Concentrating on Customers: A New Measure of Size-Based Market Power *

Alex Blumenfeld
UC Berkeley

Christina Patterson
Chicago Booth & NBER

Joseph Vavra
Chicago Booth & NBER

March 2026

Abstract

Does increasing concentration of sales at 'superstar' firms imply greater market power? In a variable-markup framework with heterogeneous customers, a firm's size-based market power depends not on its total market share but instead on its "effective" market share–how much of its customers' spending basket it captures. We use trillions of dollars of credit and debit-card transactions to measure effective shares for the near-universe of U.S. final retail and service firms. We find that firms' effective shares rise only modestly with traditional market shares, implying that the effective size distribution is markedly less skewed than suggested by past measurement. This attenuation arises in part because important sources of firm sales growth–like geographic expansion and increasing e-commerce–are associated primarily with expanding the number of customers rather than increasing within-customer dominance. This firm-level attenuation in turn attenuates trends in market-level concentration: while overall sales concentration has increased over the last decade, effective concentration has been much more stable. Our findings suggest that rising concentration of overall sales has not been associated with commensurate increases in customer dominance and size-based market power.

*We thank Violet Hamlin, Zifan Huang and Luisa Pomarici for excellent research assistance, and seminar participants at UIC, UVA, the Empirical Macro Workshop-Phoenix and the Federal Reserve Bank of Philadelphia for helpful comments.

1 Introduction

A growing literature documents rising sales concentration at "superstar" firms and debates whether this has increased market power. Much of this debate—especially in macro and trade—implicitly builds on representative household variable-markup frameworks like Atkeson and Burstein (2008), in which overall market share is a sufficient statistic for pricing power. Our starting point is simple: with heterogeneous customers, the relevant notion of firm "size" in these models is not a firm's share of total revenue, but how dominant it is in the spending baskets of the specific customers it serves. We show this distinction is empirically first-order. When we measure size using a customer-based statistic consistent with these models, we find no evidence of a broad-based rise in size-based market power.[1]

Using transaction-level credit card data covering the near universe of firms in a broad set of consumer-facing retail and service sectors, we construct what we term "effective market shares," a measure of customer dominance we define precisely below, that can diverge from traditional market shares. We use these model-consistent effective shares to document four new empirical patterns. (1) Effective shares often diverge sharply from firms' traditional market shares, yielding an effective-size distribution that is far less skewed than the distribution of sales. That is, firms that look dominant by traditional market share do not similarly dominate their customers' baskets. (2) Effective shares better predict proxies for demand elasticity than traditional shares. (3) Increasing sales from incumbent customers raises effective shares while increasing sales from new, more marginal customers tends to lower effective shares. Specifically, two prominent sources of firm growth in recent years—the expansion of multi-establishment chains and e-commerce— are associated with rising market shares but not rising effective shares. (4) Most importantly for the market-power debate, sales concentration has increased substantially over the last decade while average effective shares have barely risen. Together, these facts imply that even through the lens of variable-markup models that directly tie market power to "size", the rise of superstar firms has not produced a broad-based increase in size-based market power.

We begin with a simple framework showing how customer heterogeneity breaks the usual mapping from market share to market power. Related mechanisms appear in recent macro and IO work (e.g., Mongey and Waugh (2025)), but empirical discussions typically treat aggregate market share as the relevant notion of "size." Our setting is a standard BLP-style discrete-choice model with profit-maximizing firms, so markups are pinned down by demand elasticities. Household preferences imply size-dependent elasticities of demand: as a firm's category share rises, the relevant substitution margin for its customers shifts from within- to across-category alternatives, lowering its demand elasticity. In representative-household frameworks like Atkeson and Burstein (2008), the relevant "size" for this channel is a firm's *aggregate* market share within a category. With heterogeneous customers, elasticities instead depend on the firm's share of *each customer*'s category spending, which captures how dominant the firm is in its specific customers' baskets. We define a firm's "effective share" as the sales-weighted average of its customer-level spending shares.

[1] Market power may also change for unrelated reasons. We focus on "size-based market power," where size directly affects demand elasticity, because it is central to the variable-markup literature and concerns about distortions from concentration.

The rest of the paper measures effective shares and documents their empirical properties. We use comprehensive credit- and debit-card transaction data from a large card-payment processor with trillions of dollars of U.S. spending each year. We focus on consumer-facing retail and service categories, in which we observe spending for the near universe of U.S. firms and establishments.[2] Transactions include the dollar amount, anonymized card identifiers, merchant identifiers, establishment location for in-person transactions, and a card-present flag that we use to distinguish online from in-person transactions. Beginning in 2021, we link cards to TransUnion records to aggregate spending across multiple cards per individual.[3] Our main results use this linked 2021–2025 sample. Our analysis of long-run trends uses a less granular version of the data available back to 2014.

We observe how each customer allocates spending across firms, which lets us construct realized expenditure shares. However, like any approach that maps observed purchases into demand objects, we must relate realized expenditures to ex-ante choice probabilities.[4] We use two complementary strategies to approximate latent propensities from realized expenditures. First, we pool each customer's transactions over time and interpret longer-run purchase frequencies as revealing latent shopping propensities. Our baseline measure pools purchases within a year, but results are similar when pooling over 2021–2025. This approach allows idiosyncratic preferences even among observationally similar customers, but it can understate consideration sets when purchase counts are small.[5] Second, we construct "synthetic customers" by pooling transactions for individuals with similar demographics who live nearby each other.[6] This approach imposes common preferences within groups, but is more likely to capture customers' full choice set. In practice, the two measures are highly correlated; we use the individual pooling approach as our baseline and show robustness to the synthetic-customer alternative.

Standard market-share approaches face the same realized-to-latent issue. However, a key relative advantage of effective shares is that they avoid the need to specify geographic market boundaries. This issue is central in the empirical concentration literature because measured market shares—and their trends—depend meaningfully on the level of geographic aggregation (Rossi-Hansberg, Sarte and Trachter (2021)), yet appropriate boundaries are often unclear even within narrow sectors. Batch et al. (2024) make progress by defining "consumption zones" from observed spatial shopping flows, but this still assigns a single geographic scope to a location–sector cell rather than allowing scope to vary across firms and shoppers. For example, two grocery stores on the same block may draw customers from very different catchment areas if only one offers parking. Households near any boundary will likely shop across it. Effective shares naturally accommodate both facts: they allow geographic scope to differ across firms and impose no restrictions on where households shop. More generally, the relevant geography of competition is revealed by the firm's customer base and where else they shop, rather than imposed ex

[2] We restrict to establishments with 200+ customers; standard consumer panels typically observe too few customers per firm to measure customer-level spending shares for most firms.

[3] TransUnion data also contains additional demographic information like customer ZIP-11, age and estimates of income.

[4] These need not coincide in general. For example, with a single purchase a customer's realized share will always be one even if they considered purchasing from many firms.

[5] In most analyses we also require a minimum number of category purchases per customer to limit this concern.

[6] Our baseline synthetic customer groups individuals by census tract, age bin, income bin, and an indicator for car ownership but other groupings produce similar conclusions.

ante.

Our empirical analysis begins with descriptive analysis of the cross-section of effective shares and conventional market shares. This analysis reveals a firm "size" distribution that is substantially less skewed when size is measured by effective share rather than by traditional sales shares. Looking first at national sales for firms as a whole, we find that effective shares are meaningfully less dispersed than aggregate market shares: firms that appear to dominate industries in market-share space are far less extreme once size is measured by the share they capture within their customers' baskets. A similar pattern holds when we compare effective shares of firms in particular locations against the firms' local market shares. Although local shares are closer in level to effective shares, the relationship is strongly concave: smaller firms tend to have effective shares systematically above their local market shares, while larger firms tend to have effective shares systematically below their local market shares.

Two forces generate this concavity. First, within-market segmentation raises effective shares relative to local market shares: a neighborhood coffee shop can be a tiny share of citywide restaurant spending yet account for a large fraction of spending among its regular customers. This segmentation can potentially arise from various forces, like micro-geographic frictions (e.g., commuting patterns, local neighborhoods) and idiosyncratic preferences/loyalty. Second, out of market spending (including online) tends to lower effective shares relative to local market shares: a firm with a large share of spending in some local geographic area may not dominate its customers' spending if they shop beyond this local boundary. This is particularly important for large firms, for which the gap between in-person and remote purchasing is often largest. For example, a big-box electronics retailer may dominate local, in-person spending but have a much smaller effective share if its customers buy a substantial fraction of electronics online. Together, strong segmentation for small firms and substantial out-of-location spending for large firms imply that conventional local market shares overstate the dominance of large firms and understate the dominance of small firms.

After documenting that effective shares often diverge from traditional market shares, we ask whether they capture information about firms' demand elasticities. The model implies that higher effective shares map into lower elasticities, but even if the model is correct, two issues could weaken this link in the data. First, as already discussed, effective shares are measured from realized purchases rather than latent shopping propensities, so measurement error may remain despite our pooling and sample restrictions. Second, we observe equilibrium cross-sectional variation in effective shares rather than exogenous shifts: as emphasized by a vast IO literature, observed market shares reflect forces like prices, product characteristics, and competitive conditions as well as demand curvature. Thus, a firm with a high effective share in equilibrium need not actually face inelastic demand.[7]

Because we do not observe prices, we cannot estimate elasticities directly. Instead, we provide two pieces of evidence that effective shares have predictive content for demand elasticity beyond traditional market shares. First, we use customer retention over time as a proxy for demand elasticity: establishments with higher effective shares retain customers at systematically higher rates, even controlling for

[7]For example, in a homogeneous-good Bertrand market with highly price-sensitive consumers, the lowest-priced firm can capture nearly all sales even though its residual demand is extremely elastic.

traditional market shares and sector effects. Second, we study entry events and customer switching: among customers who shopped at the same incumbent in the prior year, those with higher effective shares are substantially less likely to switch to the entrant after it appears.[8]

Our cross-sectional results show that traditional market shares differ substantially from the notion of "size" that matters for variable markups with heterogeneous customers. We next ask why effective shares rise so slowly with market shares, and what that implies for superstar firms and size-based market power. The answer is rooted in the nature of firm growth and a trade-off between customer depth and reach: market shares reflect *how much* a firm sells, while effective share depends crucially on *who* it sells to. Market shares increase whenever a firm expands its total sales relative to a category, regardless of who those sales come from. In contrast, effective shares depend on the spending intensity of the firms' particular customers. A growing body of evidence emphasizes the important role of customer acquisition on the extensive-margin for explaining sales growth (e.g. Klenow et al. (2022)). However, we find that new customers tend to be more marginal than incumbents: on average, their effective shares are 25% lower than those of incumbent customers. The positive, but small, overall elasticity we find reflects the net of two opposing forces: as firms grow, they draw in new customers, which dilutes their effective share, but they also expand sales among their incumbent customers, which increases their effective share.[9]

Two growth channels that have been especially important in recent years tilt growth toward the extensive margin and therefore generate particularly muted changes in effective shares. First, geographic expansion by multi-establishment firms (Hsieh and Rossi-Hansberg (2023)) raises national market share by attracting customers in new markets who previously lacked access. We find that these "new-market" customers have lower effective shares than the firm's customers in established markets. Second, growth in online sales expands firm reach but also exposes customers to a broader choice set. Consistent with this, customers who purchase only online have lower effective shares than customers who purchase in person at the same firm, and firms that gain market share primarily through online growth see much smaller increases in effective shares than firms whose growth is primarily in-person.

Our results also materially reshape concentration trends and change the implications of growing "superstar" firms for market power. To show this, we complement our 2021-2025 linked customer-firm sample with a historical, but less granular version of the data that exists back to 2014.[10] We show that effective shares increased much less over this decade than the concentration of sales. Additionally, while the Nielsen data is too small to replicate most of our analysis, it can be used to measure the national spending concentration of top firms just in the grocery sector. Using this data which goes back further, to 2004, we find similar patterns of rising sales concentration but relatively stable effective shares. Thus, our broad conclusion is that even though a small share of "superstar" firms accounts for an increasing

[8] We relate our measure to customer-overlap statistics in Einav, Guido and Klenow (2026) below; their validation exercises provide complementary evidence on the link between customer dominance and demand elasticity.

[9] These forces can coexist because effective shares place more weight on high-intensity customers, whereas total sales weight customers more uniformly; the two objects are therefore driven by different parts of the customer distribution.

[10] This historical data can only be used to measure spending shares at a subset of firms, and we focus on measures of "top 4" concentration.

share of sales in the U.S., this has not led to a similar increase in the effective size relevant for size-based market power.

Our findings speak to work on customer-based firm growth and market power. Consistent with Klenow et al. (2022), Argente et al. (2025) and Afrouzi, Drenik and Kim (2025), we find that sales growth is driven primarily by the extensive margin of customer acquisition. We also echo Afrouzi, Drenik and Kim (2025)'s conclusion that extensive-margin growth has limited implications for market power, and extend this evidence from the grocery sector to a broader set of consumer-facing retail and service sectors. Relative to their accounting decomposition into customer counts and sales per customer, our contribution is to measure the theory-relevant object for size-based market power directly: firms' effective shares. While effective shares are related to sales per customer, they are not equivalent: effective shares depend on the distribution of customers' shopping and can move even when average spending per customer does not. Empirically, our proxies for demand elasticity are more strongly related to effective shares than to either customer counts or sales per customer, and incorporating this information changes the interpretation of concentration trends.

A separate literature documents rising national concentration of sales, but there is more debate about trends in local concentration.[11] As already discussed, our approach side-steps the need to choose the relevant geography for concentration. Another natural approach is to estimate markups directly, but doing so at the firm—let alone establishment—level typically requires strong assumptions about production technologies and cost measurement.[12] As a result, there remains substantial disagreement about markup trends and their connection to concentration.[13] Our approach instead uses a demand-based lens to study size-based market power without prices, quantities, or cost data, and we validate the resulting measure using empirical proxies for demand elasticity.

On the theory side, we build on a large literature on variable markups. We follow most closely Mongey and Waugh (2025), which emphasizes the role of customer dominance in shaping demand. We take this mechanism to the data by providing empirical measures of firms' effective shares. Our finding that the effective-size distribution is much less skewed than sales-based size is also related to work studying the aggregate implications of markup distortions and misallocation.[14]

Finally, our measure is closely related to Einav, Guido and Klenow (2026), who compute bilateral customer overlap between firms using the same transactions data. Their overlap measure is a more granular version of our own object: summing their measure over all other firms in the category delivers its effective share. The overlap statistic is well-suited for questions like merger analysis that focus on competition between specific firms, but their measure requires observing the full spending vector across all firms and it suffers from a curse of dimensionality.[15] Our effective share-measure reduces dimension substantially

[11] Autor et al. (2020*a*),Rossi-Hansberg, Sarte and Trachter (2021),Autor, Patterson and Van Reenen (2023).

[12] Steve Bond (2021).

[13] De Loecker, Eeckhout and Unger (2020), Benkard, Miller and Yurukoglu (2025), Traina (2018), Syverson (2019).

[14] Baqaee and Farhi (2020), Edmond, Midrigan and Xu (2023), Afrouzi, Drenik and Kim (2025).

[15] The median sector in 2024 has 25 (41) thousand firms (establishments), while the largest sector has almost 600 thousand establishments. Bilateral customer overlap is thus a 360 billion entry matrix while effective shares are a much smaller 600 thousand entry vector.

and it is also empirically feasible in settings (like our long-run analysis) where disaggregated spending is only observed for a subset of firms.[16] We view the two approaches as complements: effective shares are well-tailored for analyzing sector-level and macro questions about concentration and market power, while customer overlap provides additional fine-grained information useful for analyzing competition between specific firms.

The remainder of the paper proceeds as follows. Section 2 lays out the theoretical framework and derives effective share as the sufficient statistic for size-based market power under customer heterogeneity. Section 3 describes the transaction and credit-record data and our construction of effective shares. Section 4 documents the cross-sectional relationship between effective and conventional shares. Section 5 shows that effective shares better predict customer retention relative to traditional market shares. Sections 6 and 7 study the dynamics of effective size and explore the implications of these patterns for understanding the rise in market concentration. Section 8 concludes.

2 Motivating theoretical framework

This section motivates the measurement in the rest of the paper by showing that in models with size-based market power, "size" should be measured from a customer-focused perspective rather than by aggregate market share.[17] We illustrate the mechanism in a simple BLP-style discrete-choice environment with a nested-logit structure and a parsimonious source of persistent customer heterogeneity. The same mechanism arises more generally in models where individual demand satisfies Marshall's second law of demand, which is that the elasticity of that customer's demand for the good falls when the good becomes a larger share of that customer's spending basket.

2.1 Model Setup

Households indexed by i choose among differentiated firms indexed by j, with products grouped into broad categories $c \in C$ (e.g., grocery stores, restaurants, home furnishings). In each period t, the household either purchases one unit from one firm in one category, or chooses an outside option normalized to price 1. The utility from choosing firm j is given by

$$U_{ijt} = -\ln(p_j) + \phi_j + \mu_{ij} + \zeta_{ijt}, \qquad \mu_{ij} \equiv \lambda_i \psi_j, \tag{1}$$

where p_j is the price, ϕ_j is a common quality shifter, μ_{ij} is a permanent customer–firm match component, and ζ_{ijt} is a transitory taste shock.[18] The term $\mu_{ij} = \lambda_i \psi_j$ parsimoniously captures persistent differences across households in the attractiveness of firms, for example due to geography, shopping technology, or tastes. We assume ζ_{ijt} follows a generalized extreme value structure (nested logit), with parameter

[16] Measuring the effective share at a specific firm j requires only knowing a customer's spending at j and their total spending, while measuring customer overlap requires knowing the customer's spending at every individual firm.

[17] This setup is very similar to Mongey and Waugh (2025), and we intentionally adopt their notation where applicable.

[18] We do not observe prices, so we choose this functional form to eliminate variation in markups across firms not arising from size. This simplification is not essential for the conclusion that "size" depends on customer shares.

η governing substitution across firms within a category and parameter θ governing substitution across categories. The empirically relevant case for size-based market power is $\eta > \theta$: firms within categories are more substitutable than across categories. Because μ_{ij} is fixed over time while ζ_{ijt} is transitory, customer-level choice probabilities differ persistently across households.

On the supply side, each variety is produced by a unique firm that competes via Bertrand price competition. Optimal pricing implies the familiar constant-elasticity markup rule,

$$p_j = MC_j \cdot \frac{\varepsilon_j}{\varepsilon_j - 1}, \tag{2}$$

where MC_j denotes marginal cost of firm j and ε_j is the elasticity of demand for firm j.

2.2 The elasticity of demand

Let ρ_{ij} denote customer i's probability of choosing firm j conditional on purchasing within j's category. Since the permanent match term μ_{ij} is fixed for a given household, we suppress that conditioning in the notation. In this nested-logit environment, the customer's price elasticity for firm j can be written as

$$\varepsilon_{ij} = \eta(1 - \rho_{ij}) + \theta\rho_{ij}. \tag{3}$$

The key feature of Equation 3 is its dependence on ρ_{ij}. When $\eta > \theta$, there is size-based market power, meaning that this elasticity declines as a firm becomes a bigger share of a customer's spending in the category. The key observation is that the relevant notion of "size" is customer-specific: what matters is how big a firm is for each customer, not the firm's total market share.

In order to get the firm's total elasticity of demand, we aggregate the customer-specific elasticities ε_{ij}. The appropriate firm-level elasticity is therefore a sales-weighted average of those customer elasticities, placing more weight on customers who account for a larger fraction of the firm's revenue. Formally, let $\alpha_{ji} = \frac{\rho_{ij}}{\sum_i \rho_{ij}}$ denote the weight of customer i in firm j's sales, and define the firm's elasticity as

$$\varepsilon_j = \sum_i \alpha_{ji}\varepsilon_{ij}. \tag{4}$$

Substituting Equation 3 into Equation 4 and simplifying yields a particularly transparent expression:

$$\varepsilon_j = \eta - (\eta - \theta)\sum_i \alpha_{ji}\rho_{ij}. \tag{5}$$

Equation 5 delivers the key empirical object: in this class of models, the object that governs size-based market power is not the firm's aggregate market share, but rather the sales-weighted average of customer-level choice probabilities. We define firm j's "effective share" as

$$\omega_j^{\text{eff}} \equiv \sum_i \alpha_{ji}\rho_{ij}, \tag{6}$$

so that the pricing-relevant elasticity is

$$\varepsilon_j = \eta - (\eta - \theta)\omega_j^{\text{eff}}. \tag{7}$$

Since markups are solely a function of the price elasticity of demand, markups grow with a firm's effective share.[19]

Market shares are equivalent to effective shares only if there is a representative customer, meaning there is no heterogeneity in where customers shop. In that case, ρ_{ij} is constant across i and equals the firm's aggregate market share Ω_j. In that knife-edge case, Equation 5 collapses to the familiar Atkeson and Burstein (2008) mapping from market share to demand elasticities,

$$\varepsilon_j = \eta - (\eta - \theta)\Omega_j. \tag{8}$$

However, with customer heterogeneity, a firm's effective share ω_j^{eff} can diverge substantially from its aggregate market share Ω_j. In the present formulation, this heterogeneity is generated by the permanent match term μ_{ij}, which can capture geographic segmentation as well as preference heterogeneity and segmentation within geographic areas. The main empirical exercise of this paper is to measure this wedge and assess how it reshapes our understanding of firm size, concentration, and size-based market power.

Most of our empirical work focuses on firm-by-firm comparisons of ω_j^{eff} and Ω_j. However, it is also interesting to analyze markets as a whole and whether average market power has varied over time. The sales-weighted average elasticity of demand for a market as a whole is given by

$$\varepsilon \equiv \sum_j \Omega_j \varepsilon_j = \eta - (\eta - \theta) \sum_j \Omega_j \omega_j^{\text{eff}}.$$

In the representative-customer case, $\omega_j^{\text{eff}} = \Omega_j$ for all firms, and thus

$$\sum_j \Omega_j \omega_j^{\text{eff}} = \sum_j \Omega_j^2 = HHI.$$

This linearly declining elasticity of demand with the sales Herfindahl is one reason the past literature has focused on measuring the concentration of market shares.

However, with heterogeneous customers, the elasticity of demand depends instead on the average effective share in the market. Computing this average and rearranging terms points to the concentration of individual customers' spending as the relevant object for market power. In particular, letting y_{ij}

[19]The baseline model can be extended to include heterogeneous price sensitivity (e.g. Berry, Levinsohn and Pakes (1995) or Mongey and Waugh (2025)), by simply rescaling the relative weight the firm places on each customer. For example, in Mongey and Waugh (2025), $\varepsilon_j^i = \left(\eta(1-\rho_j^i) + \theta\rho_j^i\right) \times MU_i$. Recovering the level of each firm's elasticity then requires either information on price variation or additional structure, but measuring the distribution of ρ_j^i remains the central empirical task for understanding how relative differences in market power arise from differences in customer dominance.

denote spending of customer i at firm j, $Y_j \equiv \sum_i y_{ij}$, $Y_i \equiv \sum_j y_{ij}$, and $Y \equiv \sum_i Y_i$, we have

$$\sum_j \Omega_j \omega_j^{\text{eff}} = \sum_j \left(\frac{Y_j}{Y}\right) \left(\sum_i \frac{y_{ij}}{Y_j} \rho_{ij}\right) = \sum_i \left(\frac{Y_i}{Y}\right) \left(\sum_j \rho_{ij}^2\right) \equiv HHI^{cust}.$$

Thus, with heterogeneous customers, the relevant concentration object for size-based market power is the concentration of customers' spending baskets, not the concentration of firms' aggregate sales.

3 Data Description and Measurement Details

3.1 Data Description

Our primary data sample begins from the universe of all U.S. credit and debit card transactions of a major card-payment processor from 2018-2025. The scope of the data is large, it covers more than 600 million cards and more than 4 trillion dollars of spending each year. This corresponds to roughly one-fifth of aggregate annual personal consumption expenditures (PCE).

The unit of observation is an individual transaction between an (anonymized) credit card and a merchant. For each transaction, we observe the exact dollar amount, time and date, the unique identifier of the card that made the transaction, and the identity of the merchant, and whether the transaction was in-person (card present) or remote (card not present). 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). For in-person transactions, in addition to the identity of the firm, we observe 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. For remote purchases (e.g. online, phone, recurring purchases), we observe the firm identifier but no physical location for the transaction. For many remote transactions we further observe if it was an online purchase, a recurring payment initiated by the firm (e.g. monthly subscription fees), or a purchase made over the phone. In the analysis that follows, we refer to all transactions without cards present as "online" transactions.[20]

Each transaction in the dataset is assigned a merchant category code (MCC). This 4-digit code captures the type of good or service that the firm provides. This code is assigned at the time that a merchant registers to accept card payments, and it determines the merchants' fee structure and corresponding credit card rewards, so is a fundamental characteristic of each transaction. We aggregate very similar MCC codes into a single MCC category and subset to the MCC codes that are customer-facing with good coverage of individual firms, leaving us with 33 MCC codes in our analysis sample.[21] Some merchants have multiple MCC codes corresponding to establishments providing different products (e.g.

[20] 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.

[21] We drop some categories like airlines and hotels, where a large fractions of transactions cannot be linked to specific firms because they are made through intermediary websites. See Appendix table A2 for a detailed list of categories.

Costco wholesale vs. Costco gas stations). We assign all transactions within a firm to its MCC code with the most transactions. The largest categories in our sample are general merchandise stores (31% of our sample), Grocery Stores (18% of our sample), and Restaurants (9% of our sample) (see Appendix Table A10 for details).

The raw transactions dataset is a card-level panel, and has no links that aggregate multiple cards owned by the same customer. However, beginning in 2021, this card-level panel can be linked to TransUnion credit records, allowing us to group multiple cards belonging to the same person. The TransUnion linkage also provides a set of key consumer observables, including home location, age, and estimated income. The linked sample covers around 100 million individuals and 3 trillion of spending. 41% of the spending in the full transaction dataset is included in the merged sample. Credit records do not capture debit-cards, so this is the main source of dropped spending.[22] We show robustness of our main results to performing our analysis at the card-level instead of the person level, which allows us to use both debit and credit cards as well as the cards for people who do not appear in Transunion credit records.[23] In the linked sample, the average person has 1.6 credit cards per year and 2.2 cards over the sample from 2021-2025, with 13.76% of people having at least 3 cards within a year.[24] Lastly, since the linked sample restricts to those customers with some credit history, the resulting customers have a slightly higher median household income than the general population and are on average 3 years younger than the average American adult. Appendix Table A8 shows the full comparison.

Starting from the linked sample, we impose additional restrictions that focus our sample on consumer cross-firm substitution patterns. First, we drop business cards. Second, we restrict the sample to transactions that are either online or occur within 100 miles of the consumer's home ZIP code. This restriction excludes idiosyncratic travel spending (e.g., vacation purchases) and focuses on the set of stores plausibly relevant for regular shopping behavior.[25] Third, we remove transactions routed through payment facilitators, intermediaries, or peer-to-peer systems (e.g., PayPal, Toast, Venmo), because these records often obscure the underlying merchant identity.[26] Fourth, we drop all customers whose home address in TransUnion changes 3-digit zip codes within the year. Fifth, we focus our attention on category-zip3 pairs where we observe at least $450K in spending between 2021-2025. Sixth, to limit potential sampling error in measured effective shares, we focus only on firms with at least 200 customers observed in our linked subsample in a year. We henceforth refer to this filtered sample as our "Primary" sample.

[22] 76% of cards that are not linked are debit cards; the remaining unlinked cards arise because not all credit-issuers report to TransUnion.

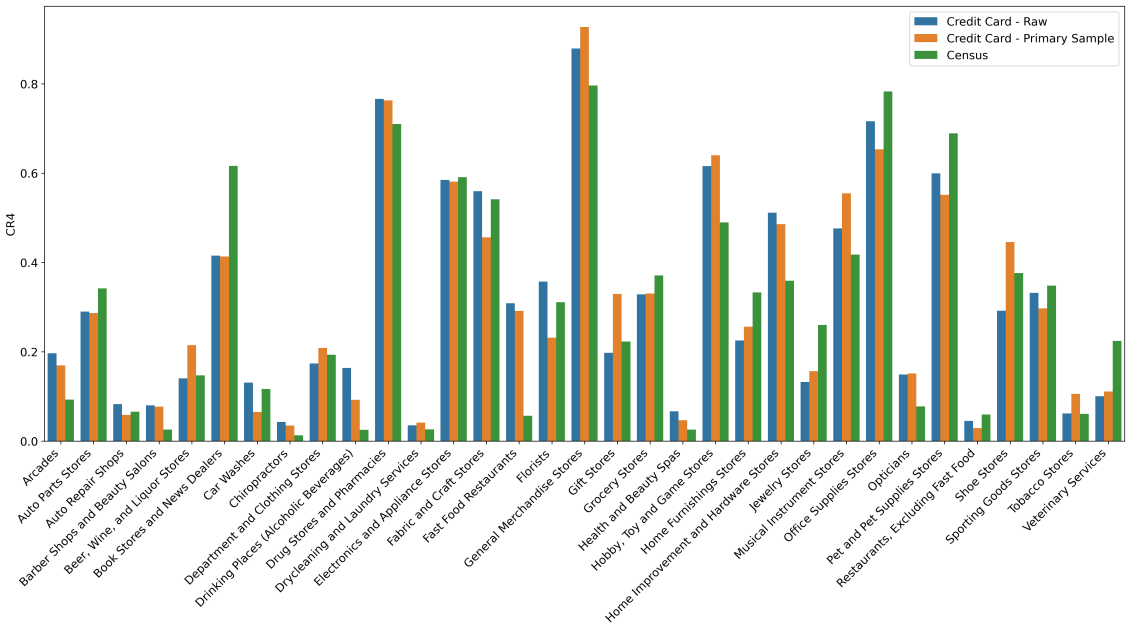
[23] Specifically, we find that the firm-level effective share calculated using all cards and the firm-level effective share using just credit cards are highly correlated in all categories. See Appendix Figure A8 for the details.

[24] Individuals with multiple cards may use certain cards for certain types of spending due to rewards incentives. Rewards incentives vary across MCC codes but not across merchants within MCC codes. Since our analysis focuses on spending concentration within categories as measured by MCC codes, cross-MCC card substitution is unlikely to matter for our conclusions.

[25] Since this is a small share of spending, this restriction is unimportant in practice.

[26] 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 33 categories, 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.

Figure 1: Benchmarking Credit Card Data to Census Data



Notes: This figure plots the 2022 four-firm concentration ratio (CR4) by category in three data sources: the raw credit card transactions sample, the primary credit card sample merged with TransUnion, and U.S. Census data. For each industry, CR4 is defined as the sum of the market shares of the four largest firms. The primary credit card sample is restricted to categories with at least 20 distinct firms in each year. To harmonize category definitions across sources, category labels are standardized prior to merging. Our sample is further restricted to industries observed in all three datasets with non-missing CR4 values. Used Merchandise Stores, Gyms and Fitness Clubs, and Gas Stations are excluded. Census NAICS code in 2022 are manually mapped and verified to our credit card category definitions.

Does this sample capture the true distribution of spending across firms in the economy? Figure 1 shows that the credit card spending we observe aligns well with firm spending data from the Census. In particular, this figure shows that the national sales share of the top 4 firms (CR4) in each category in our card data aligns well with the same measure for similar categories in the Census.[27] This is true both in the more comprehensive raw card data as well as in our more restricted primary sample. The category with the most notable deviation is "Fast-Food Restaurants", and this is because the Census treats franchises of the same chain as separate firms while our data combines all franchises in the same firm. We think our measure is likely more relevant for understanding customer demand and firm product market power, since consumer shopping choices as well as pricing and product assortment are often made at the firm rather than the franchise level. Our primary conclusions extend across categories and so are robust to dropping categories with less-aligned relationships.

In addition to our primary dataset, we also utilize a less granular "historical panel" that goes back

[27]The crosswalk we use to map NAICS codes used by Census to MCC categories in credit card data are shown in Appendix Table A2. It is worth nothing that NAICS and MCC-based categorizations of firms will not perfectly align, so we should not expect a perfect match even if spending for every individual firm exactly matches.

to 2014 to study trends in spending concentration.[28] The main benefit of this historical panel is the longer time coverage, since our primary sample starts in 2021. However, this historical panel has two important limitations. First, this historical panel data cannot be linked to TransUnion, and therefore for this analysis, our unit of observation is the card, not the person. Second, this historical panel measures total spending by individual cards in each zip code, but it only further dis-aggregates this spending within zip codes for a subset of (relatively large) "named firms". Just like in our primary data set, the spending that cards do at these named firms is separately identified, but the remainder of spending is grouped together as spending at "Other-un-named firms". Although we can see a card's *total* spending at these unnamed firms we are unable to measure its spending at any specific unnamed firm within this group. This means that the data is well-suited for studying spending trends for larger, named firms but not for studying spending trends at the smallest firms in the economy.[29] We restrict our national analysis to only categories with at least 8 named firms in every year and our local analysis to zip3 X categories with at least 4 named firms in every year.[30] The number of named firms varies substantially across categories, but the median category has 61 named firms while we only need to reliably capture the top 4 firms for our analysis.[31] Despite these limitations, we confirm that the key relationships between effective shares and market shares are similar in our primary and historical panel in the years where they overlap, which gives us confidence that the historical panel captures the same underlying patterns as our primary data. We discuss this benchmarking more in Section 6.

3.2 Measuring Effective Shares

After constructing the analysis sample described above, we measure effective shares following the formula derived in Section 2:

$$\omega_j^{\text{eff}} = \sum_i \alpha_{ij}\rho_{ij} \tag{9}$$

where α_{ij} is the (expected) share of firm j's spending that comes from customer i and ρ_{ij} is the ex-ante probability of choosing firm j, averaged over all idiosyncratic taste shocks. As we said above, we define categories using MCC codes and construct all spending shares within market.

With the transaction data, we have a good view of what the customers actually chose to spend on and can easily construct this object. However, the key empirical challenge is that ρ_{ij} is the share of a customer's *ex-ante* spending that goes to the firm (i.e. their shopping probabilities before their logit shock is realized), not the share of their realized spending. This is most clearly exemplified in the discrete choice

[28] A version of this data goes back to 2008 but its coverage is extremely sparse before 2014, so we begin analysis in 2014.

[29] Importantly, as defined in Section 3.2, the effective share of customer i at firm j depends on i's spending at j and i's *total* spending at all other firms but not how that other spending is split across specific firms.

[30] We choose a weaker threshold for local than national analysis since most firms do not operate nationally. We choose a relatively mild cutoff for both, so that we are able to study a comprehensive set of markets. Results are similar if we restrict to national (local) markets with 20+ (10+) named firms.

[31] The mean is 118, the minimum is 8 firms (Veterinary Services) and the max is 798 (restaurants, excluding fast-food)

setting, where a customer's realized purchase shares will always be zero or one ex-post, independent of their true underlying choice probabilities.

We implement several complementary approaches to inferring these underlying choice probabilities from the realized spending data. Our baseline approach addresses this by pooling within individuals over time. In the baseline, we pool all transactions made within a year across all cards associated with the customer.[32] Pooling over time increases the probability that we observe enough transactions that the observed spending shares approximate the ex-ante choice probabilities. It also preserves idiosyncratic heterogeneity in preferences and constraints, which is central to the conceptual distinction between effective shares and conventional market shares.[33] Moreover, to make sure that our results are not the result of discrete choice or infrequent purchases, we further restrict attention in each category to the subset of customers who make at least 5 transactions in that category in that year.[34] Appendix Table A3 shows the average and median number of transactions per category per customer each year. The average number of transactions per year per customer ranges from 57 in General Merchandise to 6 at opticians or Auto Repair shops.

For robustness, we also implement two alternative specifications. First, we find that all results are robust to a version where we extend the within-person pooling to a longer period of time, aggregating each individual's spending over the full 2021–2025 data. This increases the possibility that we have observed spending at the set of stores in which the customer has a non-trivial shopping probability.[35] Second, we also show robustness of the patterns to an arguably more conservative version where we pool annual spending across observationally similar customers who live in the same location. Specifically, we define "synthetic customers" as groups of shoppers living in the same census block who are demographically similar (i.e. the same age bin, income bin, and with the same (inferred) car ownership status). For each synthetic customer group g, we compute group-level spending shares ρ_{gj} analogously to Equation 9 and then construct firm effective shares by weighting group-level shares by the group's contribution to firm sales. The motivation for this approach is that similar consumers living in the same neighborhood plausibly face similar geographic constraints and may have correlated preferences and so pooling their shopping patterns can reveal the distribution of relevant shopping options for these customers. However, a disadvantage of this approach is that it assumes away any preference heterogeneity *within* these

[32] We restrict our attention to the set of customers who are active in the credit card data throughout the entire year, imposing that in the median month of the year, they have at least 5 transactions. This excludes 33.22% of people per year but only 10.67% percent of the total spending.

[33] The model framework in Section 2 also suggests that pooling over time will reveal the underlying ρ_{ij} as long as firm primitives (relative prices and relative quality) are fixed. Specifically, if the logit shocks ζ_{ijt} are i.i.d. across t, then the realized long-run purchase frequency at firm j converges to the latent choice probability ρ_{ij}, which differ systematically across households because of variation in μ_{ij}, which captures individual preferences for firm j.

[34] While this is a mild restriction in some sectors, this meaningfully restricts the number of customers in some sectors. Appendix Table A9 shows the fraction of customers and spending in each industry that are dropped with this restriction on shopping frequency. It ranges from excluding only 0.4 percent of spending in General Merchandise Stores to 80 percent of spending on Jewelry Stores. In implementing this restriction, we are implicitly making the assumption that the people who shop frequently in a given year have a similar distribution of underlying choice probabilities to those who shop infrequently in a given year.

[35] The downside of this approach is that it must implicitly assume stable preferences and market conditions over this five year period, and it also precludes any panel analysis.

groups, as all customers within the same demographic group in the same location are assumed to have the same choice probabilities. With this definition, the average synthetic customer pools together 67 actual customers and includes far more annual transactions per year (See Appendix Table A3).[36] Despite these differences, the firm-level effective shares calculated using additional pooling over time or using synthetic customers have a correlation with our baseline measure of 0.96 and 0.94, respectively.

While the decision of how to define customers is essential for the analysis, it is useful to emphasize that our effective share construction does not impose *any* constraints on the firm's geographic market. The question of what geographic boundary to draw around firms has been central to the discussion of superstar firms (Autor et al. (2020*b*), Rossi-Hansberg, Sarte and Trachter (2021), Autor, Patterson and Van Reenen (2023)). However, when calculating effective shares, we do not have to take any stand on which set of firms within a category are in the same market (e.g. if the relevant set of firms those within the zip code, the MSA, the state, etc.). Rather, the relevant set of firms is revealed by the spending of the firm's customers.

Table A1 shows summary statistics for each of the resulting samples that we consider. Each column refers to a different subsample that we consider throughout the analysis. Column 1 is the baseline approach to calculating effective shares where we pool all spending within the year for each person. Column 2 shows the sample when we pool all spending from 2021-2025, Column 3 shows the version that aggregates to synthetic customers and for comparison, Column 4 shows the same statistics on the full transaction dataset before we have imposed any of the person-level restrictions discussed above. In the baseline sample, we include almost 80 million customers, 2.5 million firms and 2.4 trillion dollars of spending. The average annual spending is $10,610.

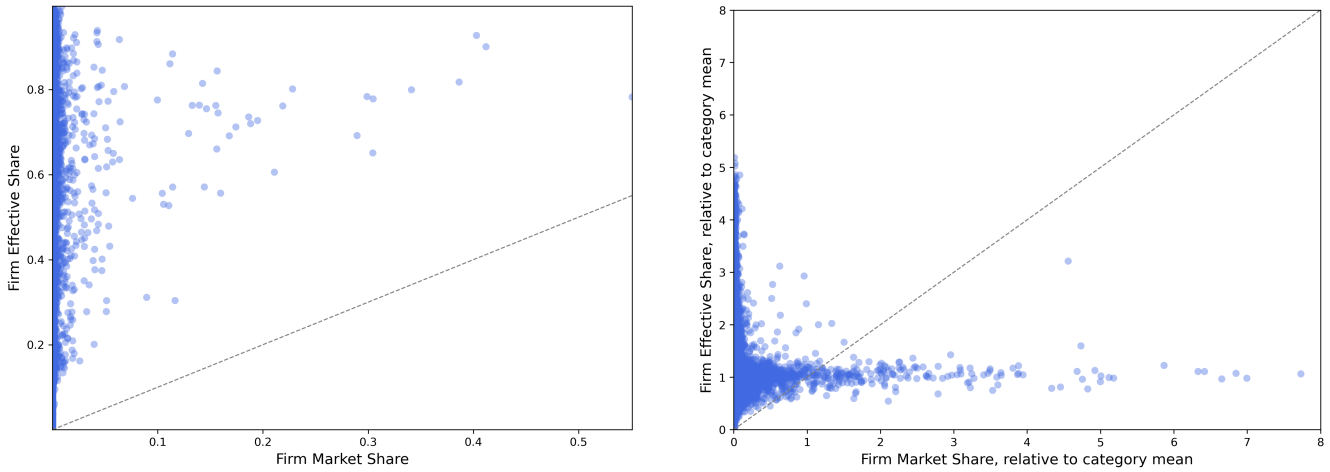
4 The Compression of Effective Shares Relative to Market Shares

In this section, we present the firm-level relationship between a firm's effective share and its market share, either nationally or locally. We show that these two shares are often very different from each other, and the relationship is such that the firm size distribution is much more compressed when size is measured from from the customer perspective (i.e. effective shares) versus from a sales-based perspective (i.e. market shares).

We begin our analysis by looking at the cross-firm relationship between national market shares and effective shares in 2025. The left panel of Figure 2 first just visualizes the raw data, showing each firm's effective share on the y-axis and its conventional national market share on the x-axis. In a closed economy, the share of spending that a firm gets among their own customers has to be at least as large as what they get in the market overall. This means that the effective shares for national firms are at least as large as the national market shares, meaning that all points will be above the 45 degree line. However, for most firms and particularly for the smaller firms, this gap is very substantial. There is also substantial

[36] Since we do not need the spending to be representative of the whole basket for any given customer in this specification, we do not impose activity restrictions on each customer before pooling across customers. This increases the total spending included in this calculation (See Table A1 for details).

Figure 2: Effective shares differ from national market shares



(a) Overall

(b) Within Category

Notes: In each panel, each dot reflects 1 firm in 2025. In panel (b), each value is divided by the sales-weighted average value in the category. Effective shares are defined using annual customer observations. In the right panel, both the x and y axis are truncated at 8.

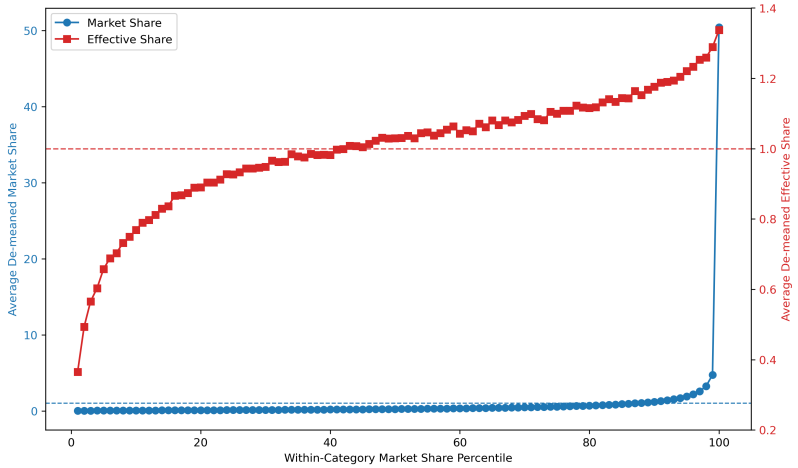
dispersion in effective shares among firms with very similar market shares. While there is clearly some positive correlation, the relationship is weak and there is substantial heterogeneity in effective shares even for firms with the same national market share.

The left panel of Figure 2 pools firms across all categories. However, categories differ substantially in the distributions of market shares, and this cross-category variation might confound within-category patterns that are of more interest.[37] Thus, in the right panel, we divide the values on both the y and x axes by their respective category means. This means that the x axis is now the firm's market share relative to the market share of the average firm in the category and the y axis is the effective share relative to the average effective share in the category. With this transformation, the relationship between the two variables looks even weaker. The firms that are large outliers relative to their category in market share space are not the firms that are outliers in their category in the effective share space.

Since the distribution of firm size is very skewed, most firms have small national market share and there is a large mass clustered near zero in both panels of Figure 2. This makes it hard to unpack the relationship for the typical firm in these raw scatter plots. Thus, in Figure 3, we explore an alternative specification that shows patterns throughout the distribution. Specifically, we order firms by their position in the market-share distribution within their category and plot the average market share relative to the category average (left axis) and the average effective share relative to the category average (right axis) for firms in that percentile. The values for market shares are weakly increasing by construction

[37] For example, firms in "General Merchandise" and in "Drug Stores and Pharmacies" have much higher national market shares than firms in "Restaurants - Excluding Fast Food".

Figure 3: Effective shares are more uniformly distributed



Notes: The x-axis reports within-category market share percentiles, constructed by ranking firms within each category by their market share and assigning them to up to 100 percentile groups. Percentile 1 corresponds to the lowest market share within a category, and percentile 100 to the highest. The left vertical axis (blue) plots the average ratio of market share within each percentile group to the category average market share, meaning that a value of 1 indicates being at the category average. The right vertical axis (red) plots the average effective share within each percentile group, normalized by the category mean effective share, so that values greater than 1 indicate above-average effective share within a category. Horizontal dashed lines at 1 show the category mean. The figure aggregates observations across categories after within-category normalization, and all firms receive equal weight within percentile groups.

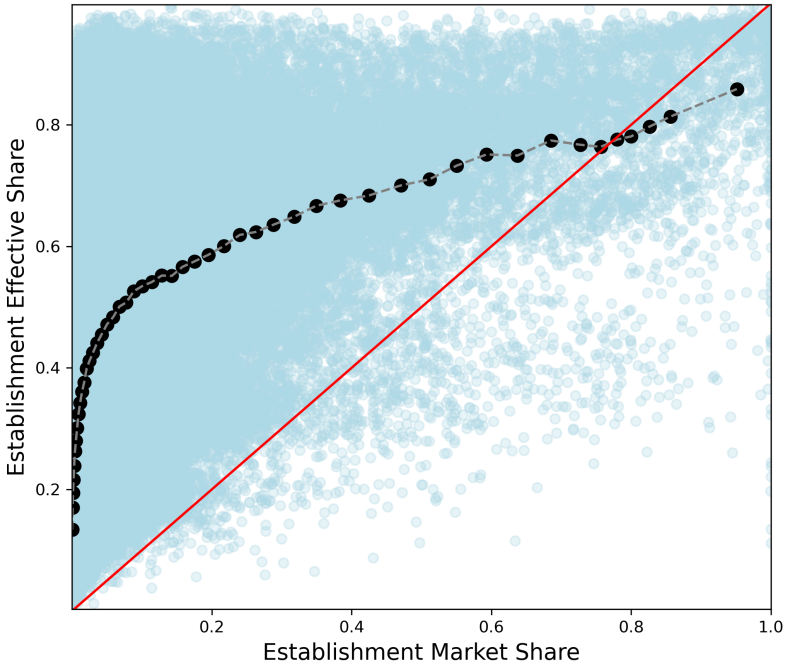
(since the percentiles are defined in terms of market shares). However, the relevant thing to note is the shape of the line – the blue circles show that conventional market shares are extremely right-skewed. The curve is nearly flat through most of the distribution and then rises sharply at the top percentiles, with the top firms having market shares that are nearly 50 times bigger than the average firm in their category. The red squares plot the average effective share for those same firms. By contrast, these shares rise much more smoothly across the distribution. The largest firms, defined in terms of their market shares, capture a larger share their customer's spending, but their effective share is only 0.3 times bigger than that of the average firm in their category.

A natural explanation for the gap between national market shares and effective shares is geographic aggregation: many consumer markets are local rather than national. In sectors such as restaurants and grocery stores, national market shares may therefore be the wrong benchmark, because they average over many distinct local markets. We test this by moving to the local level and asking whether effective shares align more closely with local market shares. If the divergence is mainly a market-definition issue, the two measures should line up much more closely once firms are compared within the local markets where they actually compete.

We explore this by moving to the local level, where an observation is now the combination of a firm identifier and a 3-digit ZIP (ZIP3), which we henceforth refer to as an establishment.[38] For each

[38]In reality, this measure sometimes combines physical establishments within a 3-digit zip code, but it will capture the rele-

Figure 4: Effective shares differ from local market shares



Notes: Each blue dot is an establishment (firm-zip3 combination). The black dots show binscatter relationships with 50 bins, where the bins are defined weighting firms by their total sales. For each bin, we plot the weighted mean effective share against the weighted mean market share, where weights are given by establishment sales.

establishment k of firm j in period t, we compute a local establishment market share, which is defined as the sales at the establishment divided by the total sales in the 3-digit zip code in that category. We also construct a local establishment effective share similar to how we construct the measure at the firm level. We continue to define effective shares over the customer's full spending basket and we account for the spending that is done at different establishments of the same firm when calculating the establishment's effective share. This means that the only difference in effective shares across locations of the same firm are the weights on customers (e.g. the Target in Houston will have different customers than the Target in Chicago).[39]

Figure 4 plots the resulting estimates. The y-axis shows the effective shares for each establishment and the x-axis shows the establishment's share of total sales in each 3-digit zip code in which it operates. Each blue dot is an establishment, while the black series traces binned averages, with each bin representing the same amount of total sales. The red line is the 45-degree benchmark, which is what we would expect the relationship to be if these two measures moved one for one. There are several observations.

vant size of the firm locally. We could instead measure each physical establishment's market share in a ZIP3 or ZIP5, but it is not clear this is economically relevant when some firms have many establishments even in these narrow geographic areas.

[39]Spending at online-only firms is included in customer spending totals even though they do not have establishment-level shares in any given geographic market. However, online only firms do not have a physical location and therefore, do not have a well-defined local market share.

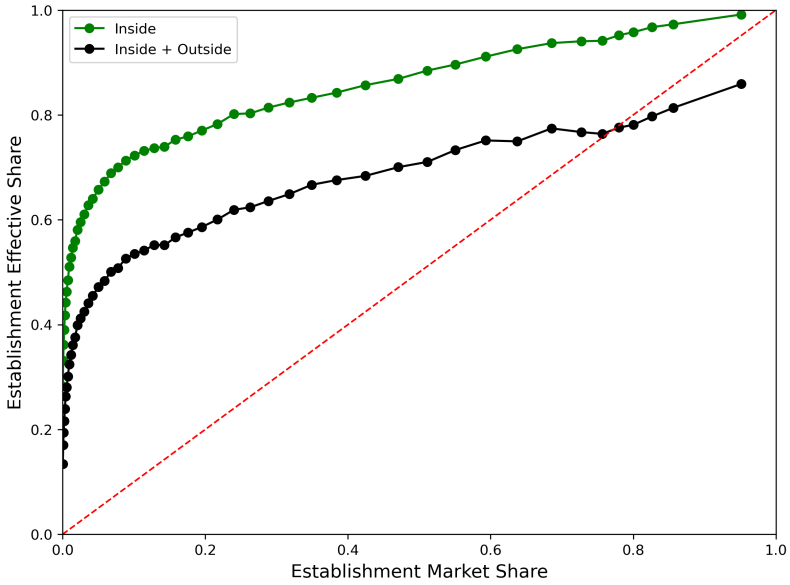
First, the levels of the effective share and market share are much more comparable here than they were when considering national firms. This arises entirely because local market shares are much higher than national market shares. Second, despite the levels being closer, there is meaningful variation in effective shares that is not explained by the firm's share of the local market – even for firms with the same market share, firms differ substantially in their effective shares. Third, the relationship is strongly concave. Establishments with small local market shares have effective shares far above their market shares – those firms are capturing a much larger share of their customer's spending than of the market as a whole. At the other extreme, establishments with very large local market shares have effective shares that are somewhat *below* their market shares. A store can appear dominant in local market shares while still being only a moderate share of its customers' total basket once one accounts for purchases made outside the local market (either in-person or online). In Appendix Figure A3, we show that this same pattern broadly holds when we consider alternative ways of aggregating customer spending, either pooling across all 5 years in our sample or aggregating to synthetic customers.[40]

Indeed, this concave relationship implies that there is generally no single geographic market definition that will make traditional market shares line up with effective shares for most firms. The reason is that firms differ systematically in how geographically "contained" their customers' spending baskets are, either because of heterogeneity in travel patterns across firms (e.g. some stores have parking and others don't) or differences in online footprints. This means that even for firms in the same location and category, any given partition will be too broad for some firms and too narrow for others. This does not preclude selecting a geographic definition that yields close alignment for a particular firm (or a subset of firms), but it implies that no common geographic partition is likely to work particularly well for all firms. Empirically, when we use 3-digit ZIP codes as the market, some firms have effective shares meaningfully above their market shares, consistent with an economically relevant market that is smaller than the 3-digit ZIP, while other firms have effective shares below their market shares, consistent with a broader relevant market. Indeed, Appendix Figure A2 shows that this tension persists under alternative geographic definitions, including states, which are larger, and 5-digit ZIP codes, which are substantially smaller. In particular, when markets are defined at the 5-digit ZIP level, traditional market shares exceed effective shares for the majority of firms, confirming that 5-digit ZIP codes are typically too narrow to capture the relevant market.

To isolate the forces that drive a wedge between effective shares and market shares, we construct an alternative effective share for each establishment using only the spending that its customers do within the region. By construction, this measure lies above the 45-degree line for the same reason that the national effective share was always above the 45 degree line – in a closed local economy, a firm's share of its customers' spending must be at least as large as its share of total market spending. The gap between this within-region effective share and the local market share therefore captures within-market segmentation. The green series in Figure 5 shows that this force is positive throughout the distribution, and especially strong for small and mid-sized establishments. Even when these establishments account for only a small

[40]Consistent with this stability, we show in various simulations that with the sample restrictions we impose, there is likely no mechanical effect from discrete choice.

Figure 5: Understanding the difference between market and effective shares



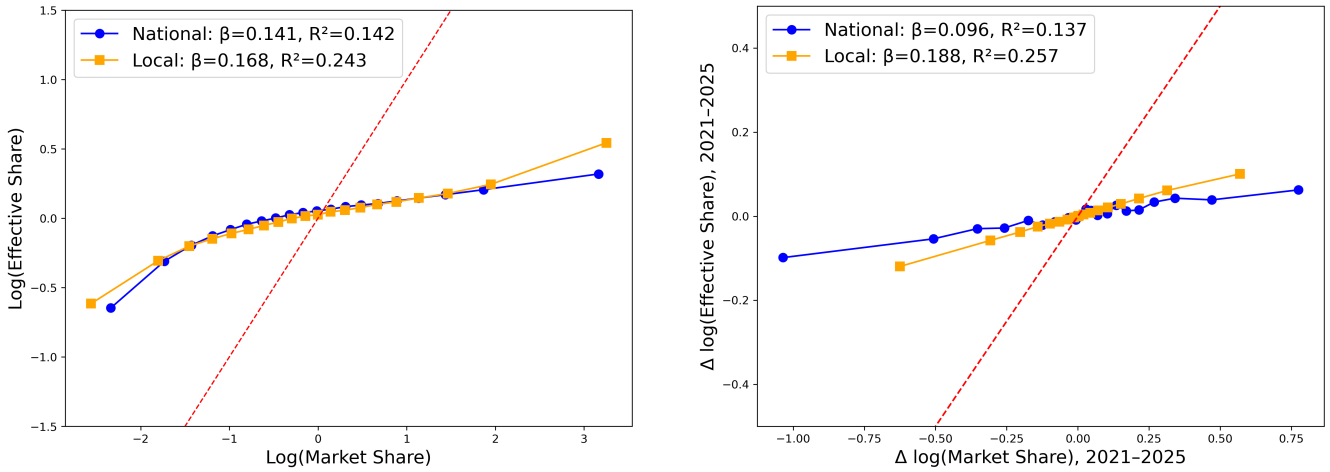
Notes: The black line shows the same binscatter as in Figure 4, where each bin has the same total sales. The green dots show the analogous binscatter but calculating effective shares only using the spending that is done within the zip3.

share of local sales, they capture a substantial share of the within-region spending of the customers they serve.

This within-market segmentation is then offset by spending that customers do outside the region, either online or in person. For single-market firms with no online presence, the overall effective share must be weakly below the within-region measure, since the firm captures none of its customers' spending outside the region. For multi-market firms, the overall effective share could in principle lie either above or below the within-region measure, as it would rise if customers were more likely to shop at the firm outside the region than within it. In practice, however, the curve shifts downward almost uniformly. When customers shop outside their region, they tend to spend less at the firms that dominate their local spending.

The concave relationship between the level of effective shares and the level of benchmark shares also suggests that the relationship between effective shares and market shares is closer to linear in levels. The left panel of Figure 6 shows exactly this, plotting a binscatter of the log of effective shares and the log of market shares. The blue circles show the relationship when the shares are defined for national firms (as in Figure 2 above) and the orange squares show the shares defined for each establishment. As above, we focus just on the cross-sectional distribution in 2025. We see that the relationship is nearly linear, with a slope of either 0.14 nationally or 0.17 locally, meaning that firms with 10 percent higher national market shares have effective shares that are only 1.4 percent higher. The right panel shows that this relationship holds not only in the cross section, but also when looking at changes within the firm between 2021-2025. There, the slope falls slightly to 0.1 when firms are defined nationally and 0.19 when firms are defined locally.

Figure 6: Elasticity between effective shares and market shares



(a) 2025 Levels

(b) Changes from 2021-2025

Notes: This figure presents binscatter relationships between market share and effective share estimated using national and establishment data. The left panel plots the relationship in levels in 2025, while the right panel plots the relationship in changes between 2021 and 2025. In both panels, the horizontal axis reports log market share and the vertical axis reports log effective share. Circles correspond to the national sample, while squares correspond to the establishment sample. In the left panel (levels), log market share and log effective share are demeaned using fixed effects: category fixed effects in the national sample and category × ZIP3 fixed effects in the local sample. The plotted points represent weighted averages within 20 bins of the residualized log market share, where the weight is the logarithm of firm sales in the national sample and the logarithm of establishment sales in the local sample. In the right panel, extreme changes in log market share are trimmed symmetrically at the top and bottom 2.5% of the distribution. As in the left panel, changes are residualized with respect to category fixed effects in the national sample and category × ZIP3 fixed effects in the local sample, using weighted demeaning. The binscatter plots weighted averages within 20 bins of residualized changes in log market share. In both panels, the coefficients and R^2 of the weighted regressions on the underlying data are reported in the legend. The dashed 45-degree line is shown for reference.

If effective shares are not highly correlated with local market shares, are there other firm-level variables that do a better job at predicting effective shares? One natural candidate is sales per customer, which cannot be calculated in traditional firm-level data (e.g. Census data, Compustat) but does not require knowledge of each customer's spending at other firms. While effective shares are clearly related to deepening of sales within customers, they do not move one for one with sales per customer for several reasons. First, sales per customer could change with no change in effective shares if firms get higher income customers but continue to get the same share of their customer's spending. Second, effective shares weight customers by their share of a firm's spending (i.e. customers who spend more at you matter more for the effective share). This means that changes in the composition of spending across customers could result in changes in effective shares while sales per customer stays fixed. Appendix Table A4 shows regression estimates that summarize the relationship between the log of effective share and the log of sales per customer or log number of customers. Perhaps unsurprisingly, we find that effective shares co-vary strongly with sales per customer within category – a 1 percent increase in sales per customer is associated with a 0.5 percent increase in effective shares. Indeed, after conditioning on sales

per customer, the effective shares does not co-vary positively with the number of customers in the firm. However, despite this strong co-movement, there is still substantial variation in a firm's effective share that is not predicted by the firm's average sales per customer, with partial R^2 from these regressions hovering around 0.5 depending on the specification. Therefore, we conclude that sales per customer is a good but not perfect proxy for a firm's effective share.

5 Effective shares predict customer retention

In this section, we provide evidence that — as predicted by variable markup models with customer heterogeneity — effective shares capture information about firms' demand elasticities not captured by their traditional market shares. This validation is important, since it shows that variation in effective shares does not merely reflect noisy measurement. It also helps confirm that the equilibrium cross-sectional variation in effective shares that we measure is actually relevant for capturing firms' demand curvature, and is not instead driven by other confounding forces that correlate with size. Unfortunately, we do not observe prices in our data, and so we cannot estimate price elasticities directly. We instead proxy for the elasticity of demand using a measure of customer retention. The idea underpinning this proxy is that across time, there are shocks to relative prices across firms. If firms with higher effective shares have lower elasticities of demand, then they should have higher customer retention over time.[41]

5.1 Average Customer Retention

We begin by showing average relationships between customer retention over time and effective shares. We define a retained customer as an individual i who purchases from firm j in both period t and period $t+1$[42]:

$$R_{ijt} = \mathbf{1}\{i \text{ buys from firm } j \text{ in both } t \text{ and } t+1\}. \tag{10}$$

Our firm-level retention measure is then the sales-weighted average of R_{ijt}, weighted by each customer's period-t spending at firm j.[43] Local retention rates are analogous averages at the firm $\times$ ZIP3. The sample-wide average retention rate for firm $\times$ ZIP3s is 61%, ranging from 44% for Department and Clothing Stores to 94% for Chiropractors. Our main specifications focus on variation across firm $\times$ ZIP3s since this maximizes the predictive power of traditional market shares and so is conservative relative to specifications using national variation. However, the patterns hold even more strongly when using variation in national firm shares, and using national specifications allows us to include online-only firms.

[41] The underlying identifying assumption is that relative price shocks are of similar magnitude across the effective size distribution. This is not directly testable without price data, but we find similar patterns when re-estimating relationships within subsets of the effective size distribution.

[42] A combination of sample attrition plus infrequent purchases mean that in some categories (e.g. Jewelry stores) many customers do not make any purchases in a category in consecutive years. Thus, in our primary specification, we restrict analysis only to customers who make some purchase (at any firm) in the category in $t+1$, but results are similar if we instead include customers who are active in the dataset in $t+1$, but make no category purchases in $t+1$ as non-retained customers.

[43] Results are similar if we instead equally weight all customers.

We measure the relationship between customer retention and effective shares using versions of the following specification:

$$\text{Retention}_{j,t} = \beta_1 \log \omega_{j,t}^{\text{effective}} + \beta_2 \log \Omega_{j,t} + \gamma_{\text{category,t}} + \epsilon_{j,t}$$

where j indexes the firm $\times$ ZIP3, t is the year, and $\gamma_{\text{category,t}}$ are category by year fixed effects. In these results, we focus on customer retention behavior from 2024 to 2025, and we standardize all regressors by category for ease of interpretation. The theoretical framework predicts that firms with higher effective shares should have more loyal customers and hence lower elasticities, so we expect $\beta_1 > 0$. Moreover, we expect that effective shares are more tightly linked to elasticities than market shares are, so we also expect that $\beta_1 > \beta_2$ and that the R-squared on $\log \omega_{j,t}^{\text{effective}}$ will exceed that on $\log \Omega_{j,t}$.

Table 1 reports results. Column (1) regresses retention on effective shares alone and shows that effective shares strongly predict customer retention. The semi-elasticity of 0.097 implies that within a category, a one-standard deviation increase in a firm's log effective share is associated with a 9.7 percentage point increase in the retention rate. Column (2) regresses retention on market shares alone and similarly shows that there is a positive association but lower R-squared. Column (3) includes both variables and reveals that once we control for effective shares, traditional market share offers little extra predictive power. Columns (4) and (5) explore whether disaggregating a firm's market share into separate intensive and extensive margins changes this conclusion. This is important, because Afrouzi, Drenik and Kim (2025) provide evidence that markups are related to sales per customer but not the total number of customers. In Column (4) we reproduce a similar finding: customer retention has a strong relationship with sales per customer but little relationship with the number of customers.[44] However, Column (5) shows that effective shares continue to have more power for predicting customer retention.[45] Finally, in Column (6) we repeat the specification in Column (3) but including a firm fixed effect to isolate variation within the same firm. This specification controls for *any* fixed firm-level confounders that might drive spurious correlations between market shares and customer retention. We again find that differences in customer retention across the markets in which a firm operates are better predicted by the firm's local effective share than by its local market share.

Table A5 shows various robustness results. In particular, similar patterns hold at the firm level and not just at the firm $\times$ ZIP3 level. Patterns are robust to using the level instead of the log of market shares, and to using equally weighted instead of sales-weighted measures of customer retention. Finally, they are also robust to using our alternative measure of effective shares based on the synthetic customers.

[44] This pattern also helps validate that our customer retention statistic indeed captures variation related to demand elasticities and markups.

[45] For example, the R-squared in Column (5) is substantially higher than in Column (4) but only slightly higher than in Column (3).

Table 1: Sales-Weighted Customer Retention vs. Log Effective Shares and Market Shares

	(1)	(2)	(3)	(4)	(5)	(6)
$\log \omega_{j,t}^{\text{effective}}$	0.097***		0.085***		0.121***	0.080***
	(0.008)		(0.006)		(0.009)	(0.003)
$\log \Omega_t^j$		0.064***	0.026**			0.029***
		(0.015)	(0.012)			(0.001)
$\log(\text{Sales / Customer})_t$				0.049***	-0.036**	
				(0.015)	(0.017)	
$\log(\text{\# of Customers})_t$				0.014	0.016*	
				(0.016)	(0.008)	
Firm FE						Yes
Observations	447,750	447,750	447,750	447,750	447,750	194,985
R^2	0.517	0.413	0.528	0.383	0.535	0.841
Within R^2	0.285	0.130	0.301	0.085	0.311	0.266

Notes: Columns (1)-(3) show that effective shares are better predictors of our retention measure than traditional market shares, in terms of both R^2 and the magnitudes of the coefficients. Column (6) demonstrates that the coefficient on effective share is still larger than the coefficient on market share when we only use variation across establishments within a firm. In columns (4) and (5), we show that effective shares are still a valuable predictor of a firm's customer retention even conditional on the firm's sales per customer and number of customers.
We examine customer retention patterns from 2024 to 2025, and restrict to establishments with at least 200 customers in both of those years. All independent variables are standardized by category. All regressions are weighted by $\log \text{Sales}_{2024}$. We cluster standard errors by industry, except in Column (6) where they are clustered by firm.

5.2 Customer Retention Conditional on Competitor Entry

In this section, we move beyond these average cross-sectional relationships and instead study the response to a specific market change: the entry of new local competitors.[46] We show that if a customer has a high effective share at an incumbent *prior* to entry, then they are less likely to switch to the new firm after it enters.

We identify a new establishment in a particular 5-digit ZIP code based on the first quarter in which a firm records a transaction in that ZIP code in our data (meaning that the firm has an establishment there, not just that it sells to a customer who lives there).[47] We require the following conditions in order to consider an establishment as an "entrant": 12 consecutive months of nonzero transaction activity, at least 25 transactions in every month after entry, and at least $50,000 of sales in its first year after entry. Since all analysis in our paper focuses on customers' annual spending, we restrict the sample to establishments that enter in the first quarter of the year.

Our outcome variable is an indicator for whether a customer i, who purchased from incumbent establishment j in year $t-1$, both (1) did not buy from j in year t and (2) made a purchase at the newly entering establishment in t. We refer to this object as $\mathbf{1}(Switch_{j,i})$. This captures a simple notion of whether customers switched to the new

[46] Importantly, firm entry itself is clearly endogenous. For this reason, all specifications compare across incumbents within a location rather than comparing across markets. This means that identification requires incumbents' customers to have equal exposure to the new entrant, but it does not require the implausible assumption that entry itself is exogenous.

[47] We focus on the first entry of a firm in a 5-digit ZIP rather than entry of additional establishments in ZIP5s where a firm already has a presence. Furthermore, we also drop from our analysis any incumbent establishments that are part of the same firm as the entering establishment (e.g. we do not analyze within-firm switching behavior across establishments).

entrant after it opened. As in the "unconditional" customer retention analysis in the previous section, to deal with attrition concerns, we only consider customers who made some purchase (at any establishment) in the incumbent's category during year t. Unlike the establishment-level variables in the previous section, this switching indicator is defined at the *individual customer* level at every incumbent establishment.

In order to increase statistical power as well as for computational feasibility, we focus this analysis on a specific set of customers who could feasibly shop at the new store if they wanted to.[48] For each entry event, we identify the 5-digit ZIP which had the most customers visit the entering firm in the year of entry, and the exposed sample for that event is then the set of *all* consumers who lived in that 5-digit ZIP in both $t-1$ and t (whether or not they shopped at the new entrant).[49] The underlying rationale is that if an individual's neighbors shop at a new entrant, then it was likely a potential shopping option for that individual as well.

Since the unit of observation is now a customer $\times$ incumbent firm pair, we use each specific customer's lagged effective share at a particular incumbent, $\omega^{\text{effective}}_{i,j,t-1}$, as our main right-hand-side variable of interest.[50] In addition, we include the incumbent's lagged market share. Our main specification is:

$$\mathbf{1}(Switch_{k,j,i}) = \beta_1 \times \omega^{\text{effective}}_{i,k,t-1} + \beta_2 \times \Omega^{\text{market}}_{j,t-1} + \gamma_{\text{entrant}} + \epsilon_{i,j,t}$$

where γ_{entrant} are fixed effects for each entry event.[51] Our estimates of β_1 and β_2 thus reveal whether a customer's probability of switching from an incumbent firm to a newly entering firm is more closely related to that customer's effective share at the incumbent or to the incumbent's overall share of local sales, respectively.

In our baseline specification in Table 2, we use a 10% sample of all identified firm-entry events. This sample corresponds to around 1,000 entry events and more than 35 million exposed customer $\times$ incumbent observations.

Table 2 shows that there is a consistent negative relationship between an individual's effective share at a particular incumbent and their probability of switching to a new firm after its entry into the market. That is, when an incumbent captures a larger share of a customer's total spending prior to entry, the customer is then less likely to leave the incumbent for the newly entering firm. Overall, a customer's effective share ($\omega^{\text{effective}}_{i,j,t-1}$) is more predictive of these shopping effects than is the traditional market share of the incumbent ($\Omega_{j,t-1}$). The baseline sample of entry events is heavily tilted towards restaurants, since this category has a large number of establishments and high churn. However, column (4) shows that similar patterns hold when we restrict to entry events outside this category. Perhaps most importantly, column (5) shows that similar patterns also hold if we include incumbent firm fixed effects. In this specification, identification comes only from comparisons of customers who all shop at the same incumbent firm, but who do so with varying degrees of intensity. This rules out concerns that segmentation might differ across incumbents and generate spurious correlations between effective shares and switching.[52] These within-firm specifications hold constant the degree of competition between the incumbent and entrant and identify only off of within-firm differences in customer shopping intensity.

[48] Computation requires narrowing the set of customers because the sample size in the regression is number of entry events $\times$ the number of exposed customers, which becomes enormous as the number of customers counted as exposed grows. We focus on a location-based measure of exposure since distant stores are unlikely to be relevant options for most customers. Including a broader set of exposed customers in our analysis reduces average switching probabilities but has little effect on the size of regression coefficients *relative* to this average and so does not change our conclusions.

[49] Note that the ZIP with the largest number of customers may or may not be the same as the ZIP of the entrant. Results are similar if we instead use the ZIP of the entrant to identify a set of exposed customers.

[50] In particular, this is the share of a customer's spending in the category in $t-1$ that went to the incumbent establishment.

[51] This fixed effect implies that the regression only identifies off of differential exposure of customers to some specific entry event and not off of variation in exposure to entry arising from firms' endogenous entry decisions.

[52] For example, perhaps high-quality firms do not compete with low-quality entering firms and high-quality firms happen to also have high effective shares. This could produce the observed correlation across firms but not across customers within the same firm.

Table 2: Regressions Relating Customers' "Switching Probability" to Market Shares

	(1)	(2)	(3)	(4)	(5)
$\omega_{j,t-1}^{\text{effective}}$	-0.020***		-0.020***	-0.023***	-0.018***
	(0.001)		(0.001)	(0.002)	(0.001)
Ω_{t-1}^{j}		-0.006***	-0.004***	-0.008***	
		(0.001)	(0.001)	(0.001)	
Including Restaurants?	X	X	X		X
Entrant FE?	X	X	X	X	X
Incumbent FE?					X
Observations	35,347,846	35,347,846	35,347,846	3,970,716	3,454,261
R^2	0.060	0.054	0.060	0.098	0.113

Notes: Columns (1)-(3) show that effective shares are a better predictor of customers' switching patterns in response to firm entry than market shares, and that market shares add virtually zero predictive power on top of effective shares. Column (4) shows that the coefficient on effective share is still larger than the coefficient on market share when we exclude restaurant firms (which make up a large share of our sample), and column (5) shows that our result for effective shares holds up when only using variation between customers of a single incumbent firm for a given entrant.
All standard errors are clustered by entry event. In Column (5), for each entering establishment, we use a 10% random sample of all of the associated incumbents.

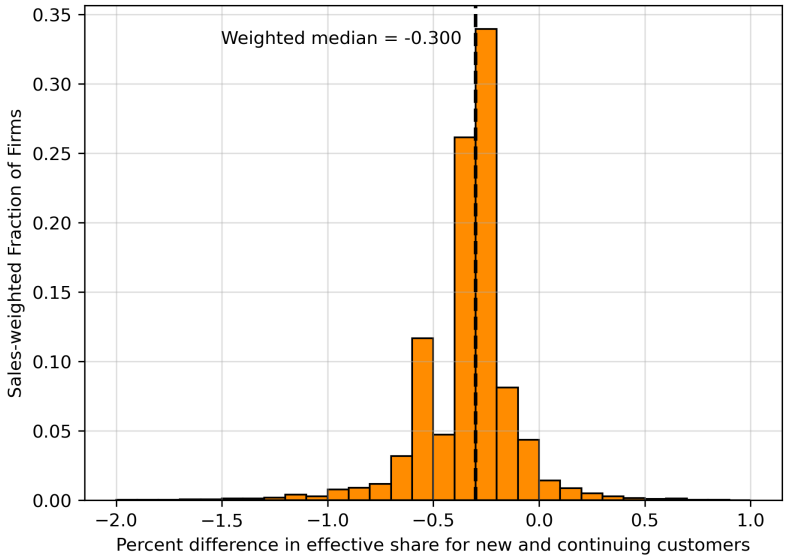
We report results for various robustness checks in Table A6. In one specification, we replace the incumbent's market share with a customer-specific analog, $\Omega_{i,t-1}^{\text{market}}$, which we define as the average (spending-weighted) local market shares of the establishments where that customer shops in $t-1$, an alternate way of thinking about the market share that is relevant for consumers' decision-making. We also run a version of this exercise that restricts attention to incumbents located in the entrant's 3-digit ZIP code, to focus only on incumbent firms that are "most exposed" to a given entry event. Across all specifications, the results remain consistent: customers with higher effective shares are less likely to switch to shopping at newly entering firms. Finally, in Table A7 we re-do our analysis for a subset of "verified" entry events and reassuringly show similar results.[53]

6 Firm Margins of Growth and Effective Shares

In this section, we show that the gap between effective and market shares depends on how firms grow. Growth that expands reach by adding new customers tends to raise market shares more than effective shares, while growth that deepens spending among incumbent customers raises both more proportionately. We then show that two prominent growth channels in recent years, namely geographic expansion and online sales, generate especially muted changes in effective shares, though for distinct reasons. Geographic expansion is associated with lower effective shares because it results in firms primarily by adding new customers in new places, while online growth is associated with lower effective shares because online spending tends to come from lower-effective-share customers.

[53]In particular, we manually confirmed a subset of the entry events identified from our card data using Yelp data as well as manual searches for the firm's entry date using various other web sources. This is important because there are various sources of potential spurious entry in the data like firm name changes (e.g. through mergers or acquisitions), re-configuration of payment terminals, etc.

Figure 7: Effective Shares for New versus Continuing Customers



Notes: This figure shows the distribution of the percent difference in firms' effective shares in 2025 between customers that did not shop at the firm in 2021 but did in 2025 and those who shopped at the firm in both 2021 and 2025. The unit of observation is a firm and firms are weighted by their total sales. The x axis is trimmed to exclude extreme values outside the interval [-200, 100]. The dashed vertical line indicate the median of the weighted distribution.

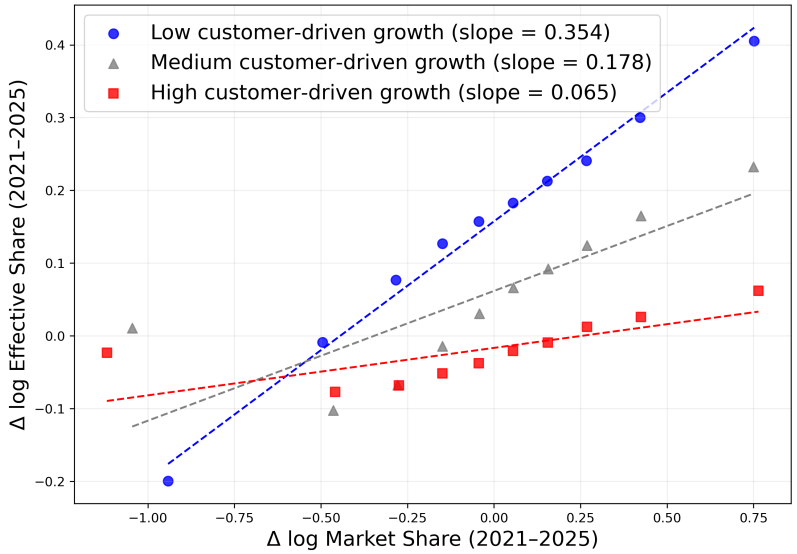
New Customers Are More Marginal Than Incumbent Customers: We begin by comparing the spending shares firms capture from continuing and newly acquired customers. For each firm, we partition customers observed in 2025 into two groups: continuing customers, who purchased from the firm in both 2021 and 2025, and new customers, who purchased from the firm in 2025 but not in 2021. We then compute the firm's effective share separately within each group. Figure 7 plots the distribution across firms of the percent difference between the effective share of new customers and that of continuing customers.

The distribution is sharply shifted to the left, indicating that for the large majority of firms, newly acquired customers allocate a smaller share of category spending to the firm than continuing customers do. The magnitude is economically meaningful: the effective share of new customers is roughly 40 percent lower than that of incumbent customers for the median firm. Thus, although customer acquisition raises total sales and therefore conventional market share, the marginal customers driving that growth tend to be relatively weakly attached to the firm.

This fact has a direct implication for the relationship between effective-share growth and market-share growth. If firms expand primarily by adding customers with relatively low effective shares, then market share should rise by substantially more than effective share. Figure 8 evaluates this implication by grouping firms into terciles based on the fraction of their 2021–2025 sales growth accounted for by net customer acquisition and then plotting the relationship between effective-share growth and market-share growth separately for each group.[54] Consistent with the customer-level evidence, the elasticity is 5 times smaller for firms whose growth is most customer-acquisition-driven than it is for the firms whose growth is least customer-acquisition-driven.

[54]Specifically, we group firms into terciles based on their value of $\frac{|\Delta \log(\text{No. Customers}_i)|}{|\Delta \log(\text{No. Customers}_i)|+|\Delta \log(\text{Sales per Customer}_i)|}$. This measure is bounded between zero and one for all firms and is well-defined even in cases where the firm has very little growth overall or these two terms offset each other.

Figure 8: Elasticity by customer acquisition



Notes: This figure plots the relationship between firms' change in log effective share and change in log market share between 2021 and 2025. The sample is constructed from the national panel and restricted to firms observed in both years. The figure trims the bottom and top 2.5 percent of the distribution of the change in log market share. Firms are divided into terciles based on the share of overall growth attributable to customer growth, defined as $\frac{|g_N|}{|g_N|+|g_S|}$, where g_N is log customer growth and g_S is growth in sales per customer. Within each tercile, points denote weighted bin means from a binscatter of Δ log effective share against Δ log market share, using common bins constructed from the full sample. Dashed lines show tercile-specific weighted linear fits. Weights are the log of firm sales in 2021.

At the same time, the elasticity remains positive even in the group that grew the most on the extensive margin. This implies that customer acquisition is not the only margin through which these firms grow: even firms that expand primarily by adding new customers also deepen sales among their set of customers. The observed elasticity should therefore be interpreted as the net result of two offsetting forces: a dilution force from customer acquisition and a deepening force from stronger sales within those customers.

Geographic Expansion: One important channel for firm growth is geographic expansion. This margin has featured prominently in recent work on the rise of large service-sector firms, which argues that national concentration has increased in part because firms have expanded the number of local markets they serve rather than becoming proportionately larger within each market (Hsieh and Rossi-Hansberg, 2023). Relatedly, recent evidence on consumer goods entrants shows that firms build market share by entering more geographic markets and placing products in more stores within those markets (Argente et al., 2025). In our framework, the natural prediction is that growth through this margin should lead to increases in market share but small increases in effective shares (i.e. a low elasticity). Indeed, Appendix Figure A1 shows that when a firm enters a new market, the effective share of the new establishment is substantially below that of the firm's existing establishments.

The left panel of Figure 9 directly shows the implications of these patterns for the elasticity between effective shares and market shares. The purple squares show the elasticity for the firms that changed the number of 3-digit zip codes in which they operated (11 percent of firms) while the green circles show the same relationship for the rest of the firms whose geographic footprint remained the same over this period. For the same increases in market

share, the rise in the effective share is almost 4 times higher for firms that achieved that rise in national market share within their local markets compared to firms that achieved that same growth by entering new markets. However, note that the elasticity is positive in both cases, meaning that geographic expansion both brings in marginal customers, and deepens relationships within markets in a way that offsets the decline in their effective share that comes from the acquisition of these new marginal customers.

Online Sales A second dominant form of growth over the last several decades is the expansion of online sales. The last decades were characterized by an enormous growth in online sales, especially among the biggest firms. Appendix Figure A4 shows that within this sample of retail industries, online sales has grown from around 10 percent of sales in 2014 to over 25 percent of sales in 2025. There was a pronounced blip in 2020 during the pandemic, but that mostly subsided by 2023 and the economy returned to a trend growth in the online share of sales.

Growth through online sales relative to in-person sales lowers effective shares to the extent that firms capture a smaller share of their customer's spending when they shop online. We find that this is the case. Customers who shop online tend to be richer (Dolfen et al., 2023) and thus spend more at any given firm. However, once you condition on the total spending of the customer, the share of the customer's spending that goes to the firm is meaningfully lower if that spending happened online than if that spending happened in-person. Specifically, Figure 10 shows the estimates of β in each category from a regression relating customer i's effective share in firm j to the customer's mode of purchase:

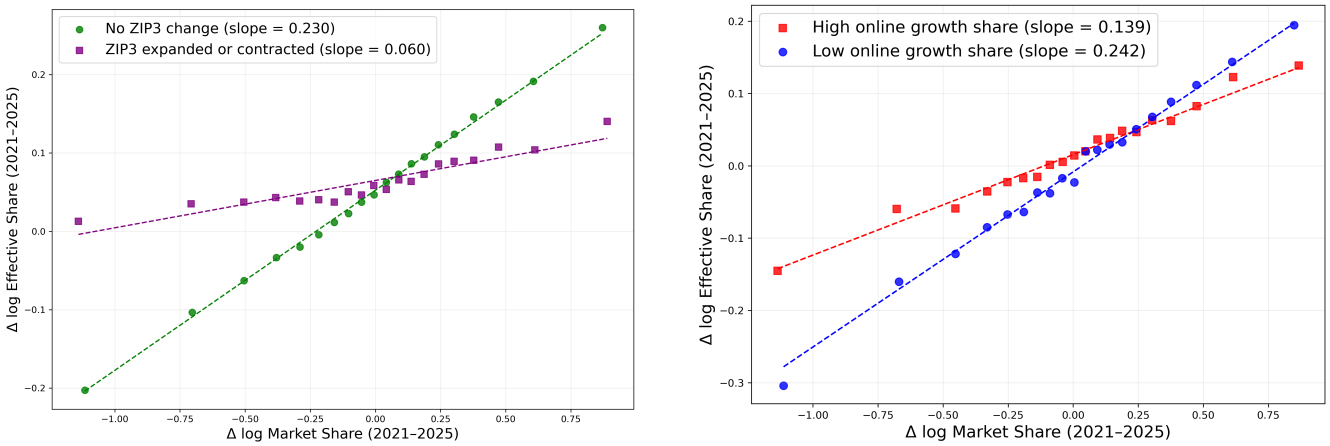
$$\omega_{ijt}^{\text{effective}} = \beta \times \text{online share}_{i,j,t} + \alpha \log(Y_{ijt}) + \gamma_{jt} + \varepsilon_{ijt} \tag{11}$$

where $\omega_{ijt}^{\text{effective}}$ is the share of customer i's spending that goes to firm j, Y_{ijt} is the customer's total spending at firm j in year t, γ_{jt} are firm fixed effects, and "online share" is the fraction of customer i's spending in firm j that is online. For online-only customers, this share is 1. Since we include firm fixed effects in the regression, we isolate variation across customers of the same firm who differ in the share of their spending that is done online. The blue bars show the estimates of β in each category, revealing the difference between the effective share of customers who shop more online versus those who shop more in-person. For the majority of categories, the coefficient is negative, meaning that, conditional on the amount that they spend, customers who shop online have lower effective shares than customers who shop in person. In some categories, like florists, the differences are very large. In other categories, like department and clothing stores, the differences are still there but smaller in magnitude. The one category in which the estimate is positive is grocery stores. Unlike most other online retail, online shopping for groceries is still essentially buying from whatever stores are local, and shopping online is more about saving time than expanding the set of available stores.

These customer-level patterns imply that conditional on growing sales by a given amount, firms would have higher effective shares if those dollars had come from in-person customers than if those dollars had come from online customers. We confirm this pattern in the right panel of Figure 9, where we estimate the elasticity of effective shares with respect to market shares separately for firms whose growth came mostly through online sales and firms whose growth came mostly through in-person sales. We find that the elasticity is meaningfully lower for firms that grew mostly through online sales than for firms that grew mostly through in-person sales, meaning that for the same change in market share, firms that grew by gaining more online customers saw a smaller increase in their effective shares.

Taken together, these results highlight that the growth margins that generate large increases in measured market size, specifically expanding into additional markets, shifting sales online, and acquiring new customers, do

Figure 9: The Nature of Firm Growth and Effective Shares



(a) Geographic Expansion

(b) Online Spending

Notes: Panel A splits firms according to whether the number of ZIP3 markets in which the firm operates changes between 2021 and 2025. Panel B restricts the sample to retail industries and splits firms by the share of total sales growth attributable to online sales growth. Online growth share is defined as $\frac{|g_{\text{Online}}|}{|g_{\text{Online}}|+|g_{\text{Offline}}|}$, where g_{Online} is growth in log online sales and g_{Offline} is the residual component of log sales growth not accounted for by online sales growth. The sample in Panel B is further restricted to firms with online growth share below 0.2 or above 0.8, labeled "Low online growth share" and "High online growth share," respectively. In both panels, the bottom and top 2.5 percent of the distribution of $\Delta \log$ market share are trimmed. Points denote weighted bin means from binscatters of $\Delta \log$ effective share against $\Delta \log$ market share, using common weighted bins constructed from the estimation sample in each panel. Dashed lines show group-specific weighted linear fits. Weights are the log of firm sales in 2021.

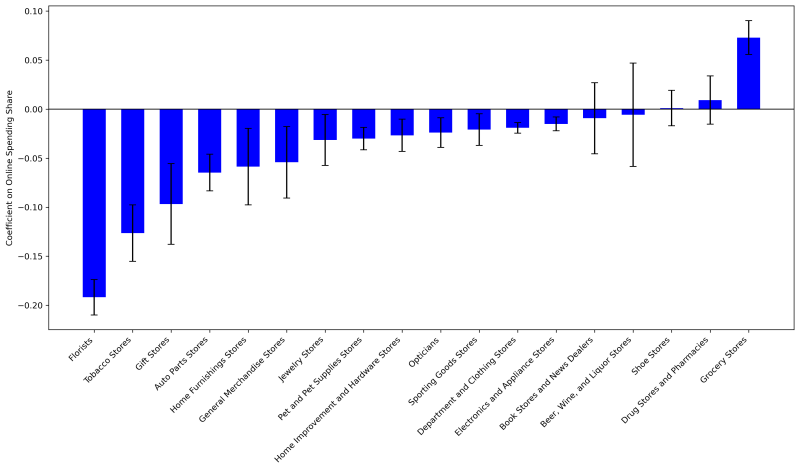
not generally translate into commensurate increases in effective size. These facts help explain why effective shares and market shares do not move together strongly.

7 Concentration Trends

What do our results imply for recent trends in size-based market power? A large empirical literature documents an increase in the concentration of national sales. However, our theoretical framework tells us that size-based market power depends not on a firm's national sales share Ω_j but instead on its effective market share ω_j. In particular, the sales-weighted average elasticity of demand in the economy is given by $\sum_j \Omega_j \varepsilon_j = \eta - (\eta - \theta) \sum_j \Omega_j \omega_j$. In a representative customer economy, $\omega_j = \Omega_j$ for all firms, so that $\sum_j \Omega_j \omega_j = \sum_j \Omega_j^2 = HHI$, and the elasticity of demand indeed declines with traditional measures of market concentration. However, our evidence thus far shows that $\Omega_j \neq \omega_j$. In this section, we show that empirical trends in effective concentration $\overline{\omega} = \sum_j \Omega_j \omega_j$ differ substantially from trends in traditional concentration measures.

For this analysis we turn to our historical panel data. As discussed in Section 3, this data covers a sample from 2014-2025. This historical panel measures total spending by individual cards in each zip code, but it only further disaggregates this spending for a subset of relatively large, "named firms". Because this historical panel does not permit constructing full customer baskets over all firms, we implement a top-firm analog of customer concentration: we compute the average effective share for customers of the top 4 firms in each category (which

Figure 10: The Relative Effective Share of Online Customers



Notes: This figure reports industry-specific estimates of β from Equation 11, estimated separately for each category. For each category, the regression specification includes firm fixed effects and controls for log total firm spending, and is estimated using data from 2025. Vertical lines indicate 95 percent confidence intervals, constructed using standard errors clustered at the firm level.

we term ES4) and compare this to the conventional sales share of these top 4 firms (CR4). In Appendix Figure A5, we provide several exercises to validate this data focus. First, to validate that despite its limitations, the historical panel captures similar information as the primary panel, we show that CR4s and ES4s align closely in the two data sets in overlapping years. Second, while we are primarily interested in $\overline{\omega}$ (the dollar-weighted average effective share of all firms), we can only measure ES4 (the dollar-weighted average effective share of the top 4 firms) in our historical panel. However, we show that in our primary data set, there is a very strong relationship between ES4 and $\overline{\omega}$ both at the national and local level, suggesting that trends in ES4 are informative for trends in $\overline{\omega}$.

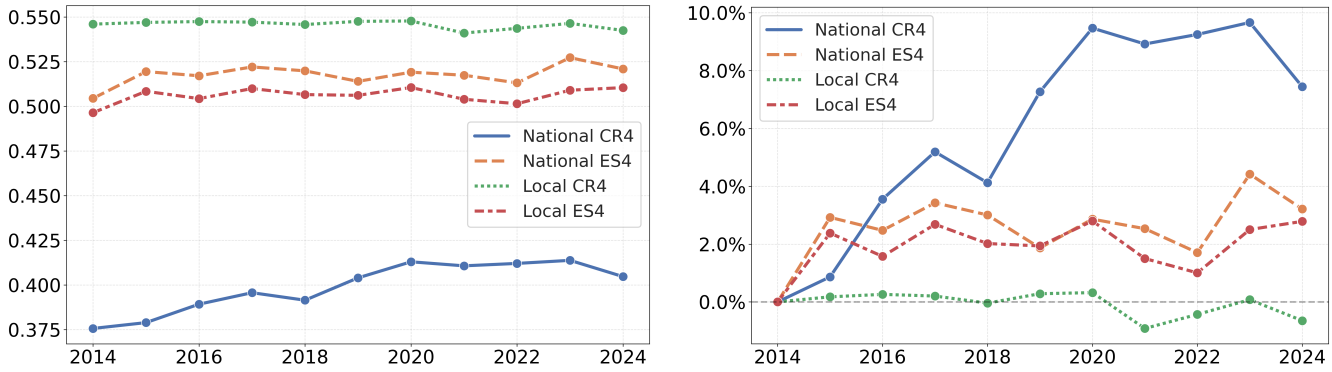
Figure 11 shows concentration trends in national and local CR4s and ES4s in this historical panel. The left figure shows the levels of concentration over time while the right panel normalizes by the values in 2014 to highlight differences in growth. Focusing first on national patterns, the blue line shows that there is substantial growth from 2014 to 2024 in the national sales share of the top 4 firms.[55] The national CR4 increases by around 10 percent, from 0.376 to 0.404.[56] However, the orange line shows that average effective shares only increase by around 3.5 percent, from 0.504 to 0.521. Thus, the growth in national sales concentration that has been extensively documented in past research overstates the actual rise in firms' effective size-based market power.

Turning to local patterns, the green line shows that local sales concentration has not increased even as national sales concentration has risen. This divergence between local and national trends has been documented in prior

[55] All of these series average across the industry CR4 using fixed category weights over time, since we are interested in changes in market power within category rather than composition changes in the sales share of different categories over time. However, Figure A6 shows that similar patterns obtain when averaging industries based on their sales share in each year. The main difference is that there is a sharp increase in concentration in 2020 since spending shifts from an unconcentrated sector (restaurants) towards more concentrated sectors (grocery stores and general merchandise stores).

[56] In addition to the validation exercises in Appendix Figure A5, we can also compute trends in CR4 from publicly available census data. When we calculate these trends for consistent categories in census data for the closest available years, we find remarkably trends. In the Census from 2012 to 2022, CR4 rises from 0.37 to 0.41. In credit card data CR4 rises from 0.39 to 0.42 from 2014 (the first available year) to 22. Note also that this 0.39 to 0.42 change is not identical to what is reported in Figure A6 since not all categories in the credit card data are available as consistent NAICs series in census data.

Figure 11: Concentration levels and growth trends in the balanced sample



(a) Concentration levels trend

(b) Concentration growth trend

Notes: The figure reports national and local concentration measures from 2014 to 2024. CR4 denotes the four-firm concentration ratio, and ES4 denotes the average effective share of the top four firms. Effective shares are constructed from the historical panel, with customers defined at the card level. National series are computed by first constructing industry-level concentration measures and then averaging across industries using fixed 2024 industry total sales weights. Local series are computed by first aggregating concentration measures across industry–ZIP3 markets using fixed 2024 ZIP3-level sales weights within industry and then averaging across industries using the same fixed 2024 industry total sales weights. The sample is restricted to a balanced panel of industries observed in every year from 2014 through 2024. Panel A plots levels. Panel B plots growth rate relative to 2014; for each measure X, the plotted value in year t is $\log(X_t) - \log(X_{2014})$.

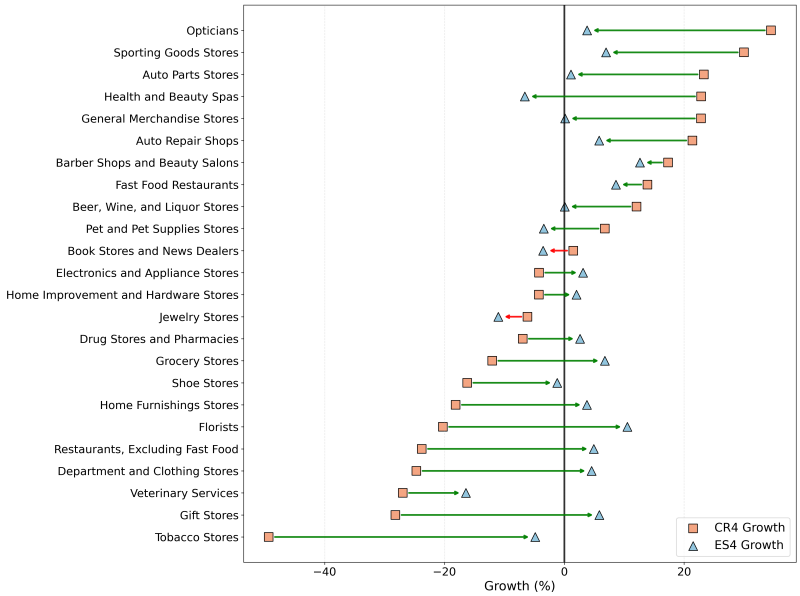
research (Rossi-Hansberg, Sarte and Trachter, 2021), and it has in turn led to questions about whether economically relevant market power has actually increased with the rise of national sales concentration. Economically relevant boundaries for competition likely vary between categories and locations (and even across firms within categories and locations), which makes it hard to interpret which of these diverging trends is more relevant for market power. In contrast, measuring effective shares requires no assumption about the geographical boundaries of competition and instead infers this directly from the data based on the revealed shopping patterns of firms' customers. Indeed, we find that average local effective shares in red are extremely similar to average national effective shares in orange, both in levels and trends.[57] Thus, our new empirical measure resolves the tension in the literature between national and local trends in average concentration trends.

It is also worth noting that the average trends shown in in Figure A6 mask substantial heterogeneity across markets. Indeed, the attenuated relationship between trends in sales shares and trends in effective shares is much more apparent when looking within particular categories and local markets. Figure 12 compares 2014 to 2024 category-level growth in traditional concentration (national CR4) vs effective shares of these same firms (ES4) from 2014 to 2024. Some categories see rising traditional CR4 (orange squares to the right of zero) while other categories have declining concentration. However, comparing to the blue triangles industry-by-industry shows that 22 out of 24 industries see more muted trends in effective shares than in sales shares. In particular, the green arrows show the 22 industries where switching from CR4 to ES4 moves the measured trend closer to zero while the 2 red arrows show those where switching from CR4 to ES4 moves the measured trend further from zero.

Similar attenuation arises also when looking at local markets. To show this, we begin by splitting the sample

[57] By construction, every individual firm has a national effective share that is the sales weighted average of its local effective shares. This means that local ES4 and national ES4 will be identically equal if the top 4 firms nationally are identical to the top 4 firms in every location, but can otherwise potentially differ.

Figure 12: Trends in National Concentration by Industry



Notes: The figure reports the change between 2014 and 2024 in the four-firm concentration ratio (CR4) and the average effective share of the top four firms (ES4) by industry at the national level. Effective shares are constructed from the historical panel, with customers defined at the card level. For each variable X, growth is defined as $100 \times \log(X_{2024}/X_{2014})$. The sample is restricted to a balanced industry panel consisting of industries observed in every year from 2014 through 2024 and with at least eight named firms. Industries are ordered by CR4 growth.

of zip3 X categories into those where the CR4 grows from 2014 to 2024 (45% of markets) and those where it shrinks from 2014 to 2024 (55%) of markets. We then compute the average growth in CR4 and ES4 separately for these rising and shrinking CR4 markets. Figure 13 shows that the average increase in CR4 for increasing markets is quite substantial while the corresponding increase in ES4 is negligible. Similar attenuation also occurs among declining markets.[58]

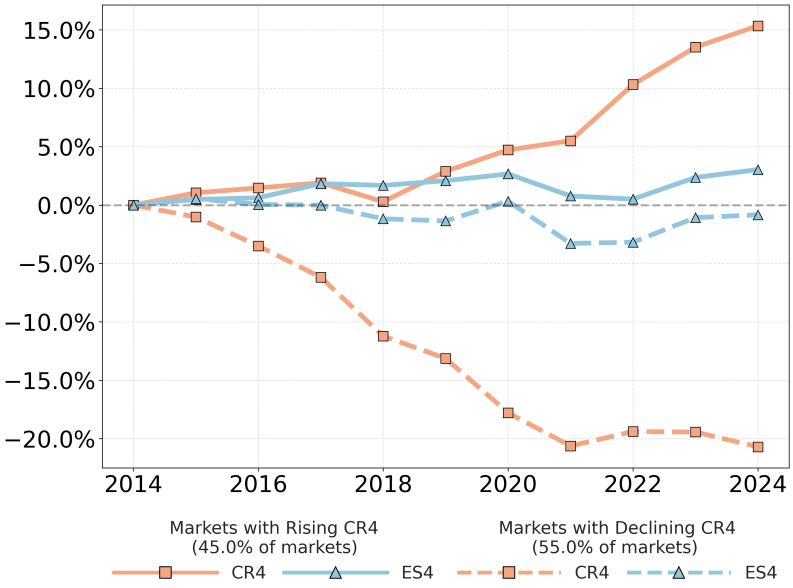
Taken together with the within-firm evidence on growth margins, these results imply that the rise of "superstar" firms has primarily reflected an expansion in reach rather than a broad-based increase in within-customer dominance, and thus, even within the variable-markup lens that maps market power to "size," there is little evidence of a broad-based increase in size-based market power.

8 Conclusion

This paper argues that in variable-markup environments with heterogeneous customers, conventional market shares are generally not the relevant notion of "size" for size-based market power. The pricing-relevant object is instead a firm's *effective share*, which captures the extent to which the firm dominates the spending baskets of the customers it serves. Using linked customer–firm transaction data covering a broad set of consumer-facing

[58] Importantly, this reflects a weak correlation between ES4 and CR4 changes rather than constant ES4 within markets. Many markets experience substantial changes in ES4, but these markets do not necessarily correspond to those with substantial changes in CR4.

Figure 13: Trends in Local Concentration by Industry



Notes: This figure plots the evolution of concentration in local industry–ZIP3 markets, separately for markets whose CR4 rises over the sample period and markets whose CR4 declines. Effective shares are constructed from the historical panel, with customers defined at the card level. Markets are classified by comparing final-year CR4 to initial-year CR4. For each group, the plotted series are weighted averages of cumulative log changes, $\log(X_t/X_0)$, where X is either the four-firm concentration ratio (CR4) or the average effective share of the top four firms (ES4), and X_0 is the market's first observed value. Weights are fixed and equal to the log of average top-4 named firm sales. Solid lines denote markets with rising CR4, dashed lines denote markets with declining CR4, and legend shares report each group's share of total fixed weight.

industries, we measure effective shares at scale and show that they differ sharply from both national and local market shares. In the cross section, effective shares imply a substantially less skewed size distribution: market shares tend to overstate the dominance of large firms and understate the dominance of small firms. We use these measurements to revisit the interpretation of "superstar" growth and concentration trends. We show that effective size responds weakly to changes in market share because the dominant growth margins of the past decade, namely geographic expansion, e-commerce, and new-customer acquisition, raise sales by expanding reach rather than by increasing within-customer dominance. As a result, customer-based concentration measures implied by the theory move substantially less than standard concentration measures based on market shares. Taken together, these facts imply that even within the variable-markup lens that maps market power to firm "size," there is little evidence of a broad-based increase in size-based market power.

References

Afrouzi, Hassan, Andres Drenik, and Ryan Kim. 2025. "Concentration, Market Power, and Misallocation: The Role of Endogenous Customer Acquisition." *NBER Working Paper 31415*.

Argente, David, Doireann Fitzgerald, Sara Moreira, and Anthony Priolo. 2025. "How Do Entrants Build Market Share? The Role of Demand Frictions." *American Economic Review: Insights*, 7(2): 160–176.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review*, 98(5): 1998–2031.

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*a*. "The Fall of the Labor Share and the Rise of Superstar Firms." *The Quarterly Journal of Economics*, 135(2): 645–709.

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

Baqaee, David Rezza, and Emmanuel Farhi. 2020. "Productivity and Misallocation in General Equilibrium." *The Quarterly Journal of Economics*, 135(1): 105–163.

Batch, Andrea, Benjamin Bridgman, Abe Dunn, and Mahsa Gholizadeh. 2024. "Consumption Zones." *Journal of Economic Geography*, 24(2).

Benkard, C. Lanier, Nathan H. Miller, and Ali Yurukoglu. 2025. "The Rise of Market Power and the Macroeconomic Implications: Comment." *NBER Working Paper 3145*.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *The Quarterly Journal of Economics*, 135(2): 561–644.

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.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2023. "How Costly Are Markups?" *Journal of Political Economy*, 131(7): 1619–1675.

Einav, Liran, Mariana Guido, and Pete Klenow. 2026. "Customer Overlap and Diversion Ratios." *NBER Working Paper 34693*.

Hsieh, Chang-Tai, and Esteban Rossi-Hansberg. 2023. "The Industrial Revolution in Services." *Journal of Political Economy*.

Klenow, Peter J, Jonathan D Levin, Raviv Murciano-Goroff, and Liran Einav. 2022. "Customers and retail growth." Working Paper.

Mongey, Simon, and Michael E. Waugh. 2025. "Pricing Inequality." National Bureau of Economic Research NBER Working Paper 33399.

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

Steve Bond, Arshia Hashemi, Greg Kaplan Piotr Zoch. 2021. "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data." *Journal of Monetary Economics*.

Syverson, Chad. 2019. "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives*, 33(3): 23–43.

Traina, James. 2018. "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements." The University of Chicago Booth School of Business, George J. Stigler Center for the Study of the Economy and the State Working Paper 272.

A1 Appendix

A1.1 Data Description Details

A1.1.1 Details of the Transaction data to Transunion Credit Records

First, the Transunion data had to be cleaned to ensure each card matched to only one person each month. This procedure was created by the credit card company, which we followed for years that had yet to be de-duplicated. Each person-card couple is noted with what type of account it is; examples include an individual account or an authorized user account. A hierarchy of account types was created and used for de-duplication. The hierarchy is as follows: an individual account, the maker of the account (i.e., the primary account holder, but has a guarantor), joint contractual liability for an account, an authorized user/guarantor/participant, an undesignated account type, and finally a terminated or deceased account.

Then, the Transunion data was merged with the transaction-level data on the card-quarter level for 2021 and 2022, and on the card-month level for 2023 onward. Transunion data was provided quarterly data for 2021-2022, and monthly data from 2023 on.

After the initial merge, a one-to-one match between cards and people was forced over the entire dataset, as some cards had varied owners over time. First, the person who appeared as the card owner for the longest period was designated the cardholder. In the case of a tie, the person with the maximum number of transactions was selected. Finally, in the case of another tie, the person who spent the most on the card was then selected as the cardholder.

Finally, some cards that appeared in the transaction data for a longer period of time than they appeared in the Transunion data. Thus, for all cards where this was the case, we did a card-level merge between Transunion's person identifier and the unmatched months of transaction level data. Since a one-to-one match of people to cards was already required, there are no concerns about the person-card match. Then, each newly matched month's demographics data was filled in with information from the person's nearest month in Transunion.

A1.2 Validation of Entry Events

We use two methods to verify the estimated dates of firm entry in our transaction data. Between the two, we identify a total of 220 "verified" entering establishments.

1. **Yelp data**: For some of the entrants identified in the transaction data, we use both exact and fuzzy matching to find the date of the store's first review in the Yelp Academic dataset, a subset of Yelp's data with information on businesses' locations and review histories for some establishments. We consider an entry date in our transaction data to be "verified" with the Yelp Academic table if the date of the first transaction activity is less than four months away from the date of the entrant's first review on Yelp.
2. **Manual searches**: For other entering firms that are not present in the Yelp Academic table, we attempt to manually confirm their opening dates with a general Google search. We look for indicators such as the date of the store's grand opening announcement on its Facebook page or other social media accounts (if one exists), the first review for that store on Yelp (the whole site, not the subsample mentioned above), or news articles mentioning the opening of the new store.

A1.3 Model Details

This appendix section provides the supporting derivations for Section 2.

Consider utility

$$U_{ijt} = -\ln(p_j) + \phi_j + \mu_{ij} + \zeta_{ijt}, \qquad \mu_{ij} = \lambda_i \psi_j, \tag{12}$$

where μ_{ij} is fixed over time and ζ_{ijt} follows the same nested-logit generalized extreme value structure as in the main text. Define

$$v_{ij} \equiv \phi_j + \mu_{ij} - \ln(p_j) = \phi_j + \lambda_i \psi_j - \ln(p_j).$$

Then the within-category choice probability is

$$\rho_{ij} = \frac{\exp\{\eta v_{ij}\}}{\sum_{k\in c}\exp\{\eta v_{ik}\}}. \tag{13}$$

The category inclusive value is

$$I_{ic} = \frac{1}{\eta}\ln\left(\sum_{k\in c}\exp\{\eta v_{ik}\}\right), \tag{14}$$

so category demand can be written as

$$Q_{ic} \propto \exp\{\theta I_{ic}\}, \qquad q_{ij} = \rho_{ij} Q_{ic}. \tag{15}$$

Because $\partial v_{ij}/\partial \ln p_j = -1$ and $\partial v_{ik}/\partial \ln p_j = 0$ for $k \neq j$, the within-category derivative is

$$-\frac{\partial \ln \rho_{ij}}{\partial \ln p_j} = \eta(1-\rho_{ij}). \tag{16}$$

Also,

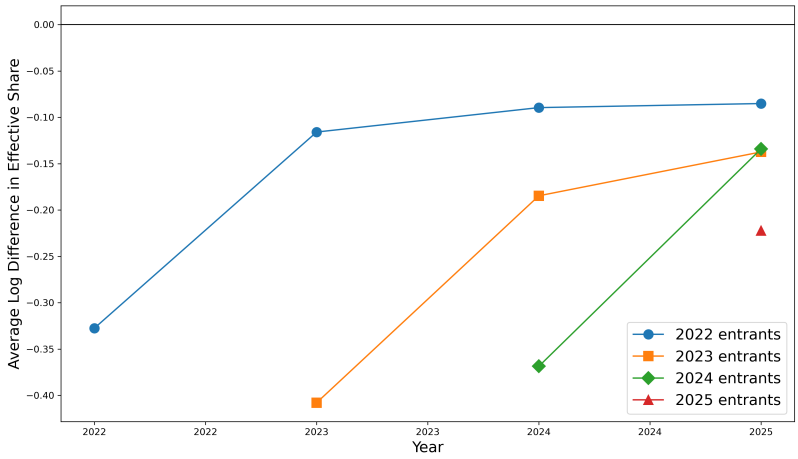
$$\frac{\partial I_{ic}}{\partial \ln p_j} = -\rho_{ij}, \qquad -\frac{\partial \ln Q_{ic}}{\partial \ln p_j} = \theta\rho_{ij}. \tag{17}$$

Therefore,

$$\varepsilon_{ij} \equiv -\frac{\partial \ln q_{ij}}{\partial \ln p_j} = -\frac{\partial \ln \rho_{ij}}{\partial \ln p_j} - \frac{\partial \ln Q_{ic}}{\partial \ln p_j} = \eta(1-\rho_{ij}) + \theta\rho_{ij} = \eta - (\eta-\theta)\rho_{ij}. \tag{18}$$

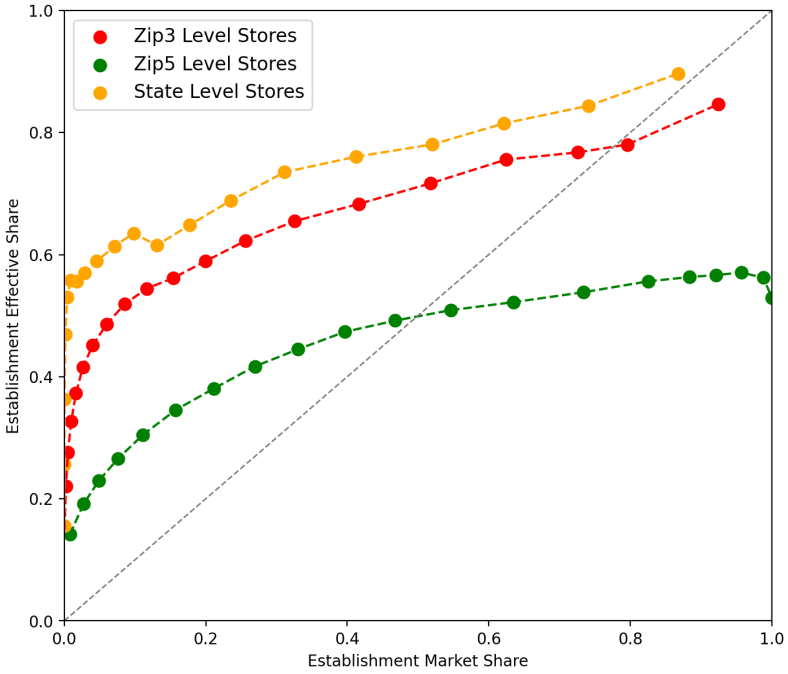
Note that the fixed match-specific term in the utility function (μ_{ij}) provides a rationale for pooling transactions over time to reveal ρ_{ij}. Specifically, fix firm primitives (p_j, ϕ_j, ψ_j) and consider repeated purchase opportunities for household i in category c. If μ_{ij} is fixed over time and ζ_{ijt} is i.i.d. across t, then the realized long-run purchase frequency at firm j converges to the latent choice probability ρ_{ij}. Hence pooling transactions within a customer over time recovers the persistent spending shares that are relevant for demand elasticities.

Figure A1: Growth through Geographic Expansion: New Establishments



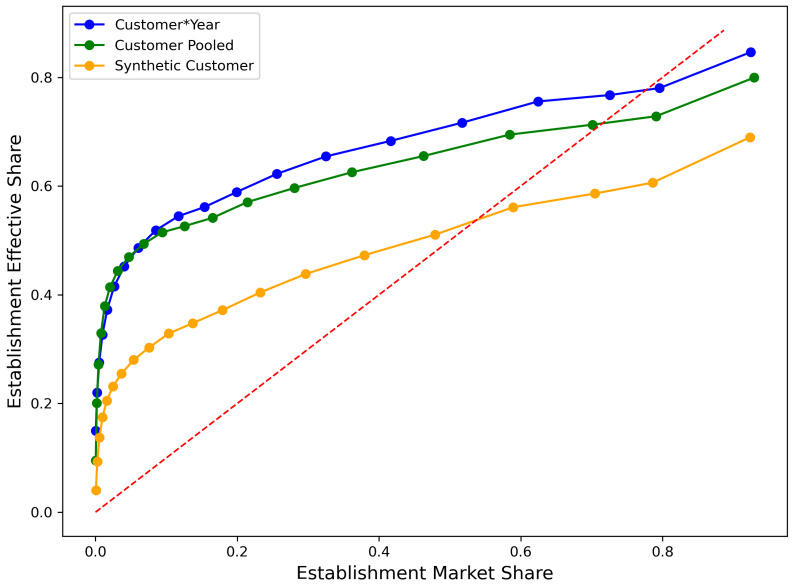
Notes: This figure presents differences in effective market share between newly entered and incumbent ZIP3 markets within the same firm. The sample consists of multi-establishment firms that operated in at least one ZIP3 market in 2021 and subsequently entered at least one additional ZIP3 market between 2022 and 2025. ZIP3 entry cohorts are defined by the first year in which a firm appears in a given ZIP3 market. For each firm and calendar year, effective market shares are averaged separately across (i) incumbent ZIP3 markets in which the firm operated in 2021 and (ii) ZIP3 markets entered in a given cohort year. The plotted outcomes are log differences between these entrant and incumbent averages, computed within firm and then aggregated across firms weighting firms by the share of sales in 2021.

Figure A2: Robustness: Alternative Geographic Market Definitions



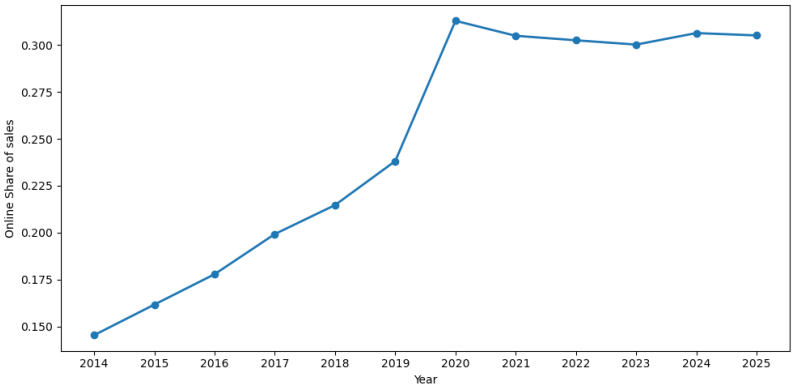
Notes: The dots are binscatter relationships with 20 bins, where the bins are defined by weighting firms by total establishment sales. Each bin plots the mean market effective share against the mean market share. Data is from 2025 and only includes in-person establishments with at least 200 customers. Zip5 Level Stores are establishments and markets defined at the zipcode level. Zip3 Level Stores are establishments and markets defined at the zip3 level. State Level Stores are defined establishments and markets defined at the state level.

Figure A3: Patterns by the level of customer aggregation



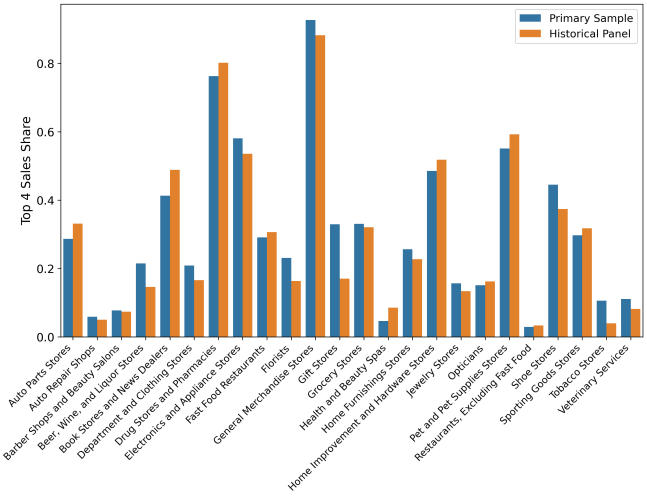
Notes: This figure shows binned scatterplots of effective snares and market shares computed using three distinct aggregation methods: pooling all spending within a year for each customer (our baseline approach), pooling all spending across the years 2021-2025 for each customer, or using our synthetic customer definitions that group together customers with similar demographics in a location. In all three cases, small establishments (as measured by their traditional market shares) tend to have effective shares that are much larger than their market shares, while the largest establishments have effective shares that are below their market shares.

Figure A4: The Growth in Online Spending in Retail Industries

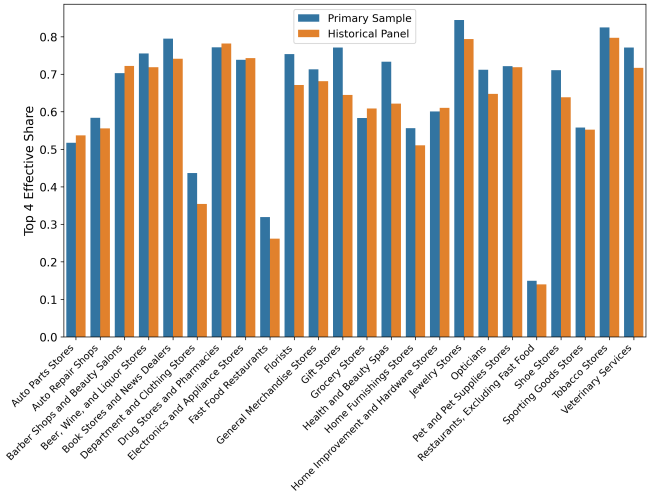


Notes: This figure plots the annual share of spending conducted online in the historical panel across retail industries from 2014 through 2025. Online spending share is measured as the card-not-present ("CNP") share of total sales, aggregated across retail industries in each year. The retail industry here includes Shoe Stores; Gas Stations; Sporting Goods Stores; Department and Clothing Stores; Electronics and Appliance Stores; Home Furnishings Stores; Drug Stores and Pharmacies; Auto Parts Stores; Grocery Stores; General Merchandise Stores; Pet and Pet Supplies Stores; Florists; Jewelry Stores; Beer, Wine, and Liquor Stores; Opticians; Home Improvement and Hardware Stores; Book Stores and News Dealers; Gift Stores; Used Merchandise Stores; and Tobacco Stores.

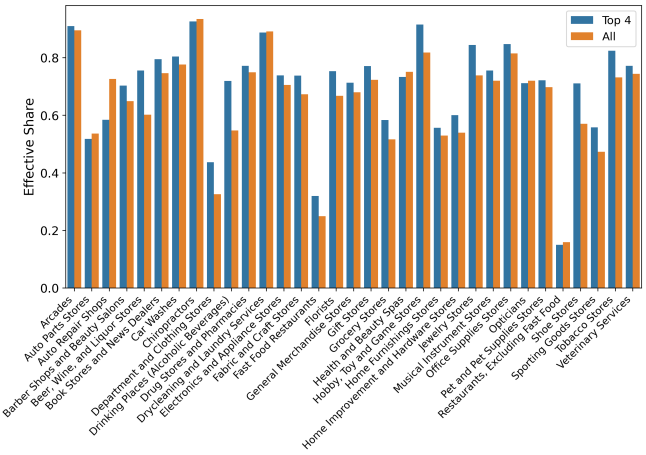
Figure A5: Validation of Historical Panel Data



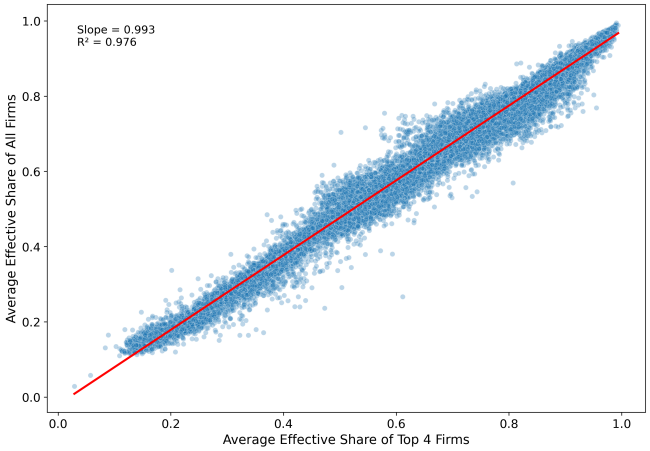
(a) Top 4 Sales Shares: Primary vs Historical Sample



(b) Average of Top 4 Effective Shares: Primary vs Historical Sample



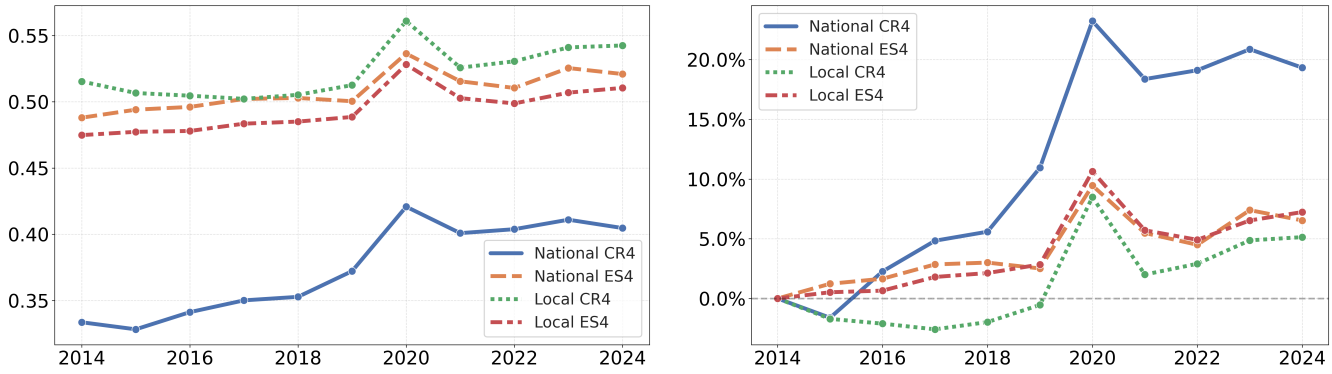
(c) Top 4 vs All Firms Average Effective Share: National



(d) Top 4 vs All Firms Average Effective Share: Local

Notes: Panel A compares the four-firm concentration ratio (CR4) across industries in 2022 for two samples: the Primary Sample and the Historical Panel. For each industry, CR4 is defined as the combined sales share of the four largest firms. Panel B compares, for the same industries and year, the average effective share of the top four firms (ES4) across the two samples. Panels A and B include only industries observed in both datasets with non-missing values for the relevant measure. Panel C compares, by industry in 2022, ES4 with the average effective share computed over all firms in the national sample. "Top 4" denotes the average effective share of the four largest firms in each industry, while "All" denotes the corresponding measure computed over the full firm distribution. Panel D plots, at the local-market level in 2022, the relationship between ES4 and the average effective share computed over all firms. Each observation is at firm x ZIP3 x category level. The solid line shows the fitted linear relationship, and the reported slope and R^2 are from a regression of the average effective share of all firms on ES4. Panels C and D include only observations with non-missing values for both measures.

Figure A6: Concentration levels and growth trends w/ time-varying category weights

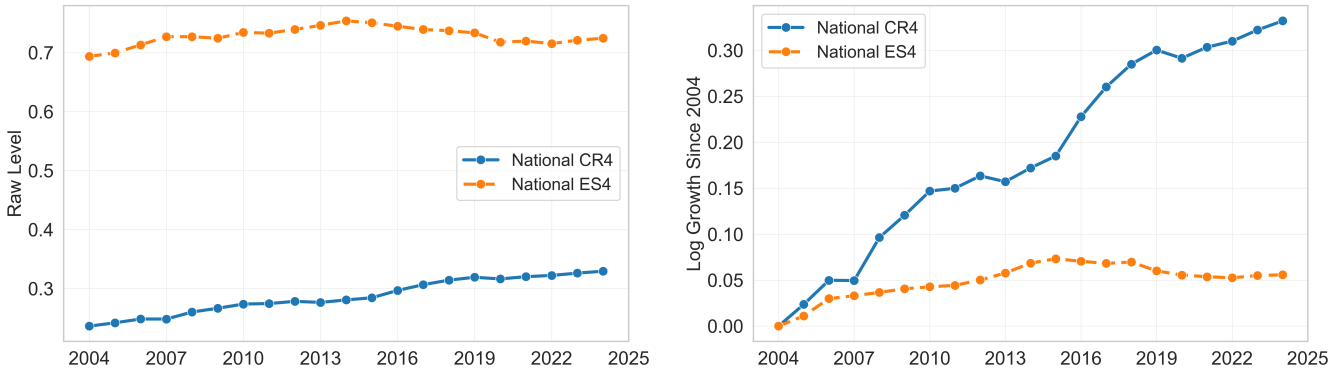


(a) Concentration levels trend

(b) Concentration growth trend

Notes: Panel A reports annual weighted averages of national and local concentration measures from 2014 to 2024: the four-firm concentration ratio (CR4) and the average effective share of the top four firms (ES4). National series are constructed by averaging industry-level concentration measures using time-varying industry total sales weights. Local series are constructed by first averaging across industry–ZIP3 markets using time-varying ZIP3-level sales weights within industry and then averaging across industries using the corresponding time-varying industry weights. Panel B reports cumulative log changes in each series relative to 2014, computed as $\log(X_t) - \log(X_{2014})$. The local series are based on the unbalanced panel.

Figure A7: Concentration levels and growth trends in Nielsen Data

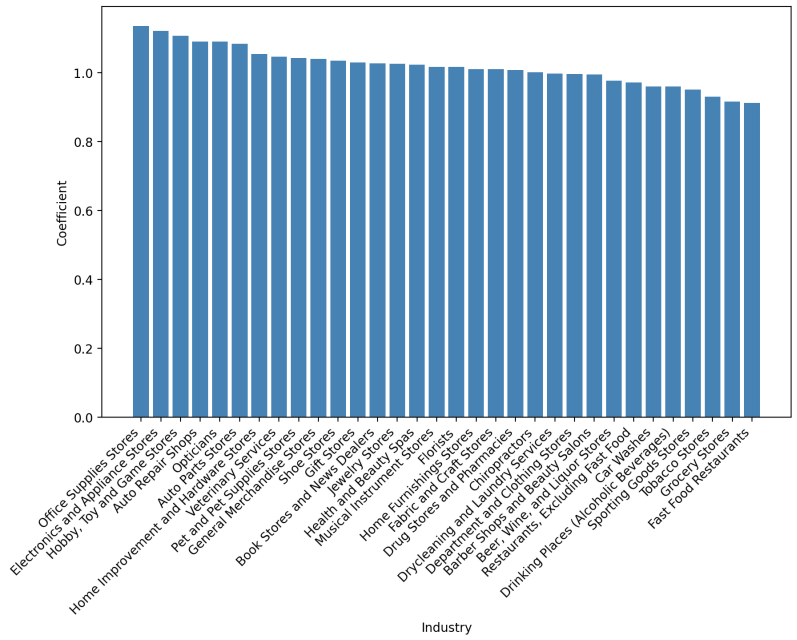


(a) Concentration levels trend

(b) Concentration growth trend

Notes: Panel A plots the levels of two national four-firm concentration measures for the grocery industry from 2004 to 2024. Panel B plots composition-adjusted growth rate in each series relative to the 2004 base year. CR4 denotes the conventional four-firm concentration ratio, defined as the combined revenue share of the four largest retailers by annual sales. ES4 denotes the average effective share of the top four firms. Panel B aggregates the top four firms using a rank-slot Tornqvist index to remove composition effects from changes in top firm identity over time. The sample covers the national grocery channel defined by Nielsen and is annual from 2004 through 2024.

Figure A8: Credit-Only vs. Debit-and-Credit Effective Shares: Coefficients by Industry



Notes: The figure visualizes by-industry slope coefficients from weighted least squares regressions of credit-only effective market share on effective market share computed using combined debit and credit transactions. Regressions are weighted by annual firm debit-plus-credit sales.A coefficient closer to 1 indicates that the credit-only effective market share closely tracks the debit+credit effective market share within the matched sample. The sample consists of matched merchant-year-industry observations from the national effective shares from 2020 to 2025, with customers defined at the card level.

A2 Appendix Tables and Figures

Table A1: Summary Statistics for Various Analysis Samples

	Customer*Year	Customer Pooled	Synthetic Customer	Transaction Data
Customer Count	80.06 million	36.15 million	8.88 million	1.21 billion
Firm Count	2.45 million	2.49 million	2.67 million	3.15 million
Establishment Count	2.65 million	2.68 million	2.75 million	3.56 million
Total Sales	2.47 trillion	2.08 trillion	2.97 trillion	873.96 trillion
Online Share	39.71%	38.71%	38.94%	34.02%
Average Annual Spend	10,610.06	11,526.48	93,466.91	3,387.61

Notes: All columns include data from 2021-2025. Customer count refers to the person count for customer by year, and customer pooled. Customer count is the synthetic customer count for the synthetic customer and is the card count for transaction data. Firm count is the number of national firms in the sample. Establishment count refers to the number of in-person establishments; it does not include online stores. Average annual spend refers to the average annual customer spend — person spend for the first two samples, synthetic customer spend for the synthetic customer, and card spend for the transaction data.

Table A2: Mapping of Merchant Category Codes (MCC) to Industry Groups and Census Industries

MCC	MCC Name	Industry Group	Census Industry
5013	MOTOR VEHICLE SUPPLY/NEW PARTS	Auto Parts Stores	Automotive parts, accessories, and tire stores
5531	AUTO/HOME SUPPLY STORES	Auto Parts Stores	Automotive parts, accessories, and tire stores
5532	AUTOMOTIVE TIRE STORES	Auto Parts Stores	Automotive parts, accessories, and tire stores
5533	AUTOMOTIVE PARTS STORES	Auto Parts Stores	Automotive parts, accessories, and tire stores
7531	AUTO BODY REPAIR SHOPS	Auto Repair Shops	Automotive repair and maintenance
7534	TIRE RETREAD/REPAIR SHOPS	Auto Repair Shops	Automotive repair and maintenance
7535	AUTO PAINT SHOPS	Auto Repair Shops	Automotive repair and maintenance
7538	AUTO SERVICE SHOPS/NON DEALER	Auto Repair Shops	Automotive repair and maintenance
7230	BEAUTY/BARBER SHOPS	Barber Shops and Beauty Salons	Beauty salons
5309	DUTY FREE STORES	Beer, Wine, and Liquor Stores	Beer, wine, and liquor stores
5921	PKG STORES/BEER/WINE/LIQUOR	Beer, Wine, and Liquor Stores	Beer, wine, and liquor stores
5192	BOOKS/PERIODICALS/NEWSPAPERS	Book Stores and News Dealers	Book stores and news dealers
5942	BOOK STORES	Book Stores and News Dealers	Book stores and news dealers
5994	NEWS DEALERS/NEWSSTANDS	Book Stores and News Dealers	Book stores and news dealers
7542	CAR WASHES	Car Washes	Car washes
8041	CHIROPRACTORS	Chiropractors	Offices of chiropractors
5311	DEPARTMENT STORES	Department and Clothing Stores	Clothing and clothing accessories stores

5611	MEN/BOYS CLOTHING/ACC STORES	Department and Clothing Stores	Clothing and clothing accessories stores
5621	WOMENS READY TO WEAR STORES	Department and Clothing Stores	Clothing and clothing accessories stores
5631	WOMENS ACCESS/SPECIALTY	Department and Clothing Stores	Clothing and clothing accessories stores
5641	CHILDREN/INFANTS WEAR STORES	Department and Clothing Stores	Clothing and clothing accessories stores
5651	FAMILY CLOTHING STORES	Department and Clothing Stores	Clothing and clothing accessories stores
5681	FURRIERS AND FUR SHOPS	Department and Clothing Stores	Clothing and clothing accessories stores
5691	MENS/WOMENS CLOTHING STORES	Department and Clothing Stores	Clothing and clothing accessories stores
5699	MISC APPAREL/ACCESS STORES	Department and Clothing Stores	Clothing and clothing accessories stores
5948	LUGGAGE/LEATHER STORES	Department and Clothing Stores	Clothing and clothing accessories stores
5965	COMBINATION CATALOG & RETAIL	Department and Clothing Stores	Clothing and clothing accessories stores
5813	BARS/TAVERNS/LOUNGES/DISCOS	Drinking Places (Alcoholic Beverages)	Drinking places (alcoholic beverages)
5912	DRUG STORES & PHARMACIES	Drug Stores and Pharmacies	Pharmacies and drug stores
5975	HEARING AID/SALES/SERVICE	Drug Stores and Pharmacies	Pharmacies and drug stores
5976	ORTHOPEDIC GOODS	Drug Stores and Pharmacies	Pharmacies and drug stores
5997	ELEC RAZOR STORES/SALE/SERV	Drug Stores and Pharmacies	Pharmacies and drug stores
7210	LAUNDRY/CLEANING/GARMENT SV	Drycleaning and Laundry Services	Drycleaning and laundry services (except coin-operated)
7211	LAUNDRIES-FAMILY/COMMERCIAL	Drycleaning and Laundry Services	Drycleaning and laundry services (except coin-operated)
7216	DRY CLEANERS	Drycleaning and Laundry Services	Drycleaning and laundry services (except coin-operated)
5044	OFFICE/PHOTO EQUIPMENT	Electronics and Appliance Stores	Electronics and appliance stores
5045	COMPUTERS/PERIPHERALS/SOFTWARE	Electronics and Appliance Stores	Electronics and appliance stores
5722	HOUSEHOLD APPLIANCE STORES	Electronics and Appliance Stores	Electronics and appliance stores
5732	ELECTRONICS STORES	Electronics and Appliance Stores	Electronics and appliance stores
5946	CAMERA & PHOTO SUPPLY STORES	Electronics and Appliance Stores	Electronics and appliance stores
5131	PIECE GOODS/NOTIONS/DRY GOODS	Fabric and Craft Stores	Sewing, needlework, and piece goods stores
5949	FABRIC STORES	Fabric and Craft Stores	Sewing, needlework, and piece goods stores
5970	ARTIST/CRAFT SHOPS	Fabric and Craft Stores	Sewing, needlework, and piece goods stores
5193	FLORIST SUPPLIES/NURSERY STOCK	Florists	Florists
5992	FLORISTS	Florists	Florists
5947	GIFT, CARD, NOVELTY STORES	Gift Stores	Gift, novelty, and souvenir stores
5300	WHOLESALE CLUBS	General Merchandise Stores	General merchandise stores

5310	DISCOUNT STORES	General Merchandise Stores	General merchandise stores
5331	VARIETY STORES	General Merchandise Stores	General merchandise stores
5399	MISC GENERAL MERCHANDISE	General Merchandise Stores	General merchandise stores
5411	GROCERY STORES/SUPERMARKETS	Grocery Stores	Grocery stores
5422	FREEZER/MEAT LOCKERS	Grocery Stores	Grocery stores
5441	CANDY/NUT/CONFECTION STORES	Grocery Stores	Grocery stores
5451	DAIRY PRODUCT STORES	Grocery Stores	Grocery stores
5462	BAKERIES	Grocery Stores	Grocery stores
7297	MASSAGE PARLORS	Health and Beauty Spas	Beauty salons
7298	HEALTH & BEAUTY SPAS	Health and Beauty Spas	Beauty salons
5945	HOBBY, TOY & GAME STORES	Hobby, Toy and Game Stores	Hobby, toy, and game stores
5712	FURNITURE/EQUIP STORES	Home Furnishings Stores	Furniture and home furnishings stores
5719	MISC HOME FURNISHING SPECIALTY	Home Furnishings Stores	Furniture and home furnishings stores
5950	GLASSWARE/CRYSTAL STORES	Home Furnishings Stores	Furniture and home furnishings stores
5039	CONSTRUCTION MATERIALS - DEF	Home Improvement and Hardware Stores	Hardware stores
5072	HARDWARE EQUIPMENT/SUPPLIES	Home Improvement and Hardware Stores	Hardware stores
5074	PLUMBING/HEATING EQUIPMENT	Home Improvement and Hardware Stores	Hardware stores
5198	PAINT, VARNISHES & SUPPLIES	Home Improvement and Hardware Stores	Hardware stores
5200	HOME SUPPLY WAREHOUSE STORES	Home Improvement and Hardware Stores	Hardware stores
5211	LUMBER/BUILD. SUPPLY STORES	Home Improvement and Hardware Stores	Hardware stores
5231	GLASS/PAINT/WALLPAPER STORES	Home Improvement and Hardware Stores	Hardware stores
5251	HARDWARE STORES	Home Improvement and Hardware Stores	Hardware stores
5261	NURSERIES, LAWN/GARDEN SUPPLY	Home Improvement and Hardware Stores	Hardware stores
5713	FLOOR COVERING STORES	Home Improvement and Hardware Stores	Hardware stores
5714	DRAPERY & UPHOLSTERY STORES	Home Improvement and Hardware Stores	Hardware stores
5718	FIREPLACES & ACCESSORIES	Home Improvement and Hardware Stores	Hardware stores
5998	TENT AND AWNING SHOPS	Home Improvement and Hardware Stores	Hardware stores
5094	PRECIOUS STONES/METALS/JEWELRY	Jewelry Stores	Jewelry stores
5944	JEWELRY STORES	Jewelry Stores	Jewelry stores
5733	MUSIC STORES/PIANOS	Musical Instrument Stores	Musical instrument and supplies stores
5111	STATIONERY/OFFICE SUPPLIES	Office Supplies Stores	Office supplies and stationery stores
5943	STATIONERY STORES	Office Supplies Stores	Office supplies and stationery stores
8042	OPTOMETRISTS/OPHTHALMOLOGISTS	Opticians	Offices of optometrists
8043	OPTICIANS	Opticians	Offices of optometrists

8044	OPTICAL GOODS & GLASSES	Opticians	Offices of optometrists
5995	PET STORES/FOOD & SUPPLY	Pet and Pet Supplies Stores	Pet and pet supplies stores
5811	CATERERS	Restaurants, Excluding Fast Food	Full-service restaurants
5812	EATING PLACES AND RESTAURANTS	Restaurants, Excluding Fast Food	Full-service restaurants
5814	FAST FOOD RESTAURANTS	Fast Food Restaurants	Limited-service restaurants
5661	SHOE STORES	Shoe Stores	Shoe stores
5655	SPORTS/RIDING APPAREL STORES	Sporting Goods Stores	Sporting goods stores
5940	BICYCLE SHOPS/SALES/SERVICE	Sporting Goods Stores	Sporting goods stores
5941	SPORTING GOODS STORES	Sporting Goods Stores	Sporting goods stores
5993	CIGAR STORES/STANDS	Tobacco Stores	Tobacco stores
742	VETERINARY SERVICES	Veterinary Services	Veterinary services

Notes: Industry Group refers to the industry assigned to an MCC in the analysis sample. Census Industry is the industry that the Census assigns to the MCC code.

Table A3: Average and Median Industry Transactions

	Person*Year		Synthetic Customer		Person Pooled	
Industry	Average	Median	Average	Median	Average	Median
Shoe Stores	7.738	6	15.250	11	11.225	8
Tobacco Stores	14.521	10	19.120	12	27.035	13
Jewelry Stores	16.299	6	11.913	7	12.852	7
Home Furnishings Stores	8.866	7	21.337	13	14.253	10
Drycleaning and Laundry Services	12.603	9	21.302	12	21.445	11
Office Supplies Stores	8.415	6	14.302	10	12.078	8
Drug Stores and Pharmacies	16.707	12	74.740	37	45.344	25
Florists	8.412	6	10.798	8	10.561	7
General Merchandise Stores	80.345	54	563.609	294	356.543	246
Restaurants, Excluding Fast Food	29.832	19	183.619	79	106.151	62
Opticians	6.600	6	11.484	9	8.628	7
Drinking Places (Alcoholic Beverages)	12.157	8	23.926	14	19.406	10
Department and Clothing Stores	16.235	10	70.005	37	41.450	22
Veterinary Services	9.852	8	20.372	14	18.498	12
Barber Shops and Beauty Salons	10.137	8	30.166	17	22.576	15
Fabric and Craft Stores	9.870	7	14.640	10	15.500	9
Car Washes	9.512	8	22.123	14	17.360	11
Musical Instrument Stores	9.677	8	11.077	8	14.863	9
Home Improvement and Hardware Stores	17.217	11	72.900	36	43.516	23
Electronics and Appliance Stores	8.570	6	17.302	12	11.562	8
Chiropractors	11.872	9	15.047	11	20.531	12
Sporting Goods Stores	10.691	8	30.228	17	20.390	12
Pet and Pet Supplies Stores	11.622	9	26.579	17	23.955	14
Hobby, Toy and Game Stores	9.392	7	16.083	11	15.074	9
Arcades	11.015	7	11.914	8	13.746	8
Gift Stores	8.692	6	14.216	10	12.121	8
Auto Repair Shops	6.993	6	16.039	12	10.407	8
Book Stores and News Dealers	10.204	7	17.848	11	15.225	9
Health and Beauty Spas	9.998	8	15.727	11	16.941	10
Beer, Wine, and Liquor Stores	16.911	10	46.294	23	36.524	16
Fast Food Restaurants	41.849	23	243.322	102	147.149	73
Grocery Stores	52.720	34	342.944	149	211.852	133
Auto Parts Stores	10.270	7	24.670	17	16.528	10

Notes: The data is from 2021-2025. For the first four columns, the values are the average and median yearly transaction counts per industry. The first four columns have an activity filter applied that only includes customers with at least 5 transactions in that industry in a year. The final two columns are the average and median transaction count in an industry over 5 years. There is an activity filter of at least 5 transactions within the industry over 5 years.

Table A4: The Relationship between Effective Shares and Sales per Customer

	No Fixed Effects			Industry Fixed Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: ZIP3 Panel						
log(SPC)	0.612		0.610	0.548		0.547
	(0.000)		(0.000)	(0.000)		(0.000)
log(Customers)		-0.056	-0.044		0.054	0.004
		(0.000)	(0.000)		(0.000)	(0.000)
R^2	0.501	0.008	0.506	0.792	0.591	0.792
Partial R^2	0.501	0.008	0.506	0.499	0.015	0.499
Observations	2538299	2538299	2538299	2538299	2538299	2538299
Panel B: National Panel						
log(SPC)	0.585		0.574	0.504		0.507
	(0.000)		(0.000)	(0.001)		(0.001)
log(Customers)		-0.141	-0.088		0.008	-0.021
		(0.001)	(0.000)		(0.000)	(0.000)
R^2	0.509	0.040	0.525	0.788	0.601	0.789
Partial R^2	0.509	0.040	0.525	0.470	0.000	0.472
Observations	1806162	1806162	1806162	1806162	1806162	1806162

Notes: All regressions are weighted by log sales. Columns (1)–(3) include no fixed effects. Columns (4)–(6) include industry fixed effects. Partial R^2 equals R^2 for no-FE specifications and measures incremental explanatory power relative to industry fixed effects otherwise. Heteroskedasticity-robust standard errors are reported in parentheses.

Table A5: Robustness Checks for Customer Retention Results

	(1)	(2)	(3)	(4)
$\log \omega_{j,t}^{\text{effective}}$	0.096***		0.050***	0.079***
	(0.007)		(0.005)	(0.007)
$\log \Omega_t^j$	-0.013*		0.040***	0.015
	(0.007)		(0.011)	(0.011)
$\omega_{j,t}^{\text{effective}}$		0.082***		
		(0.006)		
Ω_t^j		0.014**		
		(0.006)		
Unit of Observation	Firm	Firm * ZIP3	Firm * ZIP3	Firm * ZIP3
Retention Measure	Sales-Weighted	Sales-Weighted	Equal-Weighted	Sales-Weighted
Market Shares Version	Baseline	Baseline	Baseline	Synthetic Customer
Observations	309,674	447,750	447,750	447,750
R^2	0.479	0.493	0.573	0.488
Within R^2	0.233	0.249	0.286	0.242

Notes: In column (1), we repeat the exercise in column (3) of table 1 with firms as the unit of observation rather than establishments. Column (2) does the same exercise again at establishment level, but with market shares in levels rather than logs. In column (3), the dependent variable is an equally-weighted average of the "customer retention" indicator rather than a sales-weighted average. In column (4), we replace the baseline "person-based" effective and market share measures with their synthetic customer equivalents.
We examine customer retention patterns from 2024 to 2025, and restrict to establishments with at least 200 customers in both of those years. All independent variables are standardized by category. All regressions are weighted by $\log \text{Sales}_{2024}$. We cluster standard errors by industry.

Table A6: Robustness Checks for Consumer Responses to Firm Entry

	(1)	(2)	(3)	(4)
$\omega_{j,t-1}^{\text{effective}}$	-0.020***	-0.025***	-0.019***	-0.019***
	(0.001)	(0.003)	(0.001)	(0.002)
$\Omega_{i,t-1}^{\text{market}}$	-0.002***	-0.004***		
	(0.000)	(0.001)		
Ω_{t-1}^{j}			-0.010***	-0.014***
			(0.001)	(0.001)
Including Restaurants?	X		X	
Incumbents Included	All	All	Entrant's ZIP3	Entrant's ZIP3
Observations	35,347,846	3,970,716	21,420,353	2,558,048
R^2	0.060	0.097	0.054	0.082

Notes: Columns (1) and (2) use the "sales-weighted average market share of firms visited by the customer" in $t-1$ as the market share predictor, rather than the incumbent establishment's market share in $t-1$: even with this change, the coefficient on effective share is still much larger than the coefficient on market share. Columns (3) and (4) repeat the analysis in Columns (3) and (4) of Table 2 using only incumbents that are located in the same 3-digit ZIP code as the entering store.
All standard errors are clustered by entry event.

Table A7: Consumer Responses to Firm Entry, Restricting to Verified Entrants

	(1)	(2)	(3)	(4)	(5)
$\omega_{i,j,t-1}^{\text{effective}}$	-0.024***		-0.023***	-0.033***	-0.019***
	(0.002)		(0.002)	(0.004)	(0.002)
Ω_{t-1}^{j}		-0.010***	-0.006***	-0.008***	
		(0.002)	(0.001)	(0.002)	
Including Restaurants?	X	X	X		X
Entrant FE?	X	X	X	X	X
Incumbent FE?					X
Observations	7,928,255	7,928,255	7,928,255	2,549,912	792,628
R^2	0.065	0.057	0.065	0.088	0.131

Notes: In this table, we only consider entering stores whose entry dates we verified manually using Yelp reviews. The results are extremely similar to those in Table 2: the coefficient on effective share is consistently significant and larger than the coefficient on market share.
All standard errors are clustered by entry event. In Column (5), for each entering establishment, we use a 10% random sample of all of the associated incumbents.

Table A8: Comparison of Census and Credit Card Demographics

Source	Median Income	Avg Age	Weighted Avg Age	Median Age
Census	$77,540	51	47	51
Credit Card	$82,000	48	49	46

Notes: All data is from 2022. Census Median Income comes from the Annual Social and Economic Data Supplement. Age data is from the American Community Survey (ACS) Public Use Microdata (PUMS), which was then restricted to only include Americans over 18. The weight for weighted average age is income.

Table A9: Impact of Industry Transaction Filter

Industry	Percent of People Dropped		Percent of Spending Dropped	
	5 Txns	10 Txns	5 Txns	10 Txns
Shoe Stores	88.22	97.88	66.84	87.39
Pet and Pet Supplies Stores	65.33	83.36	24.17	47.40
Opticians	94.80	99.53	85.99	96.98
Musical Instrument Stores	86.74	94.94	56.90	75.86
Home Improvement and Hardware Stores	47.86	67.47	12.27	25.97
Florists	94.28	98.89	71.31	84.49
Car Washes	74.25	90.15	44.16	68.36
Home Furnishings Stores	77.11	92.91	47.87	72.72
Electronics and Appliance Stores	86.40	96.88	66.84	87.48
Auto Parts Stores	80.05	93.17	50.85	70.77
Sporting Goods Stores	70.61	88.26	32.14	56.69
Veterinary Services	69.05	87.77	38.72	65.92
Arcades	89.10	96.72	63.14	78.56
Office Supplies Stores	88.25	97.35	60.90	79.86
Drycleaning and Laundry Services	79.29	90.72	36.24	55.48
Grocery Stores	12.98	24.58	1.02	3.04
Auto Repair Shops	90.63	98.76	75.33	94.32
Drinking Places (Alcoholic Beverages)	76.89	90.50	41.30	61.93
Drug Stores and Pharmacies	45.06	65.40	11.19	25.56
Fabric and Craft Stores	85.27	95.23	48.13	68.07
General Merchandise Stores	8.32	16.12	0.39	1.37
Restaurants, Excluding Fast Food	22.85	39.33	3.07	8.87
Chiropractors	66.76	83.38	36.60	57.42
Book Stores and News Dealers	84.44	95.15	53.86	75.92
Beer, Wine, and Liquor Stores	63.19	79.73	18.73	36.07
Tobacco Stores	72.68	85.87	22.43	39.96
Department and Clothing Stores	44.55	67.36	11.09	25.80
Barber Shops and Beauty Salons	64.31	85.05	27.21	54.29
Fast Food Restaurants	21.49	35.74	2.67	6.97
Jewelry Stores	96.27	99.12	80.99	92.05
Hobby, Toy and Game Stores	83.55	94.80	49.51	71.15
Health and Beauty Spas	79.11	91.85	47.04	71.16
Gift Stores	89.91	97.65	66.73	83.84

Notes: Includes anyone who appears in the main sample from 2021-2025. "5 Txns" includes people who have more than five transactions in the industry in a year. "10 Txns" includes people who have more than ten transactions in the industry in a year. This is in comparison to the baseline sample with distance from home, industry, home zip3, and median monthly transaction filters already applied.

Table A10: Transunion Industry Spending Shares

Industry	Studied Industries	Total Spend
General Merchandise Stores	0.312932	0.277591
Grocery Stores	0.168211	0.149214
Restaurants, Excluding Fast Food	0.093096	0.082582
Home Improvement and Hardware Stores	0.071286	0.063235
Department and Clothing Stores	0.065354	0.057973
Fast Food Restaurants	0.036497	0.032375
Home Furnishings Stores	0.034021	0.030179
Sporting Goods Stores	0.023516	0.020860
Electronics and Appliance Stores	0.022924	0.020335
Drug Stores and Pharmacies	0.019696	0.017472
Veterinary Services	0.019345	0.017160
Auto Parts Stores	0.017990	0.015958
Auto Repair Shops	0.017797	0.015787
Beer, Wine, and Liquor Stores	0.012856	0.011404
Jewelry Stores	0.011647	0.010331
Barber Shops and Beauty Salons	0.010783	0.009565
Pet and Pet Supplies Stores	0.009686	0.008592
Opticians	0.007231	0.006414
Shoe Stores	0.007006	0.006215
Drinking Places (Alcoholic Beverages)	0.005176	0.004591
Health and Beauty Spas	0.004752	0.004215
Gift Stores	0.003472	0.003080
Hobby, Toy and Game Stores	0.003325	0.002949
Book Stores and News Dealers	0.003033	0.002690
Car Washes	0.002820	0.002501
Tobacco Stores	0.002539	0.002252
Florists	0.002323	0.002061
Office Supplies Stores	0.002301	0.002041
Fabric and Craft Stores	0.002273	0.002016
Drycleaning and Laundry Services	0.002008	0.001781
Chiropractors	0.001878	0.001666
Musical Instrument Stores	0.001766	0.001566
Arcades	0.000463	0.000411
Total	1.0000	0.8879

Notes: This is for the Transunion sample prior to applying any filters, including activity filters for cards, and distance from home zipcode. The left column is the share of spending of only the industries included in the final analysis sample. The right column represents the industry's share of the total TransUnion sample. Rows are only included for industries in the final sample.