Customer Churn and Intangible Capital*

Scott R. Baker† Brian Baugh‡ Marco Sammon§

March 2023

Abstract

Intangible capital is a crucial and growing piece of firms' capital structure, but many of its distinct components are difficult to measure. We develop and make available several new firm-level metrics regarding a key component of intangible capital – firms' customer bases – using an increasingly common class of household transaction data. Linking household spending to customer-facing firms that make up over 30% of total household spending, we show that churn in customer bases is associated with lower markups and market-to-book ratios and higher leverage. Churn is closely linked to firm-level volatility and risk, both cross-sectionally and over time. This new measure provides a clearer picture of firms' customer and brand capital than existing metrics like capitalized SG&A, R&D, or advertising expenditures and is also observable for private firms. We demonstrate that low levels of customer churn push firms away from neoclassical investment responsiveness and that low churn firms are better able to insulate organization capital from the risk of key talent flight.

JEL Classification: D22, E22, G32, L11

Keywords: customer base, transaction data, customer churn, intangible capital, risk, volatility

*The authors wish to thank Lorenz Kueng, Steve Davis, Pete Klenow, Jan Eberly, Francois Gourio, Nicolas Crouzet, Ye Li, Lauren Cohen and seminar participants at the Hoover Institution, San Francisco Federal Reserve and the Fed Board, the Kellogg School of Management, University of Pittsburgh, ITAM, Rice University, the University of Delaware, West Virginia University, ASU, and NBER for their helpful comments and suggestions. An earlier version of this paper was circulated with the title 'Measuring Customer Churn and Interconnectedness'.

†Northwestern University, Kellogg School of Management s-baker@kellogg.northwestern.edu.

‡University of Nebraska - Lincoln bbaugh2@unl.edu.

§Harvard Business School mcsammon@hbs.edu.

1 Introduction

Intangible capital has become an increasingly important factor of production for many industries and has been put forth as one major factor driving increases in corporate concentration and markups over the past several decades.[1] Accordingly, this increase in the amount and scope of intangible capital has also driven changes in the exposure of firms to risk embodied in their intangible capital.

This growth has made the measurement of intangible capital more important when analyzing firm investment decisions, value, and risk exposure (see e.g., Peters and Taylor (2017), Eisfeldt et al. (2020)). Measurement is made more difficult by the fact that intangible capital is composed of a number of different components such as (1) technology/patents (see e.g., Kogan et al. (2017)), (2) customer base/branding (see e.g., Gourio and Rudanko (2014), Belo et al. (2019), Fornell et al. (2016)), (3) human-resource intangibles (see e.g., Eisfeldt and Papanikolaou (2013), Edmans (2011)), and (4) organizational design (see e.g., Lev (2000), Lev and Radhakrishnan (2003) and Lev et al. (2009) for more on this decomposition).

Previous research has often relied on proxies of intangible capital such as capitalized SG&A or R&D spending (see Eisfeldt and Papanikolaou (2013) and Peters and Taylor (2017)) or 'residual methods' which attribute to intangibles all of the value that cannot be explained by tangible assets (e.g., Ewens et al. (2020)). Given the breadth of intangible capital, it's not always obvious how these imprecise metrics should be related to firm risk or decision-making. In this paper, we use new data regarding customer bases to pry open this black box and discover an important component of intangible capital that has less ambiguous effects on both firm risk and firm behavior.

In this paper, we demonstrate that household financial transaction data can create accurate customer-centric metrics that describe a wide range of firm-level attributes. This paper focuses primarily on one such metric, customer churn, but proposes, demonstrates, and makes available online others that can also be recreated using any household transaction database.[2] While we are not the first paper to suggest that stable customer bases are valued by firms, we are the first to

[1] See work such as Crouzet and Eberly (2019), Eisfeldt and Papanikolaou (2014), Eisfeldt et al. (2020), Ewens et al. (2020), Belo et al. (2019), Sim et al. (2013), and Corrado et al. (2009).

[2] Numerous researchers around the world have gained access to similar household financial transaction data for research focusing mainly on questions relating to household decision-making. Some research utilizing financial transaction data has used sources from Mexico (Bachas et al., 2019), Singapore (Agarwal and Qian, 2014), Brazil (Medina, 2020), Turkey (Aydin, 2019), Germany (Baker et al., 2020), Iceland (Olafsson and Pagel, 2018), and the United States (Balyuk and Williams, 2021). See a review of this literature in Baker and Kueng (August 2022).

directly measure customer retention and attrition across a wide range of firms.[3] This approach can be employed for both public and private firms, at high frequency and geographic granularity, and for firms without any intangible assets on their public balance sheet.

While household financial transaction data has proven invaluable to fields related to households and consumers, this paper shows that this class of data has substantial utility when applied to research regarding *firms*, as well. Across millions of households, we link transactions representing over 30% of total household spending to firms in industries as diverse as retail, grocery, restaurants, aviation, utilities and telecom. Customer-centric databases such as the Nielsen Consumer Panel cover much narrower slices of consumer spending and prohibit researchers from de-anonymizing retailer identities. Moreover, we are able to observe both public and private firms.[4] We link household transactions to firms and validate our firm-level matches against a range of external data spanning firm-level revenues, prices, and geographic locations.

We build several measures of the churn for hundreds of customer-facing firms. Our primary measure is the difference in individual customer spending shares among the customer base of firm f in year t and the customer base of firm f in year $t-1$. Other measures focus separately on intensive- or extensive-margin shifts in customer base growth. These measures are distinct from existing measures of intangible capital and can be measured at a monthly and city level.

Using these measures of firm-level customer churn, we make three main contributions. First, we turn to the drivers of churn, specifically focusing on the role of customer frictions. We argue that cross-industry differences in switching costs may drive a significant share of the variation in churn (Kleshchelski and Vincent (2009), Paciello et al. (2019)). Specifically, we show that churn is relatively lower in industries with either explicit long-term contracts (e.g., Consumer Telecom) or strong loyalty programs (e.g., Hotels) than it is in industries like General Merchandise, where many retailers offer the exact same products.

We then explore how switching costs influence the way firms attract customers. We argue that

[3] For instance, Belo et al. (2019), Fornell et al. (2016), Gourio and Rudanko (2014), and Morlacco and Zeke (2021) suggest that customer or brand capital can be a substantial portion of intangible value for some firms. Kleshchelski and Vincent (2009) show that firms limit price volatility due to sensitivity about customer attrition and turnover. Ascarza (2018) discusses the importance of targeting customers likely to churn and Lemmens and Gupta (2020) notes a range of approaches to enhance customer retention.

[4] To our knowledge, Agarwal et al. (2020) and Klenow et al. (2020) are perhaps the only other papers that propose a similar approach. They demonstrate that disaggregated spending data can provide high quality signals about consumer demand, firm growth, and equity prices for consumer-facing firms.

in low switching cost industries, firms may need to constantly spend on customer acquisition – either through advertising or lower markups (Hall et al. (1997), Fitzgerald et al. (2016)) – but still have high churn. On the other hand, in high switching cost industries, advertising may be more effective at bringing in long-term customers. Moreover, the most productive firms – which can support the stickiest customers (Afrouzi et al., 2020) – may spend the most on their customer relationships, leading to a negative correlation between spending on customer acquisition and churn. Consistent with this hypothesis, we find that there is a negative relationship between spending on customer acquisition and churn for non-retail firms, but the relationship is flipped for retailers.

As a finer test of this hypothesis, we use our transaction data to proxy for these switching costs with overlap in customer bases across firms within an industry. The logic is that low overlap is evidence of exclusive customer relationships and thus higher switching costs. We find that in industries with more overlap, there is a positive relationship between spending on customer acquisition and churn, while in industries with low overlap, the slope of this relationship is negative.

We also consider local market power (i.e., market share in a given geographic area) as another source of frictions. We argue this should decrease churn, either mechanically through giving consumers fewer choices, or through deterring entry and competition (Kalyanaram et al. (1995); Bornstein et al. (2018)). Empirically, we show that more local market power is correlated with lower churn. Local market power also interacts with switching costs, as its effect on churn is weaker in industries with long-term contracts than those with regular purchases.

Our second contribution is to provide empirical evidence that customer churn is highly predictive of firm-level valuation and decisions regarding capital structure, markups and investment dynamics. In both the cross-section and within firms, we find that lower levels of churn are associated with higher valuations, even for similarly sized customer bases. Such relationships between lower churn and higher valuations (or market-to-book values) may underpin a recent drive among firms to shift to subscription-based business models and loyalty programs that emphasize customer base stability.

We provide support for models of customer frictions where a firm's customer base acts as a state variable. These models act as one foundation for an adjustment cost model of firm investment and give predictions that firms with lower levels of customer base churn will have higher rates of investment and markups, but would respond more slowly to shocks their investment opportunity set over time (e.g., Christiano et al. (2005), Eberly et al. (2012)).[5] Empirically, using churn aligns results with model predictions more clearly than proxies like SG&A used in previous research (e.g., Gourio and Rudanko (2014)).

Finally, we demonstrate that, consistent with theoretical evidence in Gilchrist et al. (2017) and Gourio and Rudanko (2014), churn is related to systematic risk and that firms without stable customer bases are more exposed to macroeconomic fluctuations. We focus on systematic risk, a central element in many models of asset pricing and corporate finance, whose determinants are still understudied. Broadly, intangible capital is exposed to different risks than physical capital. Unlike a machine or a plot of land, employees can abscond with ideas and human capital, patents can be found invalid, and customer bases can evaporate due to changing tastes or marketing blunders.

We provide several pieces of evidence to support this linkage between systematic risk and churn. We find a strong positive relationship between churn and CAPM beta across all firms in the cross-section, within industries, and even within firms over time. OLS estimates imply that a firm in the 90th percentile of churn has CAPM beta of up to 0.4 higher than a firm in the 10th percentile and CAPM beta is monotonically increasing across value-weighted portfolios formed on churn. Furthermore, even controlling for beta, we argue that high churn firms are especially exposed to negative systematic shocks. We show that high churn firms were the hardest hit during the beginning of the COVID pandemic, even accounting for other measures of exposure to systematic risk, size, and seasonal patterns in spending across industries.

In addition to churn being intimately associated with higher systematic risk, measuring churn directly helps to also clarify the separate impacts of different elements of intangible capital on firm-level risk. If increases in intangible capital are embedded in employees (e.g., employee training), firm-level risk may be elevated due to the potential for employees to take their human capital and exit the firm (Eisfeldt and Papanikolaou, 2013). Crucially, this mechanism depends on the extent to which measured intangible capital is embodied in movable employees rather than specific to the firm (e.g., capitalized advertising, loyalty programs). We find that there is no relationship

[5]Our findings contribute to a larger debate on how firms acquire and extract value from their customer base over the business cycle. Dou et al. (2019) show that key talent drives customer capital and financial constraints may force this talent to leave in bad times. Gilchrist et al. (2017) show that firms build customer capital with low prices but charge high prices in bad times to maintain cashflow. On the other hand, Kim (2018) shows that firms decrease prices in bad times to boost cashflow. The divergences in findings may be caused by measurement issues – these papers are forced to measure customer bases indirectly.

between capitalized SG&A and systematic risk among low churn firms – consistent with SG&A being transformed into brand value rather than employee human capital among low churn firms. Among high churn firms, however, the relationship is monotonically increasing from low to high SG&A. In addition, among every tercile of organization capital, there is an increasing relationship between churn and risk.

The rest of the paper is organized as follows. Section 2 defines our measure of churn and lays out testable hypotheses. Sections 3 and 4 describe and validate our data. Section 5 illustrates some drivers of churn and demonstrates how churn relates to intangible capital and affects firm investment decisions. Section 6 details the relationship between churn and systematic risk, how this relationship may be amplified in economic downturns, and how churn can be used to distinguish between the effects of brand capital and employee capital. Section 8 concludes.

2 Measurement & Hypothesis Development

2.1 Measuring Customer Churn

To construct our measure of customer churn, we calculate the spending-share-adjusted change in firm f's customer base between two given years, t and $t-k$. We define spending share, $s_{f,i,t}$, as the share of firm f's revenue that comes from customer i in year t. This definition implies that $s_{f,i,t} \in [0,1]$ and $\sum_i s_{f,i,t} = 1$ for all f and t. We then define churn as:

$$Churn_{f,t-k} = \left(\sum_i |s_{f,i,t} - s_{f,i,t-k}| \right) \Big/ 2 \tag{1}$$

where the sum $\sum_i |s_{f,i,t} - s_{f,i,t-k}|$ is taken over all customers that shop at firm f in *either* year t or year $t-k$.[6] In words, churn is the difference in spending shares coming from each customer i between years t and $t-k$. The way it is defined, $\sum_i |s_{f,i,t} - s_{f,i,t-k}|/2$ can vary between zero and one. A value of zero would imply constant revenue shares and a constant customer base between years t and $t-k$, while a value of one implies a completely new customer base.

In order to not conflate attrition from our sample with attrition from a customer base, we require

[6] In Appendix Figure A.1, we plot histograms of this measure across all firms for $k = [1,4]$. As one would expect, churn increases over time. There is substantial spread among firms at all horizons: at the most extreme, about 10% of firms see 90% of their revenue coming from new customers.

that customer i has positive spending for at least one firm in both years t and $t-k$, although not necessarily at firm f. The sample has very low attrition and re-computing our churn measure without this restriction yields a correlation of approximately 0.98. Therefore, our churn metric captures three different types of changes in a firm's customer base: (1) the acquisition of new customers (2) the attrition of existing customers and (3) the change in spending by persistent customers.[7]

2.2 Drivers of Customer Churn

Customer switching costs are an important driver of cross-sectional variation in customer churn (Kleshchelski and Vincent, 2009; Paciello et al., 2019). For example, it is easy to try a new clothing store, but customers are likely locked into a long-term contract with their cell phone service provider. Switching costs arise not only from explicit contracts, but also from customer loyalty programs offered by hotels, for example. Our first empirical prediction focuses on industry-level differences in churn, as switching costs often stem from the type of product or contract offered, which systematically differs across industries:

Prediction 1A: *Industries with higher switching costs will have lower average churn.*

The presence of switching costs has implications for how firms acquire and retain customers. For example, because each individual client is more valuable, industries which can support a stable customer base should be willing to spend more on customer acquisition through advertising and lowering markups (Fitzgerald et al., 2016). Further, within such industries, we might expect a negative relationship between churn and spending on customer acquisition, as the most productive firms – which can support the stickiest customers (Afrouzi et al., 2020) - are willing to spend more on their customer relationships. In industries with less stable customer bases, however, firms may need to constantly advertise to get new clients in the door. This implies the following prediction for differences in the relationship between churn and spending on customer acquisition across high and low switching cost industries:

Prediction 1B: *In industries with substantial switching costs, churn should be negatively re-*

[7]In the Appendix, we discuss variants of this measure which focus on subsets of customer churn: the fraction of spending sourced from *new* customers in a given year and the fraction of spending from the previous year conducted by customers that *left* the customer base. Both variants are highly correlated with our baseline measure, with correlation coefficients of 0.91 and 0.81, respectively.

lated to spending on customer acquisition. This relationship will be less negative in industries with fewer switching costs.

As discussed in (Afrouzi et al., 2020), most of firms' non-production expenses go to acquiring new customers, rather than retaining existing customers. The relationship in prediction 1B, therefore, should be stronger when measuring churn on the extensive margin (i.e., from new customers). Further, survey evidence in Hall et al. (1997) suggests this effect may be especially strong in retail firms, which change their prices more frequently than non-retail firms.

Another driver of churn is local market power. If customers have fewer options, they are less likely to shift spending to another retailer. As an extreme example, in some geographic areas there is only one grocery store, making switching possible only by driving impractically long distances. Further, as discussed in Bornstein et al. (2018), customers have significant inertia in where they shop, which makes it hard for firms to enter new geographic markets. This may lead to self-reinforcing cycles, as once a firm has significant local market power, it may deter entry by other firms, further reducing churn (Kalyanaram et al., 1995). These findings motivate our final prediction on the drivers of customer churn:

Prediction 1C: *Churn should be negatively related to local market power.*

2.3 Churn and Intangible Capital

By its nature, intangible capital is not straightforward to measure. Some common approaches are to capitalize spending in areas that generate intangibles such as R&D and SG&A (Peters and Taylor (2017), Eisfeldt et al. (2020)). While customer base turnover is an important piece of information for a firm, unlike R&D and SG&A spending, churn is not something that is directly under a firm's control. Many dimensions of intangible capital which firms can control, however, contribute to churn including customer loyalty programs, brand value and the type of products firms sell (Belo et al. (2019), Bronnenberg et al. (2009), Bronnenberg et al. (2012), Dubé et al. (2010)).

In addition, developing superior products can lead to greater customer satisfaction, making it easier to retain customers (Fornell et al., 2016). This may come in the form of patents, which have been shown to deter competition (Argente et al., 2020) and may further reduce churn. A common thread in these channels is that churn should be a decreasing function of intangible capital, which implies the following prediction:

Prediction 2A: *Churn should be negatively related to measures of intangible capital.*

Past work has shown that leverage is negatively correlated with various metrics of intangible asset intensity across firms (see e.g., Crouzet and Eberly (2019) and Caglio et al. (2021)). This may be driven by the fact that intangible capital, unlike physical capital, cannot be used as collateral for asset-backed borrowing.[8] If Prediction 2A holds, we would expect low churn firms to be composed of relatively more intangible capital, which has the following additional implication:

Prediction 2B: *Churn should be negatively related to measures of leverage.*

Prediction 2A may also have implications for firms' pricing strategies. If low churn firms have more brand value, and if a firm's customers have a strong affinity to its brand, they will be less sensitive to prices (Smith and Brynjolfsson, 2001). Another possible explanation is that – as discussed in Prediction 1C – low churn firms may have more local market power and market power is what allows firms to charge higher markups (Thisse and Vives, 1988). In either case, we expect the following prediction to hold:

Prediction 2C: *Churn should be negatively related to markups.*

2.4 Churn and Firm Dynamics

Adjustment costs are an important feature in developing macroeconomic models which fit stylized facts on the persistence in inflation and output (Christiano et al. (2005), Eberly et al. (2012)). Capital and labor adjustment cost models have been extended to feature customer frictions (i.e., accounting for customer attachment or loyalty) when shifting across firms or products. In such models, customer bases act as state variables and thus can affect the rate of return on any given investment. As laid out in Christiano et al. (2005), these investment adjustment costs take the form:

$$k_{t+1} = (1-\delta)k_t + F(i_t, i_{t-1}) \tag{2}$$

$$F(i_t, i_{t-1}) = \left(1 - S\left(\frac{i_t}{i_{t-1}}\right)\right) i_t \tag{3}$$

where each period a share δ of capital k_t depreciates and firms purchase investment goods, i_t, to increase the capital stock. The function $F(i_t, i_{t-1})$ describes how current and past investment is

[8] Recently, however, firms have begun to use the strength of their customer base as a source of collateral. For example, during the COVID-19 outbreak, airlines borrowed billions of dollars against their loyalty programs to fund operating losses.

transformed into installed capital that can be utilized by the firm in the next period. The convex function $S\left(\frac{i_t}{i_{t-1}}\right)$ penalizes deviations from the prior level of investment, with $S(1) = 0$.

These adjustment costs shift firms' responses to investment opportunities away from a frictionless benchmark. In the limit without any frictions, markups and profits are zero and Tobin's Q is equal to one (i.e., market value is equal to book value and the marginal dollar of investment will not affect the value of the firm). In a framework that features customer or product frictions, however, firms with a higher degree of customer stickiness can be expected to have higher market-to-book values (Q) and higher levels of markups, consistent with Predictions 2A and 2C above.

Moreover, such low-churn or high-friction firms are predicted to feature an investment profile that is smoother over time (see Eberly et al. (2012)). Gourio and Rudanko (2014) pursue this general line of reasoning, building a model of product market competition that features customer attachment driven by frictions in search that prevent customers from costlessly shifting between firms. In their model, firms with slower-moving customer bases adjust more slowly to new investment opportunities. This results in weaker investment responses to changes in firms' investment opportunity set, which they measure using Tobin's Q. Firms with faster-adjusting customer bases are predicted to more closely approximate the frictionless benchmark, where increases in firm productivity drive immediate increases in firm investment.

Our measures of firm-level customer churn can be seen as identifying heterogeneity in the function $S(\cdot)$, or differences in the speed of customer-base adjustment across firms, as some customer bases are more difficult to adjust than others. They are also consistent with identifying time-series variation in $S(\cdot)$, as e.g., it may be more costly to adjust a customer base during a recession when people are hesitant to try new firms/products. This implies the following prediction for the relationship between churn and firms' investment behavior:

Prediction 3: *Low churn firms' investment should be less volatile and less sensitive to changes in their opportunity set, as measured by Tobin's Q.*

2.5 Churn and Risk

Our final set of predictions regard how churn may affect a firm's exposure to systematic risk. In Gourio and Rudanko (2014), because they face fewer adjustment costs, firms with less stable customer bases adjust investment more in response to macroeconomic shocks and are more exposed to systematic risk. A second strand of literature links churn to systematic risk through the price setting channel. For example, a common theme of Kleshchelski and Vincent (2009) and Weber (2014) is that some firms do not adjust prices in bad times because they are worried about losing customers and market share. High churn firms, therefore, may be more exposed to systematic risk because they are likely the most concerned about losing customers and market share. When linking customer-base frictions to risk, churn has the advantage of being a direct measure of the stickiness of a firm's customer base. Unlike capitalized SG&A or R&D, churn accounts for how *effective* these types of spending are for creating a stable customer base.

Considering the results in Weber (2014) and Kleshchelski and Vincent (2009), however, the empirical relationship between churn and exposure to systematic risk is not obvious, because firms may be high churn by choice (i.e., by choosing to have flexible prices) or firms may be high churn because they are in an industry with low frictions e.g., big box retail. Therefore, it is important to measure the relationship between churn and exposure to systematic risk both unconditionally and within a given industry. This implies the following prediction for the relationship between churn and systematic risk:

Prediction 4A: *High churn firms will be more exposed to systematic risk. This should be truth both within and across industries.*

The relationship between churn and systematic risk, however, might be amplified in bad times. For example, Baker et al. (2021) show that, when their income falls, customers shop at fewer stores, driven especially by retrenchment to shops they are familiar with. This means that high churn firms, which rely on a steady stream of new customers, are more exposed to negative systematic shocks. Another channel which could link negative systematic shocks to churn and risk is firms' price setting strategies. As argued in Gilchrist et al. (2017), when times are bad, firms would like to cut prices to maintain market share but may be unable to because of financing constraints. In equilibrium, firms without loyal customer bases lose market share to firms with stable customers in downturns, making them more exposed to these negative shocks. These two channels imply the following refinement of prediction 4A:

Prediction 4B: *Conditional on average exposure to systematic risk, high churn firms should be especially exposed to negative shocks.*

As with prediction 1B, this effect should be stronger when using extensive-margin-based measures of churn, as the findings in Baker et al. (2021) explicitly address customers trying fewer new shops.

As noted above in Section 2.3, churn can help to disentangle one important component of aggregated measures of intangible capital. This feature is also useful for differentiating between theories of exposure to systematic risk. Several papers quantify the stock of intangible capital using capitalized spending on R&D and SG&A. However, this kind of spending can represent investment in many different elements of intangible capital including technology/patents, brand value, human capital, and organizational design which influence firm risk in different ways.

Separately identifying customer attachment as a component of intangible capital can not only make inferences regarding risk and customer capital clearer, but can also clarify the impacts of other elements of intangible capital within a firm. For example, consider the model in Eisfeldt and Papanikolaou (2013). Firm i's output at time t, $y_{i,t}$, is a function of its initial endowment of physical capital K_i and organization capital O_i. Specifically:

$$y_{i,t} = \theta_t K_i + \theta_t e^{\epsilon_i} O_i \tag{4}$$

where θ_t is a disembodied technology shock that affects both forms of capital and follows a geometric random walk.

In this model, higher levels of organization capital make a firm riskier both in terms of total stock return volatility and CAPM beta. Organization capital is a source of risk because firm i's efficiency in using O_i, ϵ_i, is set to the level of aggregate efficiency, x_t, at the time the firm is founded. Efficiency follows a random walk, so if x_t becomes sufficiently high, it is attractive for employees to start a new firm. O_i is specific to employees, not the firm, so they can take the stock of O_i with them and use it more efficiently in the new firm. As a result, x_t shocks become a source of risk for firms, where exposure is proportional to the level of organization capital.

Our prior is that there are multiple types of organization capital that have different implications for firm riskiness. One way to formalize this thinking is to modify the baseline model of Eisfeldt and Papanikolaou (2013), splitting organization capital into two components: (1) $O_i^{employees}$ which employees can take with them if they start a new firm and (2) O_i^{brand} which is specific to the firm and thus cannot be absconded with by the employees.

In our proposed modification, $O_i = O_i^{employees} + O_i^{brand}$, so it is possible for firms to have high O_i, but not be very risky. Specifically, higher levels of O_i^{brand} do not expose firms to more x_t risk. We believe firms with low churn have relatively more of their organization capital in brand value, while firms with high churn have more of their organization capital accrue to their employees.

Eisfeldt and Papanikolaou (2013) measure organization capital using SG&A spending (Sun and Xiaolan (2019) use capitalized R&D to proxy for intangible capital embedded in firms' employees), which they argue can accrue to employees through salaries and training programs. On the other hand, SG&A may be used to build brands and more loyal customer bases, which are specific to the firm rather than the employees. Our churn measure, therefore, can be used to help determine whether SG&A is mostly accruing to the employees or the firm. We believe that low observed churn is evidence that SG&A is going to the firm, which yields our final testable prediction:

Prediction 4C: *The relationship between organization capital and systematic risk should only be present among firms where organization capital is embedded in employees i.e., high churn firms.*

Beyond the predictions tested directly in this paper, customer-base centric measures constructed from transaction-level data could shed light on new dimensions of firms that are not observable from accounting numbers. For example, models with firm heterogeneity have become increasingly important for explaining stylized facts about the distribution of firm size and market power (Hottman et al., 2016) as well as aggregate growth rates (Klenow et al., 2020). The ability to observe disaggregated sources of revenue, customer base characteristics, spatial competition, and individual-level shifts in customer loyalty promises an enhanced ability to test models across macroeconomics, IO, marketing, and asset pricing.

3 Data

3.1 Transaction-Level Linked-Account Data

Our data comes from a large online financial service provider that acts as an online aggregator of individuals' financial accounts. Online aggregation of financial accounts is a popular service that allows users to easily monitor financial activities from across multiple financial institutions using a single web-page or smart-phone app. Further, many large banks offer aggregation services as a feature on their own websites, mitigating sample selection concerns in these situations.

After linking credit card and checking accounts, the service automatically and regularly pulls data from the user's financial institutions. The data contains transaction-level data similar to those typically found on bank or credit card statements, containing the amount, date, and description of each transaction. The full dataset contains 2.7 million users from 2010 to 2015 and, though the sample grows slowly over time, there is very little attrition.

Recent work has utilized similar transaction-based sources to make inferences about the financial habits of the broader population. For instance, Baker (2018), and Kueng (2018) also utilize similar data from an online personal finance platform. They perform a multitude of validation tests comparing to data sources such as Census Retail Sales, home price data from Zillow, the Survey of Consumer Finance, and the Consumer Expenditure Survey. They find a close parallel between household-level financial behaviors and distributions in these sources relative to those found among users of the online platform. Such results point to the fact that, while these types of bank-derived sources will mechanically exclude financial activity by the unbanked, transaction-level financial data can produce accurate portrayals of aggregate economic activity and household behavior.

3.2 Firm Selection and Matching Procedure

We begin our analysis by matching credit and debit card transactions to firms in order to link to time-varying firm characteristics and financial performance. The initial universe of transaction descriptions is made up of about 25 million unique strings. This reflects not only a large number of unique firms, but also differences in description strings within firm driven by things like numeric transaction descriptions (e.g., 'txn: 491349'), establishment locations (e.g., 'walmart super center lancaster'), and how different credit and debit cards include or exclude punctuation.

Because we link transaction descriptions to particular firms, we are unable to utilize transactions without an associated merchant. For instance, ATM withdrawals, physical checks, and payment apps (e.g., Venmo or Paypal) will not be able to be matched to a merchant. This introduces some measurement error into our transaction-based pictures of firms. However, cash transactions are a fairly small and shrinking component of overall consumer spending and checks are most typically utilized for large financial payments like rent and car payments rather than for the retail goods and services purchases that we focus on.

To develop the set of firms to match to these transactions, we start with Compustat and refine to the universe of public firms in a set of industries that meet our criteria of being mostly consumer-facing. These industries include building materials and garden supply, general merchandise retailers, grocery stores, restaurants, hotels, personal and business services, utilities, home furnishings, apparel, communications, and airlines.[9] In addition, to supplement our set of public firms, we search the web for lists of large private firms in these sectors. We find lists from sources such as Business Insider, Forbes, and Wikipedia that enumerate the largest firms and retailers in a range of categories. In total, firms in these industries cover trillions of dollars of revenue per year and represent a larger portion of GDP than manufacturing.

Due to limitations of traditional machine learning algorithms in this setting, we mostly rely on manual inspection and experimentation to find descriptions that map to this set of firms. After working through our sets of large consumer-facing firms, we turn directly to the transaction data to fill in any potential holes in the data. We attempt to map any unmatched transaction descriptions within the most frequent 10,000 transaction descriptions. Generally, these descriptions refer to firms from an industry that we did not previously inspect. For instance, Lyft and Uber appear frequently in our data but are assigned a two-digit SIC industry of 41 (Local And Suburban Transit And Interurban Highway Passenger Transportation). Netflix similarly was not in one of our focused consumer facing industries according to our SIC classification (it is found with two-digit SIC of 78, which mostly contains movie producers).

Because some firms span a number of distinct brands, we must use data on subsidiaries to link brands to their parent company. For example, Yum! Brands is the corporation that owns brands like Taco Bell, Pizza Hut, and KFC. When households purchase food from one of these restaurants, their credit or debit transaction will list the merchant as ‘Taco Bell’ rather than ’Yum! Brands’. For the purposes of this paper’s analysis, we collect household transactions across all of these brands to arrive at firm-level statistics. For future work, one benefit of this class of data is that revenue and customer information can be separately identified by brand within a parent firm.

[9] These correspond to the two-digit SIC codes: 45, 48, 49, 52, 53, 54, 55, 56, 57, 58, 59, 70, 72, and 73. We end up excluding most gasoline stations as their revenue is typically combined with a large refiner or oil producer and thus, while the consumer-facing side is well observed, the consumer-facing business does not provide a good gauge of overall firm revenue or operations.

3.3 Matched Firm Sample

In the end, we are able to match 428 public and 130 private firms within our sample window. While these firms constitute a small fraction of total firms, they are also by far the largest consumer facing firms in the economy. To illustrate this, we match our public firm data to Compustat and rank firms based on their total 2014 revenue. In all industries, the average firm in our matched dataset is large relative to the average firm in Compustat. In total, we are able to observe essentially 100% of consumer spending among our sample and can assign approximately 32% of this spending to a matched firm in our data. This is substantially higher than the portion of spending matched in the Nielsen Consumer Scanner Panel. Appendix Table A.1 compares the numerical ranks (with one being the highest), and percentile ranks (with 100% being the highest) of the firms in our matched sample by industry.

For industries where we have extensive coverage, like airlines, general merchandise, and groceries, we are able to match all of the largest firms. In other industries we have only partial coverage of top firms. For example, we do not match to the Disney Corporation, one of the largest firms in the consumer telecom industry because, prior to the introduction of Disney Plus in 2019 (well after the end of our sample), households generally did not interact directly with the parent company itself. Rather, households predominantly consumed Disney products through the viewing of movies at theaters or the purchasing of toys from retailers.

Table 1 provides some summary statistics regarding our matched firm-level data. In the first row, we see the median firm in our sample receives approximately $1.6M from the linked users in our sample in a given quarter. Firm-level spending is skewed towards the largest firms, with the average firm receiving about $8.4M. The largest single firm (Walmart) has approximately $550M per quarter of observable revenue from our sample of users.

The second row in the table displays the fraction of firms' quarterly revenue that we observe among the users in our matched sample. We can only calculate this statistic for public firms with data available in Compustat. On average, we capture about 0.6% of a firm's quarterly revenue (median of 0.4%). There is substantial heterogeneity in the fraction of revenue that we observe in our data – the fraction may be impacted by the portion of a firm's revenue obtained from foreign consumers, whether a firm has substantial business-to-business revenue that is unobserved in our data, and if a firm has a large portion of transactions conducted with cash rather than cards. We think that this matching procedure suffices for illustrating the benefits of better understanding customer churn and similarities across firms. Given the amount of research that focuses solely on publicly traded firms, a limitation to large firms is not necessarily an impediment to inference regarding important drivers of firm behavior, in general. However, for researchers interested in more fully mapping out networks of competition, entry and exit, or private firms, it is possible to substantially increase the number of matches to smaller firms. Moreover, the procedure can be easily extended to similar household financial transaction datasets that have longer time horizons, allowing for a more thorough analysis of shifts in customer bases within a firm over time.

4 Validation of Firm Matching and Transaction Data

In this section, we provide evidence that our transaction data provides a meaningful view of both consumer spending in general and of firm characteristics.

4.1 Transaction Data Validation

Our sample is not drawn from a random sample of the population, but it appears to be widely representative with some exceptions. Relative to other FinTech data providers with products aimed at narrower slices of the population (e.g., lower income households or households interested in peer-to-peer lending), this data is sourced from a much broader range of households dispersed across the country. This database is also used in Baugh et al. (2018) and Baugh and Correia (2022) where the authors illustrate that the income distribution of users in the sample is comparable to that of the U.S., with substantial deviations only in the lowest income bins (see Appendix Figure A.2).

Further, similar to Baker (2018), we compare household spending in our dataset with the U.S. Census monthly retail trade report in the categories of general merchandise, groceries, restaurants, and gas. Spending for both data sources is scaled to the value of 1 in January of 2011. As shown in Appendix Figure A.3, the spending patterns in our sample closely mirror the monthly retail trade report for the given categories. The correlation between the aggregator data to the monthly retail trade report averages 89% across the four categories, with the highest correlation of 96% occurring in the category of general merchandise.

4.2 Firm Matching Validation

4.2.1 Customer Characteristics and Revenue

Our first firm matching validation test is to match total aggregated consumer spending for public firms in our sample to their quarterly Compustat revenue data from 2010 to 2015. In Figure 1, we plot both levels and changes in logged spending against levels and changes in logged Compustat revenue. While the absolute levels are different owing to the fact that we observe only a fraction of individuals in the economy (1.7 million users, out of a total U.S. adult population of 245 million), we find a strong correlation between our own spending data and the revenue reported by public firms in relative terms. We do a good job of matching relative sizes of firms as well as within-firm quarterly growth. Our measure achieves higher rates of correlation and fit when restricting to firms that do not have sizable operations overseas or business-to-business revenue.

4.2.2 Geographic Locations - Chain Store Guides

We also test whether the geographical distribution of stores revenue in our data matches the empirical distribution of firms' establishments using data from Chain Store Guide (CSG) database, which tracks the physical locations of retailer branches for large regional and national chains.

We collect CSG data from the entirety of our sample period (2010-2015) and match 58 firms from the CSG database. We then construct two measures of firm geographic dispersion from our transaction data. First, we simply calculate the fraction of consumer spending that we observe from users located in a given state at a particular firm for each year in our sample.

$$FracSpend_{ist} = \frac{\sum\limits_{i} spending_{irst}}{\sum\limits_{i}\sum\limits_{s} spending_{irst}}$$

Where i indexes users, r indexes retailers, s indexes states, and t represents a year.

We would not necessarily expect a perfect one-to-one relationship between these measures for each retailer. Especially for the fraction of spending we observe, since we do not have establishment-level sales data. While a state may have 10% of a retailer's physical stores, those stores may account for 15% of that retailer's national sales. However, on average we would expect a strong relationship between these measures.

As a second geographic metric, using the transaction-level description strings, we are able to pick out transactions at particular retailers' establishments. For instance, a transaction may be labeled as 'McDonalds (Store #391)' rather than simply as 'McDonalds'. We utilize this to construct a measure of the fraction of a retailer's locations in a state each year in both our transaction data as well as the CSG database.

In Figure 2, we display bin-scatter plots of these measures across all state-years in our sample. In the top row, we plot the relationship between the two store level measures (fraction of stores by state-year-retailer in our transaction data against fraction of stores by state-year-retailer in the CSG data). The right panel censors the plot to better highlight the fit among the smaller states. The bottom row displays the relationships between the fraction of spending that we observe for a retailer in a state-year against the fraction of stores from the CSG data in a state-year.[10] Across all specifications, there is a strong positive relationship between the geographic distribution of spending in the CSG data and in our transaction data.

5 Churn, Intangible Capital and Adjustment Costs

We now turn to our transaction-based measure of churn within a firm's customer base. Broadly, we provide evidence that customer base attributes attainable from transaction data can add significantly to the understanding of firms and cross-sectional heterogeneity among firms. We also show that low churn firms' investment is less responsive to changes in their opportunity set, consistent with them facing higher adjustment costs when launching new products.

5.1 Drivers of Customer Base Churn

As discussed in Section 2, we expect switching costs to be a key driver of customer churn. Specifically, prediction 1A argues that industries with high switching costs should have lower average churn. To explore this hypothesis, Figure 3 plots the distribution of firm-level customer churn by industry. A first clear takeaway is that much of the variation in rates of customer churn is coming from cross-industry (rather than within-industry) variation. In addition, firms in high switching

[10] Appendix Figure A.4 breaks this down by state. In all cases, we see a strong relationship lying close to the 45-degree line, suggesting that we are getting an unbiased sample of the geographic distribution of spending, on average.

cost industries – like utilities and telecommunications – tend to have highly persistent customer bases. In contrast, the customers in low switching cost industries – such as clothing retailers and general merchandise – tend to be much less persistent across years. So, consistent with prediction 1A, Figure 3 provides evidence that a significant share of the variation in churn is driven by the nature of contracts and competition within these industries. Consequently, churn tends to be fairly consistent within firms over time. For instance, regressing churn on lagged churn yields a coefficient of 0.85.[11]

As outlined in prediction 1B, the presence of these switching costs may affect firms' strategies for obtaining and retaining customers, which can ultimately drive further differences in customer churn. To test prediction 1B, we use three different measures of spending on customer acquisition. The first two are the ratio of SG&A to sales and the ratio of advertising to sales. One could also imagine that because it takes time to build a loyal customer base, the cumulative stock of SG&A – rather than its period-by-period flow – matters, so we also leverage the organization capital measure of Eisfeldt and Papanikolaou (2014), which captures the capitalized stock of current and past SG&A spending.

Figure 4 displays the correlation between our firm-year customer acquisition measures and our measure of firm-level customer churn. Following the logic in prediction 1B and the findings in Hall et al. (1997), we first split our sample into retail – i.e., low switching cost firms – and non-retail – i.e., high switching cost firms – using the SIC categorization (retail firms are those with a primary SIC-1 digit code of 5). We find that the relationship between customer churn and SG&A (or advertising expenditures) is negative for non-retail firms but positive for retail firms.[12] These results are consistent with prediction 1B, which outlines that in high switching cost industries, SG&A should be negatively related to churn, while this relationship should become more positive in low switching cost industries.

[11] While firms exhibit large differences in their average levels of customer churn when compared to each other, some firms see substantial changes in churn over time. JC Penney undertook a drastic change in pricing at the retailer in Q1 2012, doing away with most of the "sale" and coupon-based pricing. JC Penney's customer base reacted strongly and negatively to this change. In Appendix Figure A.5, we show the rate of quarter-on-quarter customer churn for JC Penney around this change, normalized by average churn for that quarter within the 1-digit SIC industry. We see a large and persistent increase in churn, approximately 1.5 standard deviations, following this change.

[12] Given that our sample of non-retail firms are also selected to be consumer-facing, these differences are unlikely to be driven by differences in measurement error across firm types. As a test for differential measurement error, splitting our sample into retail and non-retail and performing the same validation tests seen in Figure 1 yields very similar magnitudes and explanatory power across samples.

Rather than just splitting on retail vs. non-retail, a more precise test of prediction 1B would measure switching costs within finer industry groups. While we cannot observe such costs directly, we can proxy for switching costs by looking at the overlap between firms in the same industries' customer bases. The logic is that in industries with significant overlap or low exclusivity, spending on customer acquisition may be high, but churn is likely also high. Therefore, in these industries, we might expect this slope to be relatively more positive. Alternatively, in low overlap or high exclusivity industries, we expect the slope between customer acquisition and churn to be relatively more negative. In Figure 5 we show this is indeed the case, as low overlap industries have a negative relationship between SG&A and churn, while the opposite is true for high overlap industries.

As discussed in Section 2, we might expect the slope of the relationship in Figure 5 to be steeper using churn calculated only from new customers. Leveraging the flexibility of the customer base centric data, we are able to define a modified version of our churn metric to test this prediction:

$$Churn_{new,f,t-k} = \left(\sum_i |s_{f,i,t} - s_{f,i,t-k}| \right) \Big/ 2 \tag{5}$$

where the sum $\sum_i |s_{f,i,t} - s_{f,i,t-k}|$ is taken over all customers that shop at firm f *only* in year t *and not* year $t - k$. In words, $Churn_{new,f,t-k}$ is the sum of spending shares across all new customers in year t.

In the left panel of Figure 5, the slope is 10.7 while in the right panel, the slope is 30.1. Using $Churn_{new,f,t-1}$ instead of $Churn_{f,t-1}$, these slopes become 14.4 and 42.3, respectively. While these larger slopes are directionally consistent with our prior, the differences are not statistically significant. With only 9 data points (1 for each industry group), however, this test may be underpowered to detect even an economically large difference.

In addition to switching costs, prediction 1C argues that churn should be inversely related to a firm's local market power. To understand the relationship between local competition and churn, Figure 6 plots within-city churn against local categorical sales shares for two groups of firms: one composed of firms who generally have longer-term contracts (Utilities and Telecom firms i.e., high switching cost industries) and the other composed of firms that interact with customers through one-off purchases (Restaurants, Convenience Stores, General Merchandise, Groceries, and Entertainment i.e., low switching cost industries). In each group, and consistent with prediction

1C, we find that increases in local competition predict increases in local churn in customer bases.[13] Despite having the same directional relationship between local competition and churn, there are notable differences between these two categories of retailers. For one-off purchase firms, cities in which a firm tends to have fewer major competitors see much lower levels of churn. In contrast, for long-term contract firms, the local sales shares have substantially smaller effects on churn. That is, even with ample local competition, customers are often locked into a given firm for a number of years through contractual provisions.

5.2 Customer Bases as a Component of Intangible Capital

Many papers have discussed the rise in intangible capital over the past several decades and how this rise can lead economists to mismeasure things like productivity growth, competition, and markups.[14] The overall stock of intangible capital held by a firm is often measured by means of acquisition premia (e.g., Ewens et al. (2020)) or through a perpetual inventory method which aggregates flows of SG&A or R&D spending (e.g., Peters and Taylor (2017), Eisfeldt et al. (2020)).

Intangible capital, however, is not an undifferentiated concept: it reflects an amalgamation of a number of components such as R&D and patent holdings, advertising or brand capital, knowledge capital held by workers, business practices such as software utilization or novel supply chains, customer capital, and organization capital. Independent measurement of these pieces is important as they may not be highly or even positively correlated with one another. While overall productivity may hinge on aggregate intangible capital, other elements of firm-level risk or decision-making may depend on only a subset of these types. Therefore, utilizing aggregate intangible capital may yield biased estimates when examining the impacts on firm-level outcomes.

For example, past work (e.g., Traina (2021), Ayyagari et al. (2019) and De Loecker et al. (2020)) has discussed whether SG&A can be treated as an intangible investment or as a marginal operating expense. Returning to Figure 4, we see that differences in customer switching frictions

[13]Local spending shares are defined as $\frac{Spending_{icjt}}{\sum Spending_{cjt}}$ where i indexes firms, c indexes categories of spending, j indexes cities, and t indexes years. In Appendix Table A.2, we show the association between levels of local categorical spending share and churn, conditional on a range of fixed effects. Higher levels of local categorical sales dominance tend to drive significantly lower levels of customer churn. Moreover, this local sales dominance produces large increases in R^2.

[14]A small sample of papers include: Corrado et al. (2009), Sim et al. (2013), Eisfeldt and Papanikolaou (2014), Belo et al. (2019), Crouzet and Eberly (2019), Eisfeldt et al. (2020), Ewens et al. (2020).

can flip the sign of the relationship between spending on customer acquisition and churn. Our results show that using advertising expenses or SG&A as a proxy for customer attachment or brand capital will lead researchers to substantially different conclusions in different industries. Our measure of customer churn speaks directly to this element of intangible capital: higher levels of customer attachment to a firm and/or brand manifest in lower levels of churn within a firm's customer base.

More broadly, Prediction 2A outlines several reasons why we expect high churn firms to have less intangible capital. To test this hypothesis we run the following regression:

$$\text{Intangible Capital}_i = \alpha + \beta\overline{Churn_i} + \theta\overline{Sales_i} + \psi_j + \epsilon_i \tag{6}$$

where Intangible Value$_i$ are various measures of customer-related intangible capital. $\overline{Churn_i}$ is the average churn for a firm in our sample across all years it is present[15], $\overline{Sales_i}$ is average annual sales in Compustat, and ψ_j is a set of industry fixed effects.

Table 2 contains the results, highlighting a strong association between customer churn and intangible capital both within and across industries. Column 1 examines the relationship between customer churn and firms' book-to-market ratios, finding that firms with lower levels of churn command higher market values relative to their book value. This suggests high churn firms have relatively more of their value coming from assets in place, rather than intangible capital e.g., growth options (Kogan and Papanikolaou, 2014). Column 2 shows that this effect persists once we include industry fixed effects.

The next measure of intangible capital we examine is Brand Value, which we obtain for a subset of firms from Brand Finance's 'Brandirectory'. Their methodology examines components such as emotional connection, financial performance and sustainability and then applies royalty rates to calculate a capitalized brand value.[16] Columns 3 and 4 show that brand value is highly correlated with levels of customer churn.

The last measure of intangible capital we examine is market value per customer. We compute this by first estimating the number of customers for a firm by calculating average spending per

[15]We use average churn as annual churn to reduce measurement error concerns and increase power when performing these cross-sectional regressions. Results are similar using annual churn each year instead of average annual churn.

[16]Appendix Figure A.6 displays the relationship between brand value rankings and churn across a range of industry categories.

customer at a firm in our data, then dividing total firm sales (from Compustat) by that average customer spending. Then, we divide the firm's market capitalization by the estimated number of customers. Columns 5 and 6 show that average market value per customer is negatively related to churn. In terms of magnitudes, a one standard deviation decrease in churn is associated with an increase in per-customer market value of 25-40%, an effect that persists even after controlling for markups and brand value. Markup data are obtained from Loualiche (Forthcoming) who calculates markups using the method in De Loecker et al. (2020). Overall, these results are consistent with Prediction 2A, showing that lower churn in a firm's customer base tends to manifest as a more valuable customer base, a form of intangible capital.

As discussed in Predictions 2B, 2C and 3, the nature of the intangible capital held by low churn firms may affect their financing , price setting and investment behavior, which we examine in Table 3. The regression specification is the same as in Equation 6, except we further control for the ratio of SG&A to Sales to account for its relationship to various intangible proxies documented in Figure 4.

In Columns 1 and 2, we examine the relationship between customer churn and a firm's capital structure. As discussed in Prediction 2B, because they are composed of relatively more intangible capital – which is difficult to use as collateral – we might expect low churn firms to have relatively less leverage. Consistent with this, we find that high churn firms tend to hold less cash and are much more highly levered than low churn firms. This dynamic holds even when controlling for other proxies of intangible capital like levels of SG&A or R&D within a firm. In Column 3, we examine the relationship between churn and markups. As discussed in Prediction 2C, by nature of having more brand value or local market power, we might expect low churn firms to be able to charge higher markups. In line with this logic, we find that high churn firms have lower markups on average.

Finally, Column 4 shows the relationship between churn and the volatility of a firm's investment rate, defined as the ratio between capital expenditures and lagged assets.[17] As discussed in Prediction 3, low churn firms could face higher adjustment costs, and therefore may have a smoother investment profile over time. Consistent with this, we find that high churn firms have more volatile investment. In the next subsection, we aim to better understand this relationship

[17]To account for differences in the volatility of firms' investment opportunity sets, we normalize this quantity by the standard deviation of a firms Q (Hayashi, 1982).

under the framework of customer acquisition as a form of adjustment costs.

5.3 Churn and Adjustment Costs

Having shown that churn is related to intangible capital and firm behavior, we seek to apply our measure of customer churn to the Gourio and Rudanko (2014) framework, which highlights the importance of this specific component of intangible capital and produces predictions for firms' investment behavior. To this end, we turn to prediction 3, which argues that low churn firms should respond less to changes in their investment opportunity set, measured by Q shocks, and whether these predictions hold for SG&A as well. That is, we test whether the neoclassical model, in which firm investment responds immediately to changes in productivity, is a weaker fit for firms with high levels of customer attachment (and low customer churn). Explicitly required for SG&A to perform well as a proxy for customer capital is that SG&A is highly linked to firms or industries that have high barriers/frictions in their markets.

The model in Gourio and Rudanko (2014) is not necessarily about overall churn, but more about a firm's ability to attract new customers. Again, leveraging the richness of our data, we use the modified version of churn from new customers defined in Section 5.1 to better align with the parameters of the model. Specifically, prediction 3 argues that low churn firms should be less responsiveness to changes in their investment opportunity set, as measured by Tobin's Q. To test this hypothesis, we run the following regression:

$$\begin{aligned} Investment_{i,t} = a + \beta_1 Q_{i,t-1} + \beta_2 Q_{i,t-1} \times 1_{\text{Low SG\&A}_{i,t}} + \beta_3 Q_{i,t-1} \times 1_{\text{High New Churn}_{i,t}} \\ + \beta_4 Size_{i,t} + \phi_t + \theta_i + \epsilon_{i,t} \end{aligned} \tag{7}$$

where $Investment_{i,t}$ is the investment rate, defined as the ratio between a firm's capital expenditures and lagged assets. $1_{\text{Low SG\&A}_{i,t}}$ is an indicator for whether firm i has below median SG&A in year t, while $1_{\text{High New Churn}_{i,t}}$ is an indicator for whether firm i has above median $Churn_{new}$. The key coefficient of interest is β_3, which measures the difference in how high and low churn firms respond to Q shocks.

The results are in Table 4. In Column 1, we show that firms with low levels of SG&A do appear to be more like 'classical' no-adjustment-cost firms who respond more strongly to shocks to Q than firms with higher levels of SG&A spending. In Column 2, we repeat this regression, restricting the sample solely to firms in the retail sector (SIC-1 code of 5). Here, the coefficient on SG&A switches sign and is significantly different than zero, producing an effect opposite to our prediction. We assert that this change in sign is not due to this conceptual model failing among such firms, but because SG&A is not a good predictor of customer stickiness within the retail industry. Firms in this industry with the highest levels of customer attachment tend to be those that actually spend only small amounts on SG&A, as seen in Figure 4.[18]

Finally, Columns 3 and 4 include an interaction of lagged Q with an indicator for a firm having higher than median levels of annual customer churn alongside the low SG&A indicator. Here, consistent with prediction 3, the interaction terms on the high churn indicator are of the predicted sign and significant when examining all firms and when restricting to retailers. Specifically, high churn firms tend to respond about 50% more strongly to changes in Q than do low churn firms. Moreover, controlling for firm-level churn renders the coefficients on the SG&A interaction term near-zero in magnitude and statistically insignificant.

In short, our measure of customer churn consistently demonstrates the impacts of firm-level customer search frictions and provides further empirical support for an adjustment cost model proposed by Christiano et al. (2005). Moreover, we find that using SG&A can yield substantially biased estimates of the effects of customer capital on firm markups and investment behavior.

6 Customer Churn and Firm Risk

Churn in firm-level customer bases over time is a key metric with which to assess customer-facing firms. Higher levels of churn in customer bases can be a source of risk and volatility across firms who rely on such customers for their sales. More specifically, as discussed in Section 2, we expect churn to be positively related to exposure to *systematic risk*. This could be either because high churn firms will adjust investment more in the face of aggregate shocks (Gourio and Rudanko,

[18]More specifically, consider the model of Afrouzi et al. (2020), where there are differential benefits to acquiring customers based on a firm's productivity. High productivity firms which by nature can support stickier customers might spend more on S&GA precisely because they know each acquired customer is more valuable. As discussed in Section 2, however, retail firms may need to spend money on SG&A as part of constant business stealing, rather than because of differences in productivity. This may explain why SG&A fails to capture customer base adjustment costs for retail firms.

2014) or because high churn firms are less able to adjust prices (Kleshchelski and Vincent (2009), Weber (2014)). To test Prediction 4A, we run the following regression:

$$Risk_{i,t} = \alpha + \beta Churn_{i,(t-1,t)} + \gamma ln(Revenue_{i,t-1}) + \text{FE} + \epsilon_{i,t} \tag{8}$$

$Churn_{i,(t-1,t)}$ is measured based on each year's customer base, relative to the previous year's customer base, $Revenue_{i,t-1}$ is the firm's total revenue in Compustat last year, and FE represent industry- or firm-level fixed effects. To measure $Risk_{i,t}$, we focus on CAPM beta and idiosyncratic stock volatility, but find similar results using total stock volatility or revenue volatility.

Column 1 in Table 5 shows that a strong positive correlation between our measure of churn and CAPM beta. In terms of magnitudes, the point estimate implies that a firm in the 90th percentile of churn (0.87) has a 0.39 higher CAPM beta than a firm in the 10th percentile (0.30). Columns 2-3 show that including fixed effects for 2-digit SIC industries or Hoberg and Phillips (2010)-50 industries shrinks the point estimate by about 50%, but the relationship between churn and CAPM beta remains statistically significant.

Column 4 shows that the relationship between churn and CAPM beta survives including firm-level fixed effects. This is a high bar, as we only have 4 years of data for each firm. Finally, columns 5-8 show the results for idiosyncratic volatility, which mirror those for CAPM beta. Overall, our results are consistent with prediction 4A, and suggest that firms which have more churn in their customer bases are more exposed to systematic risk and are more volatile than other firms. Further, the robustness of our results to including firm fixed effects implies that *changes* in churn over time predict changes in risk exposure and volatility within a firm. [19]

[19]Two potential issues arise from Equation 8. This estimation approach mask a non-linear or non-monotonic relationship between churn and risk and it may be heavily influenced by small firms. To rule out these issues, we form value-weighted portfolios of firms based on the churn in their customer base and test the relationship between average churn and firm-level equity price volatility between 2010 and 2019. In Appendix Table A.3, we find that CAPM betas monotonically increase from low churn to high churn portfolios but that the positive relationship between churn and CAPM beta is mostly coming from the extreme portfolios. Another concern is that these results are driven by the relationship between firm size and churn. We perform a double sort, first forming 3 groups based on a firm's total revenue in the previous year and then forming 3 sub-groups based on churn. Appendix Table A.4 shows that the monotonic increasing relationship between churn and CAPM beta holds within each tercile of firm revenue.

6.1 Firm-Specific Revenue Declines During COVID-19

In Section 2, we argue why high churn firms may be especially exposed to economic downturns. COVID-19 presented an opportunity to perform an out of sample test of how having low customer-base attachment (high churn) can drive demand-side risk for firms. Baker et al. (2021) noted that the tendency for households to visit new retailers declines as income declines. This may manifest during a recession with households retrenching into their usual retailers/restaurants and not trying out somewhere they have not visited before. High churn firms may also lack the ability to adjust prices as freely during downturns to retain customers and market share (Gilchrist et al., 2017). To test prediction 4B, we examine whether firms relying on a steady stream of new customers (i.e., high churn) are more strongly impacted by the recent COVID-19 outbreak and recession.

During March 2020, city and state governments began unprecedented efforts to halt the spread of COVID-19 by dramatically limiting the ability of retail businesses to remain open and to operate normally. Many businesses were virtually halted or else mandated to operate only remotely. For instance, restaurants were often required to allow only take-out or delivery orders, and many other retail establishments were forced to operate only online, using delivery services or curbside pickup.

In Table 6, we utilize data from the SafeGraph Data Consortium to examine the impact of these events on consumer spending at a range of retail establishments and how these changes in spending are linked to rates of churn measured at those retailers in earlier years. SafeGraph uses data from a range of debit cards to track aggregated levels of daily consumer spending across merchants. We use daily spending data from January 2019 through the end of March 2020 and can observe hundreds of millions of transactions at retailers linked to our measure of customer churn.

Column 1 shows that firms, on average, saw 30% reductions in customer spending during March 2020 as compared to March 2019. In Column 2, we see that firms with high levels of customer churn (estimated using the 2010-2015 data) saw much larger declines in customer traffic and spending than those with low levels of churn: a firm in the top quartile of churn saw a decline in spending about three times larger than those in the bottom quartile. In Column 3, we retain a strong negative impact of churn above and beyond controls for firm-level CAPM betas and firm size interacted with indicators for March 2020. Given that we are controlling for *average* exposure to systematic risk by including a firm's CAPM beta on the right-hand-side, these results support prediction 4B that high churn firms should be especially exposed to negative economic shocks.

These effects are mostly driven by firms with high levels of extensive margin churn.

While there are substantial concerns about differential treatment across different sectors of the economy during COVID (e.g., some types of retailers faced more legal restrictions than others), these correlations between revenue declines and customer churn are not driven by differences across industries. Using both industry and industry by month fixed effects in columns 4 and 5, we see that the effect persists with a similar magnitude.

As discussed in Section 2, we might expect the relationship in Table 6 to be stronger when using the new-customer-only version of churn in Equation 5. We find that this prediction holds empirically, with coefficients twice as large using $Churn_{new,f,t-1}$ in place of $Churn_{f,t-1}$.

6.2 Churn and Organization Capital

Separately identifying customer attachment as a component of intangible capital can not only make inferences regarding risk and customer capital clearer, but can also clarify the impacts of other elements of intangible capital within a firm. One application of this is discussed in Prediction 4C, where we can use our measure of churn to distinguish between SG&A spending which accrues to the firm vs. to employees. This is motivated by the model in Eisfeldt and Papanikolaou (2013), where organization capital becomes a source of exposure to systematic risk for firms because it accrues to employees, who can abscond with it when their outside option is high enough. We argue that in low churn firms, SG&A is mostly accruing to the brand/firm, and therefore the relationship between organization capital and exposure systematic risk should only be present among high churn firms.

To test this hypothesis, each month we perform a 3×3 sort on churn and (Organization capital)/(Total book assets plus organization capital), hereafter OK/AT. Organization capital is measured by capitalizing SG&A in a perpetual inventory method (see e.g., Eisfeldt and Papanikolaou (2013), Eisfeldt and Papanikolaou (2014), Eisfeldt et al. (2020)).[20] To this end, we first sort firms into 3 terciles of churn. Then, within each of these 3 buckets, we form 3 sub-terciles based on OK/AT. To reduce the influence of small firms, within each month, observations are value-weighted within each of the 9 portfolios.

[20] We obtain data on organization capital scaled by total assets from the authors' GitHub repository. Following Eisfeldt et al. (2020), we remove all observations were OK/AT is 0 because SG&A is missing/zero in Compustat, or where OK/AT is less than zero because book assets are less than zero.

Table 7 contains the results. Consistent with prediction 4C, we see that there is a monotonic increasing relationship between OK/AT and CAPM beta among high churn firms, but the relationship is essentially flat among low churn firms. We believe that this is because if a firm has both low churn, but high organization capital, SG&A is accruing to the firm through the creation of brand capital. The fact that this type of organization capital is sticky means that these firms are not riskier than low churn firms with less organization capital.

The opposite is true for the firms with high organization capital and high churn: their investments in intangible capital (i.e., SG&A) are accruing to employees. Because employees can leave the firm at any time, this stock of organization capital makes these firms riskier.[21]

7 Conclusion

With the importance of intangible capital among firms growing substantially in the past few decades, it is imperative to have metrics that clearly identify its components. These measures can help to illustrate the drivers of heterogeneity across industries and firms when it comes to risk, investment, and markups. Intangible capital is generally described as an amalgamation of a number of components such as brand or customer capital, organization capital, business practices, and applied R&D/patent activity but is often measured in an undifferentiated manner.

Using household financial transaction data, this paper demonstrates that it is possible to construct accurate pictures of firm characteristics at a highly granular level. We develop measures of firm-specific churn in customer bases that aim to provide a tool to disentangle important elements of intangible capital across firms.

Having developed our churn measures, we start by exploring the determinants of churn, highlighting the role of customer frictions. Specifically, churn is higher on average in industries with low switching costs and for firms with less local market power. We then show that customer churn is important for understanding both firm financial and economic outcomes. Churn correlates highly with a range of metrics of firm-level risk and volatility and outperforms typical measures in predicting revenue declines during the COVID-19 pandemic. We demonstrate that churn uniquely captures elements of customer and organization capital that are unobserved when using a proxy

[21] Appendix Table A.5 performs a triple sort on size, churn and OK/AT to rule out that these results are driven by size. The relationship between OK/AT and CAPM beta is strongest among high churn firms in each firm size bin.

like SG&A spending, better explaining cross-sectional variation in markups, investment behavior, and equity returns.

In addition, this paper highlights the broader potential for further customer centric measures to be constructed with household transaction data for use by policymakers and researchers, making several firm-level measures available publicly. We would encourage other researchers in areas that focus on firm behavior and asset prices to leverage transaction data in order to answer questions regarding consumer-facing firms.

8 Data Availability

Code replicating the tables and figures in this article can be found in the Harvard Dataverse: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/1ILIGD (Baker et al., 2023)

References

Hassan Afrouzi, Andres Drenik, and Ryan Kim. Growing by the masses-revisiting the link between firm size and market power. 2020.

Sumit Agarwal and Wenlan Qian. Consumption and debt response to unanticipated income shocks: Evidence from a natural experiment in singapore. *American Economic Review*, 104(12):4205–4230, 2014.

Sumit Agarwal, Wenlan Qian, and Xin Zou. Disaggregated sales and stock returns. *Working Paper*, 2020.

David Argente, Salomé Baslandze, Douglas Hanley, and Sara Moreira. Patents to products: Product innovation and firm dynamics. 2020.

Eva Ascarza. Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 2018.

Deniz Aydin. Consumption response to credit expansions: Evidence from experimental assignment of 45,307 credit lines. *Working Paper*, 2019.

Meghana Ayyagari, Asli Demirguc-Kunt, and Vojislav Maksimovic. The rise of star firms : Intangible capital and competition. *Policy Research Working Paper Series 8832*, 2019.

Pierre Bachas, Paul Gertler, Sean Higgins, and Enrique Seira. How debit cards enable the poor to save more. *Working Paper*, 2019.

Scott R. Baker. Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data. *Journal of Political Economy*, 126(4):1504–1557, 2018.

Scott R. Baker and Lorenz Kueng. Household financial transaction data. *Annual Review of Economics*, 14:47–67, August 2022.

Scott R. Baker, Lorenz Kueng, Steffen Meyer, and Michaela Pagel. Measurement error in imputed consumption. *Working Paper*, 2020.

Scott R. Baker, Brian Baugh, and Lorenz Kueng. Income Fluctuations and Firm Choice. *Journal of Financial and Quantitative Analysis*, 2021.

Scott R Baker, Brian Baugh, and Marco Sammon. Replication data for customer churn and intangible capital. *Harvard Dataverse*, 2023. URL https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/1ILIGD.

Tetyana Balyuk and Emily Williams. Friends and family money: P2p transfers and financially fragile consumers. *Available at SSRN*, 2021.

Brian Baugh and Filipe Correia. Does paycheck frequency matter? evidence from micro data. *Journal of Financial Economics*, 143(3):1026–1042, 2022.

Brian Baugh, Itzhak Ben-David, and Hoonsuk Park. Can taxes shape an industry? evidence from the implementation of the amazon tax. *Journal of Finance*, 73(4):1819–1855, 2018.

Frederico Belo, Vito Gala, Juliana Salomao, and Maria Ana Vitorino. Decomposing firm value. *NBER Working Paper 26112*, 2019.

Gideon Bornstein et al. Entry and profits in an aging economy: The role of consumer inertia, 2018.

Bart J Bronnenberg, Sanjay K Dhar, and Jean-Pierre H Dubé. Brand history, geography, and the persistence of brand shares. *Journal of political Economy*, 117(1):87–115, 2009.

Bart J Bronnenberg, Jean-Pierre H Dubé, and Matthew Gentzkow. The evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, 102(6):2472–2508, 2012.

Cecilia R Caglio, R Matthew Darst, and Sebnem Kalemli-Ozcan. Risk-taking and monetary policy transmission: Evidence from loans to smes and large firms. Technical report, National Bureau of Economic Research, 2021.

Lawrence J Christiano, Martin Eichenbaum, and Charles L Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45, 2005.

Carol Corrado, Charles Hulten, and Daniel Sichel. Intangible capital and us economic growth. *Review of income and wealth*, 55(3):661–685, 2009.

Nicolas Crouzet and Janice Eberly. Understanding weak capital investment: The role of market concentration and intangibles. *Working Paper*, 2019.

Jan De Loecker, Jan Eeckhout, and Gabriel Unger. The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644, 2020.

Winston Dou, Yan Ji, David Reibstein, and Wei Wu. Customer capital, financial constraints, and stock returns. *Financial Constraints, and Stock Returns (February 7, 2019)*, 2019.

Jean-Pierre Dubé, Günter J Hitsch, and Peter E Rossi. State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*, 41(3):417–445, 2010.

Janice Eberly, Sergio Rebelo, and Nicolas Vincent. What explains the lagged-investment effect? *Journal of Monetary Economics*, 59(4):370–380, 2012.

Alex Edmans. Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial Economics*, 2011.

Andrea Eisfeldt and Dimitris Papanikolaou. Organization capital and the cross-section of expected returns. *Journal of Finance*, 68(4), 2013.

Andrea L Eisfeldt and Dimitris Papanikolaou. The value and ownership of intangible capital. *American Economic Review*, 104(5):189–94, 2014.

Andrea L Eisfeldt, Edward Kim, and Dimitris Papanikolaou. Intangible value. Technical report, National Bureau of Economic Research, 2020.

Michael Ewens, Ryan Peters, and Sean Wang. Measuring intangible capital with market prices. *Working Paper*, 2020.

Doireann Fitzgerald, Stefanie Haller, and Yaniv Yedid-Levi. How exporters grow. Technical report, National Bureau of Economic Research, 2016.

Claes Fornell, Forrest Morgeson, and Tomas Hult. Stock returns on customer satisfaction do beat the market: Gauging the effect of a marketing intangible. *Journal of Marketing*, 2016.

Simon Gilchrist, Raphael Schoenle, Jae Sim, and Egon Zakrajšek. Inflation dynamics during the financial crisis. *American Economic Review*, 107(3):785–823, 2017.

Francois Gourio and Leena Rudanko. Customer Capital. *Review of Economics Studies*, 81, 2014.

Simon Hall, Mark Walsh, and Anthony Yates. How do uk companies set prices? 1997.

Fumio Hayashi. Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica: Journal of the Econometric Society*, pages 213–224, 1982.

Gerard Hoberg and Gordon Phillips. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies*, 23(10):3773–3811, 2010.

Colin J Hottman, Stephen J Redding, and David E Weinstein. Quantifying the sources of firm heterogeneity. *The Quarterly Journal of Economics*, 131(3):1291–1364, 2016.

Gurumurthy Kalyanaram, William T Robinson, and Glen L Urban. Order of market entry: Established empirical generalizations, emerging empirical generalizations, and future research. *Marketing science*, 14(3_supplement): G212–G221, 1995.

Ryan Kim. The effect of the credit crunch on output price dynamics: The corporate inventory and liquidity management channel. *Available at SSRN 3163872*, 2018.

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

Isaac Kleshchelski and Nicolas Vincent. Market share and price rigidity. *Journal of Monetary Economics*, 56, 2009.

Leonid Kogan and Dimitris Papanikolaou. Growth opportunities, technology shocks, and asset prices. *The journal of finance*, 69(2):675–718, 2014.

Leonid Kogan, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, 132(2):665–712, 2017.

Lorenz Kueng. Excess Sensitivity of High-Income Consumers. *Quarterly Journal of Economics*, 133(4):1693–1751, 2018.

Aurelie Lemmens and Sunil Gupta. Managing churn to maximize profits. *Marketing Science*, 2020.

Baruch Lev. *Intangibles: Management, measurement, and reporting*. Brookings institution press, 2000.

Baruch Lev and Suresh Radhakrishnan. The measurement of firm-specific organization capital, 2003.

Baruch Lev, Suresh Radhakrishnan, and Weining Zhang. Organization capital. *Abacus*, 45(3):275–298, 2009.

Erik Loualiche. Asset pricing with entry and imperfect competition. *Journal of Finance*, Forthcoming.

Paolina C. Medina. Selective attention in consumer finance: Evidence from a randomized intervention in the credit card market. *Working Paper*, 2020.

Monica Morlacco and David Zeke. Monetary policy, customer capital, and market power. *Journal of Monetary Economics*, 131, 2021.

Arna Olafsson and Michaela Pagel. The liquid hand-to-mouth: Evidence from personal finance management software. *Review of Financial Studies*, 31(11):4398–4446, 2018.

Luigi Paciello, Andrea Pozzi, and Nicholas Trachter. Price dynamics with customer markets. *International Economic Review*, 60(1):413–446, 2019.

R. H. Peters and L. A. Taylor. Intangible capital and the investment-q relation. *Journal of Financial Economics*, 123 (2), 2017.

Jae Sim, Dalida Kadyrzhanova, and Antonio Falato. Rising intangible capital, shrinking debt capacity, and the us corporate savings glut. Technical report, Society for Economic Dynamics, 2013.

Michael D. Smith and Erik Brynjolfsson. Consumer decision-making at an internet shopbot: Brand still matters. *Journal of Industrial Economics*, 49, 2001.

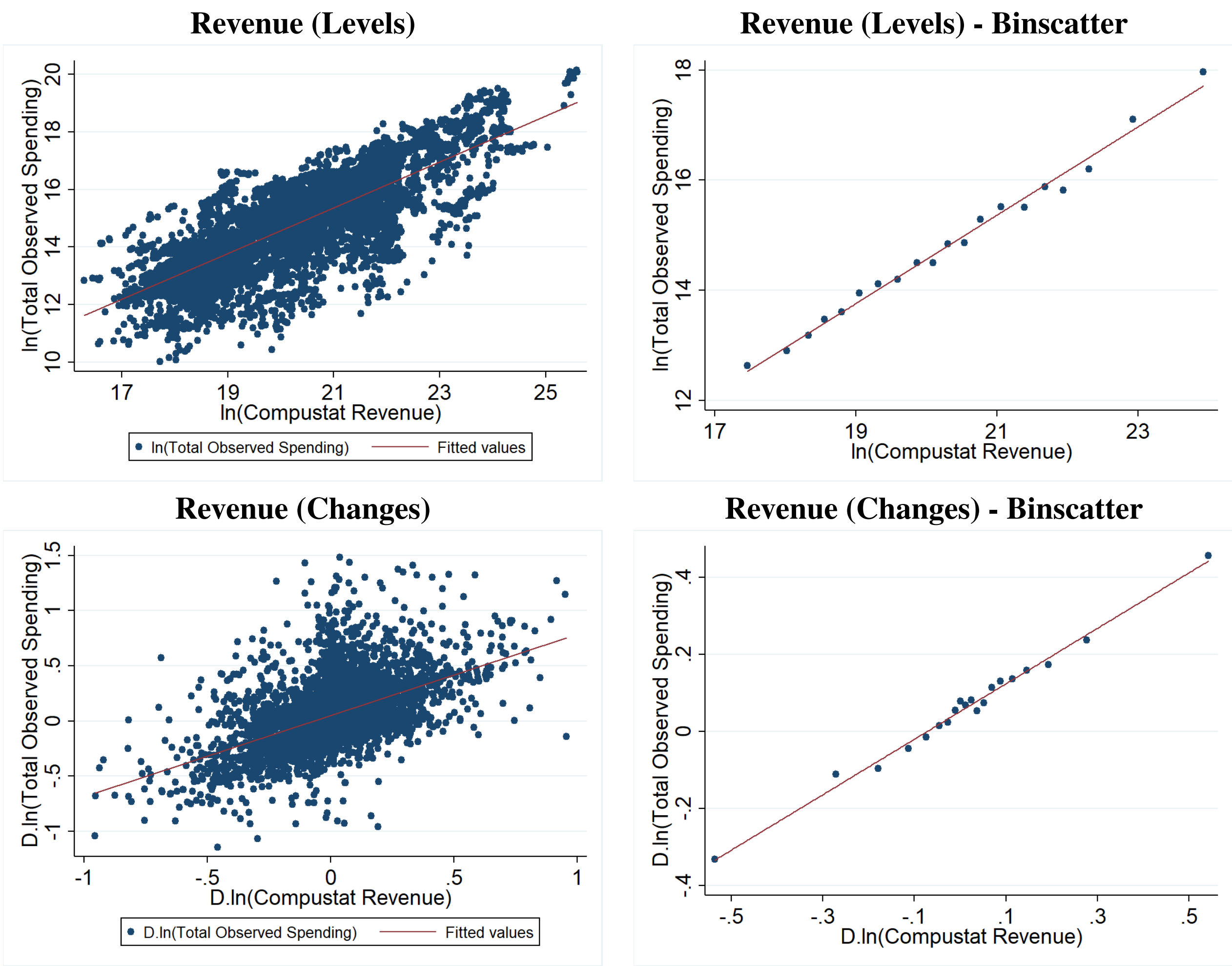
Qi Sun and Mindy Zhang Xiaolan. Financing intangible capital. *Journal of Financial Economics*, 133(3), 2019.

Jacques-Francois Thisse and Xavier Vives. On The Strategic Choice of Spatial Price Policy. *American Economic Review*, 78, 1988.

James Traina. Is aggregate market power increasing? production trends using financial statements. *Stigler Center Working Paper Series*, 2021.

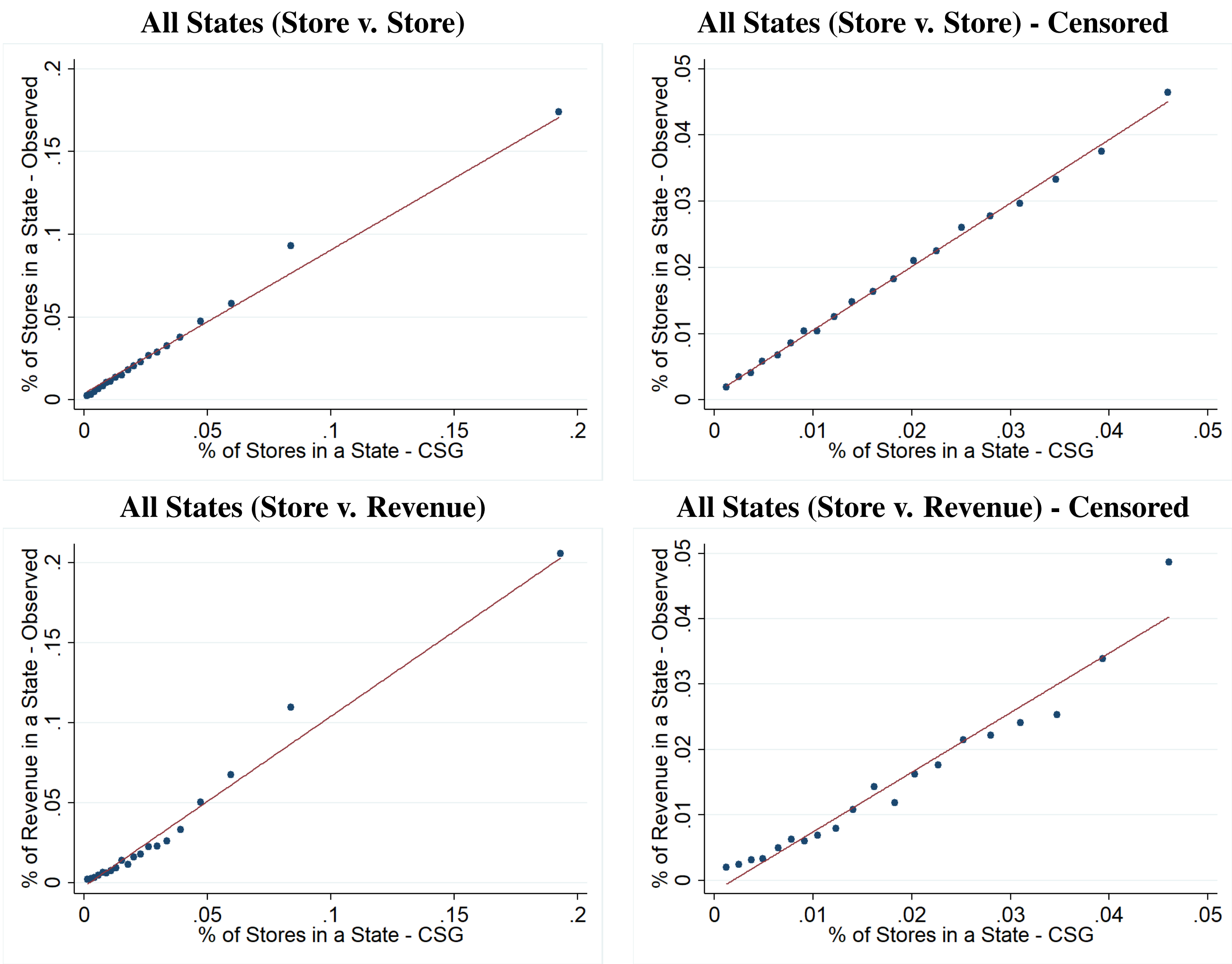
Michael Weber. Nominal rigidities and asset pricing. *UChicago Working Paper*, 2014.

Figure 1: Comparison Between Reported Revenue and Observed Spending



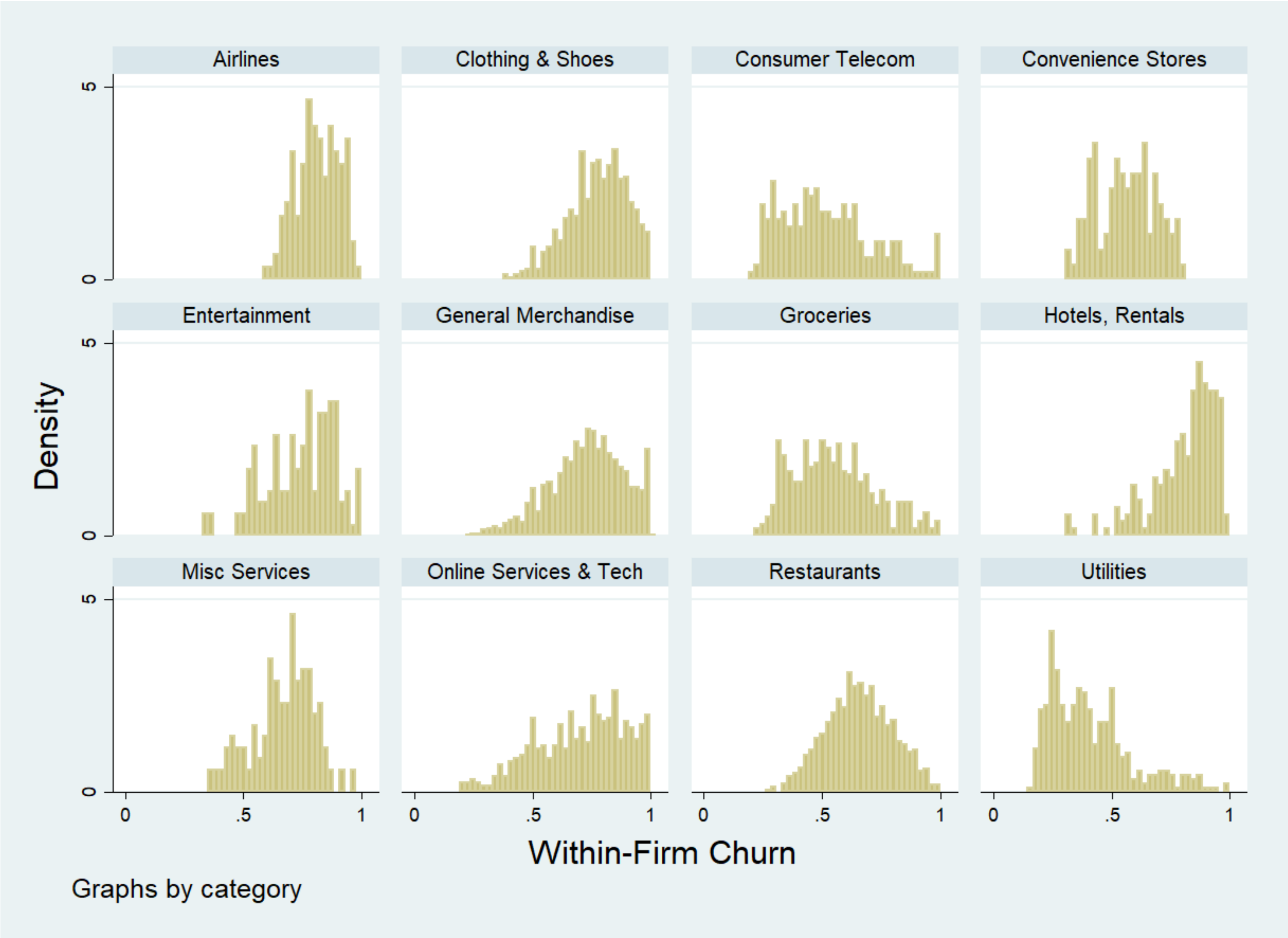
Notes: These graphs show the relationship between firm-level revenue measured in two ways: through Compustat and as observed in our transaction data. Each dot denotes a firm-quarter observation. Along the x-axis, we measure $ln(Revenue_{it})$ obtained from Compustat. Along the y-axis, we measure the total spending observed at a firm in a quarter within our transaction database. The top two panels examine levels of revenue and observed transaction spending. The bottom two panels examine changes in revenue and observed transaction spending.

Figure 2: Geographic Concentration - Transaction Revenue Data and Chain Store Guide Data



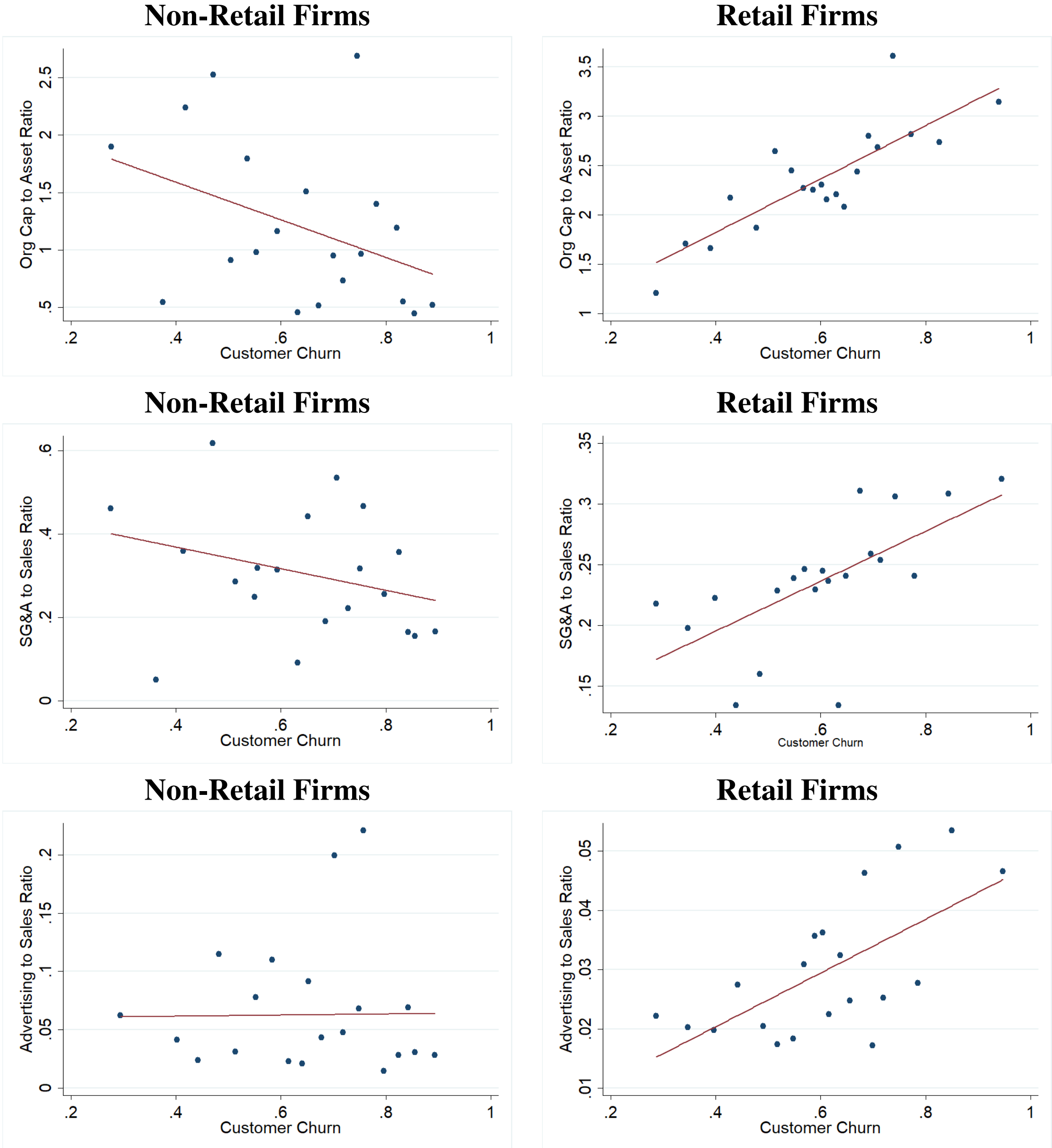
Notes: The graphs demonstrate the relationship between geographic concentration within a firm in two different ways. The first, measured on the x-axis, uses data from Chain Store Guide data and limits our sample primarily to retail firms. The x-axis measures the fraction of a firm's stores that are in a given state in a year (an observation is a firm-state-year). The y-axis measure uses data from our transaction data base and measures the fraction of spending at a retailer that is conducted by users living in a given state. Data covers all retailers able to be matched between samples and spans all 50 states, 2011-2014.

Figure 3: Customer-Base Annual Churn, By Industry



Notes: Each panel denotes the distribution of customer base churn over time across all firms in a given industry grouping in our sample. In this figure, churn is measured as the dollar-weighted overlap between the customer base of a firm f in year t and the customer base of firm f in year $t-1$. Overlap is scaled between 0 and 1 where 1 is an identical customer base and 0 is no overlap between customer bases across years.

Figure 4: Organization Capital, S,G&A, Advertising, and Customer Churn

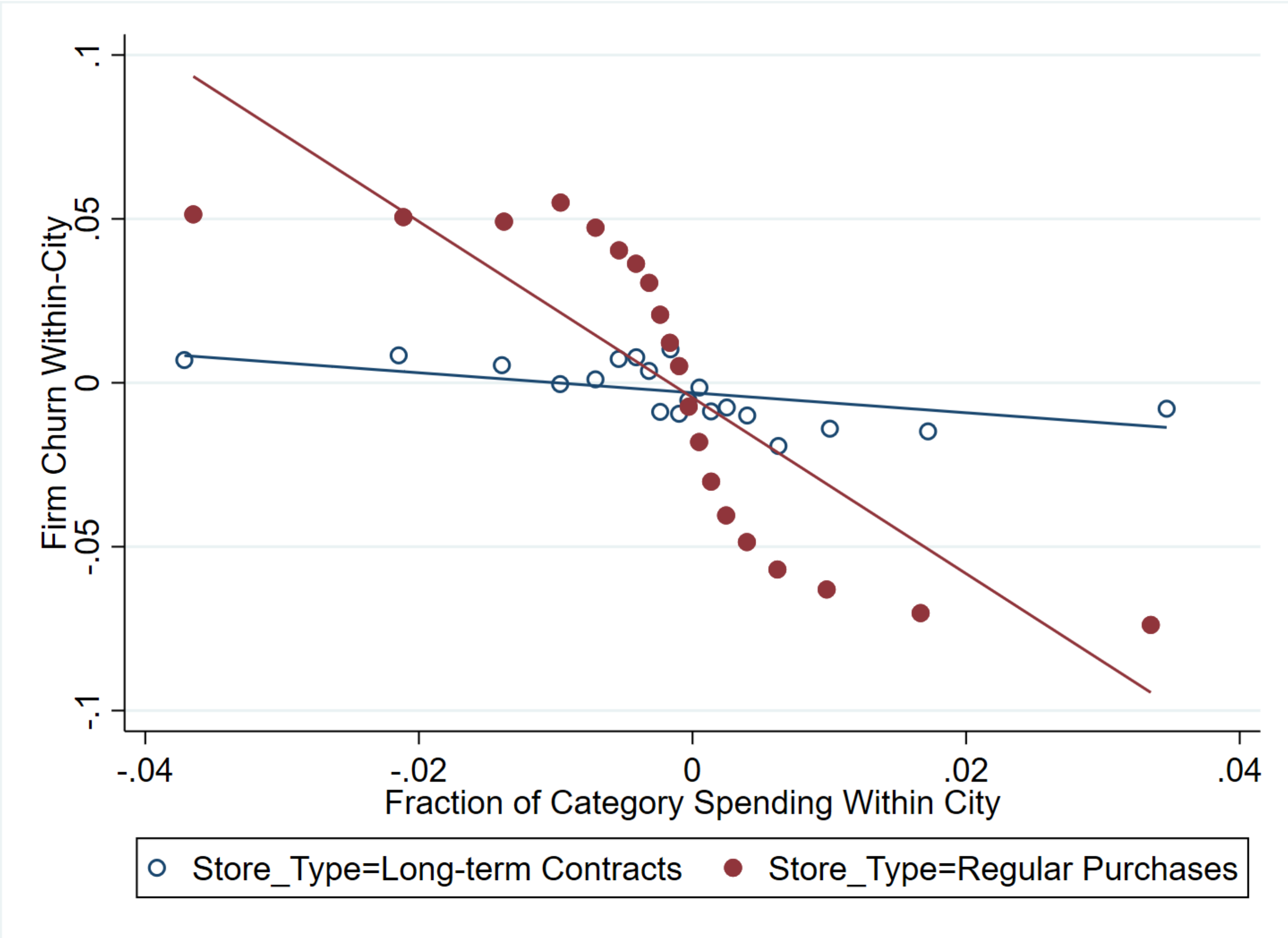


Notes: Retail firms defined as public firms in our sample with a one-digit SIC code of ‘5’. Organization Capital defined as in Eisfeldt and Papanikolaou (2013). SG&A expenses and Advertising expenses obtained for all firms with non-missing data in Compustat. Customer churn scaled between zero and one and is measured as the similarity of a firm’s customer base at time t relative to the customer base at time $t-1$, weighted by customer spending. Observations in the underlying data are firm-year. Plotted data cover 2011-2014 to exclude partial-year observations.

Figure 5: Spending on Customer Acquisition, Churn and Customer Overlap

Notes: For each public firm in our sample, we compute the average SG&A to sales ratio and year-over-year churn using data between 2012 and 2014. Then, within each category, we run a regression of the SG&A to sales ratio on churn, controlling for each firm's average log sales. For each firm, we also compute the equal weighted average overlap between their customer base, and all other firms in that firm's category (including private firms) each year. Then, we compute the equal weighted average of this within-category overlap between 2012-2014. The left panel includes all broad categories, while the right panel excludes restaurants.

Figure 6: Churn and Local Sales Shares Within Category



Notes: Pictured are bin-scatter plots of churn against the fraction of spending in a category done at a given retailer. Local spending shares are defined as $\frac{Spending_{icjt}}{\sum Spending_{cjt}}$ where i indexes firms, c indexes categories of spending, j indexes cities, and t indexes years. Churn is also measured at a city-firm-year level. Both variables are residuals of regressions on year and firm dummies. Firms are split into two categories. The first is composed of Utilities and Telecom firms (Long-term Contracts). The second is composed of Restaurants, Convenience Stores, General Merchandise, Groceries, and Entertainment (Regular Purchases).

Table 1: Summary Statistics, by Firm-Quarter

Variable	# Obs.	Mean	10%	25%	50%	75%	90%
Observed Spending	10,528	$8,368,492	$51,955	$439,811	$1,616,576	$5,324,263	$16,539,201
$\frac{ObservedSpending}{CompustatRevenue}$	6,751	0.0061	0.0002	0.0013	0.0041	0.0076	0.0127
Number of Transactions	10,528	204,425	734	6,964	39,472	131,970	423,665
Unique Users	10,528	66,317	353	4,082	19,969	64,603	171,473

Notes: Table reports basic summary statistics regarding the 558 matched firms in our sample. Compustat revenue data are only available for the subset of public firms in our sample. An observation is a firm-quarter. Quarters with no observed transactions for a given firm are dropped.

Table 2: Customer Churn, Valuation and Brand Value

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	B-to-M	B-to-M	ln(Brand Val)	ln(Brand Val)	$\frac{MktValue}{Customer}$	$\frac{MktValue}{Customer}$
Annual Churn	0.435***	0.440**	-4.756***	-6.413***	-1.722***	-1.674**
	(0.154)	(0.185)	(1.595)	(2.293)	(0.560)	(0.648)
Observations	354	354	87	87	250	250
R^2	0.110	0.183	0.408	0.536	0.159	0.434
Industry FE	NO	YES	NO	YES	NO	YES
Firm Controls	YES	YES	YES	YES	YES	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Annual Churn is average annual churn within a firm computed over all years the firm is present in our sample. Book-to-Market (B-to-M) comes from the WRDS financial ratios suite. Number of customers for a firm is estimated by first calculating average spending per customer at a firm in our data, then dividing total firm sales (from Compustat) by that average customer spending. Value per customer is then measured as the market value of the firm divided by the estimated number of customers. Brand values are calculated by Brand Finance's Brandirectory which looks at components such as emotional connection, financial performance and sustainability and then applies royalty rates to calculate a capitalized brand value. Market value per customer and brand value are invariant over time within a firm. Firm level controls include sales, firm level R&D, firm level R&D to sales ratio, firm level advertising, firm level advertising to sales ratio, and HP measures of firm-level HHI and differentiation. Heteroskedasticity robust standard errors in parenthesis.

Table 3: Customer Churn and Firm Characteristics

VARIABLES	(1) Cash	(2) Leverage	(3) Markup	(4) SD(Invest Rate)
Annual Churn	-0.0867*	1.571**	-0.448**	0.0299*
	(0.0495)	(0.748)	(0.218)	(0.0152)
SG&A to Sales	0.270***	-0.684	3.412***	0.0181
	(0.0843)	(0.962)	(0.657)	(0.0226)
Observations	235	235	220	207
R^2	0.516	0.211	0.582	0.268
Industry FE	YES	YES	YES	YES
Firm Controls	YES	YES	YES	YES

*** p<0.01, ** p<0.05, * p<0.1

Notes: Annual Churn is average annual churn within a firm computed over all years the firm is present in our sample. Cash is defined as the ratio of cash and short term investments to total assets. Leverage is defined as the ratio of long term debt plus debt in current liabilities to book equity. Q is the inverse of the book-to-market ratio from the WRDS financial ratios suite. SD(Invest Rate) is the standard deviation of a firm's investment rate (the ratio between capital expenditures and lagged assets) between 2000 and 2019, normalized by the standard deviation of a firm's Q between 2000 and 2019. Markup data are obtained from Loualiche (Forthcoming) who calculates markups using the method in De Loecker et al. (2020). Firm level controls include average annual sales, firm level R&D, firm level R&D to sales ratio, firm level advertising, firm level advertising to sales ratio, and HP measures of firm-level HHI and differentiation. Robust standard errors in parenthesis. Heteroskedasticity robust standard errors in parenthesis.

Table 4: Customer Churn and Firm Investment Dynamics

VARIABLES	(1) All Firms	(2) Retail	(3) All Firms	(4) Retail
Q_{t-1}	0.491***	1.10***	0.796***	0.797***
	(0.0303)	(0.150)	(0.152)	(0.192)
Q_{t-1} *Low SG&A	0.347***	-0.227	0.153	0.189
	(0.0533)	(0.189)	(0.217)	(0.263)
Q_{t-1} *High New Churn			0.352*	0.599**
			(0.211)	(0.246)
Observations	44,116	5,741	3,252	2,412
R^2	0.656	0.657	0.631	0.629
Firm FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry-Year FE	YES	NO	YES	NO
Firm Controls	YES	YES	YES	YES

*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is investment rate, defined as the ratio between a firm's capital expenditures and lagged assets, multiplied by 100. Tobin's Q is measured as the inverse of book-to-market ratio from the WRDS financial ratios suite. Customer churn is calculated at a firm-year level and then averaged across all years in the sample (2010-2015). Sample includes investment data from 2000-2019. Low SG&A (High New Churn) is a firm-level indicator for being in the bottom (top) half of the SG&A (new customer churn) distribution. Retail firms are those with the one-digit SIC code of 5. Firm level controls include sales, firm level R&D, firm level R&D to sales ratio, firm level advertising, firm level advertising to sales ratio, and HP measures of firm-level HHI and differentiation. Standard errors clustered by firm.

Table 5: Customer Churn and Volatility

	CAPM β (1)	CAPM β (2)	CAPM β (3)	CAPM β (4)	Idio. Vol. (5)	Idio. Vol. (6)	Idio. Vol. (7)	Idio. Vol. (8)
Churn	0.625***	0.296***	0.591***	0.288***	1.301***	0.439**	1.070***	0.593***
	(0.086)	(0.101)	(0.100)	(0.098)	(0.218)	(0.222)	(0.266)	(0.214)
ln(Lagged Revenue)	-0.0217*	-0.0234*	-0.00996	0.0695	-0.215***	-0.277***	-0.217***	-0.402**
	(0.012)	(0.014)	(0.013)	(0.079)	(0.029)	(0.026)	(0.030)	(0.170)
Observations	992	992	992	992	992	992	992	992
R-squared	0.14	0.282	0.187	0.735	0.275	0.413	0.294	0.833
Specification	Univar	SIC2 FE	SIC1 $\times$ Year	Firm FE	Univar	SIC2 FE	SIC1 $\times$ Year	Firm FE

Notes: The level of customer churn is calculated at a firm-year level (2011-2014), and it is the churn from last year's customer base. "CAPM β" is the beta from a regression of a stock's daily excess returns on the excess returns of the market in a given year. "I. Vol." is idiosyncratic volatility, 100 times the standard deviation of daily CAPM residuals in that year. "Ln(Lagged Revenue)" is the natural logarithm of last year's total revenue from Compustat. To be included, a firm must have non-missing, non-negative total lagged revenue. All regressions are equal weighted. Standard errors are clustered at the firm level. All LHS variables Winsorized at the 1% and 99% level.

Table 6: Customer Churn and Revenue Decline During COVID-19 Outbreak

VARIABLES	(1) ln(Spend)	(2) ln(Spend)	(3) ln(Spend)	(4) ln(Spend)	(5) ln(Spend)
March 2020	-0.307***	-0.00933	0.258		
	(0.0580)	(0.0658)	(0.416)		
Mar 2020*Churn		-0.519***	-0.935**	-0.494***	-0.837*
		(0.107)	(0.349)	(0.115)	(0.411)
Observations	141,363	141,363	42,306	141,363	42,306
R^2	0.910	0.910	0.920	0.916	0.924
Month/Day/DoW FE	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES
Month*Beta Control	NO	NO	YES	NO	YES
Month*Size Control	NO	NO	YES	NO	YES
Industry*Month FE	NO	NO	NO	YES	YES

*** p<0.01, ** p<0.05, * p<0.1

Notes: The level of customer churn for each firm is calculated at a firm-year level and then averaged across all years in our sample. 'March 2020' is an indicator equal to one in March of 2020. It is interacted with the continuous measure of churn, ranging from roughly 0.33 - 0.9. Regression sample in this table is daily data from January 1, 2019 to March 31, 2020. Standard errors clustered by firm.

Table 7: Double Sort on Churn and Organization Capital

Churn OK/AT	Low Low	Low 2	Low High	2 Low	2 2	2 High	High Low	High 2	High High	HML Low	HML 2	HML High
Mkt. Excess Ret.	0.817***	0.968***	0.733***	1.053***	0.900***	1.211***	1.088***	1.332***	1.406***	0.271**	0.364***	0.672***
	(0.066)	(0.070)	(0.078)	(0.083)	(0.083)	(0.118)	(0.080)	(0.076)	(0.114)	(0.104)	(0.105)	(0.110)
Alpha	0.00794***	0.00274	0.00537*	0.00184	-0.00142	-0.00548	0.00406	-0.00128	-0.0127***	-0.00388	-0.00402	-0.0181***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)
Observations	120	120	120	120	120	120	120	120	120	120	120	120
R-squared	0.547	0.656	0.491	0.571	0.497	0.494	0.585	0.732	0.535	0.052	0.103	0.205
St. Dev.	0.143	0.155	0.136	0.181	0.166	0.224	0.184	0.202	0.249	0.154	0.147	0.193

Notes: Each month, we form 3 value-weighted portfolios based on average churn at the GVKEY level between 2011 and 2015. We then form 3 sub portfolios based on organization capital over assets from the Eisfeldt et al. (2020) replication file. We then regress the excess returns of these portfolios on the excess return of the market factor from Ken French's data library using data from 2010 to 2019. The HML columns represent a long-short portfolios, which go long high churn firms, and short low churn firms, within each OK/AT tercile. Robust standard errors in parenthesis. The last row reports the standard deviation of each portfolio over the whole 2010-2019 sample.

A Detail on Firm Matching Procedure

A.1 Transaction Description Cleaning

Our first step is to reduce this count of unique strings by removing capitalization, numeric characters, punctuation, and common components (e.g., 'inc'). We are then left with approximately 1.5 million unique cleaned strings. Appendix Table A.6 displays some samples of the transaction descriptions in our dataset. For each of these unique cleaned descriptions, we display the number of times that transaction is observed in our data from 2010-2015, the average transaction amount, the fraction of transactions that are debited from an account (instead of credited), and the fraction of transactions that are similar to a previous transaction with that description within a user.

Some transactions are much more commonly observed than others. This reflects both the relative size of retailers but also the degree to which a given retailer has different descriptions for different locations or types of transactions. For instance, we estimate that Walmart Inc. (and its subsidiary Sam's Club) is associated with approximately 15,000 unique description strings that span different types of Walmart stores (e.g., 'Neighborhood Market', 'Super Center'), different locations, and differences in whether debit or credit cards were used.

A.2 Firm Selection and Matching

Given our sample of 1.5 million unique cleaned strings, we set out to develop a set of firm names to match with these strings. Our goal is to match our transaction data to all major firms that directly transact with households and for whom we have a relatively complete picture of revenue.

Using the set of public and large private firms from consumer-facing industries, we then manually search our database of unique transaction strings for transactions that mention the firm name precisely or a range of potential abbreviations and variants of a firm's name. Most firms have a multitude of distinct strings associated with them across different establishments, payment types, and brands (e.g., 'wal mart', 'walmart', 'wm super center', 'sams club', 'walmart sacramento', 'walmart joliet', etc.).

Using regular expressions to define our match criteria, our goal is to capture as many true positives as possible while not flagging excessive amounts of false positives. For instance, the term 'subway' will match sandwich purchases at a Subway restaurant but also transactions made at any number of public subway systems around the world or any of the hundreds of small businesses whose name includes the word 'subway'. For this reason, we also often employ limitations in our matching procedure based on retailer category (which is noted in our transaction database) as well as transaction sizes. As one example, when attempting to match Subway sandwich stores, we limit the retailer category of the transaction description to restaurants and the *average* transaction size for the transaction description to under 20 dollars.

Unfortunately, traditional machine learning algorithms are not well suited to the task of mapping these transaction descriptions to firms. Given the huge set of firms in the transaction data (everything from large national retailers to single-establishment stores), automated methods that rely on string-similarity measures tend to produce extremely high rates of false positives. Moreover, many firms' descriptions are varied and are dissimilar to their official firm name (e.g., 'tgt' may refer to 'Target Corporation'). The mean number of unique text descriptions associated with a given retailer is 176 and the median number is 41. For this reason, we mostly rely on manual inspection and experimentation to find descriptions that map to firms.

B Other Transaction-based Measures of Customer Base Characteristics

B.1 Customer Income and Firm-Level Prices

Another aspect of firm customer bases that can be easily surmised from transaction-level data is that of the average income of any given consumer-facing firm. We construct a quarterly index of the average user income of a store's clients, weighted by the amount they spend at that retailer:

$$Quality_{rt} = \frac{\sum\limits_{i} spending_{irt} * income_{it}}{\sum\limits_{i} spending_{irt}}$$

Where r identifies a retailer, i indexes users, and t refers to a calendar year. Firms in our sample exhibit large differences in this measure, aligning well with an ex-ante notion of the firm's quality. This measure correlates strongly with other indicators of retailer quality and prices. From Yelp.com, we are able to obtain indicators of how expensive the average product at a particular firm is for about two thirds of our sample of firms. For each matched firm, we record a rating between $ and $$$$ that indicates low to high prices, respectively. We regress our measure of firm quality on indicators for these price rankings and report the results in Appendix Table A.7.

Unsurprisingly, we find that firms that have higher income customer bases in our data tend to be those selling higher priced goods, on average. This is both true overall and in all subcategories of firm that we examine. For instance, relative to the average customer of the lowest priced restaurants ($), the average customer of the highest priced restaurants in our sample ($$$$) tends to have a $24,016 higher annual income.

B.2 Customer Income Distributions and Concentration

Appendix Figure A.7 shows a selection of customer income distributions for pairs of firms in the same industry. For instance, the bottom right panel displays the distribution of customer income (weighted by spending at the firm) within two grocery stores: Save-a-Lot and Whole Foods. We sort income into $1,000 bins and censor the histogram at $300,000 for visibility. We can see that Whole Foods customers tend to be substantially richer than those of Save-a-Lot, indicating a higher quality firm.

Another illustration of the benefit of linking users to firms using this class of transaction data is the ability to get information not only about levels of spending at a particular firm, but the distribution of spending (i.e. revenue) within a firm across its customers. In Table A.8, we display statistics that illustrate how concentrated firm revenue is within its customer base. Looking across broad industry categories, we show that there is a substantial amount of variation in revenue concentration. For instance, the top 5% of customers for a given Utility firm provides approximately 15% of a firm's revenue.[22] In contrast, revenue for hotels and airlines is much more concentrated within their customers, with the highest spending 5% of customers making up almost 30% of their revenue in our sample. This variation in concentration is maintained down the distribution of customers, with the top 20% of customers making up around 40% of revenue in low customer concentration industries and over 75% in high customer concentration industries.

[22] Here, we mean the percent of revenue in our matched dataset. In this example, the top 5% of customers make up 15% of the revenue *we can see in our matched dataset*, not 15% of the revenue in Compustat. This will naturally exclude any sales to businesses or governments that are absent from our consumer panel.

B.3 Market Value Per Customer

Although our dataset only covers about 0.8% of the US population, it is still useful for estimating the total number of customers at a given firm. To do this, we start by calculating spending per customer at the firm-year level: total spending divided by the number of unique households that shopped at the firm that year.[23] Then, to get an estimate of the number of customers, we divide total sales (SALE) in Compustat by spending per customer.

An alternative method would be to scale the number of customers at the firm-year level in our sample by our coverage of the US population. With an average coverage of 0.8%, the total number of customers at each firm should be about 1/0.008=125 times as large as the number in our sample. This gives similar estimates to the 'spending per customer' method for many large retail firms e.g., Saks and Nordstrom. It also gives similar estimates for national restaurant brands e.g., Bloomin' Brands (owner of Outback Steakhouse) and Red Lobster.

This method, however, leads to substantially different estimates for firms with a significant amount of sales outside the US e.g., Tim Hortons. While most Tim Hortons locations are in Canada, they do have several hundred US locations. This means that while their customers appear in our sample, scaling up the number of customers by a factor of 125 will likely understate the true total number of customers. If the average customer, however, is similar in the US and Canada, then our 'spending per customer' method will yield accurate estimates despite our lack of Canadian coverage.

The next step is to calculate the market value per customer: the total market capitalization at the end of the year divided by the estimated number of customers in that year. Common-sense intuition suggests that market value per customer should be higher for low-churn firms. From a present value perspective, a customer should be more valuable to a firm if they are likely to continue spending there for a long period of time. Figure A.9 plots logged average market value per customer vs. average churn. There is a statistically significant and economically large negative relationship between market value per customer and churn. The standard deviation of churn is $\approx$0.16, so a 1 SD increase in churn would decrease market value by about 30-40% per customer.

This result is driven mostly by differences across industries: Some of the firms with the highest

[23]From both the numerator and the denominator we exclude household-firm-quarter observations with less than $1 of total spending.

market value per customer are utility companies like Dominion Energy and Duke Energy as well as Telecom companies like AT&T and Verizon. Some of the firms with the lowest market value per customer firms are struggling brick-and-mortar retailers like Barnes & Noble and Sears. While the relationship is still negative and significant when including industry fixed-effects, the magnitude of the slope is only about 1/2th as large.

B.4 Measuring Components and Variants of Customer Churn

One advantage of the utilization of this class of disaggregated transaction data is that many variants of customer base characteristics can be constructed. We construct a number of alternate measures of customer churn to complement our headline index. Below we note the calculation underpinning our baseline index as well as several of these variants. Overall, these measures are highly correlated with one another, featuring correlation coefficients between 0.81 and 0.97. The one exception is $Churn_{walletshare}$ which is negatively correlated with other churn metrics. That is, while overall customer base churn tends to be driven by extensive margin movements of customers (gaining new customers and attrition of existing customers), intensive margin fluctuations are actually negatively correlated with extensive margin changes.

1. $Churn_{baseline,f,t-k} = \left(\sum_i |s_{f,i,t} - s_{f,i,t-k}|\right)/(2)$ where the sum $\sum_i |s_{f,i,t} - s_{f,i,t-k}|$ is taken over all customers that shop at firm f in *either* year t or year $t-k$.

2. $Churn_{old,f,t-k} = \left(\sum_i |s_{f,i,t} - s_{f,i,t-k}|\right)/(2)$ where the sum $\sum_i |s_{f,i,t} - s_{f,i,t-k}|$ is taken over all customers that shop at firm f *only* in year $t-k$ *and not* year t.

3. $Churn_{new,f,t-k} = \left(\sum_i |s_{f,i,t} - s_{f,i,t-k}|\right)/(2)$ where the sum $\sum_i |s_{f,i,t} - s_{f,i,t-k}|$ is taken over all customers that shop at firm f *only* in year t *and not* year $t-k$.

4. $Churn_{walletshare,f,t-k} = \left(\sum_i |s_{f,i,t} - s_{f,i,t-k}|\right)/(2)$ where the sum $\sum_i |s_{f,i,t} - s_{f,i,t-k}|$ is taken over all customers that shop at firm f *both* in year $t-k$ *and* year t.

5. $Churn_{existingcustomers,f,t-k} = \left(\sum_i |s_{f,i,t} - s_{f,i,t-k}|\right)/(2)$ where the sum $\sum_i |s_{f,i,t} - s_{f,i,t-k}|$ is taken over all customers that shop at firm f *only* in year $t-k$.

6. $Churn_{growthadjusted,f,t-k} = \left(\sum_i |s_{f,i,t,t-k} - s_{f,i,t-k,t-k}|\right)/(2)$ where the sum $\sum_i |s_{f,i,t} - s_{f,i,t-k}|$ is taken over all customers that shop at firm f *only* in year $t-k$. Note that here the denominators are modified such that the spending shares are always taken as a share of total firm spending at $t-k$ rather than t and $t-k$.

Decomposing churn into its constituent components (eg. $Churn_{old,f,t-k}$, $Churn_{new,f,t-k}$, and $Churn_{walletshare,f,t-k}$) also allows us to examine the path that these variables take over time within a firm. On average, we see that $Churn_{old,f,t-k}$, and $Churn_{walletshare,f,t-k}$ are fairly stable over time within a firm, while $Churn_{new,f,t-k}$ gradually decreases. Total baseline customer churn, $Churn_{baseline,f,t-k}$, is also fairly stable over time during out sample period. Splitting firms into quintiles according to their drift in churn, we find that the top four quintiles all shift churn by less than 0.1 (about 25% of a standard deviation) over three years. Only the bottom quintile shifts by more, with a drop in churn of approximately 0.3.

C Customer Base Overlap

Another aspect of firms' customer bases that we can capture with our data is the similarity of firm i's customer base to that of firm j. Again, we define $s_{f,i,t}$ as the share of firm f's revenue in our matched sample that comes from customer i in year t. We define similarity between firms f and j in year t as:

$$Similarity_{(f,j),t} = -\left(\sum_i |s_{f,i,t} - s_{j,i,t}|\right)/(2) + 1 \tag{A.1}$$

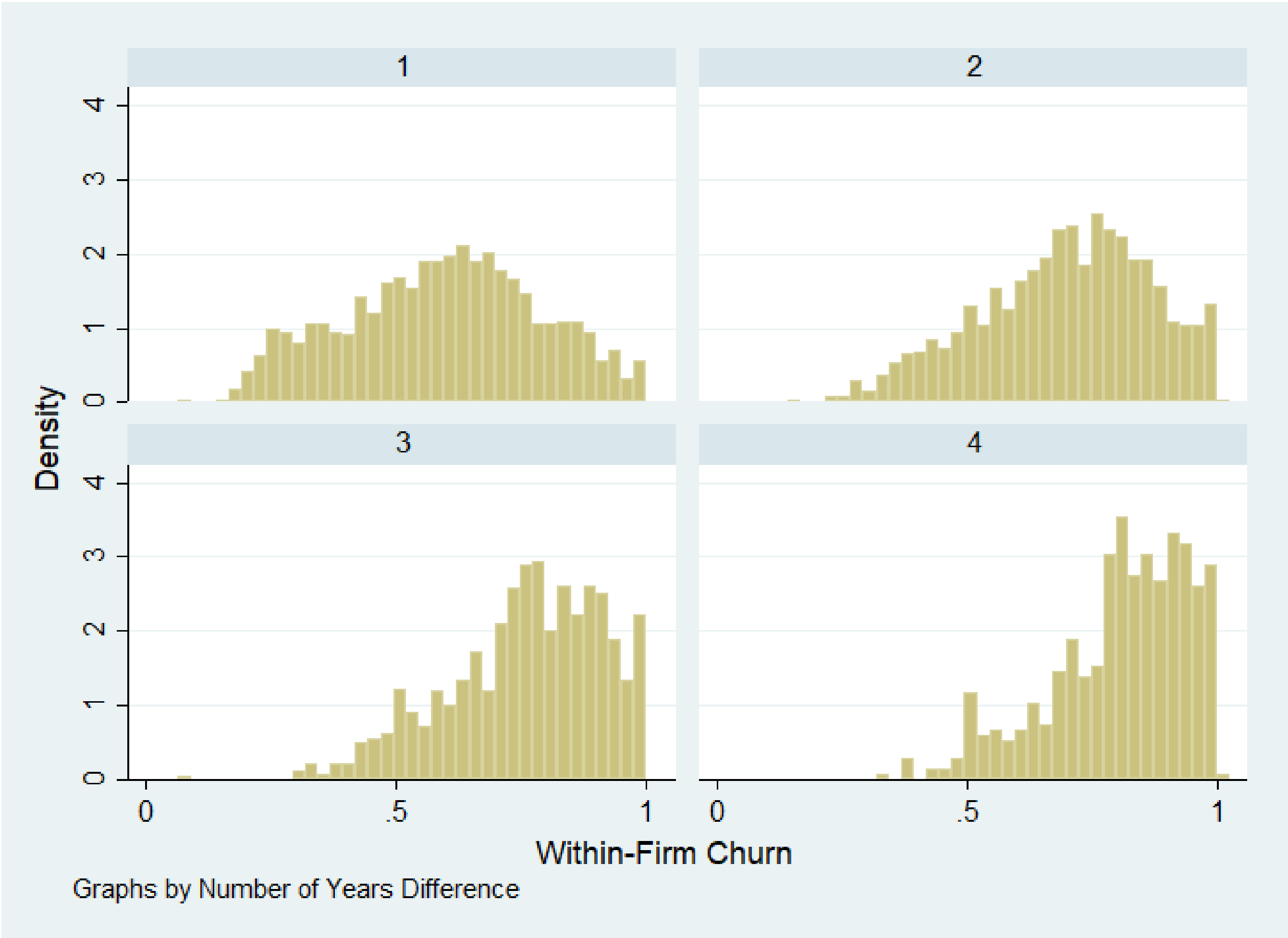
where the sum $\sum_i |s_{f,i,t} - s_{j,i,t}|$ is taken over all customers that shop at *either* firm f or j in year t. As with our churn measure, this sum can vary between zero and two. We multiply by $-1/2$ and add 1 so that a similarity score of one would imply that the firms have the exact same revenue share from each customer, and a value of zero would imply no overlap in customer bases. We calculate this measure for all firm-firm pairs in our sample at an annual frequency.

Figure A.10 displays the average level of customer base similarity within a broad industry group for all firm-firm pairs in that industry. As with the customer base churn metric discussed above, there exists substantial variation in cross-firm similarity across industries. Firms within the Utility industry are the most dissimilar to other Utility firms – which is to be expected as most customers have only a single utility provider and do not vary in their provider much over time. In contrast, restaurants have the highest amount of within-industry cross-firm similarity – over 5 times higher than that of Utility firms. This reflects the fact that many users tend to spend large amounts of money eating out but spread their spending across multiple restaurants rather than focusing on a single restaurant.

We note that, on average, within-industry customer base similarity is higher than that across industries. That is, many users tend to disproportionately weight their spending towards a particular industry, not simply a particular firm within an industry. However, for both within- and cross-industry firm-firm pairs we see some that are highly dissimilar and some that are highly similar. Moreover, the set of most similar firms for a given firm tends to span industries.[24]

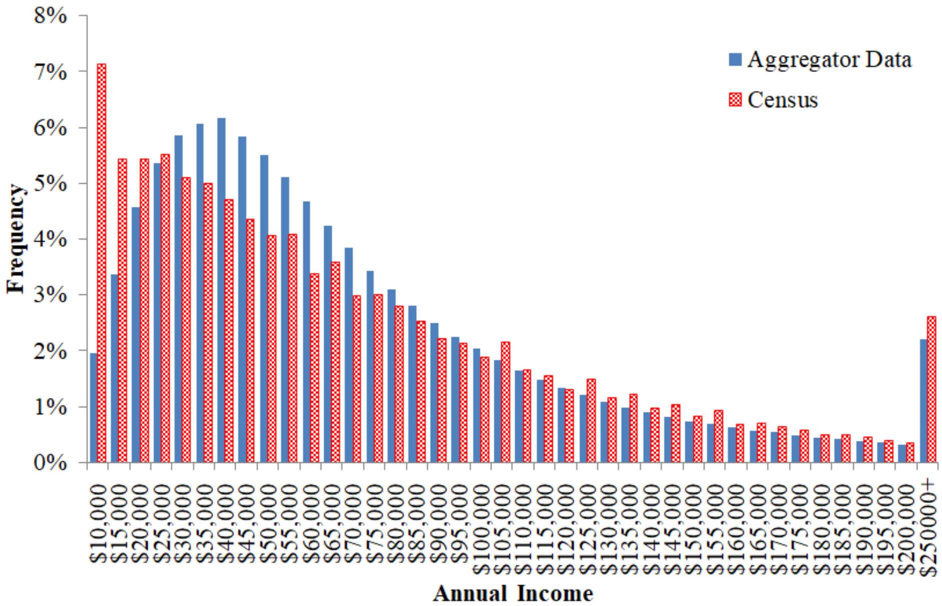
[24]For instance, the ten firms with the most similar customer bases to Walmart are: Yum Brands, Dine Brands, Darden Restaurants, Sonic Corp, Netflix, Amazon, Kohl's, Dollar Tree, Domino's, and Papa Johns. Among retailers, the ten firms with the most similar customer bases to Walmart are: Amazon, Kohl's, Dollar Tree, Bed Bath and Beyond, Autozone, Sally Beauty, Gamestop, Office Depot, Big Lots, and Dicks Sporting Goods.

Figure A.1: Customer-Base Similarity Within Firm Over Time



Notes: Each panel denotes the distribution of customer base churn over time across all firms in our sample. Churn is measured as the dollar-weighted overlap between the customer base of a firm f in year t and the customer base of firm f in year $t-x$ where x is between 1 and 4 and is labeled above each panel. Overlap is scaled between 0 and 1 where 1 is an identical customer base and 0 is no overlap between customer bases across years.

Figure A.2: Income Distribution - Aggregator Data vs. U.S. Census

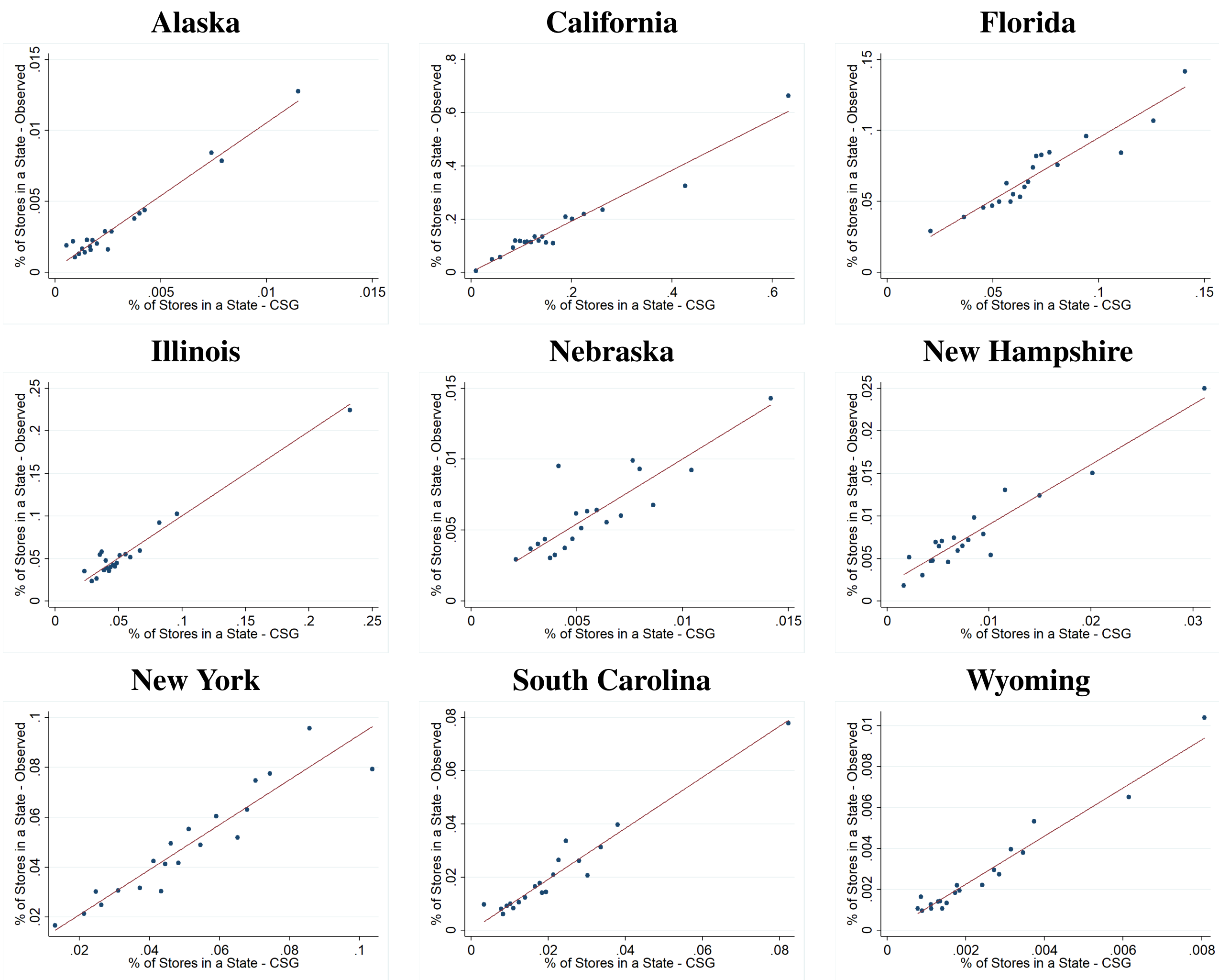


Notes: This figure compares the distribution of 2014 income of the account aggregator and the U.S. Census. The Census data uses the variable *HINC-06* and is available for download at census.gov. The difference in distributions at the bottom end of the income distribution is due to censoring of zero income users in our dataset. See Section 3 for more details.

Figure A.3: Consumption - Aggregator Data vs. U.S. Census Monthly Retail Trade Report

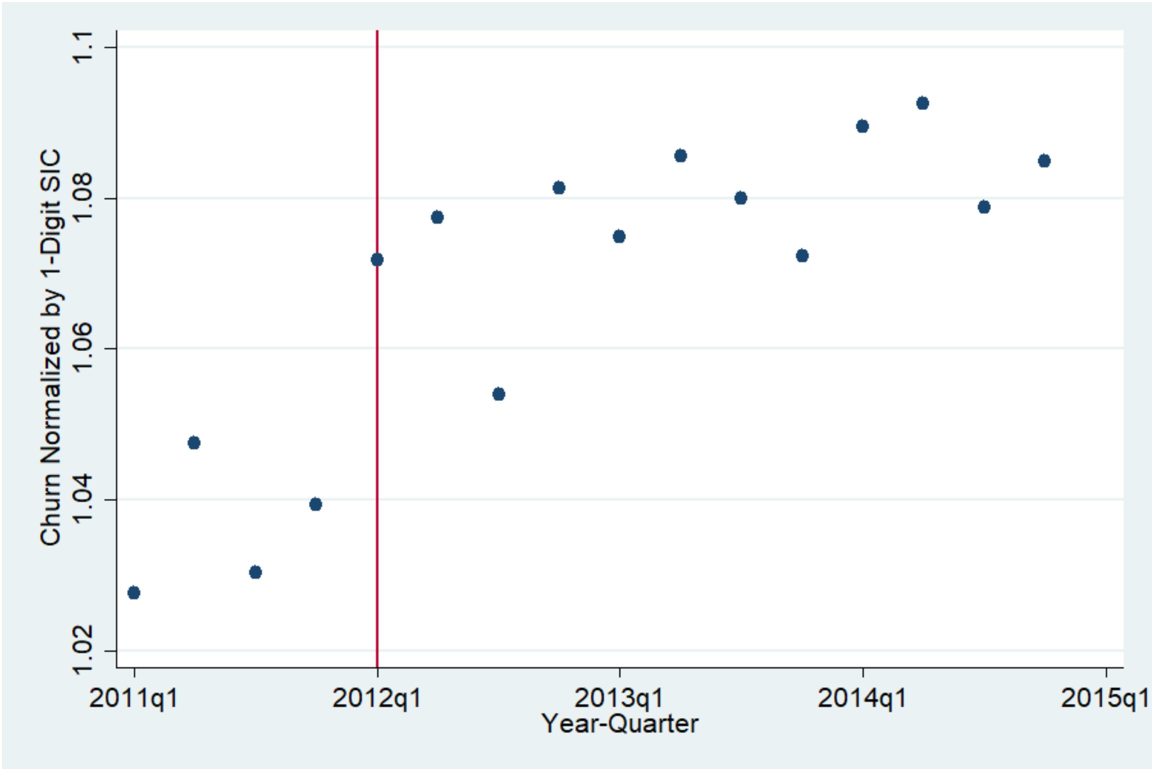
Notes: This figure compares the level of spending observed in the aggregator data to the U.S. Census monthly retail trade report (https://www.census.gov/retail/index.html). We plot spending from January 2011 to April 2015 and scale consumption by the Jan 2011 values to the value of 1 for both data sources.

Figure A.4: Geographic Concentration - Transaction Store Data and Chain Store Guide Data, Selected States



Notes: The graphs demonstrate the relationship between geographic concentration within a firm in two different ways. The first, measured on the x-axis, uses data from Chain Store Guide data and limits our sample primarily to retail firms. The x-axis measures the fraction of a firm's stores that are in a given state in a year (an observation is a firm-state-year). The y-axis measure uses data from our transaction data base and measure the fraction of spending at a retailer that is conducted by users living in a given state. For each graph, the data spans all retailers operating in the listed state in our matched sample, 2011-2014.

Figure A.5: Customer Churn at JC Penney

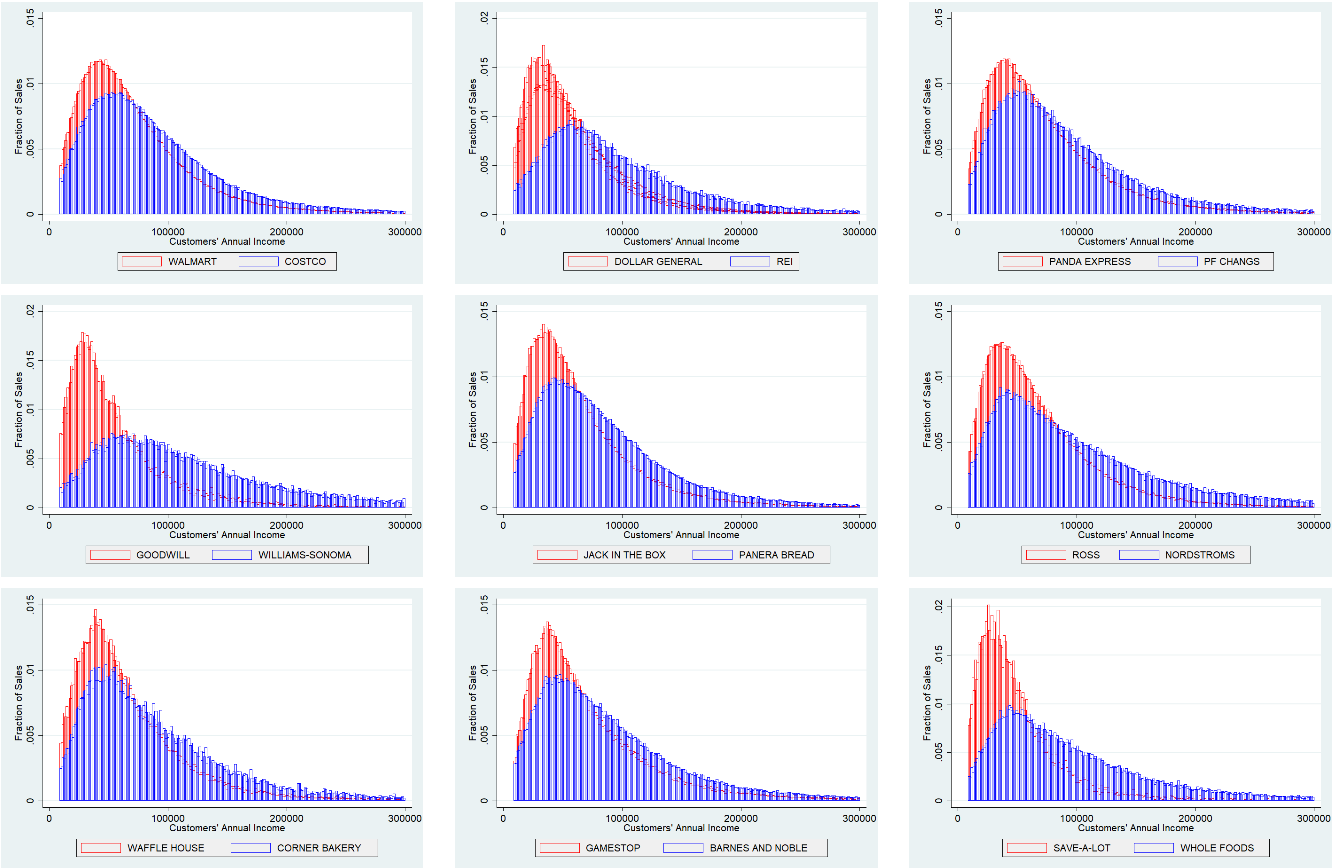


Notes: Plotted is the level of quarter over quarter customer churn at JC Penney normalized by the average level of quarter over quarter churn within the industry (one digit SIC code). A red line denotes the quarter (Q1 2012) in which JC Penney instituted a radical new pricing strategy.

Figure A.6: Brand Value and Churn, by Industry

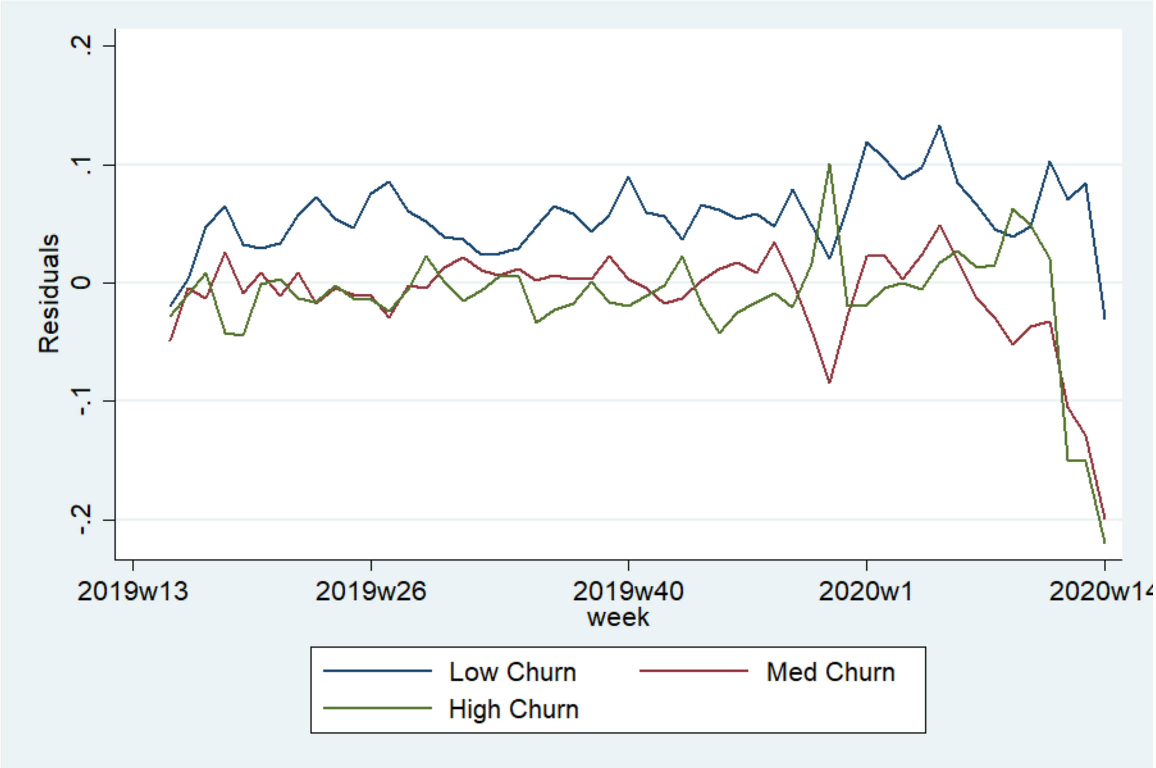
Notes: Churn denotes average annual customer churn within a firm across our sample period. Brand value rankings calculated by Brand Finance's Brandirectory which looks at components such as emotional connection, financial performance and sustainability and then applies royalty rates to calculate a capitalized brand value.

Figure A.7: Income Distribution of Customer Base, Firm-level Comparisons



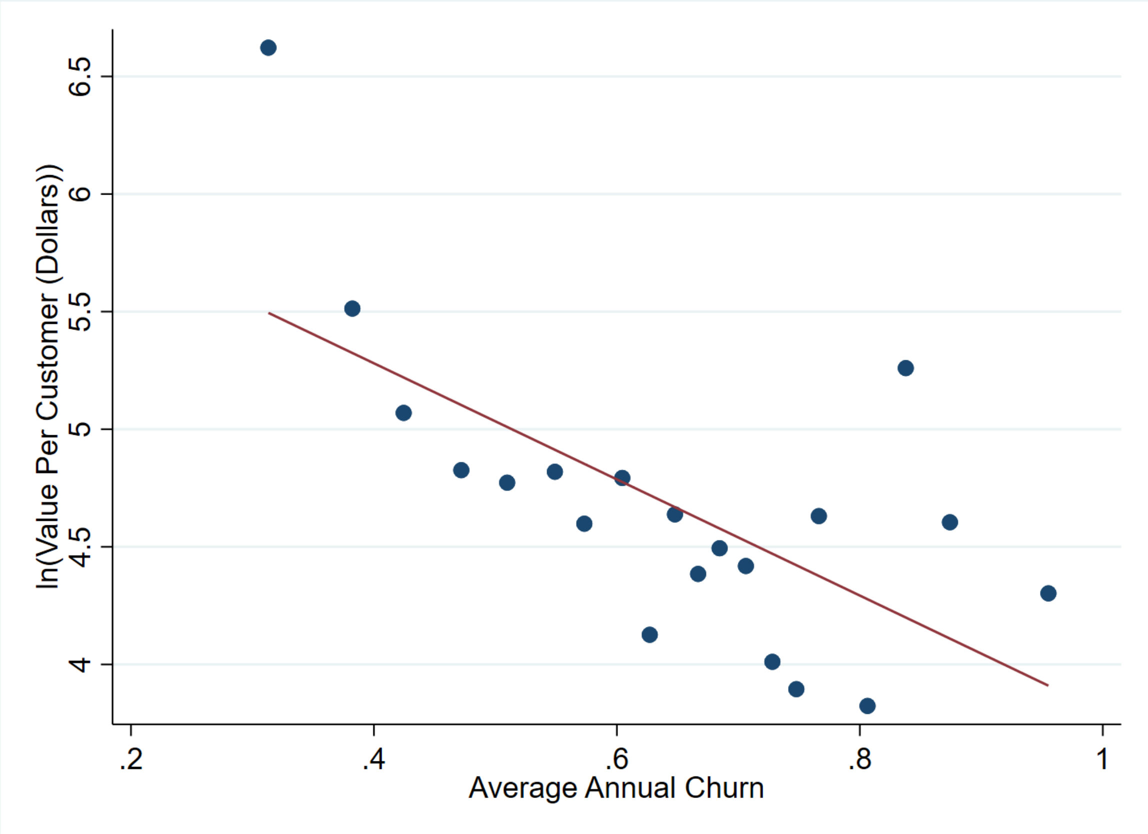
Notes: Figures demonstrate the distribution of income among customers for a selected sample of firms. Customer's are dollar-weighted by sales at a firm, so a user spending $500 at a firm will have double the weight in the histogram as a user spending $250. Annual income is binned in $1,000 increments and is censored at $300,000 for illustrative purposes. In each panel, two firms of similar types are compared. Data spans 2010-2015.

Figure A.8: COVID Firm Sales by Churn, Event Study



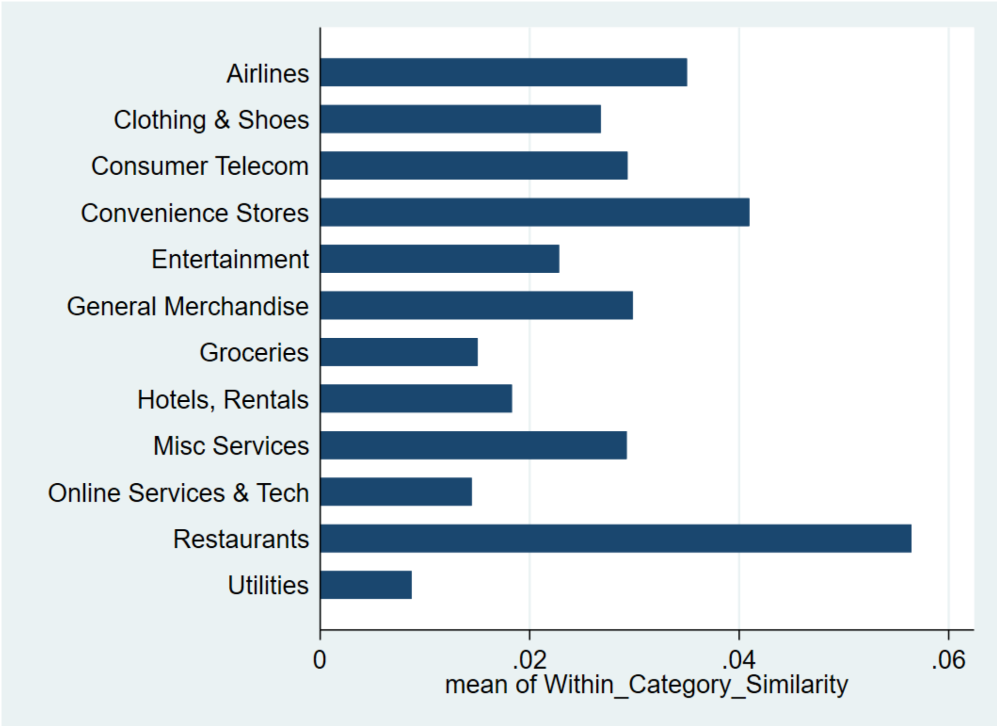
Notes: Plotted are firm sales by week across terciles of churn. Each firms' sales are normalized to 1 for the entire period. Residuals of a regression on week fixed effects are plotted to remove seasonality.

Figure A.9: Market Value per Customer vs. Churn



Notes: Y-axis is logged average market value per customer between 2011 and 2015. X-axis is average churn between 2012 and 2015 i.e., using data from 2011-2015. Estimates of market value per customer are Winsorized at the 1% and 99% level.

Figure A.10: Similarity of Firm Customer Bases Within Category, by Category



Notes: Bars denote the average cross-firm similarity within the listed industries. That is, the similarity between firm i and firm j who are both operating in broad industry classification x.

Table A.1: Matching to Largest Firms by Industry

	Avg. Rank		Avg. Percentile Rank		% of Top 5	
Industry	Matched	Unmatched	Matched	Unmatched	Matched	Unmatched
Airlines	6	15	73%	32%	100%	0%
Clothing & Shoes	19	21	52%	48%	100%	0%
Consumer Telecom	20	66	84%	45%	80%	20%
Entertainment	11	24	77%	45%	40%	60%
General Merchandise	69	103	59%	39%	100%	0%
Groceries	6	10	58%	18%	100%	0%
Hotels, Rentals	16	32	73%	43%	60%	40%
Others Services & Tech	95	195	74%	47%	20%	80%
Resturants	30	82	76%	34%	100%	0%
Utilities	23	77	83%	43%	60%	40%

Notes: We rank Compustat firms based on their total revenue in 2014. We then compare the numerical ranks (with one being the highest), and percentile ranks (with 100% being the highest) of the firms in our matched sample, with Compustat at large by industry. We then keep the 5 largest firms in each industry by revenue, and count how many of those firms are in our matched dataset. When matching to Compustat, and calculating the ranks, we restrict the sample to U.S. firms, with a traded common stock, non-missing revenue and non-missing NAICS industry.

Table A.2: Customer Churn and Local Categorical Sales Shares

VARIABLES	(1) Churn	(2) Churn	(3) Churn	(4) Churn	(5) Churn	(6) Churn
Fraction of Category Spending in City		-0.742***		-0.566***		-0.553***
		(0.00288)		(0.00285)		(0.00285)
Observations	311,264	311,264	311,264	311,264	311,256	311,256
R^2	0.076	0.241	0.350	0.422	0.701	0.762
Year FE	YES	YES	YES	YES	YES	YES
City FE	YES	YES	YES	YES	YES	YES
Category FE	NO	NO	YES	YES	YES	YES
Firm FE	NO	NO	NO	NO	YES	YES

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: The level of customer churn is calculated at a firm-city-year level (2011-2014), and it is the churn from last year's customer base. Fraction of local categorical spending is computed as $\frac{Spending_{icjt}}{\sum Spending_{cjt}}$. City-firm-years are excluded if they feature fewer than 50 customers.

Table A.3: Single Sort on Customer Churn

	Low	2	3	4	High	5 - 1
Mkt. Excess Ret.	0.627***	0.958***	0.983***	1.028***	1.198***	0.571***
	(0.051)	(0.069)	(0.073)	(0.057)	(0.083)	(0.099)
Alpha	0.00526***	0.00729**	0.000721	-0.00177	0.00388	-0.00138
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)
Observations	120	120	120	120	120	120
R-squared	0.595	0.568	0.609	0.715	0.621	0.215
St. Dev.	0.105	0.165	0.163	0.158	0.197	0.16

Notes: Each month, we form 5 value-weighted portfolios based on average churn at the GVKEY level between 2011 and 2015. We then regress the excess returns of these portfolios on the excess return of the market factor from Ken French's data library using data from 2010 to 2019. The column "5-1" represents a long-short portfolio, which goes long high churn firms, and short low churn firms. Robust standard errors in parenthesis. The last row reports the standard deviation of each portfolio over the whole 2010-2019 sample.

Table A.4: Double Sort on Firm Size and Customer Churn

Revenue		Low			2	
Churn	Low	2	High	Low	2	High
Mkt. Excess Ret.	0.990***	1.144***	1.185***	0.732***	1.003***	1.339***
	(0.089)	(0.082)	(0.113)	(0.074)	(0.073)	(0.075)
Alpha	0.00143	-0.0028	-0.00680*	0.00683***	-0.00364	-0.00143
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Observations	120	120	120	120	120	120
R-Squared	0.507	0.532	0.512	0.525	0.567	0.762

Revenue		High		Low	2	High
Churn	Low	2	High	HML	HML	HML
Mkt. Excess Ret.	0.470***	0.977***	1.013***	0.195*	0.607***	0.543***
	(0.059)	(0.068)	(0.057)	(0.104)	(0.086)	(0.084)
Alpha	0.00527**	0.00654***	0.002	-0.00823*	-0.00826**	-0.00327
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)
Observations	120	120	120	120	120	120
R-Squared	0.38	0.655	0.724	0.024	0.32	0.276

Notes: Each month, we form 3 portfolios based on the previous year's total revenue in Compustat. Then, within each of these 3 portfolios, we form 3 sub-portfolios based on average churn at the GVKEY level between 2011 and 2015. We then regress the excess returns of these value-weighted portfolios on the excess return of the market factor from Ken French's data library using data from 2010 to 2019. The columns labeled "HML" represent a long-short portfolio, which goes long high churn firms, and short low churn firms. Robust standard errors in parenthesis.

Table A.5: Triple Sort on Firm Size, Churn and Organization Capital

Size	Small					
Churn	Low		High		HML	
OK/AT	Low	High	Low	High	Low	High
Mkt. Excess Ret.	0.888***	0.945***	1.156***	1.407***	0.268***	0.462***
	(0.072)	(0.119)	(0.099)	(0.125)	(0.081)	(0.133)
Alpha	0.00252	0.00328	-0.0011	-0.0139***	-0.00363	-0.0172***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)
Observations	120	120	120	120	120	120
R-Squared	0.553	0.381	0.581	0.567	0.077	0.081
Size	Big					
Churn	Low		High		HML	
OK/AT	Low	High	Low	High	Low	High
Mkt. Excess Ret.	0.871***	0.718***	1.085***	1.176***	0.214**	0.458***
	(0.061)	(0.061)	(0.064)	(0.093)	(0.087)	(0.082)
Alpha	0.00652***	0.00437*	0.00214	-0.00522	-0.00438	-0.00959***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	120	120	120	120	120	120
R-Squared	0.651	0.541	0.688	0.624	0.056	0.216

Notes: Each month, we first split firms into two groups based on whether they had above or below median total revenue in Compustat the previous year. Then, within each of these two groups, we form 3 sub-portfolios based on average churn at the GVKEY level between 2011 and 2015. Finally we form 3 further sub-portfolios based on organization capital over assets from the Eisfeldt et al. (2020) replication file. We then regress the excess returns of these value-weighted portfolios on the excess return of the market factor from Ken French's data library using data from 2010 to 2019. The HML columns represent a long-short portfolios, which go long high churn firms, and short low churn firms, within each total revenue bucket and OK/AT tercile. Robust standard errors in parenthesis.

Table A.6: Examples of Transaction String Data

Description	Count of Txns	Average Txn Amount	Frac Debit	Avg Loose Recurring
home depot	11,002,662	74.31	0.911	0.001
starbucks corpx	8,676,113	7.14	0.999	0.007
jack in the box	3,035,066	8.91	1.000	0.005
aeropostale	327,696	41.53	0.948	0.001
duane reade th ave new	160,318	18.72	1.000	0.004
bos taxi med long island cny	46,648	17.68	1.000	0.002
sbc phone bill ca bill payment	22,248	83.07	1.000	0.132
golden pond brewing	2,385	38.98	1.000	0.001
cross bay bagel	1,542	15.46	1.000	0.000
lebanese taverna bethe	1,542	68.44	0.999	0.005
racetrac purchase racetrac port charlot	1,357	31.32	1.000	0.007
trader joes rch palos vr	1,273	41.91	1.000	0.000
chevys fresh mex aronde	956	36.83	1.000	0.000
graceys liquor	113	15.99	1.000	0.018

Notes: Table denotes sample transaction descriptions from our database of financial transactions. Each panel displays the cleaned description string (e.g., removing numerics), the number of observations of that string in our data, the average transaction amount for that description string, the fraction of transactions that are debited from an account (instead of credited), and the fraction of transactions that are similar to a previous transaction to that description within a user.

Table A.7: Firm Quality Index and Yelp Ratings

VARIABLES	(1) All Stores	(2) Restaurants	(3) General Stores	(4) Clothing	(5) Groceries
Yelp - $$	11,845***	8,176***	11,364***	18,135***	8,240***
	(402.7)	(622.9)	(833.4)	(1,023)	(1,355)
Yelp - $$$-$$$$	32,677***	24,016***	39,666***	32,214***	28,858***
	(685.9)	(2,128)	(1,458)	(1,430)	(1,502)
Year FE	YES	YES	YES	YES	YES
Observations	3,808	918	1,054	796	364
R^2	0.482	0.356	0.567	0.329	0.510

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Observations are individual retailers from our sample able to be matched to Yelp. Independent variables are indicators for a firm's price range in Yelp, where the excluded category is Yelp '$'. Coefficients denote the average difference in firm 'quality' corresponding to different Yelp price categories. Firm 'quality' is determined by the dollar-weighted average income of customers at a given retailer.

Table A.8: Customer Base Concentration, by Industry

Category	# Obs.	HHI	Top 5% Share	Top 10% Share	Top 20% Share
Clothing & Shoes	207	0.57	24.8%	37.7%	55.1%
Consumer Telecom	59	0.62	17.9%	30.3%	49.3%
Convenience Stores	44	0.70	40.6%	56.5%	73.2%
Entertainment	56	1.50	25.2%	37.7%	55.1%
General Merchandise	462	0.81	29.1%	43.1%	61.2%
Groceries	166	1.51	42.8%	59.9%	77.3%
Hotels, Rentals, Airlines	96	1.16	29.2%	42.5%	60.7%
Misc Services	59	0.57	24.8%	37.7%	55.8%
Online Services & Tech	126	1.12	24.7%	36.9%	53.9%
Restaurants	369	0.38	27.9%	41.1%	57.9%
Utilities	116	0.83	15.5%	26.7%	44.6%

Notes: Table reports summary statistics across firms in a range of industry groupings. An observation is a firm-year. HHI is within-firm concentration in customer dollars. HHI is measured as the sum of squared fractions of revenue obtained from each customer, multiplied by 10,000. In this table, we equally weight firm-years but remove firms with fewer than 7,500 observed customers in a year.