution of spending based on firm location, in the rest of the paper we use our data to document new facts that instead focus directly on customer-based spending patterns. In particular, we construct a novel *customer location*-based statistic that measures the concentration of spending across firms of all customers living in a particular location. This statistic captures information about the overall set of shopping options utilized by customers living in a location, even if individual customers only shop in a select set of firms.[4] This statistic reflects the idea that even if an individual only shops in a single firm, they could have alternatively shopped at the set of firms where their neighbors shop.[5] Measuring the distribution of spending based on the location of customers rather than the location of firms means that we do not have to make any assumptions about the geography of shopping: the data directly reveals which firms customers travel for and which they do not, and this measure can directly account for online shopping options without any changes in methodology.

We document five facts about this customer-location based measure of spending options. First, the spending concentration of customers located in particular geographies is much smaller than the spending concentration of firms located in those same geographies. That is, the set of options utilized by customers living in particular places is much larger than the set of firms located in those same places. This is driven by customers shopping (both in person and online) beyond fixed geographic boundaries, and these shopping effects are stronger for products which are purchased infrequently. While this qualitative pattern is unsurprising, the quantitative magnitude is large. Importantly, while these shopping effects are strongest across small geographic boundaries like 5-digit zip codes, they are also economi-

[3]In principle, one could obviate these issues by directly estimating the cross-price elasticity of demand between all firms directly without having to take any stand on defining "markets". However, such broad substitution patterns are infeasible to estimate, so some market boundaries are widely used in practice.

[4]We focus on concentration of spending rather than counts of firms with spending by local customers since this better captures intensity of spending and focuses on more important spending options, but similar patterns hold when instead using firm counts.

[5]While this measure overcomes many limitations of the past literature, we note that it still measures the equilibrium distribution of spending under the current distribution of prices rather than the change in market shares in response to changes in prices that is more directly relevant for market power. We discuss this in more detail below.

cally large when looking at 3-digit zip codes (roughly cities) and even when looking across states. This implies that the spending concentration of firm sales within arbitrary borders overstates the spending concentration of the customers living within those same borders.[6]

Second, customer-location based concentration varies less over space than firm-location based concentration. That is, shopping patterns and the concentration of spending by customers living in different locations are much more similar than suggested by looking at how the concentration of firm sales varies across locations. Together, facts 1 and 2 suggest that markets based on the locations of firms are likely a poor proxy for more relevant markets based on the locations of customers and where these customers actually shop. Customers shop in wide geographic areas and regularly shop across neighborhoods, across cities and even across states to take advantage of firms that are not located directly next to them.

Third, online shopping plays an important role in these first two patterns. It is not obvious ex-ante whether online shopping would increase or decrease the concentration of spending, since this depends on whether online shopping is more or less dominated by big firms than local shopping is. Online shopping might lower concentration if it leads customers to purchase at smaller, niche firms, but could instead increase concentration if it is dominated by a small number of large sellers with massive scale. To the best of our knowledge, we are the first paper to look at the relationship between online shopping and the concentration of spending, and we find that in general online shopping lowers the concentration of spending. Furthermore, these online shopping effects are strongest in locations where customers have fewer local options, as measured by the concentration of in-person spending. This means that in addition to lowering overall concentration, online shopping also reduces heterogeneity in concentration across space.

Since our data covers a short time-series, we focus on describing cross-sectional patterns. However, the fact that online shopping lowers concentration combined with the large increase in online shopping over the past decade may mean that local concentration growth has been overstated in census data since online spending is generally not captured in these measures.

Our first three facts focus on describing variation in customer-location concentration across products and space. As already discussed, customer-location concentration measures the distribution of spend-

[6]One could alternatively fix a customer geography and then ask what firm geography yields the same concentration of sales. However, even if two alternative geographies yield the same customer and firm concentration, the larger firm geography will almost always exclude some firms which are shopped at by relevant customers and include some firms which are not. Furthermore, constructing this equivalence in the first place requires the customer data that is unavailable in other data sets.

ing by everyone living in a location and thus captures information about the set of potential shopping options available, even if individual customers only shop in a small set of firms. Ultimately, we are interested in understanding whether more potential options among customers reflects higher price elasticities, and therefore, more competitive markets. While we do not measure price elasticities directly, we explore whether individuals with more potential options spread their spending across more stores, which would be consistent with more competitive markets. Our fourth fact is that the concentration of individual spending is indeed highly correlated with potential options measured by customer-location concentration but is uncorrelated with firm-location concentration. Put differently, having neighbors which only spend in a few firms predicts that an individual customer will only spend in a few firms. In contrast, only having a few firms located nearby does not predict that an individual customer will only spend in a few firms. Thus, customers who have more potential shopping options – as measured by the behavior of their neighbors – do indeed have more dispersed spending in a given year.

The final part of our paper links these individual concentration measures back to particular firms. In particular, we explore how much firms vary in their exposure to customers with different levels of individual concentration. To the extent that past customer concentration proxies for customer loyalty, firms whose customers have more concentrated spending may in turn have customers who are less likely to switch to competitors in response to changing market conditions. Our fifth fact is that there is substantial variation across firms in the concentration of their customers' spending and that this concentration is increasing in firm size. Put differently, large firms tend to have customers that shop at only a few firms while small firms tend to have customers that shop at many firms. Notably, this variation holds across firms in the same category and location and after controlling for customer-location concentration. This suggests that this variation in customer concentration and its correlation with firm size is unlikely to be driven by heterogeneity in shopping options available to local customers and is instead driven by customers with different preferences sorting into different firms.[7] Overall, the spending patterns we document suggest that large and small firms within a market have customers with systematically different price elasticities.

[7]This size relationship could arise through many potential channels. For example, large firms might become large by having lower prices, higher quality at a given price, different product assortment or more/more effective advertising. Alternatively, they might have other technologies that encourage customer lock-in even when other options are available like annual membership fees or fixed contracts, although those specific examples are unlikely to apply for most firms and categories in our data.

While our current analysis focuses on simple descriptive analysis, even these simple results show that the geography of markets and spending patterns across firms captured by linked customer-firm data differ in many ways from what one would infer from firm sales data alone. Some of these differences are driven by online shopping, which is difficult to capture in traditional data, while others arise from divergence between arbitrary geographies and the locations where households actually choose to shop.

Of course, these descriptive patterns come with some important caveats. First, our current analysis focuses on measures of spending concentration. While concentration is widely measured in practice and often interpreted as a proxy for market power, the IO literature has long recognized that this interpretation may be misleading since the relationship between concentration and market power depends on assumptions about market structure and the nature of competition.[8] While our measure improves on the traditional literature by defining markets based on actual shopping patterns and by looking at the relationship between market concentration and individual concentration, it is still subject to the same caveat. In particular, we capture the concentration of spending given observed prices rather than the change in market shares in response to changes in prices that are more relevant measures of market power.

In ongoing research we hope to make progress on this front by combining our customer shopping data with measures of firm entry and exit. The comprehensive coverage of our data allows us to measure firm entry and exit, and to look at customer switching patterns around these market changes. In particular, we will explore the relationship between customer concentration and customer substitution around firm entry and exit. By looking at these substitution patterns, we can then provide more direct measures of market power and relate these to our descriptive facts.

Our research contributes to a large related literature. An active macro literature has documented striking increases in firm concentration over the last several decades (see for example Autor et al. (2020)). A subset of this literature has explored whether national and local concentration exhibit similar trends (Autor, Patterson and Van Reenen, 2023; Rossi-Hansberg, Sarte and Trachter, 2021 ; Smith and Ocampo, 2022). Relative to this literature, we contribute a more nuanced view of market definition and individual shopping patterns. While we have a short time-series and so cannot directly speak to these long run trends, we find that in the cross-section the patterns arising from our customer-oriented perspective

[8]See Syverson (2019) for more discussion of the relationship between the macro and IO literature. See also Miller et al. (2022) and Berry, Gaynor and Scott Morton (2019) for additional discussion of concerns with interpretation of concentration relationships

often differ substantially from those using a firm-location perspective.

The most closely related paper is Batch et al. (2023). This paper uses credit card data on county spending flows to construct 'consumption zones' that measure the size of spending markets. While we share some similar motivation, our data measures individual card spending at individual firms and allows us to speak to a richer set of spending patterns. Using our data we can study individual customer and firm spending patterns within and across counties, which is impossible with the aggregated county spending data that they use in their analysis. Indeed, our data shows that there is substantial heterogeneity across firms even in the same location and category in how far their customers travel and on the concentration of their customers' spending.

Klenow et al. (2020) uses the same dataset as we use in our analysis to show an important role of customer counts in driving variation in firm sales. While related to our analysis of customer spending distributions across firms, these are distinct facts with different applications.[9] They focus primarily on understanding implications for growth while we focus on understanding market definitions and how these vary across products and space.

We are also not the first paper to use this data to document patterns related to online shopping. Our results showing the relationship between online shopping and firm concentration complement results in Dolfen et al. (2023) quantifying the welfare gains from e-commerce. While both papers use the same data and both document facts related to online shopping patterns, they are otherwise quite distinct since they document different facts and focus on different questions.

Baker, Baugh and Sammon (2023) also use customer-firm linked transaction data to document an important relationship between customer churn and firm performance. While obviously also focused on customer-firm spending relationships, we focus on a somewhat different set of questions. In addition, our data covers 250 million cards and we analyze spending at more than 1 million firms while their data covers 2.7 million users but only analyzes spending at a set of 558 specific firms. These 558 firms are large and so cover a sizeable share of spending for these users but this data is not comprehensive enough to measure the types of market wide concentration statistics that we focus on.

Our paper also relates to a large IO literature studying market power and market definition. As

[9]For example, these scenarios produce the same value for their customer count statistic but have different implications for customer concentration: 1) Customer A spends $1 at firm 1 and 2 and Customer B shops at neither firm. 2) Customer A spends $1 at firm 1 and $0 at firm 2 while Customer B spends $0 at firm 1 and $1 at firm 2.

already discussed, our data does not include price variation, and we do not currently attempt to measure market power. However, our focus on customer shopping patterns and substitution across firms moves our analysis closer to approaches taken by IO. For example, Conlon and Mortimer (2022) argue that diversion ratios (if the price of j rises, what fraction of customers who switch from j switch to k) should play an important role in merger analysis. While we do not have price variation to measure diversion ratios, we do measure the degree of switching of individual customers across firms given current pricing behavior and how this relates to the overall distribution of customers across firms in different locations. In addition, unlike traditional concentration measures which must take a stand on whether two firms in different places are "in the same market", our customer-location based measures are agnostic about the geography of firm competition. Thus, our analysis moves closer to more theoretically grounded approaches to measuring market power while still being widely applicable across a broad set of markets and products.[10]

2 Data Description

Our primary dataset for this analysis is the universe of all credit and debit card transactions run through a major credit card processor within the United States from 2018-2022. The scope of the data is large - in this time period, we observe more than 600 million cards in each year and observe more than 4 trillion dollars of spending in each year.

The unit of observation in our dataset is the individual transaction between an (anonymized) credit card and a merchant. For each of these transactions, we observe the exact amount of the transaction, the time and date of the transaction, the unique identifier of the card that made the transaction, and the location and identity of the merchant. For in-person transactions, we observe not only the identity of the firm, but also the 5-digit zip code (and typically the exact address) of the store where the transaction was made, allowing us to identify purchases in different establishments of the same firm. The firm identifier constructed by the data provider generally links franchised branches with common names (e.g. McDonalds), but does not link all subsidiaries within parent companies (e.g. Whole Foods would be separate from Amazon).

[10] Our results that customer and firm-based concentration statistics often differ sharply also suggests yet another concern with Horizontal Merger Guidelines by the Department of Justice and Federal Trade Commission (2023) focused on concentration measures based on firm-location.

Each transaction in the dataset is assigned a merchant category code (MCC). This 4-digit code captures the type of good or services that the firm provides. We aggregated very similar MCC codes into a single MCC category. Some merchants have multiple MCC codes corresponding to establishments providing different products (e.g. Costco wholesale vs. Costco gas stations). Our baseline analysis assigns all transactions within a firm to its MCC code with the most transactions, but results are robust to treating establishments in different MCC codes as different firms. We additionally observe whether the transaction was in-person (card present) or remote (card not present). If it was remote, for many transactions we further observe if it was an online purchase, a recurring payment initiated by the firm (e.g. monthly gym membership dues), or a purchase made over the phone. In the analysis that follows, we refer to all transactions without cards present as "online" transactions.[11]

While we observe many details on each transaction, we observe relatively little info about the customer making the purchase. Although we can link all transactions over time that are made on a given card, we cannot link multiple cards owned by the same customer together. Therefore, throughout the paper the term "customer" will always refer to the card, although we impose some additional sample restrictions to eliminate some cases where this distinction may be particularly relevant. The fact that cards are not linked to customers means that we do not observe demographic information on customers, except what can be inferred from their credit card spending. For each card, we impute the home location of the customer as the zip code with the most transactions in a given year.[12] Appendix Figure A1 shows that the population distribution of locations that we recover using this method tracks the distribution of population in the Census, raising confidence that we are indeed capturing home locations accurately.[13]

It is worth noting that households with multiple cards may use certain cards for certain types of spending due to rewards incentives. However, these rewards incentives vary across MCC codes but not across merchants within MCC codes. Since we measure cards but not households, we cannot measure this type of card substitution. However, since our analysis focuses on spending concentration within categories as measured by MCC codes, this type of cross-MCC card substitution driven by rewards

[11]We use this broad definition since some transactions do not have this breakdown and because phone and recurring payments are not obviously economically distinct from online payments. Nevertheless, conclusions about the role of online sales are similar if we re-do our analysis focusing on smaller subset of data explicitly labeled as online.

[12]In instances where these differ, we are capturing a notion of the "home" shopping zip code rather than the actual actual home, but since we are interested in measuring the geography of shopping, we think this is not an important limitation.

[13]We note that the presence of people with multiple cards will introduce some noise into these comparisons even if shopping zips and home zips always perfectly aligned.

incentives is unlikely to be important for our conclusions. However, store-specific credit cards which can only be used at a specific store are more likely to distort measures of firm concentration. To exclude these specialty cards, we keep only card-by-month pairs with transactions at 5+ distinct merchants.

We make several additional sample restrictions for our baseline analysis. We drop all transactions that are through payment facilitators (e.g. Paypal or DoorDash) since these transactions do not record the firm providing the good or service.[14] We also drop peer-to-peer transactions (e.g. Venmo). We further restrict attention to 29 categories across both retail and services, which are customer facing and have wide geographic representation: these categories all have firms present in at least 90% of 3-digit zip codes. To drop businesses that are not covered adequately in the transaction data, we also restrict our attention to establishments with 100 transactions in 2021.

Column 1 of Table 1 shows summary statistics for the resulting sample that we consider. Column 1 shows summary statistics for all observations meeting the sample selection criteria described above. These 29 categories have over 1.4 trillion dollars of spending in the average year in the sample, which comes from over 300 million cards spending at over 1 million firms. Each card spends an average of $4,743 in each year, and almost 25% of spending was done online. Since we focus in the descriptive analysis on geographic heterogeneity, in our main analysis sample, we restrict our attention to those zip codes where customers living there collectively spend at least $25k per year in each of the 29 selected categories. This restriction drops zip codes where there is little spending in many or all of these categories and results in a balanced panel of zip codes and categories for our analysis. Column 2 shows the same summary statistics for this subsample. Even though this restriction drops 77 percent of zip codes, since this restriction is specifically dropping those zip codes with little spending, we only lose around 14 percent of spending and retain over 250 million cards per year. The median annual spending per card and the share of spending done online remains similar, while the average firm size in this subsample falls slightly.

[14] We are able to include transactions made through Apple and Google Pay, as that affects the mode of transaction but not the firm identifier. In our final subsample of 29 industries, 5.4% of all transactions are at payment facilitators, a number that rose from around 3% in 2018 to 8% in 2022. Payment facilitators are most prevalent in barber shops and restaurants. Excluding these transactions will bias our results only to the extent that these transactions are unevenly distributed across firms or locations.

Table 1: Summary Statistics for Main Analysis Sample

	All ZIPs	Analysis Sample
Total spending	Over $1.4 trillion	Over $1.2 trillion
Number of cardholder ZIP5s	29,002	6,795
Number of firms	Over 1 million	Over 1 million
Number of cards	Over 300 million	Over 250 million
Mean card spend	$4,743	$4,894
Mean firm sales	$1,420,504	$1,242,398
Online share	24.4%	24.8%

Notes: This table shows annual averages for key summary statistics in the transaction data, covering 2018 through 2022. "Cardholder ZIP5s" are ZIP codes for which at least one card has been assigned that ZIP as its "home" location, defined as the ZIP where the card made the most in-person transactions in a given year.

3 Firms in the Same Location & Category Draw Customers from Different Distances

The existing literature typically defines markets based on the location of firms.[15] Under this approach, some boundary is drawn around some group of firms and then market shares are computed within this boundary. All firms within this boundary are then assumed to compete with each other while firms outside this boundary are outside of the market. The size of this boundary could potentially differ across space or across categories: one could use a small geographic boundary in Manhattan and a large geographic boundary in rural Kansas, and one might assume a smaller geographic boundary for everyday purchases like grocery stores than for infrequent purchases like furniture. The literature has focused primarily on debating what the "right" level of geography is for defining a market based on firm location. For example, should local market shares be based on zip codes or counties or some other geographic boundary?

In this section, we use our linked-customer shopping data to show that this empirical question is actually ill-posed and that in reality there is no "right" geography that can be applied to all of the firms in a particular location and category. Specifically, we show that even among firms located in the same narrow locations and selling in the same categories, there is substantial heterogeneity in the geographic scope of customer bases. This means that there is no boundary that can be drawn around all of the firms

[15]While we use the term "firm", for this and other firm-location based analysis we separately measure firm sales at establishments in each zip code, so location refers to the zip code where sales take place and not to the firms' headquarters.

in a location which will include all of the relevant competitors for some firms without also including irrelevant competitors for some other firms.

We illustrate this heterogeneity in customer distance using a simple summary statistic: what fraction of firm $j's$ sales come from customers who live in the same zip code as firm j?[16] For each firm j located in zip code z we compute its in-zip customer share:

$$IN_{j,z(j)} = \frac{\text{Total Spend on firm } j \text{ by customers living in } z(j)}{\text{Total Spend on firm } j}.$$

We then calculate the deviation between the in-zip customer share for firm j from the mean in-zip customer share of all firms in the same zip code and product category as j:

$$\widehat{IN}_{j,z(j)} = IN_{j,z(j)} - \underset{j \in c(j), j \in z(j)}{\text{mean}} \{IN_{j,z(j)}\}.$$

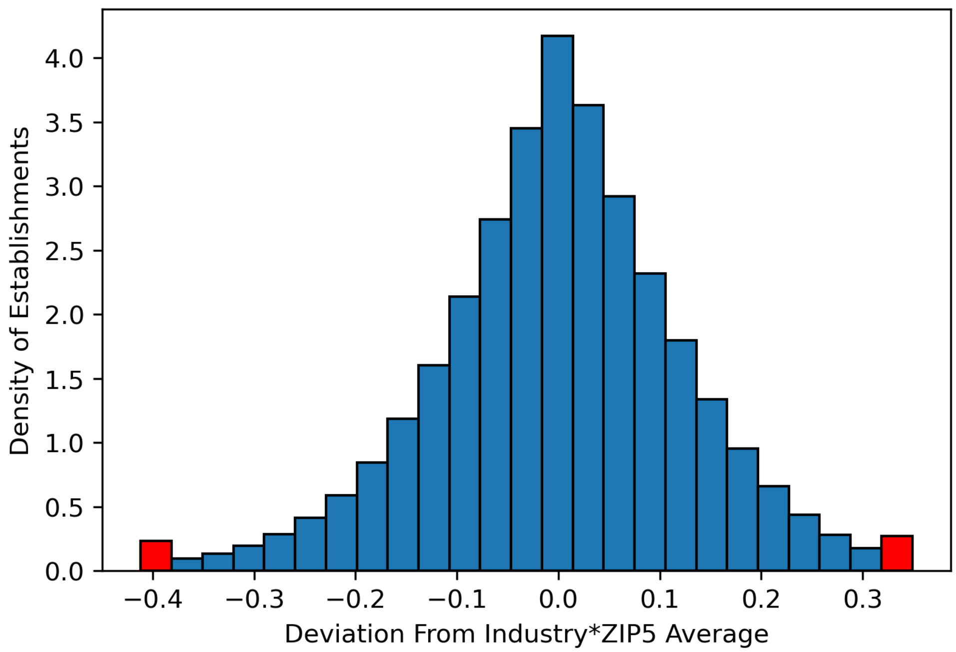
This statistic captures how much firm $j's$ in zip-customer share differs from the average firm operating in the same category and zip code as j. Figure 1 plots a histogram of $\widehat{IN}_{j,z(j)}$, pooling across all industries. This shows that there is substantial variation across firms in their in-zip customer shares even within these narrow industry x zip5 cells. Averaging across these cells, the mean within-cell standard deviation of the in-zip share is 11.2 percentage points. Firms at the 90th percentile of $\widehat{IN}_{j,z(j)}$ have in-zip shares 15.4 percentage points higher than the mean while firms at the 10th percentile of $\widehat{IN}_{j,z(j)}$ have in-zip customer shares 15.4 percentage points below the mean. Appendix Figure A2 shows $\widehat{IN}_{j,z(j)}$ separately for each industry group. Heterogeneity across firms in the scope of customer markets is pervasive and generally similar across categories.[17]

These results imply that defining markets based on firm location is problematic: even within narrow product categories, the geographic scope of customers and thus the relevant set of competitors for two firms located in the same place often differs substantially. Defining a market geography based on firm location requires that firms in the same location sell to customers at similar distances, and our results

[16] We focus on measuring heterogeneity in the in-zip customer share rather than on measures of customer distance since we do not observe the exact address of customers and only have information on the zip code where they most frequently spend. However, Appendix Figure A3 shows that despite substantial attenuation bias there is nevertheless heterogeneity when we measure customer bases using the geographic distance between firm and customer zip codes (imposing a distance of zero for same zip transactions.)

[17] The standard deviation of the in-zip share is largest for grocery stores (15.6 percentage points) and smallest for shoe stores (7.1 percentage points) but there are no obvious systematic patterns across categories.

Figure 1: Firms in the same location and category differ in the share of customers located in the same zip



Notes: This figure plots the distribution of $\widehat{IN}_{j,z(j)}$, the deviation of the share of a firm's customers who live in the same zip5 from the Industry*Zip5 average of this share. We include only observations in Industry*Zip5 cells that contain at least 2 firms, since $\widehat{IN}_{j,z(j)}$ is mechanically zero when there is only a single firm in a location.

show that this is not true even within narrow product categories. While the literature has debated the "right" market size for particular categories and firm locations our results show that this question is actually ill-posed and is not actually something that could have been answered with better data, because heterogeneity implies that no market size can be imposed uniformly across firms.

4 Customer Location-Based Concentration Definitions

In this section, we define an alternative measure of markets based on customer-location which overcomes these inherent challenges in firm-location based market definitions. Specifically, we explore customer-location based markets which measure the concentration of spending across firms among customers located in particular locations rather than measuring the concentration of spending across firms that are located in particular locations. While individual customers may only shop in a select set of stores, the set of stores shopped at by other customers living in the same location (who we assume face the same geographic constraints) reveals the set of relevant options in that location. Measures of concentration based on customer location capture the distribution of firms that customers in these locations actually choose from, without needing to take any stand on the geographic scope of purchases.[18] Unlike the firm location-based approach, this customer location-based approach does not require any uniformity in distance and is completely flexible on the nature of shopping patterns. For example, the customer-based shopping measure has no issue dealing with situations where customers mainly shop at firms in the same zip code but also purchase at some particular firms located far away without shopping at any firms in between. In contrast, these type of shopping patterns would be problematic for the firm-location based approach.

More formally, for each category and location z we measure the local market share of firm j:

$$m_{j(z)} = s_{j(z)} / \sum_{j(z)} (s_{j(z)}) \tag{1}$$

where $j(z)$ is the set of firms shopped at by all of the customers living in a particular location z. This definition of market share captures the fraction of all spending in a given category from all customers

[18]For some results, we restrict to some (large) maximum distance for inclusion of spending in order to exclude shopping that is done while on vacation, but this is not important for our conclusions.

living in region z that goes to firm j.[19] Using these market shares for each firm $m_{j(z)}$, we then calculate the Herfindahl Indexes (HHI) as

$$HHI_z = \sum_{j(z)} m^2_{j(z)} \tag{2}$$

This concentration measure captures how all customers living in the same location distribute spending across firms.[20] Importantly, unlike past research using data collected by firm location, this approach requires no decision on the geographical boundaries of markets. This allows for the possibility that customers might shop primarily nearby but also purchase from some more distant firms. Whether households shop nearby or at great distances for particular products is determined entirely by the data. Furthermore, this approach also lets us quantify the importance of online spending for concentration, since the location of the customer for online shopping is well-defined even if the location of the associated firm is ambiguous.

While we do not need to take a stand on where customers shop, this measure does require a decision on which households are likely to face a similar set of potential shopping options. In most of our results, we focus on households living in the same 5-digit zip code when defining our groups of households with similar shopping options. The median zip code is only 17 square miles and contains 28,000 people, so it is plausible that households living in these small geographic areas have similar access to all stores.

As a benchmark for how the HHIs in Equation 2 based on customer location differ from those based on the location of firms, we show comparisons to HHI measures for the firms located in the same places (i.e. all firms physically located within a given zip code).[21] Note that since we want to measure spending concentration in a particular place, we include only the sales for firm j which occur in zip code z in this measure.[22] However, it is important to again emphasize that this firm-location HHI changes with the size of the border one draws around firms. For this reason we also show comparisons to firm-concentration at other levels of geography, but note that one of the disadvantages of this traditional measure is exactly

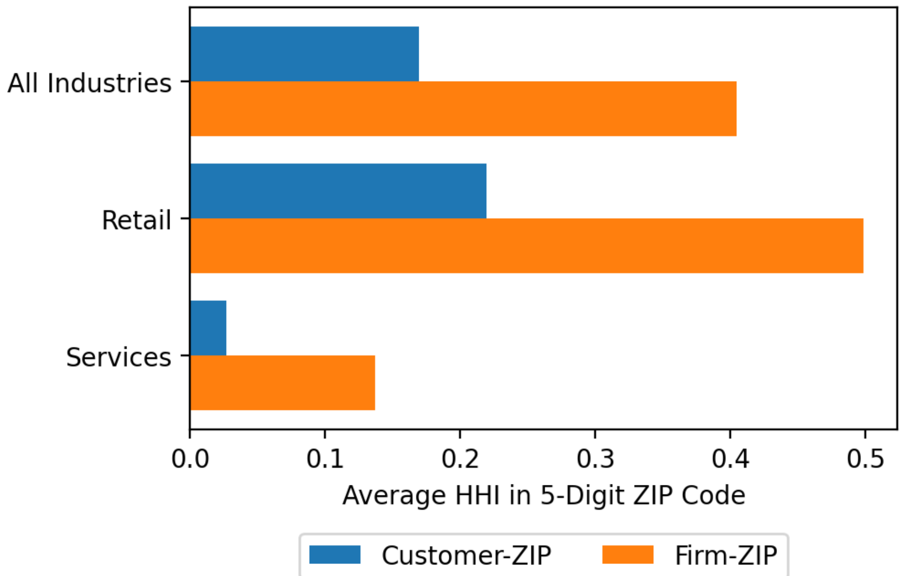
[19] For the customer-based measure, we add spending across all establishments for a firm j.

[20] All of our conclusions are very similar when using alternative measures of concentration.

[21] Given the construction of the sample, there are no zip codes where there are no customers spending in that category. However, there are several zip codes where there are no firms in a given category. For these cells, we impute an HHI of 1, but results are very similar if we exclude those zip codes and categories without any firms.

[22] The difference in the geographic nature of the underlying data means that the firm-location measure is closer to establishment concentration while the customer-location measure is closer to firm concentration. This choice of aggregation is conservative for the facts that we emphasize, since it drives up customer-concentration relative to firm-location concentration.

Figure 2: Customer-based concentration is lower concentration of firms in same location



Notes: The blue bars show HHI values using the "customer ZIP" as the unit of observation, where market shares of firms are computed using all transactions made by consumers who live in a given ZIP. The orange bars show HHI values based on the "firm ZIP" definition, where market shares are computed using all spending at establishments located in a given ZIP (note that this excludes online spending from the firm-zip measure, since online spending has no associated firm location. We discuss the role of online spending in Section 5.3). We arrive at these values by first computing unweighted ZIP-by-industry averages of annual HHI values, then aggregate across these ZIP-by-industry pairs with a spending-weighted average.

the sensitivity of these results to this arbitrary choice.

5 Customer Concentration Facts

5.1 Fact 1: Customer Concentration is Lower than Firm Concentration

Our first finding is that customer-based concentration is much smaller than the concentration of firms within the same zip code. Figure 2 shows that the typical category has a herfindahl of 0.17 when measured using all customers in a given 5-digit zip code vs. 0.40 when looking only at firms within that 5-digit zip code. Generally, spending in services is less concentrated, with an average customer-zip HHI of 0.03 compared to a value of 0.19 for retail.[23]

This discrepancy between customer- and firm-location based concentration arises because house-

[23]This comparison weights sectors by their spending. Since restaurants are the highest spending sector and the lowest customer concentration, they drive down the average concentration for the service sector. However, these comparisons also hold when weighting every zip code and industry equally in which case the customer HHI for retail is 0.16 and for services is 0.09.

holds shop at firms (both in person and online) located beyond their home zip code boundaries. Indeed, we find that in the average category and zip code, only 30% of the spending is within the zip code. To provide another sense of these magnitudes, Appendix Figure A5 shows that in the average category, customers in a given zip shop at 253 distinct firms. In comparison, each 5-digit zip code only has 16 distinct firms.

Since 5-digit zip codes are a small geographic unit, this is perhaps unsurprising, but the exact magnitude of the difference is striking. We focus primarily on 5-digit zip codes since using small geographic units helps to ensure that customers in these locations actually face the same shopping options.[24] However, Appendix Figure A4 shows that we still find large differences between firm and customer concentration for 3-digit zips (i.e. when we compare the concentration of spending of all customers residing in a given 3-digit zip code to the firms located in that 3-digit zip) and even non-trivial differences when comparing within states.[25] As another point of comparison, the difference in magnitude between customer and firm concentration at the state level is similar in magnitude to the long run increase in concentration in Autor et al. (2020), and the difference at the zip code level is much larger.

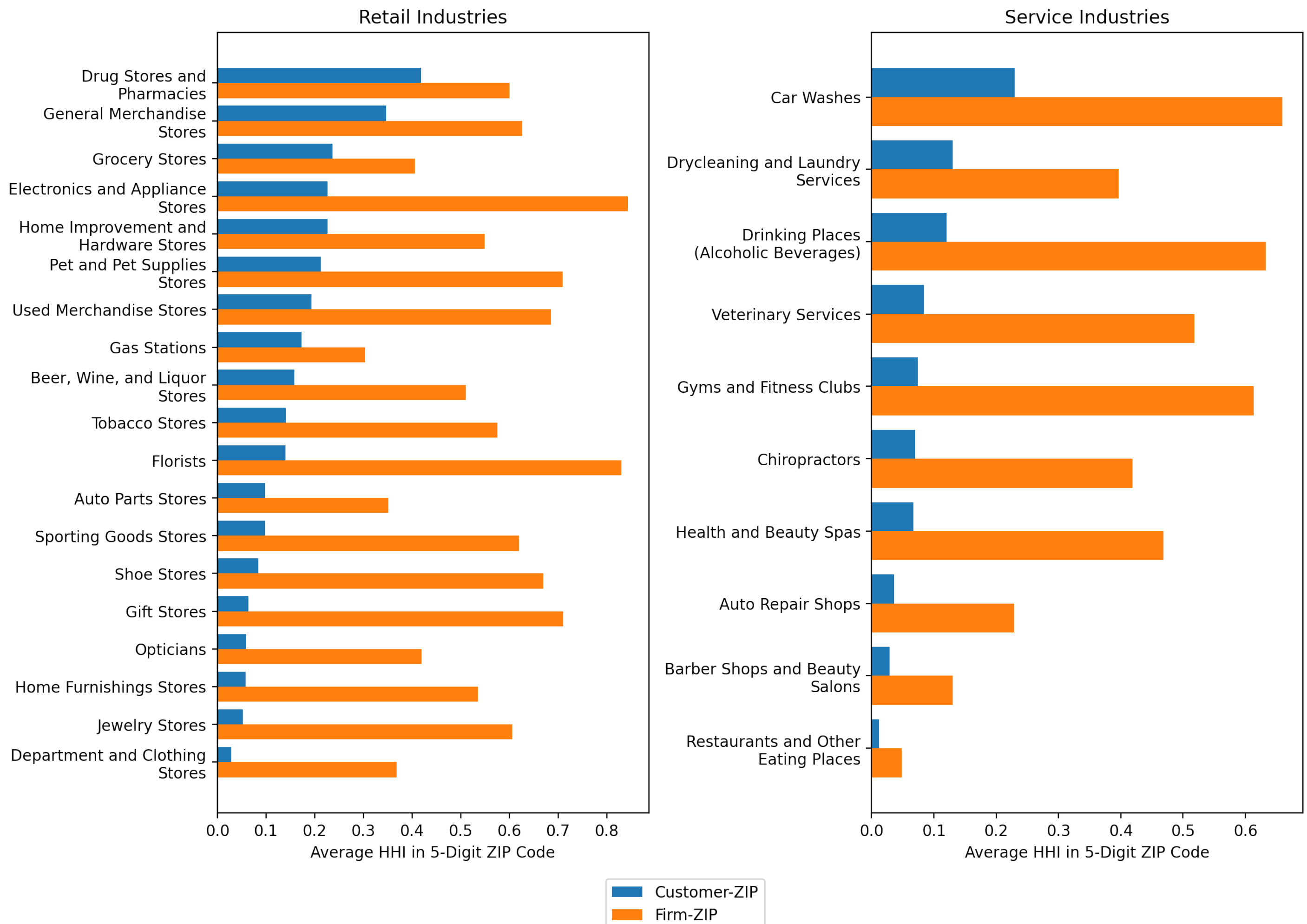
Figure 3 breaks down this aggregate number and shows the estimates for each of the 29 categories that we consider. While customer-based concentration is lower than firm-based concentration in every category, both the levels of customer-based concentration and the difference between customer-based and firm-based concentration vary substantially across categories. What predicts these cross-category patterns? Appendix Figures A6-A7 show several cross-category patterns. Unsurprisingly, the difference between customer-and firm concentration depends on the share of spending which is done on that category within a zip code. This is itself correlated with the frequency of purchases, since shopping at greater geographic distances is easier for products which are purchased infrequently.[26] This means that the largest gaps occur in categories with infrequent purchases, like appliances, and the smallest gaps occur in categories with frequent purchases, like grocery stores. For retail categories, we also find that average customer-location based HHI is highly correlated with national HHI, while we find a much

[24]One could alternatively ask what firm-geography generates the same concentration as 5-digit customer concentration. However, as previously discussed, this cannot be done without customer shopping data, and even choosing borders with the same concentration does not ensure that the actual firms inside these specific borders actually overlap. These two concerns limit the practical benefit of this approach.

[25]3-digit zip codes are roughly similar in size to the median county.

[26]This pattern is consistent with the result in Batch et al. (2023) that "shopping zones" are geographically larger for products which are purchased infrequently.

Figure 3: Customer-based concentration varies by category of spending



Notes: This figure shows that in an average ZIP code, firm-ZIP concentration is higher than customer-ZIP concentration in all the sectors we consider, and concentration is generally higher in retail industries relative to services. These HHI values are the category-level observations that were used to produce Figure 2: they are computed using a spending-weighted average across ZIP codes within each industry.

smaller correlation for service sectors. However, we now show that these category averages mask substantial heterogeneity across space.

5.2 Fact 2: Less Spatial Variation in Customer- than Firm- Concentration

Our second finding is that customer-based concentration varies much less over space than firm-based concentration. Thus far, we have focused on variation in customer concentration across industries. We next turn to an analysis of heterogeneity in concentration across space. This is particularly relevant for assessing how much shopping patterns vary across different markets with different distributions of local firms. The left panel of Figure 4 first describes the raw distribution of customer-based HHIs across 5-digit zip codes, averaged across all categories of spending.[27] As a comparison, we show the distribution of the HHIs of firms in a given 5-digit zip code. We see that there is substantially less variation in customer-based concentration across locations than there is in the distribution of firms across space.

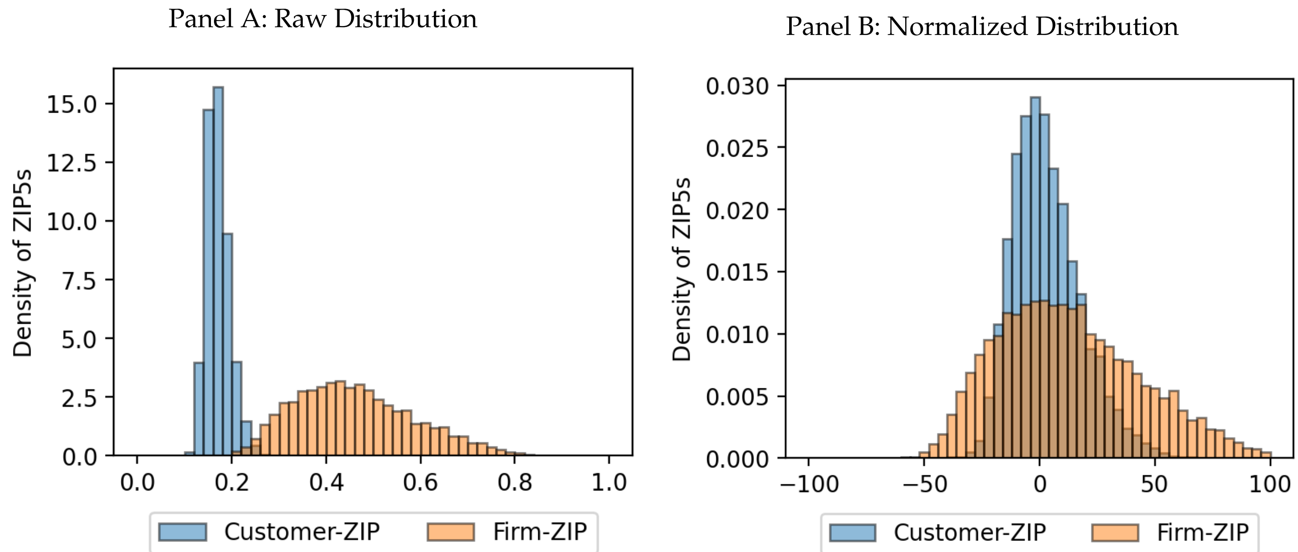
The right panel of Figure 4 shows that the greater spatial variation in firm concentration is not driven just by the fact that the average level of firm-based concentration is higher. Firm concentration also exhibits greater percentage deviations from the mean and not just larger absolute deviations across space. Appendix Figure A9 shows that these patterns hold both within cities across neighborhoods and across cities – specifically, customer-based concentration is less dispersed than firm-based concentration across 5-digit zip codes within a 3-digit zip-code (i.e. across 5-digit zip codes within greater Chicago area (606)) and across 3-digit zip codes (i.e. Chicago area (606) to Houston area (770)). Moreover, roughly half of the variation in customer concentration arises within zip3s and half across zip3s.

Together, facts 1 and 2 imply that market shares based on the locations of firms are a poor proxy for market shares based on where customers actually shop. Customers shop in wide geographic areas and often cross neighborhoods, cities and even states to take advantage of firms that are not located directly next to them.

While spatial variation in customer concentration is low compared to firm concentration, it is still substantial in its own right. It is thus interesting to describe how this variation correlates with characteristics of locations. Appendix Table A1 shows descriptive relationships between customer concentration

[27]Each observation is an individual zip code. These zip code observations average that zip codes separate category level concentration values using national spending weights on each category. This means that variation over space is driven entirely by differences in concentration over space within categories rather than reflecting different local weights on particular categories. However, using local spending weights produces similar results.

Figure 4: Uniformity of customer-location concentration across space



Notes: Panel A shows the distributions of customer-ZIP and firm-ZIP HHI across ZIP5s, while Panel B shows percentage deviations of the two measures from their respective spending-weighted averages. Concentration values at ZIP level are computed using national spending weights for our 29 industries, in order to remove variation that is solely due to differing industry composition across locations. This figure shows that firm-ZIP HHI has significantly more variation across space than customer-ZIP HHI, in addition to having a higher mean.

and various observable zip code characteristics and Appendix Table A2 shows patterns for firm concentration. We show bi-variate correlations in columns 1 through 9 and conditional correlations, including all characteristics jointly, in column 10. We note that all of these patterns are merely descriptive since we do not have exogenous variation in this covariates and many are correlated with each other. All explanatory variables are standardized so the coefficients can be interpreted as one standard deviation effects.

Focusing on column 10, shows that customers residing in zip codes with a 1 standard deviation higher firm HHI have customer-based HHIs that are 1.1 percentage point (6 percent) higher. Thus, while having concentrated local options is associated with more concentrated spending by local customers, the magnitude is relatively weak and column 9 shows that firm-zip HHI on its own explains only a modest share of overall variation in customer-based HHIs, with an R-squared of 0.17. Thus there is substantial variation in customer-based HHIs even for customers that face similar distributions of local firms. These means that even though households living in some locations face a much more concentrated set of nearby firms, this does not imply that these households have much more concentrated spending.

Appendix Table A1 also shows that there is a negative relationship between log(population density) and customer HHI, although this effect is also fairly modest. Other relationships are even weaker: customer HHI is very slightly lower in areas with higher home prices, more educated and higher income residents but the marginal effects and predictive power are low.[28]

5.3 Fact 3: Online Shopping Lowers Concentration and Reduces Variation Over Space

Another particularly important pattern that affects customer-location based concentration is online shopping. It is not obvious whether online shopping should increase or decrease the concentration of spending, since this depends on whether online shopping is more or less dominated by big firms than local shopping is. To the best of our knowledge, ours is the first paper to look at the relationship between online shopping and the concentration of spending: we find that online spending is strongly associated with a decline in concentration. Figure 5 shows that including online spending substantially lowers the HHI of products in the retail sector, bringing the average HHI from 0.2 to 0.17 and lowering the concentration of spending by upwards of 50% in sectors such as Electronics or Department stores.[29]

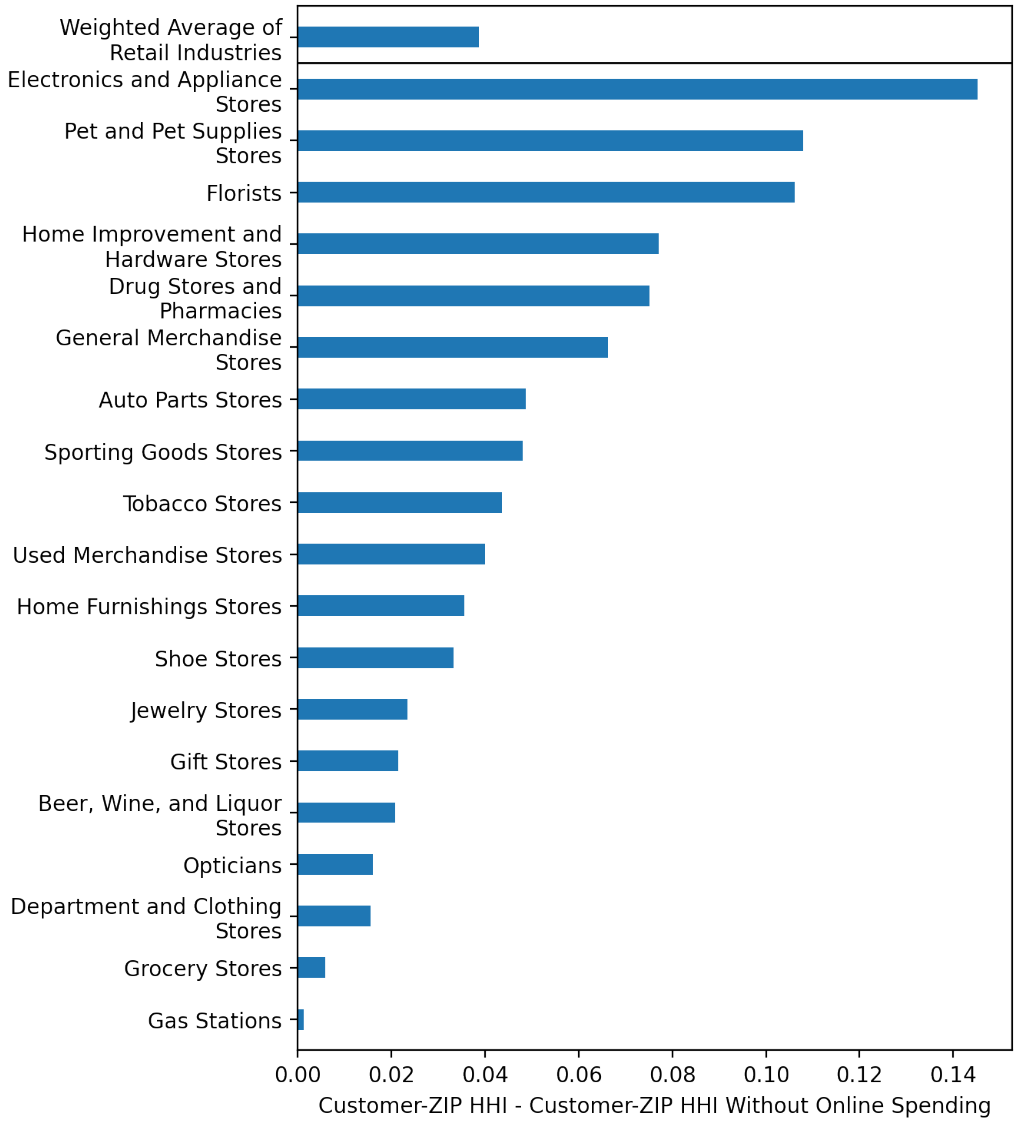
Online shopping is most important for lowering concentration for certain sectors, like electronics, where the online spending share is high. Customers in many locations might have access to only a single electronics store that they can shop at in-person (e.g. Best Buy), but when they shop online, they can purchase from both other big stores (e.g. Apple or Walmart) or a wide array of smaller online sellers. Since customers in this sector tend to access online stores that are different than the ones that they have access to in-person, we find that online shopping leads to a substantial decline in concentration. Conversely, online spending barely changes concentration for grocery stores, since most online spending in this category is at stores that already have a local presence.

Figure 6 shows that online shopping also reduces the variation across space in customer-based concentration. For each zip code we calculate the concentration of spending when including both in-person and online shopping and when including only in-person shopping. Figure 6 shows the variation across zip codes in these two different concentration measures. The presence of online shopping reduces spatial

[28]Interestingly, Appendix Table A2 shows that median household income is more predictive of firm-based concentration in a zip code, with higher income areas having more concentrated firms. However, this pattern does not result in those customers having more concentrated shopping.

[29]We focus on the retail sector only for this analysis. Although card transactions for some services, like gym memberships, can be processed online, these products are typically actually consumed in person and so true online spending shares are low.

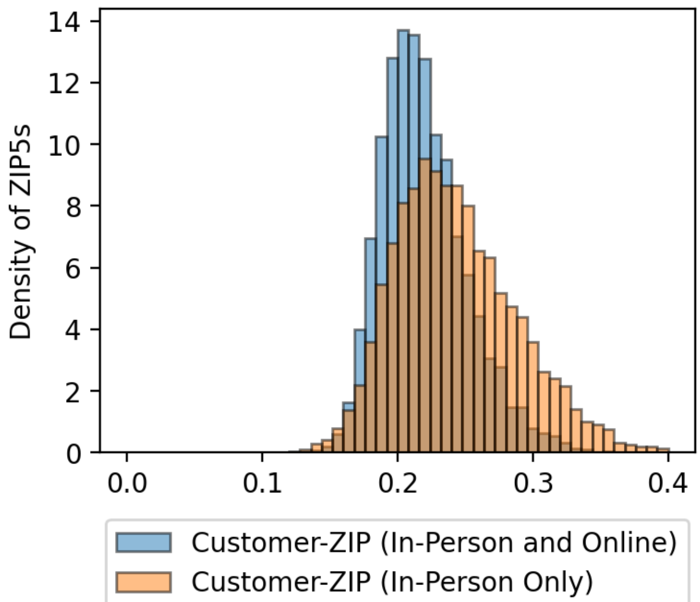
Figure 5: Online Shopping Lowers Customer-based concentration in all spending categories



Notes: This figure shows the difference between customer-ZIP HHI computed using only in-person transactions and customer-ZIP HHI using both in-person and online spending. Accounting for online shopping lowers concentration in all retail industries, and unsurprisingly, online shopping has the smallest impact for gas stations, where almost all transactions are in-person.

variation. This effect is particularly important for the right tail of zip codes: zip codes with the highest in-person concentration tend to be those where online shopping lowers concentration the most. This suggests that online shopping provides an important set of options in locations which might otherwise face little competition.

Figure 6: Online Spending Reduces Spatial Variation in Concentration



Notes: This figure shows the shows the distributions of customer-ZIP concentration across different zip codes when including or excluding online transactions. Concentration values at ZIP level are computed using national spending weights for all retail categories, in order to remove variation that is solely due to differing industry composition across locations. Service spending is excluded since these transactions are not online.

Finally, it is interesting to note that online spending is not uniformly less concentrated than in-person spending – indeed, online spending in categories like General Merchandise stores (which includes online behemoths like Amazon) is substantially more concentrated than in-person spending in the same category (See Appendix Figure A8).[30] However, even in sectors in which online spending is very concentrated, the addition of large online-only firms always lowers overall concentration. Thus, even though online shopping is itself often highly concentrated, it tends to be concentrated in a different set of firms than local spending.

[30]It is worth noting that for certain firms like Amazon which operate in part as online marketplaces, effective concentration may itself be overstated since we cannot observe the separate sellers in these market places.

5.4 Fact 4: Individual Spending Patterns Correlate with Customer but Not Firm Based Options

The previous section used the shopping patterns of those customers facing similar shopping constraints to capture the set of potential shopping options available to customers in a given location, as measured by the overall HHI of spending added across all customers in the zip code. In this section, we compare these potential options to the actual options chosen by individual customers. We find that individual customers' spending tends to be highly concentrated, but that individual concentration declines with the number of potential shopping options, implying that these measures of potential shopping options indeed have predictive power for individual shopping patterns.

Specifically, for each category and customer i, we measure the share of that customer's spending at any given firm j as

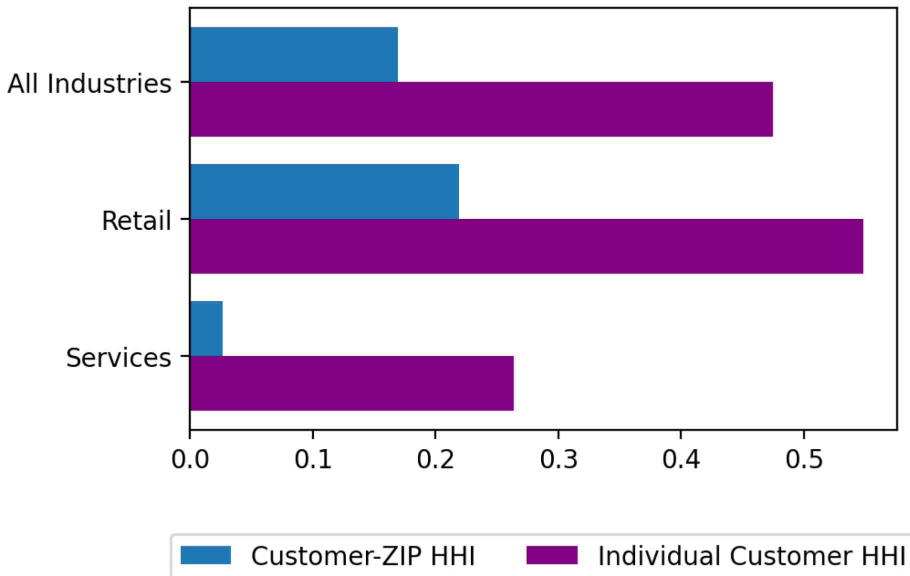
$$m_{j(i)} = s_{j(i)} / \sum_{j(i)} (s_{j(i)})$$

where $j(i)$ is the set of firms that customer i shops at in any given year. We then use these market shares to calculate an individual HHI of spending for each customer to get a measure of how concentrated spending is for specific customers (HHI_i). This individual customer HHI captures the extent to which specific customers spread their spending across different firms. In this sense, HHI_i can also be interpreted as the realized loyalty of a particular customer for specific firms: HHI_i will be high when a customer concentrates their purchases in a small number of firms while it will be low if they spread their purchases across a large number of firms.[31] Importantly, we restrict the analysis of HHI_i to customers with at least 10 transactions within a given category of spending because customers with a tiny number of transactions mechanically have high individual concentration, independent of any preferences or loyalty for particular firms.

The purple bars in Figure 7 show the average individual customer HHI across our sample, weighting each card i by its share of spending in that category. Perhaps unsurprisingly, spending on individual cards is far more concentrated across firms than the spending of all cards within a zip code (i.e. the

[31] Of course, realized loyalty as measured by HHI_i may be driven by forces beyond preferences. For example, if there were only a single option for purchasing in a category, then anyone who shopped in that category would have a high HHI_i even if they disliked this option and would prefer to shop somewhere else.

Figure 7: Individual customer spending is highly concentrated



Notes: This figure shows that spending at the level of a typical card is much more concentrated than the combined spending of all customers living in a typical ZIP. The individual customer HHI measure is computed in each year using card-by-category observations with at least 10 transactions. We take averages within customer ZIP-by-category cells by weighting each card by its total spending, then use another spending-weighted average to aggregate across ZIP-by-category pairs. See Appendix Figure A10 for the disaggregation by detailed category.

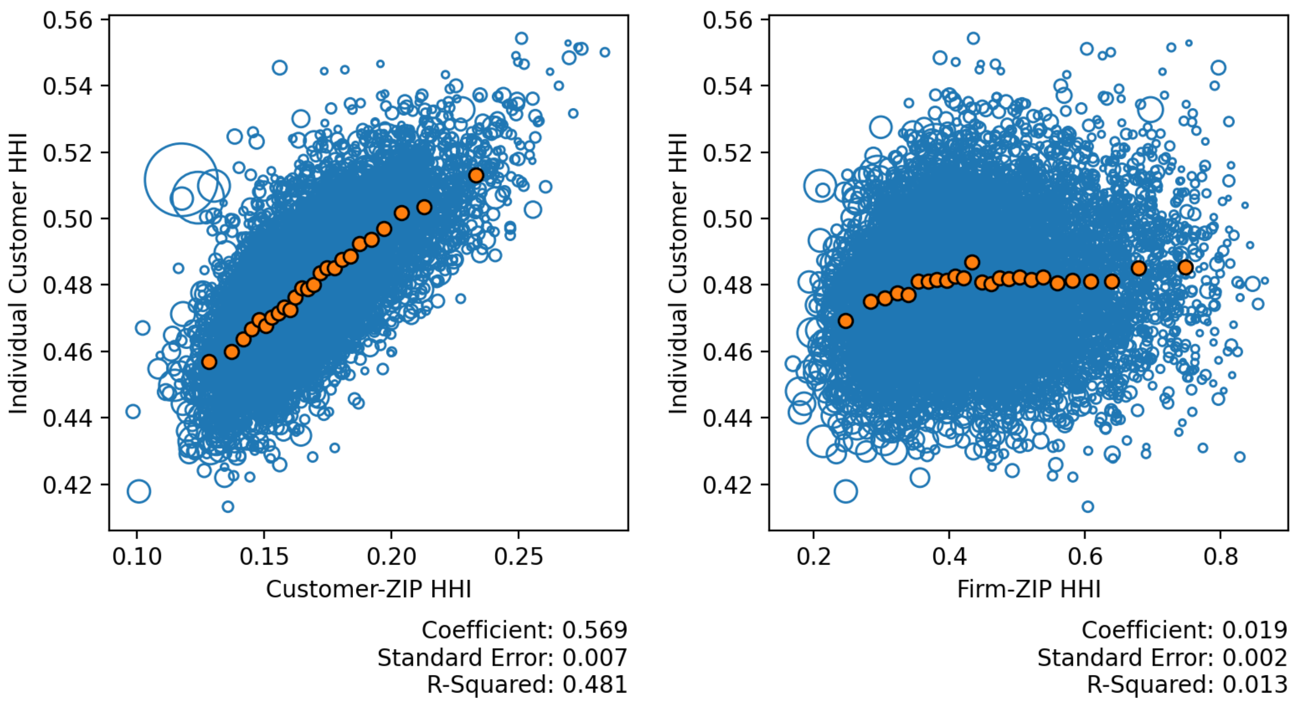
customer-zip HHI). The differences are substantial – within the retail sector, the average customer shops at 6 stores within a category within a year while all customers within a zip code shop at an average of 84 stores. Moreover, Appendix Figure A10 shows that individual customer HHI is very high (above 0.7) in 14 of 29 sectors, with by far the lowest individual customer HHI in restaurants and the highest in select service categories like chiropractors and dry cleaning and laundry.

We previously argued that the HHI of spending among all customers (i.e. customer-zip HHI) can proxy for the set of options available to customers in a particular location. Since the individual customer HHI captures the extent to which customers spread their spending across different firms, it is thus interesting to explore the relationship between individual customer HHI and customer-zip HHI. The relationship between individual customer HHI and the customer-zip HHI is not mechanical – if customers within a zip code have very heterogeneous preferences for firms within a category, it is possible that spending of all customers in a zip code is not very concentrated while each individual customer shops in only one firm and has a card-level HHI of 1. However, the left panel of Figure 8 shows that in the data,

Figure 8: Predictors of Individual Customer HHI

Panel A: Correlation with Customer-zip HHI

Panel B: Correlation with Firm-zip HHI



Notes: Panel A shows the relationship between average individual customer HHI in a ZIP (on the y-axis) and customer-ZIP HHI (on the x-axis), while Panel B has firm-ZIP HHI on the x-axis. Individual customer HHI is positively correlated with customer-ZIP HHI, but has barely any relationship with firm-ZIP HHI. The blue dots represent individual ZIP codes, and the size of each dot is related to total customer-location spending in that ZIP, while the orange dots depict the corresponding binscatter for each scatterplot.

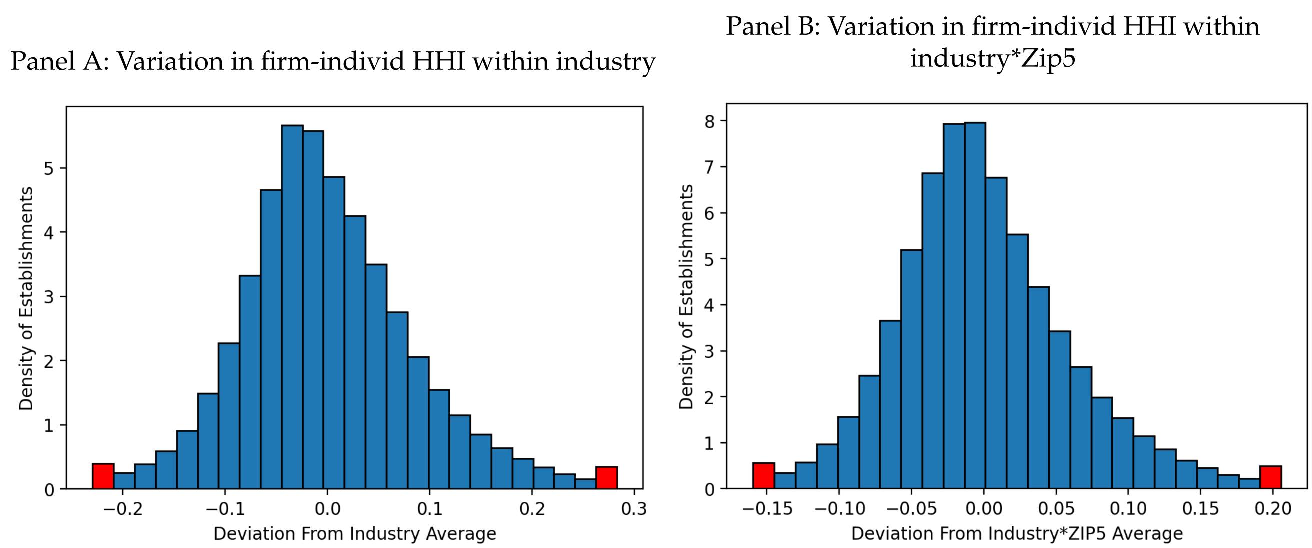
these two measures are highly correlated. Each blue dot in the scatterplot shows the average individual customer HHI for a zip code in a given category against the customer-based concentration for that zip code and category. The orange dots summarize the relationship between the two variables by showing the various percentiles of the distributions. We clearly see that customers living in zip codes with many potential options, as measured by the HHI of spending of all of the customers living in that zip code, themselves shop in more firms in a given year.

This contrasts with the right panel of Figure 8, which shows that individual customer spending concentration is poorly predicted by the concentration of firms located within the customer's own zip code. Together, these patterns suggest that customer-zip HHI measures capture not only the number of potential options, but also predicts the extent to which individual customers are shopping at many stores within a category. Moreover, Appendix Figure A11 shows that individual customer spending is more persistent across years in categories of spending with higher individual customer HHI and higher customer-zip HHI. This correlation further suggests that these measures are capturing the extent to which customers within that area are switching across firms over time.

5.5 Fact 5: Firms Vary in their Customers' Concentration and this Concentration Increases with Firm Size

In Section 3 we showed that firms located in the same place sell to customers at different distances. This motivated our study of shopping patterns based on customer location rather than firm location, and Section 5 documented a number of facts about spending patterns of individual customers or groups of customers living in particular zip codes. In this section we return to an analysis of firm heterogeneity and link this to our analysis of customer shopping patterns. In particular, we compute how much firms vary in their exposure to customers with different individual HHI and and how this relates to those customers' access to different shopping options (as measured by their customer-zip HHI).

Figure 9: Firms Vary in the Concentration of their Customers



Notes: Panel A shows deviations in firms' "average individual customer HHI" from their corresponding industry averages, and Panel B shows deviations from the average in an industry*ZIP. We restrict to firms with at least 50% of their sales going to customers living in the locations in our final analysis sample, and we also require that firms have at least $50,000 in annual sales to those customers. Both panels only show industry*Zip pairs with at least 2 firms.

Concretely, every customer in a particular category has a well-defined individual-customer HHI as well as a customer-zip HHI. For each firm j, we average the individual-customer HHI and the customer-zip HHI of the i customers that shop at firm j to construct a firm-specific measure of exposure to customers with different individual concentration (henceforth firm-indiv HHI) and to customers with different customer-zip HHI (henceforth firm-custzip HHI). Figure 9 shows that there is substantial variation within industry (Panel A) and even within industry*Zip5 (Panel B) in firm-indiv HHI.[32] That is,some firms have customers with very concentrated individual spending while other firms have customers that spread their individual spending across many firms.

Since this variation in firm-indiv HHI holds even across firms in the same places, it is unlikely to be driven by variation in the shopping options available to potential customers. Instead, this firm heterogeneity most likely reflects sorting of customers with different shopping preferences into different firms. To more formally assess this, we regress firm-indiv HHI on firm-custzip HHI first with only industry fixed effects and then again with industry*Zip5 fixed effects. The first regression produces a within R^2 of 0.106 while this falls to 0.003 when including industry*Zip5 fixed effects. This implies that firm-custzip HHI explains roughly 11% of the overall variation in firm-indiv HHI within industry. However, it only explains 0.3% of the variation in firm-indiv HHI within industry and location.[33] Since firms operating in the same zipcode generally draw customers with access to the same set of potential shopping options, it is not surprising that the explanatory power of firm-custzip HHI with industry*Zip5 is low (both in absolute terms and compared to the explanatory power across zip codes within industry).

What explains this variation in firm's individual card HHI? Our transaction data contains little information about firm characteristics beyond their location and category, which are already controlled for in Figure A12. However, we can measure firm size since we observe sales in our data. Figure 10 shows that there is a strong positive relationship between firm sales and firm-indiv HHI within industryXzip.[34] Appendix Figure A15 shows that this positive relationship holds separately in most categories.[35]

This result means that large firms tend to have customers that shop at only a few firms while small firms tend to have customers that shop at many firms. We again emphasize that this relationship holds

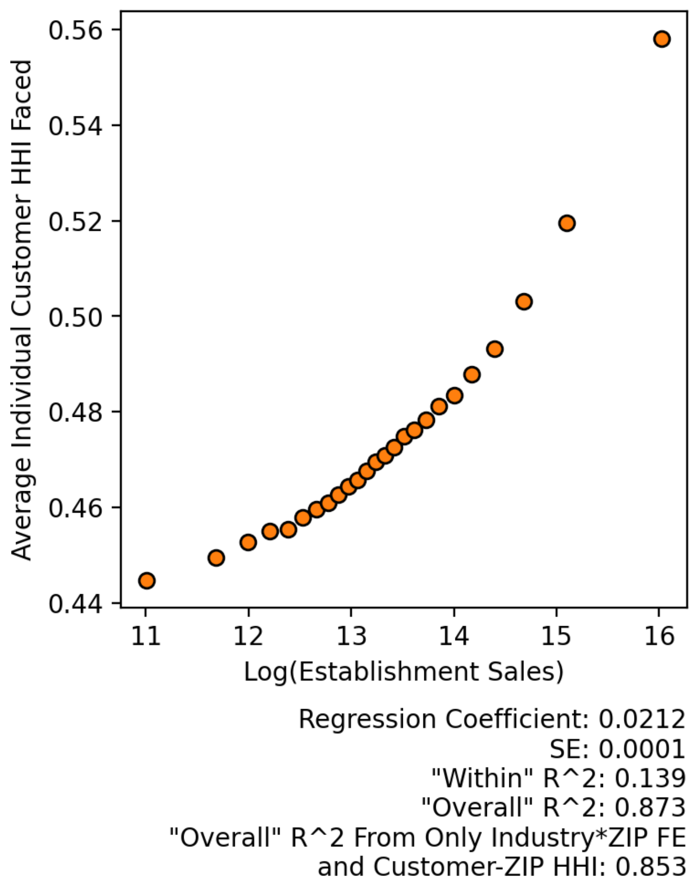
[32] Appendix Figure A12 shows this firm variation separately for each category. We exclude restaurants from all of the baseline figures in this section since the data has so many more restaurants than other firm counts and so would otherwise dominate all of the figures (which intentionally do not weight by sales). However Appendix Figure A13 shows these same figures when including restaurants.

[33] These values are 10% and 0.2% if we re-run the analysis including restaurants (see footnote 32).

[34] Similar relationships also hold if we look at relationships only within industry instead of industryXzip.

[35] Restaurants and drinking establishments are two important exceptions, although it is worth noting that these categories have much smaller variation in firm-indiv HHI in the first place.

Figure 10: Large Firms Have More Concentrated Customers



Notes: This figure shows a binscatter of firm-indiv HHI vs. log(sales), after including industry*Zip fixed effects and an additional control for the level of firm-custzip HHI. Most of the overall variance in firm-indiv HHI is explained by cross-industry*zip variation, but within a market, we can still see that larger firms tend to have customers whose spending is more concentrated.

within category and location and so it is not driven by heterogeneity in the shopping options available to local customers. It instead implies that even though small and large firms are both options, they attract different types of customers with different spending patterns. In particular, large firms tend to have customers that do most of their shopping in the large firm while small firms do not have customers with the same degree of concentrated purchases. This greater concentration for large firms could in turn arise through many different channels, which would be interesting to explore in future work. For example, large firms might become large and attract a more concentrated set of customers by having lower prices or higher quality at a given price. Alternatively these firms might have different product assortment or more effective advertising.[36] Alternatively, large firms might better lock-in customers in some way

[36] A larger assortment of products at large firms might allow customers to engage in one-stop shopping rather than having

even though other options are available. For example, some large firms like Costco and Amazon charge membership fees, which might in turn encourage a greater intensity of shopping. However, this specific mechanism seems unlikely to explain these patterns for most firms and categories in our data.

6 Conclusion

In this paper, we use data covering trillions of dollars of spending from the universe of transactions of a major U.S. credit/debit card processor to document new facts about the distribution of firm spending across individual customers. Since most data sets measure spending at the firm but not the identity of customers, a large existing literature looks at the distribution of spending based on the location of firms. However, we begin by using our customer shopping data to show that defining markets based on grouping together firms in particular locations is problematic: even within the same location and category, different firms draw customers and compete with firms at different distances. This means that any geographic market based on firm-location must either group together some firms which do not compete for customers or exclude some firms that do.

We instead propose an alternative definition of geography based on customer location that obviates these issues. After grouping together customers that live in a particular location, we can then measure the distribution of their spending across firms without having to take any stand on the geography of spending: customers can shop on some mix of nearby firms, faraway firms or any permutations in-between and we do not need to take any ex-ante stand on the relevant distance.

We document a number of new facts about customer-based spending concentration. Customer-location concentration is lower than firm-location concentration in the same areas and varies less over space. Online shopping combined with shopping beyond arbitrary fixed distances are both important forces in explaining this result. Customer-location concentration can capture information about the options available for customers to purchase even if they do not shop at particular stores. Indeed, we show that the concentration of individual customers' spending is correlated with customer-location concentration but not with firm-location concentration. Finally, we show that individual firms vary substantially in their exposure to customers with different concentration. Notably, large firms have customers with more concentrated spending than small firms, and this appears to be driven by different consumers with

to shop at many firms to get the products they need.

the same options sorting differently into large and small firms rather than by a lack of local shopping options.

Our results show that the geography of markets revealed by actual customer spending often diverges from that implied by firm sales data without any linked customer information. Our data thus simultaneously provides a reason for caution in interpreting local firm market shares and a promising alternative tool for understanding actual consumer substitution patterns. However, we again emphasize that our current analysis is purely descriptive and simply documents equilibrium spending patterns given the current options and prices available in the economy. To measure market power, we are ultimately more interested in how spending patterns change with prices or firm entry. For example, a customer with a high level of individual concentration in one firm might have a strong preference for that firm and thus a low elasticity of substitution, or they might have a high elasticity of substitution but that firm happens to simply have the lowest price. Observational data cannot distinguish these very different situations.

However, in ongoing work we hope to make progress on this question by looking at the relationship between firm entry and exit and changes in customer shopping patterns. While this work is too early to draw strong conclusions, preliminary evidence suggests that customer concentration patterns indeed have strong predictions for substitution patterns. In particular, preliminary evidence suggests that customers with high concentration in incumbent firms are less likely to later switch to new entrants than are customers with low concentration in these same incumbent firms.

References

Autor, David, Christina Patterson, and John Van Reenen. 2023. "Local and national concentration trends in jobs and sales: The role of structural transformation." National Bureau of Economic Research.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *The Quarterly Journal of Economics*, 135(2): 645–709.

Baker, Scott R., Brian Baugh, and Marco Sammon. 2023. "Customer Churn and Intangible Capital." *Journal of Political Economy Macroeconomics*, 1(3): 447–505.

Batch, Andrea, Benjamin Bridgman, Abe Dunn, and Mahsa Gholizadeh. 2023. "Consumption Zones."

Berry, Steven, Martin Gaynor, and Fiona Scott Morton. 2019. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives*, 33(3): 44–68.

Conlon, Christopher, and J Mortimer. 2022. "The Role of Diversion Ratios and Surveys in Horizontal Mergers."

Department of Justice and Federal Trade Commission. 2023. "Final 2023 Merger Guidelines."

Dolfen, Paul, Liran Einav, Peter J Klenow, Benjamin Klopack, Jonathan D Levin, Larry Levin, and Wayne Best. 2023. "Assessing the gains from e-commerce." *American Economic Journal: Macroeconomics*, 15(1): 342–370.

Klenow, Peter J, Jonathan D Levin, Raviv Murciano-Goroff, and Liran Einav. 2020. "Customers and retail growth." Working Paper.

Miller, N., S. Berry, F. Scott Morton, J. Baker, T. Bresnahan, M. Gaynor, R. Gilbert, G. Hay, G. Jin, B. Kobayashi, F. Lafontaine, J. Levinsohn, L. Marx, J. Mayo, A. Nevo, A. Pakes, N. Rose, D. Rubinfeld, S. Salop, and M. Wosinska. 2022. "On the Misuse of Regressions of Price on the HHI in Merger Review." *Journal of Antitrust Enforcement*, 10(2): 248–259. 2022b.

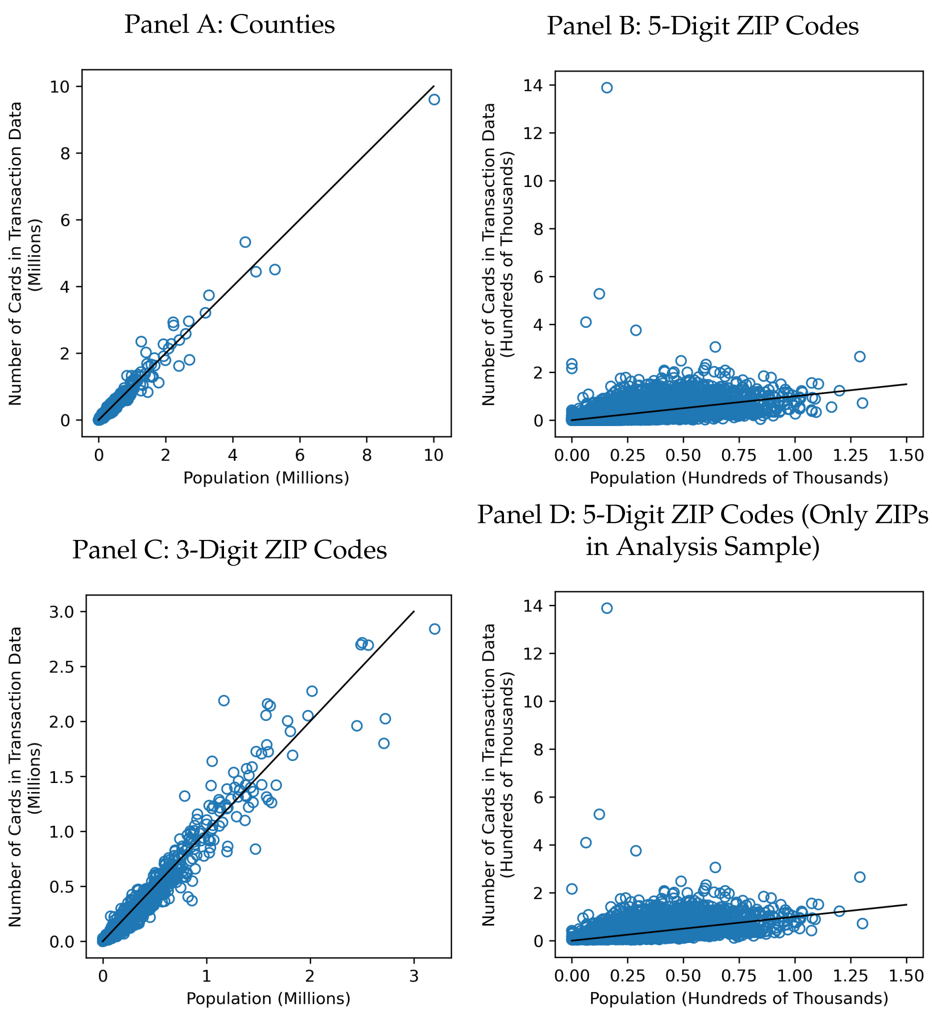
Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter. 2021. "Diverging trends in national and local concentration." *NBER Macroeconomics Annual*, 35(1): 115–150.

Smith, Dominic A, and Sergio Ocampo. 2022. "The evolution of US retail concentration." Working paper.

Syverson, Chad. 2019. "Macroeconomics and market power: Context, implications, and open questions." *Journal of Economic Perspectives*, 33(3): 23–43.

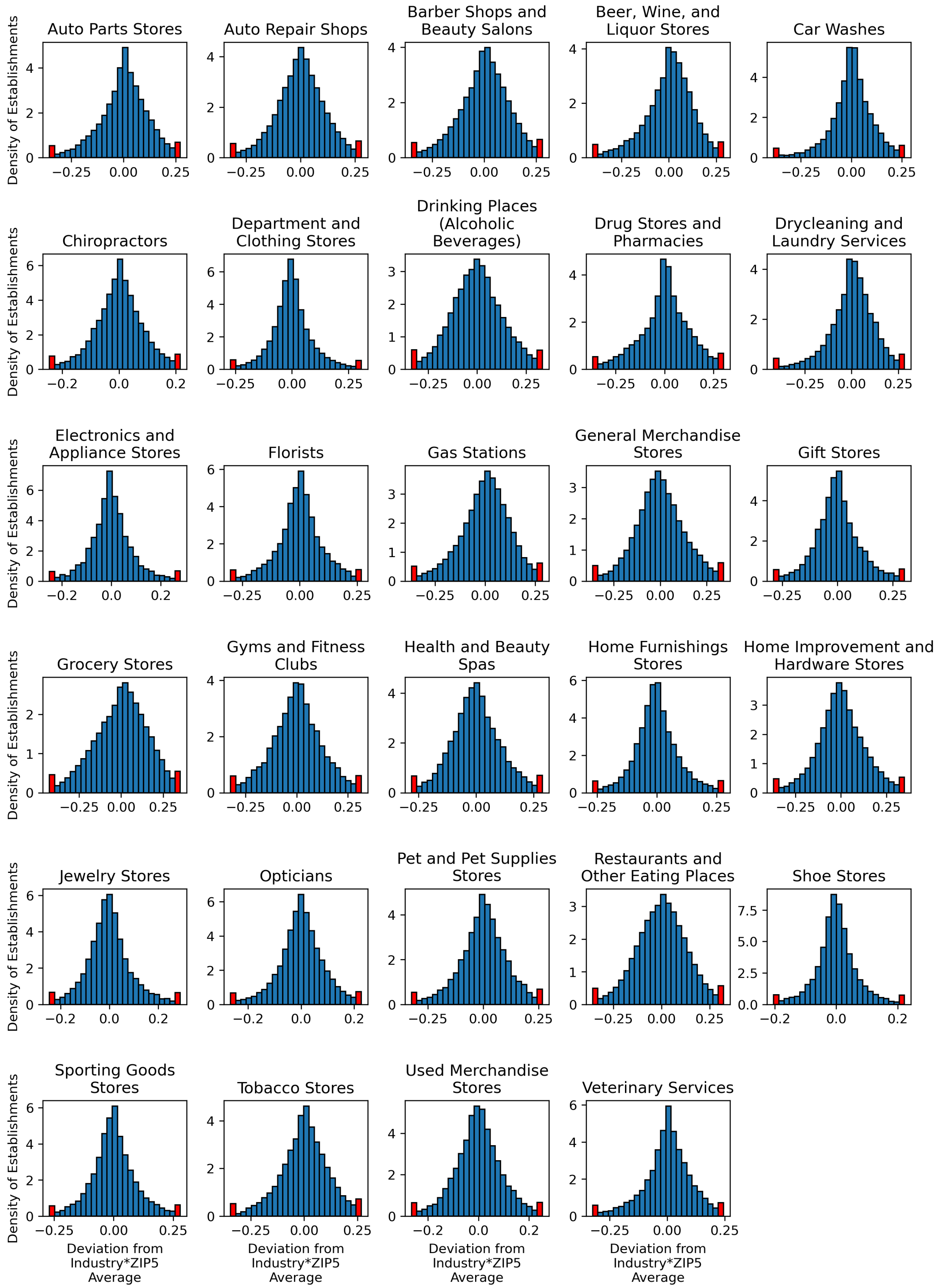
A1 Appendix

Figure A1: Number of cards assigned to home locations vs. Census population estimates



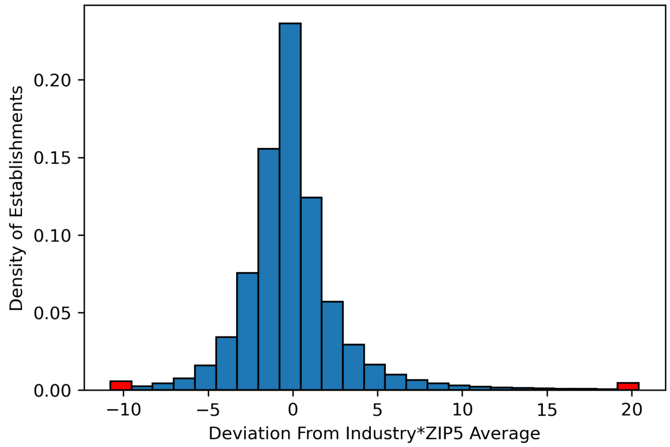
Notes: This figure shows that the geographic distribution of cards in our analysis sample is remarkably similar to the true distribution of population across counties. There is still a positive correlation when looking at the distribution across ZIP5s, but the relationship is much less precise. The y-axis of each plot shows the number of cards assigned to each geographic unit as their "home location" in 2021. The x-axis shows population data for 2021 from the American Community Survey. The 45-degree line is included on each plot for ease of interpretation.

Figure A2: Firms in the same location and category differ in their in-zip customer share



Notes: For each industry, this figure shows the distribution of firms' "in-ZIP customer share" within a ZIP code. To avoid cases where the in-ZIP share is mechanically zero, we only consider firms located in one of the final "customer ZIPs" in our analysis sample. We also screen out spending related to vacations or other travel by dropping transactions made more than 100 miles away from a customer's home ZIP.

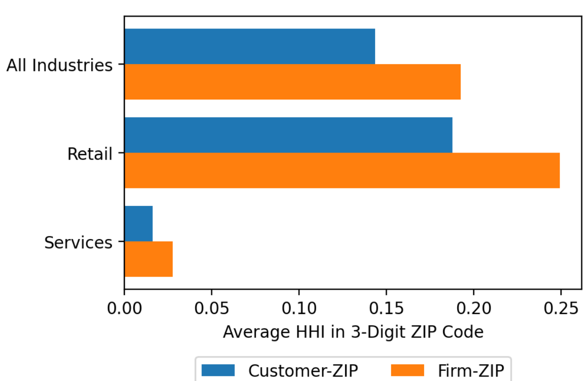
Figure A3: Firms in the same location and category sell to customers from zip codes with different average distance



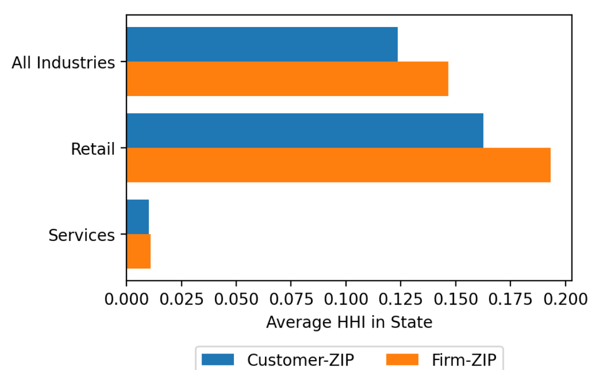
Notes: This figure shows the distribution of firms' "in-ZIP customer share" within an industry*Zip. To avoid cases where the in-ZIP share is mechanically zero, we only consider firms located in one of the final "customer ZIPs" in our analysis sample. We also screen out spending related to vacations or other travel by dropping transactions made more than 100 miles away from a customer's home ZIP.

Figure A4: Concentration for Different Geographic Aggregations

Panel A: ZIP3

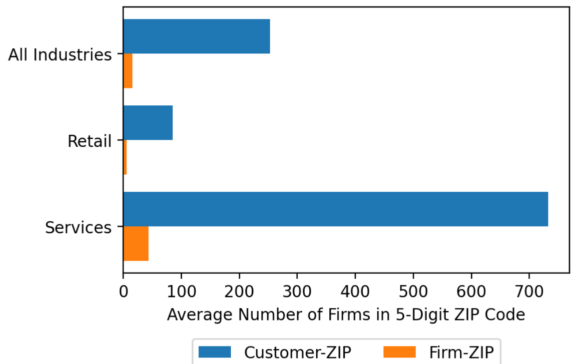


Panel B: State



Notes: Panel A shows the average customer-ZIP HHI in ZIP3-by-category cells, while Panel B shows average customer-ZIP HHI in state-by-category cells. Even in these more aggregated geographic units, there are still meaningful differences between customer-ZIP and firm-ZIP concentration.

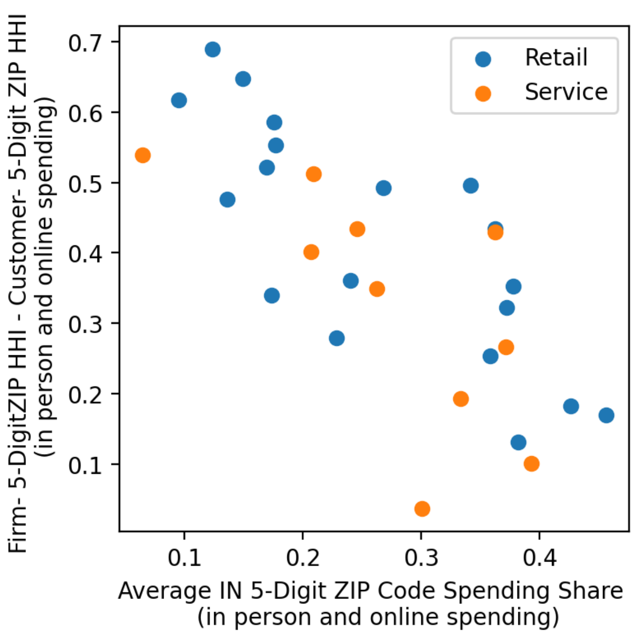
Figure A5: Number of firms available in a category within a ZIP



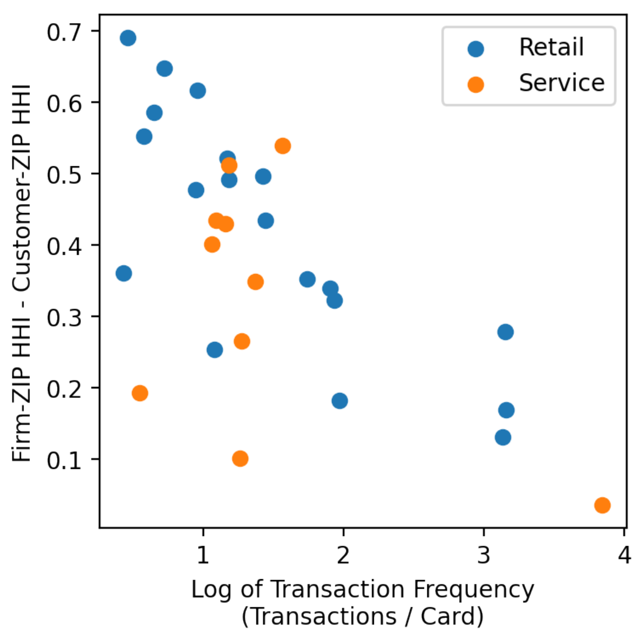
Notes: In a typical category, the number of firms visited by customers living in a ZIP code is far greater than the number of firms with establishments physically located in a ZIP code. For customer-ZIP firm counts, in order to eliminate noise from firms visited by only one or two customers due to travel or incorrectly imputed customer locations, we only count firms visited by at least 0.1 percent of all customers living in a ZIP in a given year. We compute the average number of firms visited in a category for an average ZIP using an unweighted average across ZIP-by-year pairs, then aggregate across industries using national spending weights.

Figure A6: Category-level correlates of customer-based concentration

Panel A: Within 5-digit Zip Shopping Share

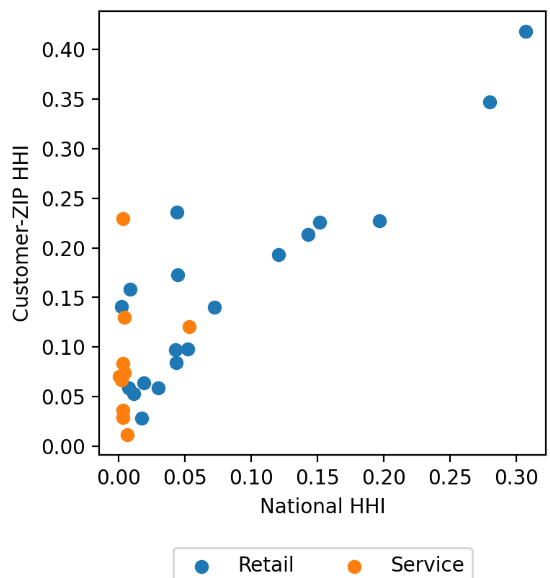


Panel B: Transaction Frequency



Notes: Panel A compares the difference between average firm-ZIP and average customer-ZIP HHIs in each industry (on the y-axis) to the average share of all spending that takes place at physical establishments within the customer's home ZIP code (on the x-axis). This panel indicates that the "wedge" between firm- and customer-ZIP concentration is higher for industries where less consumption is done locally. In Panel B, the y-axis shows the difference between average firm-ZIP HHI and average customer-ZIP HHI for each category, while the x-axis shows log(transactions per card), which is an unweighted average of transactions per card in ZIP-by-category cells. For retail industries, the difference in concentration between the customer-ZIP and firm-ZIP measures is greater in categories where consumers make purchases less frequently.

Figure A7: Customer-ZIP concentration is correlated with national concentration



Notes: This figure compares the average customer-ZIP HHI across ZIPs within an industry (on the y-axis) to the "national HHI" for that industry (on the x-axis), which is computed using all spending in our sample without grouping by a geographic unit. It shows that retail industries with higher national concentration also tend to have higher local concentration, but the relationship is weaker for services.

Table A1: Relationship between Zip Characteristics and Customer-Zip HHI

	Dependent variable: Customer-ZIP HHI									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	0.170***	0.170***	0.170***	0.170***	0.170***	0.170***	0.170***	0.170***	0.170***	0.170***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Median Age	0.001									-0.003***
	(0.001)									(0.001)
Share with College Degree		-0.005***								-0.001*
		(0.000)								(0.001)
Share of Black Residents			-0.002***							-0.002***
			(0.000)							(0.000)
Zillow Home Values Index				-0.005***						-0.001*
				(0.001)						(0.001)
Median Household Income					-0.003***					-0.005***
					(0.000)					(0.001)
Log(Population Density)						-0.010***				-0.007***
						(0.001)				(0.001)
Single-Family Home Share							0.005***			-0.001
							(0.001)			(0.001)
Homeownership Rate								0.005***		0.003***
								(0.001)		(0.001)
Firm-ZIP HHI									0.010***	0.011***
									(0.000)	(0.000)
SE Clustering	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3
Observations	6738	6738	6738	6738	6738	6738	6738	6738	6738	6738
R^2	0.001	0.036	0.006	0.039	0.014	0.153	0.045	0.048	0.173	0.362
Adjusted R^2	0.001	0.036	0.006	0.039	0.014	0.153	0.045	0.048	0.173	0.362

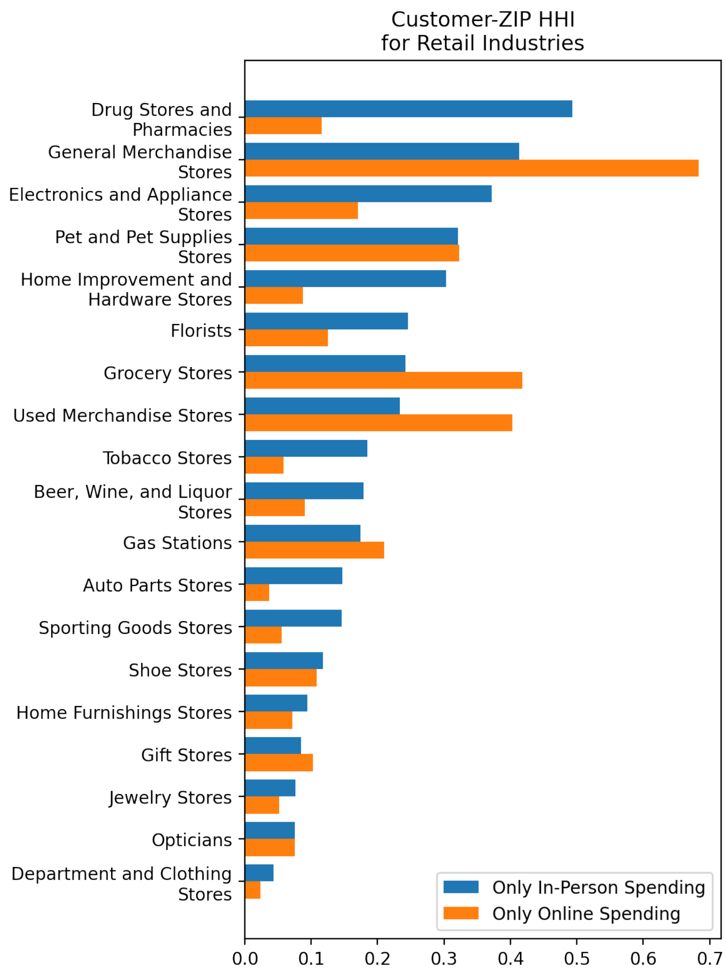
Notes: This table shows results from regressing customer-ZIP HHI in ZIP5s on several ZIP-level covariates. All independent variables have been standardized, so each coefficient can be interpreted as the effect of a one-standard-deviation increase in the corresponding independent variable. The two variables with meaningful predictive power in this regression are population density and firm-ZIP HHI – denser locations have lower customer-ZIP concentration, while ZIPs with high firm-ZIP HHI also tend to have high customer-ZIP HHI. All of the independent variables except the Zillow Home Values Index and firm-ZIP HHI come from the 2021 American Community Survey , and the home values index uses Zillow data from December 2022.

Table A2: Relationship between Zip Characteristics and Firm-Zip HHI

	Dependent variable: Firm-ZIP HHI								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.461***	0.461***	0.461***	0.461***	0.461***	0.461***	0.461***	0.461***	0.461***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
Median Age	0.020***								-0.000
	(0.002)								(0.002)
Share with College Degree		0.029***							0.007**
		(0.002)							(0.003)
Share of Black Residents			-0.019***						-0.001
			(0.002)						(0.002)
Zillow Home Values Index				0.020***					-0.005*
				(0.003)					(0.003)
Median Household Income					0.043***				0.035***
					(0.002)				(0.004)
Log(Population Density)						-0.009***			-0.011***
						(0.002)			(0.002)
Single-Family Home Share							0.021***		-0.016***
							(0.003)		(0.005)
Homeownership Rate								0.034***	0.027***
								(0.003)	(0.005)
SE Clustering	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3	ZIP3
Observations	6738	6738	6738	6738	6738	6738	6738	6738	6738
R^2	0.025	0.054	0.022	0.026	0.117	0.005	0.028	0.073	0.148
Adjusted R^2	0.025	0.054	0.022	0.026	0.117	0.005	0.028	0.073	0.147

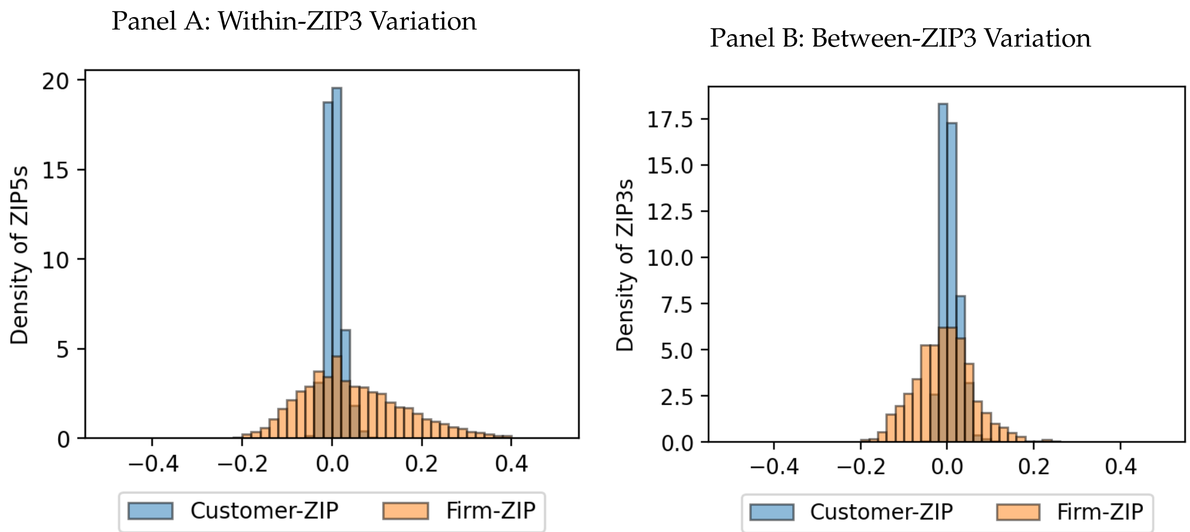
Notes: This table shows results from regressing firm-ZIP HHI in ZIP5s on several ZIP-level covariates. All independent variables have been standardized, so each coefficient can be interpreted as the effect of a one-standard-deviation increase in the corresponding independent variable. The ZIP-level variable most closely related to firm-ZIP HHI is median household income, with an R^2 of 0.12: on average, spending at establishments within a ZIP tends to be more concentrated in higher-income areas. All of the independent variables except the Zillow Home Values Index come from the 2021 American Community Survey, and the home values index uses Zillow data from December 2022.

Figure A8: Concentration of Online Spending by Category



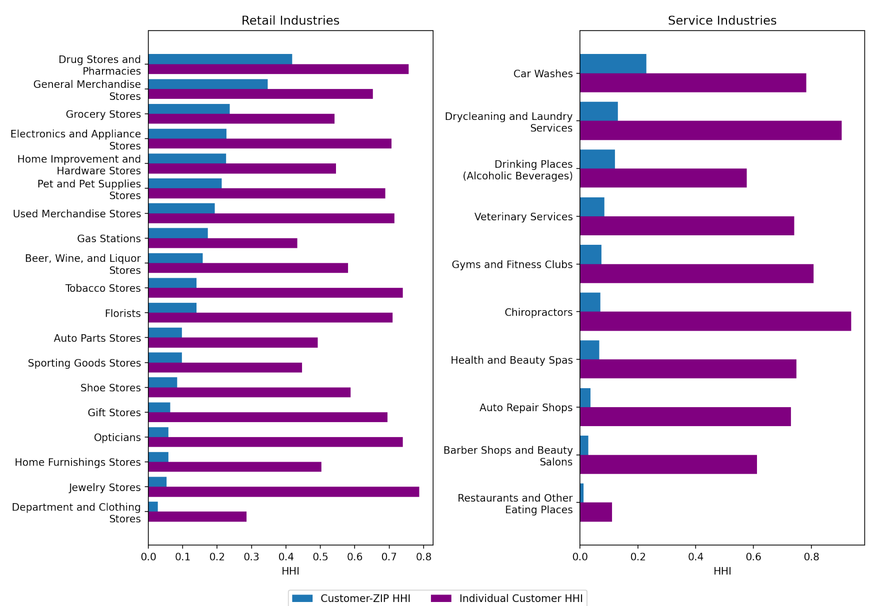
Notes: This figure compares customer-ZIP concentration measured using only in-person transactions (blue bars) to customer-ZIP concentration measured using only online spending (orange bars). We focus on retail industries for this comparison because services generally have very low online spending shares. Online-only concentration is not consistently higher or lower than in-person-only concentration.

Figure A9: Distribution of 5-digit zip concentration within and between region



Notes: Panel A plots the difference between HHI in a 5-digit ZIP and the average HHI in the corresponding 3-digit ZIP, which gives us a measure of within-region variation in concentration. Panel B plots the difference between average HHI in a ZIP3 and the overall average level of HHI across all ZIP5s. All averages are spending-weighted. We see that firm-ZIP HHI has more variation within ZIP3s than between ZIP3s, while customer-ZIP concentration has similar amounts of variation within and across regions.

Figure A10: Individual customer spending is highly concentrated in all sectors

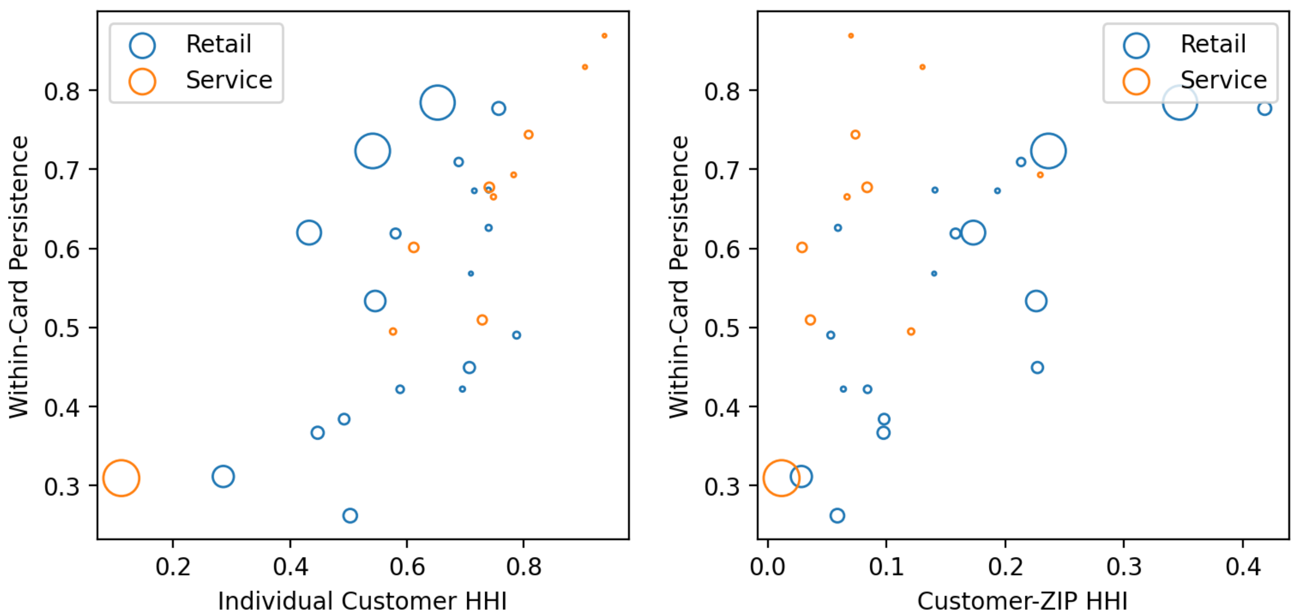


Notes: The blue bars show average customer-ZIP HHI by category as depicted in Figure 3, while the purple bars show average individual customer HHI by category, which is aggregated across ZIPs with the same spending weights used to produce the customer-ZIP HHI numbers. Across all categories, individual customer HHI is much higher than customer-ZIP HHI.

Figure A11: Concentration Measures and Spending Persistence

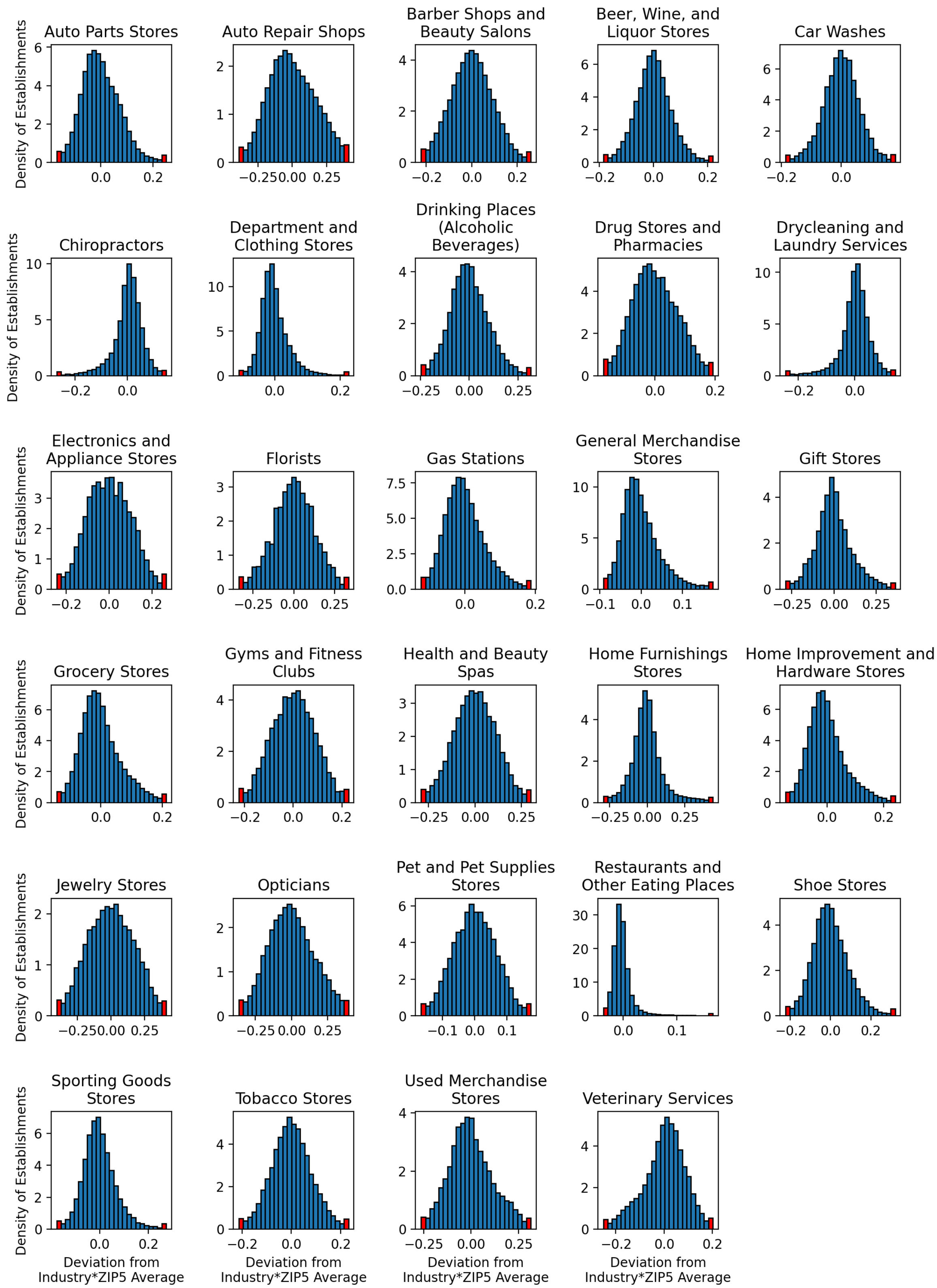
Panel A: Individual Customer HHI

Panel B: Customer-Zip HHI



Notes: The y-axes of Panels A and B plot a measure of persistence of spending for individual cards at industry level. Specifically, within each category, we take all cards with nonzero spending in years t and $t-1$ and construct an indicator for whether the firm that the card spent the most at in year t is the same as the firm they spent the most at in year $t-1$. We take a spending-weighted average of this measure across cards within a customer ZIP, in order to gauge the extent to which consumers change which firms they visit over time. We compare the persistence measure to individual customer HHI in Panel A and to customer-ZIP HHI in Panel B: consumers change their top firms less often (meaning higher persistence) both for industries where individual cards visit fewer firms on average and for industries where consumers in a ZIP have fewer options in general. The size of each dot is related to total spending in that industry.

Figure A12: Firms in the same location and category differ in their exposure to customers with different concentration

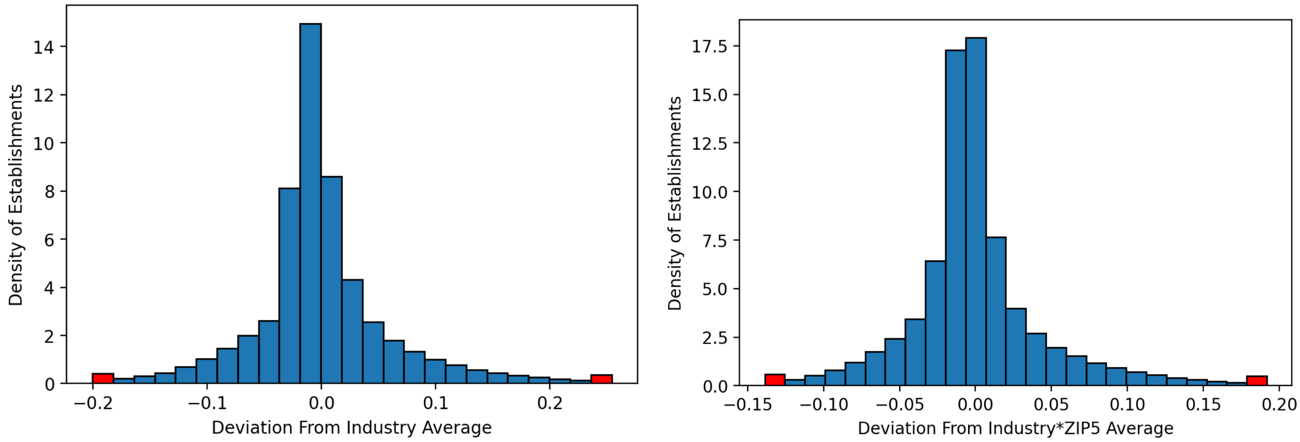


Notes: For each industry, this figure shows variation in firm-indiv HHI within a ZIP code. We restrict to firms with at least 50% of their sales going to customers living in the locations in our final analysis sample, and we also require that firms have at least $50,000 in annual sales to those customers. We only include industry*Zip pairs with at least 2 firms.

Figure A13: Firms Vary in the Concentration of their Customers: Robustness to Including Restaurants

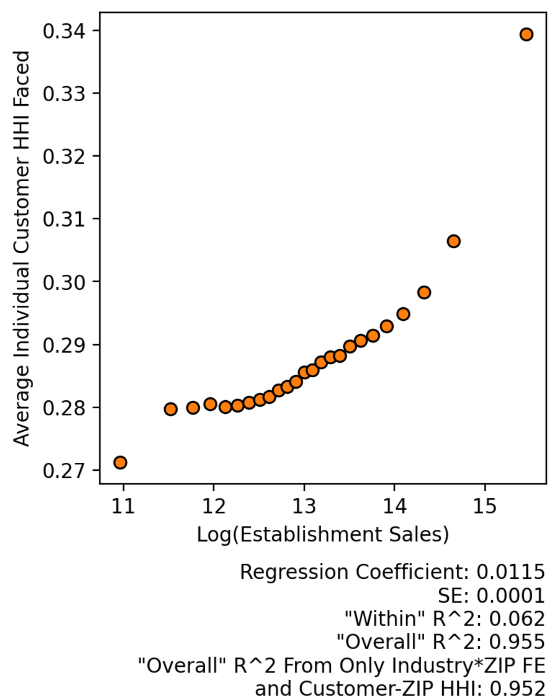
Panel A: Variation in firm-individ HHI within industry

Panel B: Variation in firm-individ HHI within industry*Zip5



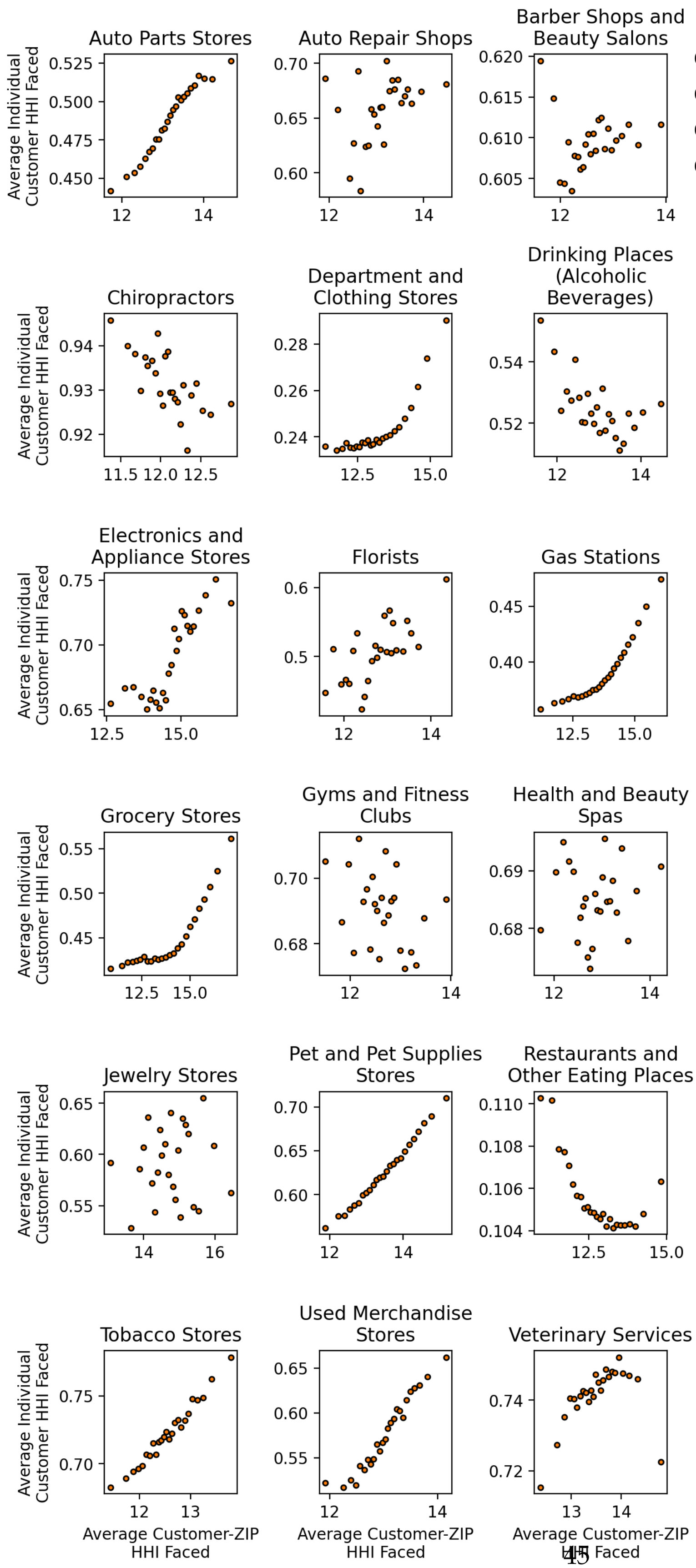
Notes: Notes: Panel A shows deviations in firms' "average individual customer HHI" from their corresponding industry averages, and Panel B shows deviations from the average in an industry*ZIP. We restrict to firms with at least 50% of their sales going to customers living in the locations in our final analysis sample, and we also require that firms have at least $50,000 in annual sales to those customers. Both panels only show industry*Zip pairs with at least 2 firms.

Figure A14: Large Firms Have More Concentrated Customers: Robustness to Including Restaurants



Notes: This figure shows a binscatter of firm-indiv HHI vs. log(sales), after including industry*Zip fixed effects and an additional control for the level of firm-custzip HHI. Most of the overall variance in firm-indiv HHI is explained by cross-industry*zip variation, but within a market, we can still see that larger firms tend to have customers whose spending is more concentrated

Figure A15: Concentration vs. Firm Size By Industry



Notes: For each industry, this figure shows a binscatter of firm-indiv HHI vs. log(firm sales), after including ZIP code fixed effects and an additional control for the level of firm-custzip HHI. Within most industries, larger firms tend to have customers whose spending is more concentrated.