Investing in Customer Capital*

Bianca He
Chicago Booth

Lauren Mostrom
Chicago Booth

Amir Sufi
Chicago Booth & NBER

February 2025

Abstract

Firms invest heavily in customer capital, and such investment is a main source of intangible capital value. Investment in customer capital is measured using sales and marketing expense from income statements, information on salaries paid to workers in sales and marketing, and text from annual 10-K SEC filings describing firms' sales and marketing strategies. There is large and persistent variation across industries in customer capital investment; industries investing the most are growing as a share of aggregate revenue and enterprise value. Industry-level variation in sales and marketing expense and R&D expense explains a large amount of the variation in the value of intangible capital, a result shown using both publicly-traded companies and prices paid in acquisitions. Residual sales, general, and administrative expense after subtracting sales and marketing expense is uncorrelated with intangible capital value. Industries focused on platform business models, online sales, and the production of high tech manufactured goods invest most heavily in customer capital.

*Carlo Geat, Freddy Pinzon, and Sixun Tang provided excellent research support. We are grateful to Michael Alter, Nicolas Crouzet, Andrea Eisfeldt, Sam Hanson, Steve Kaplan, Simone Lenzu, Yueran Ma, Quinn Maingi, Dimitris Papanikolaou, Leena Rudanko, Chad Syverson, Rob Vishny and seminar participants at Columbia GSB, Harvard Business School, the Federal Reserve Bank of Richmond, the Mitch Daniels School of Business at Purdue University, and USC Marshall for helpful comments. We are grateful to the University of Chicago Booth School of Business and the Fama Miller Center for financial support. Contact information: He: bianca.he@chicagobooth.edu; Mostrom: lmostrom@chicagobooth.edu; Sufi: amir.sufi@chicagobooth.edu.

Firms invest to create value, but what precise investments do they undertake? Recent research focuses on the growing importance of investment in *intangible capital*—non-physical capital that is more difficult to quantify and therefore omitted from firm balance sheets. By its nature, investment in intangible capital is challenging to measure; existing research measures investment in *knowledge capital* using research and development (R&D) expenditures, and it measures investment in *organizational capital* using a fraction of sales, general, and administrative (SG&A) expense.[1]

This study provides evidence that investment in *customer capital* is a quantitatively large component of investment in intangible capital. Specifically, it presents a novel data set containing comprehensive measures of investment in customer capital, it seeks to explain the determinants of this investment, and it provides evidence that firm spending on acquiring and maintaining a customer base should in fact be treated as an investment, and not as a variable input cost or as overhead.

Conceptually, an investment in customer capital centers on the idea that firms spend resources to build and maintain a customer base; this customer base is valuable because firms are able to capture part of the surplus associated with the relationships formed with existing customers.[2] Such investment encompasses, for example, spending on a sales force, on customer service, on boosting brand value, on advertising, and on acquiring and using data on customers.

Investment in customer capital is measured in this study using three data sources. The first source is the income statement of firms; around half of U.S. publicly traded non-financial firms directly report their spending on sales and marketing in their income statements as a sub-component of sales, general, and administrative expense (SG&A). Compustat does not systematically collect or report these data; however, another data set produced by S&P, Capital IQ, does. To the best of our knowledge, this is the first study to use the Capital IQ sales and marketing expense subcomponent of SG&A, a data item which is available for half of publicly-traded firms from 2007 onward. The second source of data comes from Revelio Labs, which uses LinkedIn, job postings, and other sources to estimate the jobs and salaries at firms. The Revelio Labs data set contains firm-year level information on the salaries paid to workers engaged in activities that fall within the sales and marketing function.

The third data set comes from text reported by firms in annual 10-K Securities and Exchange Commission (SEC) filings. Firms often detail their investment in customer capital, including the

[1]The seminal study using this approach in the finance literature is Eisfeldt and Papanikolaou (2013); See also Peters and Taylor (2017) and Crouzet and Eberly (2023). There is also a literature, cited below, that estimates the value of customer capital from brands and trademarks, but there is less research on the investments firms make to create such value.

[2]There are a variety of theoretical mechanisms through which existing customer relationships generate value for a firm. For example, branding efforts may produce loyal customers, lowering volatility and risk, as in Bronnenberg, Dubé and Gentzkow (2012) and Larkin (2013). Or there may be significant search frictions in switching products once a customer is with a certain firm, as in the model of Gourio and Rudanko (2014). These mechanisms are discussed in more detail in Section 1.

underlying goals of their spending. This text can be efficiently processed with the advent of large language models; we use Google's Gemini 1.5 Flash to produce quantitative data from the textual descriptions of sales and marketing efforts reported by firms in the Item 1 and Item 7 sections of their 10-Ks. Taken together, these three data sources allow us to build a comprehensive data set measuring investment in customer capital for U.S. publicly-traded firms from 2007 to 2022.

The initial part of the study shows a series of basic facts uncovered from this new data set. Investment in customer capital is large. The sales-weighted average annual sales and marketing expense to revenue ratio across U.S. publicly traded firms is 4.1%, which is higher than research and development (R&D) expenses and two-thirds of capital expenditures.

Previous research focuses on advertising expenditures—a sub-component of overall spending on sales and marketing—to measure firm investment in customer capital (e.g., Bagwell (2007)). Advertising expenses are a small fraction of sales and marketing expenses, especially for firms with high sales and marketing to revenue ratios. For example, for firm-year observations with above median sales and marketing to revenue ratios, advertising expense is less than a quarter of the overall sales and marketing expense. Furthermore, textual analysis of business descriptions shows that a narrow focus on advertising misses the majority of what firms discuss when outlining their sales and marketing strategy. Only 17% of firm-year observations discuss prominently a focus on advertising in their business descriptions; it is more common that firms discuss their efforts in building a sales force (59%), maintaining good customer service (52%), increasing brand value (48%), and building and utilizing data sets on customers (28%).

There is a striking amount of variation across industries in the amount of investment in customer capital. Firms in agriculture, mining, and petroleum and coal product manufacturing spend almost nothing on sales and marketing, whereas the median firm in the information industry—companies specializing in software, digital platforms, and web search portals—spends more than 20% of revenue. Firms in professional service industries invest heavily in customer capital, as do firms in high tech manufacturing, such as those producing medical equipment and computer and peripheral equipment. Industries with the highest amount of investment in customer capital experience the largest increase in the share of both revenue and enterprise value over the sample period.

Variation across industries in the level of investment in customer capital is robust across time and across different measures. The variation in the sales and marketing to revenue ratio from income statements is highly correlated with the variation in the ratio of salaries of sales and marketing employees to revenue from Revelio Labs. The variation across industries is persistent over the 15 years of data we have in the analysis. The evidence suggests that the industry-level variation in investment in customer capital reflects "primitive" differences across industries in how firms generate revenues and profits. As such, the rest of the study focuses on both the implications and

the determinants of this industry-level variation.[3]

Sales and marketing expense translates into valuable assets, a result shown using this across-industry variation. An empirical test motivated by the Gourio and Rudanko (2014) model shows that industries in which firms have a higher ratio of sales and marketing expenses to revenue have a higher ratio of enterprise value to physical capital (V/K^{PH}), the latter of which is referred to as Q in the literature and which we refer to as Q^{PH}. Industries with more R&D expenses also have higher Q^{PH}. These two variables alone explain more than 70% of the variation across industries in Q^{PH}. However, industries with higher residual SG&A expenses once sales and marketing expenses are removed do not have higher Q^{PH}. That is, differences across industries in SG&A expenses that are not due to sales and marketing expenses do not predict differences across industries in Q^{PH}. The evidence suggests that residual SG&A expenses, at least when measured as a whole, do not represent an investment in intangible capital. This illustrates the importance of separating out sales and marketing expense from the broader SG&A line item of income statements.

Theory predicts a strong relationship between the ratio of intangible book value of assets to tangible book value of assets and Q^{PH}, a point made clear by Crouzet and Eberly (2023). This prediction is confirmed in the data with remarkable statistical power. The stock of intangible capital is estimated using the perpetual inventory method, as in Eisfeldt and Papanikolaou (2013) and Peters and Taylor (2017) , with sales and marketing expenses and R&D expenses capitalized and externally acquired intangible capital also included. Using this new measure of the book value of intangible capital, an industry-level cross-sectional regression of Q^{PH} on the ratio of intangible book value to tangible book value yields a large positive coefficient, and the R^2 of the univariate regression is 0.80. As a placebo test, residual SG&A after removing sales and marketing expense is capitalized; such capital has no additional predictive power for Q^{PH}.

An analysis of the sources of value in acquisitions from purchase price allocations (PPA) confirms these results.[4] These data are useful in assessing the value implications of investment in customer capital, given that they represent value paid in actual transactions for intangible assets such as customer lists and customer relationships. At both the industry level and at the individual target level, a higher ratio of sales and marketing expense to revenue is associated with a higher value of customer-related intangible capital such as brands, trademarks, customer lists, and customer relationships. The ratio of R&D expenses to revenue is associated with a higher value of non-customer-related intangible capital, such as research and technology. Residual SG&A expenses after removing sales and marketing expenses are uncorrelated with the value paid for intangible

[3]The focus on industry-level variation follows much of the empirical literature on intangible capital, including Eisfeldt and Papanikolaou (2014), Gourio and Rudanko (2014), Peters and Taylor (2017), and Crouzet and Eberly (2023).

[4]The PPA data have been previously used in He (2022), Kepler, Naiker and Stewart (2023), and Ewens, Peters and Wang (2024).

capital. Both sets of analyses suggest that sales and marketing and R&D expenses are the main expenditures that firms undertake to build valuable intangible capital.

The language firms use in their 10-K filings is also consistent with the view that sales and marketing expenses is an investment, not simply a cost of doing business. When Gemini is asked to read the sections of the 10-Ks focused on sales and marketing strategy, it classifies the language to be consistent with the investment view over the cost view in 80% of the observations. For the 20% for which Gemini classifies the language to be more consistent with cost, the level of sales and marketing expenses is much lower. Furthermore, anecdotal evidence from the entrepreneurship literature shows that the idea that sales and marketing expense is an investment in customer acquisition is accepted and taught widely. A common framework used in venture capital, for example, compares the lifetime value of a customer (LTV) to the customer acquisition cost (CAC). This framework is almost exactly the same as the standard Net Present Value (NPV) formula taught in corporate finance to evaluate an investment that produces future cash flows.

The existing literature hypothesizes that sales and marketing efforts translate into valuable assets through a variety of channels, such as customer loyalty, search frictions, and network effects in demand. To elucidate these channels, the final section explores the factors that drive the variation across industries in the amount of customer capital investment. Three variables explain a large amount of industry-level variation. The most powerful of these variables is the fraction of firms in the industry that employ a platform business model in which the platform is designed to bring buyers and sellers together. Prominent examples of such firms include Ebay Inc; Uber Technologies, Inc; and Zillow Group, Inc; all of which have sales and marketing expense to revenue ratios above 20%. A second powerful variable is whether firms in the industry sell their products online. Finally, industries producing more technical products, as measured by salaries paid to engineers at the firms in the industry, invest more in customer capital. These three factors explain 70% of the variation across industries in investment in customer capital.

Related literature

A large body of research establishes the importance of intangible capital in production (e.g., Lev (2005); Corrado, Hulten and Sichel (2009); Eisfeldt and Papanikolaou (2013); Belo, Lin and Vitorino (2014), Eisfeldt and Papanikolaou (2014); Gourio and Rudanko (2014); Peters and Taylor (2017); Alexander and Eberly (2018); Crouzet and Eberly (2019); Crouzet and Eberly (2021); Corrado, Haskel, Jona-Lasinio and Iommi (2022); Crouzet, Eberly, Eisfeldt and Papanikolaou (2022); Crouzet and Eberly (2023)). Studies have emphasized the importance of intangible capital in the determination of markups and profits (e.g., Covarrubias, Gutiérrez and Philippon (2020); Crouzet and Eberly (2023)), firm investment (Gutiérrez and Philippon (2017); Alexander and Eberly (2018); Crouzet and Eberly (2019)), firm valuation and financial policies (e.g., Eisfeldt and Pa-

panikolaou (2013); Dell'Ariccia, Kadyrzhanova, Minoiu and Ratnovski (2021); Dou, Ji, Reibstein and Wu (2021); Belo, Gala, Salomao and Vitorino (2022); Falato, Kadyrzhanova, Sim and Steri (2022)), employee compensation (e.g., Sun and Xiaolan (2019)), productivity growth (e.g., McGrattan (2020); Crouzet and Eberly (2021)), and the transmission of monetary policy (e.g., Morlacco and Zeke (2021); David and Gourio (2023)).

However, measuring intangible capital is a challenge; the majority of the existing literature measures firm-level intangible investment as R&D expenses plus a fraction of overall SG&A expenses, often 30% of SG&A. A contribution of this study is to measure explicitly sales and marketing expense—a sub-component of SG&A—and to show that this sub-component is a statistically powerful determinant of the ultimate value associated with intangible capital. Residual SG&A after removing sales and marketing expense does not have statistical power in predicting intangible capital value. Recent studies measure the value of intangible capital using estimates by accountants of the price paid for different types of intangible capital in acquisitions (e.g., He (2022); Ewens et al. (2024); Kepler et al. (2024)). This study uses these data to show a high correlation between sales and marketing expenses and customer-related intangible asset value, and a high correlation between R&D expenses and non-customer-related intangible asset value.

With regard to measurement, the most closely related studies are those that measure customer capital using alternative data sets. Larkin (2013) uses estimates of brand value from Brand Asset Consulting; Belo et al. (2014) use advertising expenditures; Feng, Morgan and Rego (2015) and Nath and Bharadwaj (2020) use measures of the presence and power of marketing executives at firms; Bronnenberg, Dubé and Syverson (2022) use information on workers in sales and marketing occupations from the Occupational Employment and Wage Statistics and brand value estimates from BrandFinance; and Baker, Baugh and Sammon (2023) use a measure of customer churn based on credit card transaction data. There is also a literature that measures sales and marketing efforts in financial products such as mutual funds (e.g., Hastings, Hortaçsu and Syverson (2017), Roussanov, Ruan and Wei (2021).

With regard to the specific data set constructed in this study, the two closest articles are Ptok, Jindal and Reinartz (2018) and Markovitch, Huang and Ye (2020). The latter study collects sales and marketing expense from 10-K filings for a sample of 1300 firms from 2007 to 2009. The former study uses data from two sources: Advertising Age and Selling Power. Advertising Age contains information on marketing expenses and the Selling Power contains information on the size of the sales force. The sample size in the Ptok et al. (2018) study is approximately 500 firm-year observations. The underlying data collected from Capital IQ and Revelio Labs in this study is similar in spirit to the data collected in these two studies, but the sample sizes in this study are larger and cover a longer time series.

A related literature focuses on the determinants of a firm's market share in its various product

markets (e.g., Foster, Haltiwanger and Syverson (2008); Khandelwal (2010); Foster, Haltiwanger and Syverson (2016); Hottman, Redding and Weinstein (2016); Eslava, Haltiwanger and Urdaneta (2024); Fitzgerald, Haller and Yedid-Levi (2024)). Across a number of different settings, this literature finds that "product appeal" or "idiosyncratic demand," as opposed to differences in technical efficiency, is the most powerful determinant of product market shares across firms. This "product appeal" is modeled in this literature as a primitive in the consumer utility function: consumers tend to like some products in a given product market more than others. This is also related to the idea that acquiring new customers is important to the determination of a firm's market share, a point made in Argente, Fitzgerald, Moreira and Priolo (2021), Beaumont and Lenoir (2019), and Einav, Klenow, Levin and Murciano-Goroff (2021). This study shows that firms spend substantial resources attracting new customers and maintaining a customer base, and therefore developing product appeal and attracting new customers is endogenous to firm actions.[5]

Finally, the industrial organization literature explores concepts related to customer capital such as switching costs (e.g., Cabral (2016)), network effects in demand (e.g., Katz and Shapiro (1985), Jullien and Pavan (2019); Jullien, Pavan and Rysman (2021)), and platform economics (e.g., Rochet and Tirole (2003)). However, there is less research on how sales and marketing strategies may interact with these concepts. One notable exception is Jullien and Pavan (2019), who explore theoretically optimal marketing strategies in two-sided platform markets. This study shows empirically that industries characterized by network effects in demand have the highest investment rates in customer capital, a finding we believe is new to the literature.

The rest of this study proceeds as follows. The next section discusses the theory of customer capital. Section 2 describes the data, and Section 3 presents important facts gleaned from the constructed data set. Section 4 presents evidence that sales and marketing expense should be treated as an investment. Section 5 shows evidence on the variables that explain the cross-sectional variation across industries in customer capital investment, and Section 6 concludes.

1 Customer capital: A theory

What exactly is meant by the term "customer capital?" In order to fix ideas and guide the empirical analysis, this section discusses the model of Gourio and Rudanko (2014), which specifies a precise notion of investment in customer capital. In the model, in addition to a standard physical investment

[5]This literature hints at this endogeneity: Foster et al. (2008) discuss the importance of customer-supplier relationships; Khandelwal (2010) notes that product "quality" can result from advertising; Foster et al. (2016) build a model in which customers learn over time about a firm's products, and they state that such learning "could include customer learning through 'word of mouth', the firm's own advertising efforts, the blossoming of producer-customer relationships through repeated interactions or several other possibilities." Eslava et al. (2024) show that idiosyncratic demand for a firm's products is correlated with advertising expenditures.

decision, firms also decide how much to spend on a sales force. The product market is characterized by search frictions that require a buyer to meet with a sales representative in order to become a new customer of the firm. Firms spend up front on a sales force and offer discounts to win new customers. Once there is a match between a firm and a new customer, the customer continues to buy the firm's product as long as the present value of per-period prices does not exceed the present value of the utility flow the consumer expects to get from the goods purchased. This latter feature follows from the nature of search frictions in the product market.

In the model, s reflects the efficiency units of the sales force, $\kappa(s)$ is the (increasing and convex) cost function of the sales force, ε is the initial (discounted) price provided to a new customer, $\eta(\theta)$ is the matching function for a given s which specifies the probability of a match given a queue of $\theta \equiv \frac{b}{s}$ where b is the number of buyers, and $v_n(k, n, z|u)$ is the marginal value to the firm of an additional customer given physical capital k, the total number of customers n, and physical capital productivity shock z. The firm charges price p to its existing customers, the marginal cost of producing output is l_y, the depreciation of the customer base (which is assumed to be exogenous) is δ_n, and the discount factor is β.

As noted in Gourio and Rudanko (2014), the marginal value of acquiring a new customer is forward looking:

$$v_n(k, n, z|u) = p - l_y + \beta(1 - \delta_n)E_z v_n(k', n', z'|u) \tag{1}$$

In this case, given the search frictions in the product market, the firm is able to charge a price p that is equivalent to the flow utility that a customer has from consuming the product, which is u. In other words, the firm captures all of the surplus associated with the relationship after the customer has been acquired. The markup is $p - l_y$. Acquiring new customers allows firms to earn markups going forward on the customer base, at least until the customer relationship ends.

A key optimality condition of the model implies that firms choose the size of the sales force to equate the marginal cost of hiring more sales workers and the marginal benefit of the customers acquired by the increase in the sales force. The condition is presented in equation 2.8 of the study. Re-arranged slightly, the condition is:

$$\kappa'(s) = \eta(\theta)(v_n(k, n, z|u) - \varepsilon) \tag{2}$$

The size of the sales force, and the amount the firm spends on a sales force, is closely linked to the marginal value of an additional customer to the firm, which is forward looking given the markups that will be earned once the customer is acquired. Equations 1 and 2 together illustrate the notion of *investing in customer capital*; spending on a sales force allows the firm to acquire new customers and earn potentially large markups from the ongoing relationship. Hence, spending on a sales force is a dynamic, forward-looking decision. In the words of the authors: "Product market frictions turn

the customer base into a form of intangible capital, which manifests itself in increased firm value, profits, and markups."

The Gourio and Rudanko (2014) study also includes assumptions on industry output demand, which then allows for a characterization of a stationary competitive search equilibrium. The key comparative static evaluated in the study is variation across industries in the degree of search frictions, which is modeled as variation across industries in a multiplicative factor on the matching function. Gourio and Rudanko (2014) show that the ratio of expenditures on a sales force relative to total revenue is monotonically increasing in the degree of search frictions. They use this relationship to then explore differences across industries in profits and valuation based on the ratio of SG&A to revenue.

To the best of our knowledge, the Gourio and Rudanko (2014) model is the most direct demonstration of the concept of customer capital investment in the literature. Firms earn markups in their framework because of search frictions in the product market. However, it is possible for firms to earn markups on existing customers for a variety of reasons. Branding may produce loyal customers (e.g., Bronnenberg et al. (2012), Larkin (2013)), there may be poor product substitutes (Dixit and Stiglitz (1977)), or a lot of difficulty switching to an alternative seller (e.g., Cabral (2016), Menzio (2024)). Customers may have limited attention or inertia, leading them to continue to pay even when their flow utility falls below the value of the product (e.g., Della Vigna and Malmendier (2006), Einav, Klopack and Mahoney (2023)). There may be network effects in demand, allowing the firm that "wins" the market to earn substantial profits (e.g., Katz and Shapiro (1985), Rochet and Tirole (2003)).

One of the main goals of this study's empirical analysis is to provide descriptive evidence on the industries that most heavily invest in customer capital (see Section 5). Such evidence, we believe, informs the theoretical literature on the reasons that building a customer base is valuable to the firm.

2 Data

2.1 Compustat and Capital IQ

The baseline sample for the analysis includes U.S. based firms in the Compustat data set from 2007 to 2022. The start point of the sample in 2007 is dictated by the availability of sales and marketing information from Capital IQ; while there is some data collected by Capital IQ prior to 2007, it is sparse compared to afterward. The Compustat sample includes all firm-year observations with a few standard exceptions. We exclude financial firms (3-digit NAICS codes from 520 to 533) and firm-year observations with missing information on total assets, revenue, end of year stock price, or operating income before depreciation. We also exclude firm-year observations with a negative

value of either revenue or total book assets. Finally, given the importance of matching with SEC filings, we drop any firm observation with no central index key (CIK), which is the main identifier used by the SEC. The beginning sample covers 55,100 observations, as shown in Line 1 of Table 1. This sample represents almost all non-financial publicly-traded firms headquartered in the United States.

The data from Capital IQ come from the interactive website which offers an Excel plug-in to easily download data. The match between Capital IQ and Compustat identifiers is excellent, which is perhaps unsurprising given that both are produced by S&P. Line 2 of Table 1 shows a successful match for almost all firm-year observations. Revenue from Capital IQ and Compustat are almost identical, with a regression of one on the other giving a coefficient of 0.999 and an R^2 of 0.999. In addition to revenue, the data items retrieved from Capital IQ include costs of goods sold (COGS), SG&A, R&D, depreciation and amortization, general and administrative expense, net rental expense, and the constituent items comprising total sales and marketing expenditures: sales and marketing expense, advertising expense, and marketing expense, where available. Comparing the Capital IQ data to the underlying 10-K SEC filing reveals excellent coverage if the information is in the 10-K filing.[6]

As discussed in Markovitch et al. (2020), U.S. Generally Accepted Accounting Principles (GAAP) do not require firms to decompose their SG&A expenses into separate sub-categories. There may be a variety of reasons that firms choose to report their sales and marketing expense; these reasons are discussed in Markovitch et al. (2020). As shown in Line 3 of Table 1, the average size of firms that report sales and marketing expenses is similar to the total sample, and the median size of firms that report is slightly smaller. Given the lack of specific guidelines in GAAP, it is up to the discretion of the firm what exactly is included in the sales and marketing expense line.[7] Firms are then expected to describe these choices verbally in the text of the filing. Data collected from the text of these descriptions is discussed further in Section 2.2.

While it is difficult to know for certain, and there is likely variation across firms, sales and marketing expense generally includes both expenditures to acquire new customers and expenditures to maintain a customer base. An example of the former would be compensation for a sales force that is meeting new potential customers; an example of the latter would be expenditures on customer service for existing customers. Conceptually, one can interpret spending to maintain the customer base as spending to make up for customer capital depreciation, similar to the fact that a part of capital expenditures make up for depreciation of physical capital.

[6]While the revenue reported in Capital IQ and Compustat is almost identical, reported SG&A is not. Appendix Section A discusses the discrepancy in more detail, and it explains why the Capital IQ measure of SG&A is preferred.

[7]When Google's Gemini is given a simple prompt to explain what is in this item reported on the income statement, its answer focuses on four categories: (1) advertising and promotion, (2) sales force compensation and operations, (3) market research and analysis, and (4) customer relationship management.

The specific Capital IQ sales and marketing expense variable is non-missing for approximately 45% of the firm-year observations. If the Capital IQ sales and marketing expense variable is available, it is always non-zero. For approximately 10% of the total sample, sales and marketing expense is missing but there is information on advertising expense, marketing expense, or both. These are sub-categories of sales and marketing expense. For these observations, we add advertising and marketing expense together, and we include the sum as sales and marketing expense. This yields a sales and marketing expense variable for 55% of the firm-year observations based on the Capital IQ data, as shown in Line 3 of Table 1. It is useful to compare this coverage to two variables from Compustat that are widely used in research: the variable for advertising expenditures is available for 42% of the sample, and R&D expense is available for 61% of the firm-year observations.

2.2 Text from 10-K SEC filings

Firms often include extensive discussion of their sales and marketing and R&D efforts in their 10-K SEC filings; this is true even for firms that do not separately report line items for R&D and sales and marketing in their income statement. A main measurement exercise of this study is to use these textual descriptions to construct variables measuring the sales and marketing efforts of firms.

Appendix Section B contains a detailed discussion of how text from 10-K filings is used to construct a variety of variables; the main points are summarized here.[8] We start with a manual reading of the entire filing for a random sample of 150 firm-year observations. The manual reading is done with a particular focus on passages related to sales and marketing. To the degree that firms provide a detailed discussion of their sales and marketing strategy, it is almost always contained in the Item 1 Business Description section of the filing. It is often detailed in a sub-section called "Sales and Marketing," or "Marketing Strategy" under Item 1. Firms sometimes provide more limited information on their sales and marketing in the Item 7 Management Discussion and Analysis section of the 10-K, but this is less common. The Appendix contains examples of these descriptions.

To systematically quantify the information discussed by the firm in the text, Google's Gemini 1.5 Flash is used to process the Item 1 and Item 7 sections of the 10-K filings for all the firm-year observations in the sample.[9] The specific prompts given to Gemini are described in Appendix Table A1. The main prompt used to augment the sales and marketing expense data is the following:

> We are economists conducting research on the spending done by firms on sales and marketing. Your task is to read the following document and determine the extent to

[8] Appendix Section B also includes examples of firms describing sales and marketing efforts in their 10-K.

[9] We were able to feed Gemini the text from 10-K filings for 51,297 firm-year observations, which represent 93% of the sample. The reason for missing matches is most often because the text layout of the 10-K is not easily parsed given variation in section headings.

> which the firm spends resources on marketing, advertising, product promotion, branding, customer service, sales force, and other closely related activities. Based on your reading of the document, please use your best judgment to classify the extent of their spending on such activities into one of three categories: minimal, moderate, or substantial. Please limit your answer to one word from the following three: minimal, moderate, or substantial. Here is the document:

Ultimately, the answers provided by Gemini allow us to augment the sales and marketing and R&D variables by imputing zeros for a subset of the sample for which the raw data are missing. In particular, if Gemini indicates from its reading of both Item 1 and Item 7 that spending on sales and marketing expense for a given firm-year observation is "minimal", then we impute a zero for that observation. The logic of the exercise is the following: if a firm does not itemize sales and marketing expense on the income statement **and** a reading of the Item 1 and Item 7 section of the 10-K filing reveals no text that indicates moderate or substantial spending on sales and marketing, then we can safely assume that actual spending is zero. The same exercise and logic apply to R&D expenses as well.[10]

The use of Gemini allows for the reclassification of sales and marketing expenses from missing to zero for approximately 6 thousand firm-year observations. This increases the sample of firm-year observations for which sales and marketing information is available to 36,350, which is 66% of the total sample. For R&D, this leads to an increase in the sample size from 33,611 to 45,300, which is 82% of the sample. These counts are shown in Lines 5 and 6 of Table 1. If the line item is missing and Gemini indicates that spending on the item in question is moderate or substantial based on the text in either Item 1 or Item 7, then the variable remains missing in the final sample.

The Appendix reports details on the relationship between Gemini's answers and the quantitative data from Capital IQ. It also presents evidence that missing information on sales and marketing is often inconsistent with true sales and marketing being zero. Therefore, it is inaccurate to impute zeros for all observations for which sales and marketing expense is missing. Line 7 of Table 1 shows that capital expenditure information is available for almost the entire sample.

The detailed description of sales and marketing efforts in the 10-K filings also allows for the text to be used to describe the various sales and marketing strategies implemented by firms. Based on lessons learned from the manual reading of 10-K filings, five strategies are measured from the text: (1) building brand value, (2) advertising, (3) employing a sales force, (4) providing customer service, and (5) using customer data to acquire and maintain a customer base. For each firm-year observation, Gemini is used to obtain a $\{0,1\}$ variable if the firm describes using one of these

[10]Many researchers impute zeros for missing values of Compustat variables such as advertising expense, but we are unaware of research carefully justifying this decision, especially for variables for which GAAP do not require disclosure. This study imputes zero for missing values only if the text suggests minimal spending.

strategies. The exact prompts given to Gemini to obtain these variables are listed in Appendix Table A1.

Finally, Gemini is also used measure "primitives" of the business model for each firm-year observation. These measures include (1) whether the primary customers of the firm are households, other businesses, or the government, (2) whether the business model of the firm involves providing a platform for buyers and sellers to interact, and (3) whether the firm sells its product online. As before, the exact prompts given to Gemini are listed in Appendix Table A1. These variables are discussed in more detail in Section 5.

2.3 Purchase price allocations

Information from purchase price allocations (PPA) is used to estimate the value of intangible assets. A PPA is an allocation of the purchase price of a business into assets and liabilities during business combinations, and it is part of the intermediate stage in M&A transactions to combine a target's balance sheet with that of the acquirer. The valuation is conducted by third-party valuation and accounting professionals and is subject to audit. The purchase price allocation dataset comes from Business Valuation Resources' (BVR) DealStats database, which tracks M&A transaction records. BVR collects information on transactions related to public firms from SEC filings, including 10-K, 10-Q, 8-K(A), S-1, and S-4(A), and private firm transactions from various national and regional brokerage associations. BVR has a team of financial analysts to verify the database's accuracy. Similar data sets have been used in He (2022), Ewens et al. (2024), and Kepler et al. (2024) to measure the value of intangible assets. More details on the data are in Appendix Section C.

For this study, the intangible assets valued in the PPA are placed into groups following the methodology in He (2022). Each group of intangible assets is then grouped into two broad categories: those that are customer-related and those that are not customer-related. The customer-related category includes the value assigned to: "Customer Relationships", "Customer Lists", "Brands", "Trademark/Trade Names", "Domain," "Customer Contracts," and "Business relationships." The non-customer-related includes a large number of groups; the two largest are intangible assets associated with "Research" and "Technology."

Along with valuing intangible assets, the PPA process requires accountants and valuation professionals to estimate each intangible asset's useful life, which in turn determines its annual depreciation. Assets can be assigned a definite useful life, where a specific number of years is provided, or an indefinite useful life, indicating that there is no foreseeable limit to the period over which the asset is expected to generate cash inflows. In our data, trademarks and brands are typically the customer-related assets most often designated as having an indefinite useful life; however, only about 20% of these are actually classified as indefinite. The remaining intangible assets—especially

those categorized as customer relationships, and customer contracts—all have useful life estimates. As described in the next sub-section, these useful life estimates are used to calculate depreciation rates under straight-line depreciation. For assets recorded with an indefinite useful life, we adopt a baseline assumption of a 10-year useful life. These depreciation rates on customer capital are then used in the capitalization of sales and marketing expenses.

2.4 Capitalization

Investment in customer capital is the main variable examined in this study. However, for one key test in Section 4.1, it is necessary to measure the stock of customer capital. To do so, a modified version of the capitalization methodology of Eisfeldt and Papanikolaou (2013), as implemented by subsequent work by Peters and Taylor (2017), is followed. A crucial difference is that Peters and Taylor (2017) use R&D expense and 30% of overall SG&A to measure internally generated intangible capital, whereas this study uses R&D expense and sales and marketing expense. As shown below, a central finding of this study is that residual SG&A once sales and marketing expense is removed does not appear to be related to accumulated value associated with the intangible investment.

A main disadvantage of using sales and marketing expense relative to 30% of SG&A is that sales and marketing expense is less likely to be available, especially historically. In order to capitalize sales and marketing expense into a measure of customer capital, we use all available data on sales and marketing expense for firms, including the data reported prior to 2007 and the data reported by firms from before they became publicly traded.[11]

The sample for the capitalization is limited to firms that have at least 5 years of sales and marketing expense available. This restriction is made to ensure that any projection of sales and marketing expense backward in time (which is necessary for the capitalization) is based on enough data to ensure a reasonable level of confidence in the projection. Once this restriction is in place, the methodology follows Peters and Taylor (2017) closely. Using this methodology, it is possible to estimate at the firm-year level the amount of knowledge capital (based on R&D), the amount of customer capital (based on sales and marketing expense), and the amount of externally purchased intangible capital (based on the balance sheet item "intangible assets").

A critical set of parameter assumptions in conducting the capitalization is the assumed depreciation rates. The assumed depreciation rate for R&D follows the estimates provided in Ewens et al. (2024) which vary at the industry level but are generally in the 20 to 35% range.

The baseline approach to estimate the depreciation rate of customer capital uses the "useful life" estimates on customer-related intangibles from the PPA transactions described above. This produces estimates of depreciation of customer capital that vary at the two-digit NAICS level. The

[11]The initial 10-K SEC filing or the S-1 initial public offering filing for firms that recently become public often includes estimates of key income statement variables from before the firm went public.

average depreciation rate using the PPA data across industries is 14%, with a standard deviation of 5%.[12] Line 8 of Table 1 shows the number of firm-year observations for which an estimate of the book value of intangible capital is available. Overall, an estimate of the book value of intangible capital is available for about 35 thousand firm-year observations.

2.5 Revelio Labs data

According to the description on the Wharton Research Data Services website, Revelio Labs collects data from "publicly available professional profiles, job postings, employee sentiment reviews, and layoff notices." According to Cai, Chen, Rajgopal and Azinovic-Yang (2024), Revelio Labs "further uses proprietary algorithms to correct for the under-representation of lower-tier workers." The specific data set from Revelio Labs used here is the Workforce Dynamics data set, which contains estimates of both the number and salaries of workers at the firm-month level. These estimates are provided at different levels of seniority, for different geographies, and for different job categories.

The job category estimates are most important for this study. At the firm-year level, the Revelio Labs data is used to construct the salaries of workers who work in a sales or marketing capacity. Specifically, workers are grouped into the sales and marketing function if they work in the job category of "Marketing," or if they work in the job category of "Sales" and have a role as "Customer Service," "Product Manager," "Sales Associate," or "Sales Representative." The Revelio Labs data is also used to quantify salaries paid to workers in the job category of "Engineer," with the purpose of measuring whether the company sells a technical product.

In terms of data availability, the Revelio data set is available only for large firms. Using the match between the Revelio Labs firm identifier and the CIK identifier that is provided by Revelio Labs, we were able to match almost 29 thousand firm-year observations from the main data set to the Revelio Labs data set. As line 9 of Table 1 shows, the matched firms tend to be larger. Despite more limited coverage, this alternative source of data on investment in customer capital is useful, as it helps to test the robustness of the results shown using the income statement measure.

3 Investing in customer capital: the facts

The data set described above allows for the establishment of a number of important facts that we believe can help guide future research on intangible capital. We present 5 such facts in this section.

[12] More details on the calculation the customer capital depreciation rate are in Appendix Section C. As discussed there, customer capital depreciation can also be estimated using annual customer churn estimates from various industry reports and from Baker et al. (2023). Gourio and Rudanko (2014) use a flat depreciation rate of 15% across all industries. The key test of Table 5 using the capitalized value of customer capital is robust to the use of alternative customer capital depreciation rates, as shown in Appendix Table A5.

Fact 1: Investment in customer capital is large: It is on the same order of magnitude as capital expenditures, and it is larger and broader than spending on R&D.

Panel A of Table 2 shows a revenue-weighted average sales and marketing expense to revenue ratio of 4.1%, which is higher than R&D and almost 2/3 of capital expenditures.[13] The median sales and marketing expense to revenue ratio (I^{SM}/Rev) is 2.9%, which is close to the median amount spent by firms on capital expenditures. Both the R&D and sales and marketing distributions have more mass in the right tail; at the 90th percentile of the distribution, sales and marketing is 36% of revenue and R&D is 74%. In general, compared to sales and marketing expense, R&D expense is heavily concentrated in a relatively small set of firms and industries. Panel B of Table 2 shows these same statistics when isolating the sample to firm-year observations for which all three types of investment are available. Some of the facts are slightly different quantitatively, but qualitatively the patterns are similar. Figure 1 shows the densities of the three measures of investment for the sample of firms for which all three variables are available and the amount of all three variables relative to revenue is below 0.5. The patterns are similar to those found in Panels A and B of Table 2.

As mentioned above, a common measure of investment in organizational capital in the literature is 30% of SG&A. The revenue weighted residual SG&A to revenue ratio is 0.14, which implies a sales and marketing expense to SG&A ratio of $(0.041/(0.041 + 0.140) =)$ 23%. Table 2 includes summary statistics for residual SG&A, which is SG&A from Capital IQ less sales and marketing expense. Figure A1 in the appendix presents a bin-scatter of residual SG&A against sales and marketing expense; there is no monotonic relationship across firms. This highlights that sales and marketing expense and residual SG&A capture different variations across firms.

Panel C shows the distribution of salaries paid to workers in the sales and marketing function from the Revelio data, along with the corresponding sales and marketing expense for the same sample. The patterns are broadly similar. Figure A2 in the appendix shows a bin-scatter of sales and marketing expense against salaries paid to workers in sales and marketing; there is a strong and precise positive correlation. This is notable given that the two measures come from different data sources.

The choice to scale investment by revenue instead of capital follows from theories of customer capital. For example, Gourio and Rudanko (2014) shows that the sales and marketing expense to revenue ratio is monotonic in search frictions, and they therefore use this measure in the calibration of their model. Alternatively, the static framework of Bond, Hashemi, Kaplan and Zoch (2021) explores optimal spending on a "demand shifting" input, which is measured in this study using sales and marketing expense. As shown in Appendix Section D, the key first order condition of this

[13] All ratios of variables to either revenue or book capital are winsorized at the 1% level to minimize the influence of outliers.

framework relates the sales and marketing expense to revenue ratio to two underlying primitives of consumer demand: namely, the elasticity of revenue with respect to demand (fixing the price) and the elasticity of demand with respect to the demand-shifting input.[14]

Panel D of Table 2 shows that the stock of customer capital is also large. The revenue-weighted average share of intangible capital in total capital is 44%, and one-third of intangible capital is the capitalized value of sales and marketing expenses in particular. The median values show that the unweighted distribution is tilted even more toward intangible capital and capital based on sales and marketing. The median intangible capital to total capital ratio is 0.675, and the median sales and marketing-based capital to total capital ratio is 0.142, higher than both median R&D-based capital and externally acquired intangible capital.

Fact 2: Advertising expenses are a small part of overall sales and marketing expenses; advertising expenses do not accurately capture either the level or the distribution of investment in customer capital across industries.

Figure 2 plots $^{I^{SM}}/_{Rev}$ across the distribution along with the advertising expense to revenue ratio. The sample is limited to firm-year observations for which both variables are available. For firms below the median, advertising and sales and marketing expense are tightly linked. However, above the median, advertising is a small fraction of overall spending on sales and marketing expense. In particular, above the median, advertising represents less than a quarter of overall sales and marketing expense.

Figure 3 displays the sales and marketing strategies employed by firms, according to the text in their 10-Ks as processed by Gemini. Almost 60% of firms emphasize the importance of a sales force, followed by customer service (52%), brand value (48%), and the acquisition and use of customer data (28%). Advertising is mentioned in only 17% of the observations.

Appendix Figure A3 shows an industry-level scatter plot of the median advertising expense to revenue ratio against $^{I^{SM}}/_{Rev}$. There are several industries that spend large amounts on sales and marketing but modest amounts on advertising. Such industries include high tech manufacturing, professional services, publishing industries, data processing and hosting, and information services.

Fact 3: There is a striking amount of variation across industries in customer capital investment. The pattern across industries is robust to the use of a variety of measures, and it is persistent over time.

Figure 4 shows a striking amount of variation across industries in the median ratio of sales and

[14] Appendix Table A3 shows the distribution of investment scaled by total capital; the patterns are qualitatively similar.

marketing expense to revenue ($I^{SM}/_{Rev}$).[15] The top three industries are all in the broader Information sector; specifically, web search portals, libraries, archives, and other information services (519); publishing industries (513 – which includes many software companies); and computing infrastructure providers, data processing, web hosting, and related services (518). The largest companies as of 2022 in each of these industries are Meta Platforms (519), Microsoft Corp (513), and Alphabet Inc (518).

Other notable industries that spend heavily on sales and marketing include transit and ground transportation (485, which includes both Uber Technologies Inc and Lyft Inc), personal and laundry services (812, which includes Weight Watchers, now known as WW International Inc) and educational services (611, which includes firms such as Duolingo Inc and Coursera Inc). The broad sectors that spend the least on sales and marketing include mining (211, 212, 213) and rail and water transportation (482, 483).

The manufacturing sector shows a large amount of variation. Manufacturing firms that primarily sell products to households have high ratios (for example, beverage and tobacco product manufacturing (312) and leather manufacturing (316)), as do manufacturers of high-tech products (for example, medical equipment and supplies manufacturing (339) and computer and electronic product manufacturing (334)). In contrast, petroleum and coal products manufacturing (324) and primary metal manufacturing (331) have low ratios. These findings caution against treating manufacturing as a monolith when investigating the importance of customer capital.

The cross-sectional variation across industries in investment in customer capital is robust across time and alternative measures. Table 3 presents univariate regressions at the 3-digit NAICS industry level of various measures of customer capital investment on industry-median $I^{SM}/_{Rev}$. Columns 1 and 2 show a high correlation of median $I^{SM}/_{Rev}$ with the revenue-weighted average ratio or the simple average ratio. Our preference for the industry-level median ratio is due to the fact that it is more broadly representative of the firms in the industry compared to the weighted average, and it is not influenced by extreme outliers (and therefore less sensitive to decisions about how to winsorize) compared to the simple average.

Column 3 shows the correlation of the $I^{SM}/_{Rev}$ with the ratio of salaries paid to workers in the sales and marketing function to revenue $(WL)^{SM}/_{Rev}$. Recall that the numerator of the left hand side variable is calculated using salary data from Revelio Labs.[16] There is a high correlation between

[15] Industries are defined as the set of firms in the same 3-digit NAICS code. The codes have changed slightly over the sample period, and so they are harmonized over time. The analysis excludes 3-digit NAICS codes that have fewer than 5 firms over the sample period with sales and marketing data available. Appendix Table A2 shows the mapping from NAICS code to industry name.

[16] The sample for the regression in Column 3 is not limited to firms that have both sources of data available. All firms that have *either* the sales and marketing data from the income statement *or* the salary data from the Revelio Labs are included. The high degree of correlation across industries is notable given this fact. The availability of two separate measures of sales and marketing expense allows for further assessment of the robustness of these patterns, which is

the two ratios across industries, despite being calculated using different data sets and for different underlying samples of firms.

The industry-level variation is also robust over time. For the regression specification reported in column 4, the industry level medians are calculated for the first five years of the sample (2007 to 2011) and the last five years of the sample (2018 to 2022). Column 5 shows that the industry-level variation is highly correlated whether sales and marketing expense is scaled by revenue or the book value of capital, where the latter is estimated using the methodology described in Section 2.4.

The robustness of the industry-level variation across time and measures suggests that the importance of customer capital in firm profit functions is determined by "primitives" reflecting either the nature of demand or supply across industries. The idea that industries differ on such underlying primitives is central in the seminal studies on intangible capital (e.g., Eisfeldt and Papanikolaou (2014), Gourio and Rudanko (2014), Peters and Taylor (2017), and Crouzet and Eberly (2023)). As such, much of the rest of the analysis of this study focuses on this industry-level variation.

Fact 4: There is a negative correlation across industries in physical capital investment and customer capital investment; there is a positive correlation between R&D investment and customer capital investment.

Figure 5 presents bin-scatter plots of the industry-level median capital expenditures to revenue ratio ($I^{CX}/_{Rev}$, left panel) and the industry-level median R&D expense to revenue ratio ($I^{RD}/_{Rev}$, right panel) against the industry-level median $I^{SM}/_{Rev}$. The plot is based on the 67 industries shown above in Figure 4.[17]

As the left panel shows, industries with the lowest amount of $I^{SM}/_{Rev}$ have the highest amount of $I^{CX}/_{Rev}$. These include the oil and gas industry (211), other mining (212), utilities (221), and rail transportation (482). Of the 18 industries with $I^{SM}/_{Rev}$ above the revenue-weighted average of the entire sample, only one has $I^{CX}/_{Rev}$ higher than the revenue-weighted average. That one industry is Telecommunications (517). Air transportation (481), and amusement, gambling, and recreation (713) also invest via both sales and marketing and capital expenditures. However, they are the exception rather than the norm. A lesson from the left panel of Figure 5 is that a focus on capital expenditures as the only type of firm investment ignores many industries with high investment in customer capital but low physical capital investment.

In contrast, the right panel of Figure 5 shows that $I^{SM}/_{Rev}$ and $I^{RD}/_{Rev}$ are more complementary

reported in Appendix Section B.2. For the specification reported in column 3 of Table 3, only industries with at least five firms with Revelio data are included, which explains the smaller sample size.

[17] We prefer showing bin-scatter plots in the main text even though there are only 67 observations because a large number of observations are clustered on the left side of the plot, making patterns for low levels of sales and marketing expense more difficult to see in a scatter-plot. The labeled scatter-plots for all bin-scatters shown in the main body of the text are in the Appendix.

in production for several industries. There is one exception that explains the wide standard error on the second dot in the bin-scatter: chemical manufacturing (325) which includes bio pharmaceutical companies. This one industry has by far the highest $^{I^{RD}}/_{Rev}$ but low $^{I^{SM}}/_{Rev}$.[18]

Industries that have both high $^{I^{SM}}/_{Rev}$ and $^{I^{RD}}/_{Rev}$ include computer and electronic product manufacturing (334), electrical equipment, appliance, and component manufacturing (335), medical equipment manufacturing (339), and three industries that belong to the broader information sector (513, 518, 519). However, there are a large number of industries that have high sales and marketing expense and almost no R&D expenses. These include non-store retailers (454), ground transportation (485), educational services (611), and personal and laundry services (812). This matches the pattern in Figure 1 that $^{I^{SM}}/_{Rev}$ is more broadly distributed across the sample than $^{I^{RD}}/_{Rev}$.

Fact 5: Industries investing heavily in customer capital grow in importance over time.

For Figure 6, industries are sorted into three groups based on the industry's median $^{I^{SM}}/_{Rev}$. The size of each group is weighted in order to have the same approximate shares of revenue and enterprise value as of 2007 for the left and right panel, respectively. The evolution of the shares for each group over time is shown for revenue and enterprise value in the left and right panels of Figure 6, respectively. Industries with the highest $^{I^{SM}}/_{Rev}$ experience the largest increase in the share of both revenues and enterprise value over time.[19] This evidence is complementary to research showing that customer-related intangibles and expenditures on sales and marketing are rising over time (e.g., Bronnenberg et al. (2022) and He (2022)).

4 Sales and marketing expense is an investment

Spending on sales and marketing generates firm value, and it should therefore be considered an investment. To support this view, this section presents an analysis of the market value of publicly-traded firms, an analysis of the value of assets purchased in acquisitions, evidence from the text of

[18]The bio pharmaceutical industry (3254) is an interesting case study: revenue-weighted $^{I^{SM}}/_{Rev}$ is quite high even though median $^{I^{SM}}/_{Rev}$ is low. The largest companies in this industry, such as Pfizer Inc and Merck & Co, spend large amounts on sales and marketing, but smaller start-ups, many of whom have not yet received FDA approval and can therefore not legally engage in a heavy sales and marketing campaign, do not.

[19]Firms are sorted into industries for Figure 6 based on the NAICS code reported by the firm in their SEC 10-K filings. This is not directly comparable to industry-level data from the Bureau of Economic Analysis on value-added over time. The issue is that BEA data are constructed from establishment-level data, and it is difficult to know what industry is assigned by the BEA to the underlying establishments where sales and marketing efforts are conducted. For example, if the domestic value added for a manufacturing firm comes mostly from its design and sales and marketing efforts done at non-manufacturing establishments, then this value added will not be attributed to the manufacturing sector. This point is made clearly in Fort (2023), who shows that the value added by manufacturing firms in the United States is often attributed to establishments that are in industries other than manufacturing.

10-K filings, and evidence from the field of entrepreneurship.

4.1 Evidence from publicly-traded firms

As noted in Section 1, a central prediction of the Gourio and Rudanko (2014) framework is that firms in markets with greater search frictions have higher ${}^{I^{SM}}/_{Rev}$. Furthermore, firms in markets with greater search frictions also have a higher value of Tobin's Q, where Q (which we refer to as $Q^{PH} = {}^{V}/_{K^{PH}}$) is measured as the ratio of enterprise value to the stock of physical capital. As a result, there should be a positive correlation across industries between Q^{PH} and ${}^{I^{SM}}/_{Rev}$.[20]

Table 4 tests this prediction. As column 1 shows, the industry-level median ${}^{I^{SM}}/_{Rev}$ is strongly correlated with Q^{PH}. Column 2 shows that the industry median ${}^{(WL)^{SM}}/_{Rev}$ using the Revelio Labs data has similar explanatory power. The explanatory power shown in column 1 is augmented in column 3 when industry-level median ${}^{I^{RD}}/_{Rev}$ is added to the specification; the R^2 using these two variables is above 70%. Column 4 shows that inclusion of residual SG&A adds almost no explanatory power beyond these two variables. The results suggest that sales and marketing and R&D expense have explanatory power for Q^{PH}, and that residual SG&A does not. Column 5 presents a placebo test to ensure that there is nothing mechanical about including some types of expenses in the regression; the ratio of COGS to revenue adds no explanatory power when predicting Q^{PH}.

An alternative specification comes from the model in Crouzet and Eberly (2023), which relates Q^{PH} to the stock of tangible and intangible capital. Their model augments the classic Hayashi (1982) investment framework with two additional ingredients: (1) firms may earn rents and (2) firms may utilize intangible capital in addition to physical capital in production. The analysis here uses the balanced growth steady-state relationship between traditional measures of Q^{PH} and the underlying drivers of value for a company developed in the Crouzet and Eberly (2023) model.

Specifically, suppose there are two types of capital: physical (PH) and intangible (IT). Then, in the Crouzet and Eberly (2023) framework, there is the following relationship between Q^{PH} and intangible capital on the balanced growth path:

$$Q^{PH} - q^{PH} = \frac{\mu - 1}{r - g} R^{PH} + q^{IT} \frac{K^{IT}}{K^{PH}} + \frac{\mu - 1}{r - g} R^{IT} \frac{K^{IT}}{K^{PH}} \tag{3}$$

where q^n is the marginal effect of an additional unit of investment in capital n on enterprise value,

[20] Gourio and Rudanko (2014) test this prediction by showing that industries with higher SG&A have higher Tobin's Q. See Table 4 in their study. The main contribution of Table 4 relative to their result is to use sales and marketing expense as the measure of customer capital investment instead of broader SG&A, and to show that residual SG&A after removing sales and marketing does not add incremental power in predicting Q^{PH}. The conclusion we draw is that sales and marketing expense is the more precise measure of the Gourio and Rudanko (2014) notion of customer capital investment.

μ is a measure of rents derived from the profit function, r is the external cost of capital (ignoring depreciation and adjustment costs), g is the steady state growth rate, K^n is the replacement value of capital type n, and R^n is the required rate of return on capital type n which includes depreciation and the adjustment cost.

Re-arranging terms yields:

$$Q^{PH} = [q^{IT} + \frac{\mu - 1}{r - g} R^{IT}] \frac{K^{IT}}{K^{PH}} + q^{PH} + \frac{\mu - 1}{r - g} R^{PH} \tag{4}$$

which can then be used to motivate an across-industry regression specification:

$$Q_j^{PH} = \alpha + \beta \left(\frac{K^{IT}}{K^{PH}} \right)_j + \varepsilon_j \tag{5}$$

The intuition of the Crouzet and Eberly (2023) model is that industries have high Q^{PH} relative to q^{PH} for three potential reasons: they may have higher rents, they may have more intangible capital in production, and there may be an interactive effect of the two. In the context of the empirical analysis here, the goal is to see whether measures of intangible capital estimated using sales and marketing and R&D expense are correlated with Q^{PH}.[21]

Figure 7 presents the industry-level bin-scatter of Q^{PH} against $^{K^{IT}}/_{K^{PH}}$. As noted in Section 2.4, the book value of intangible capital (K^{IT}) includes the capitalized value of sales and marketing expense, R&D expense, and externally acquired intangible capital. The remarkable statistical power of $^{K^{IT}}/_{K^{PH}}$ is evident. The industries in the lowest quintile of the $^{K^{IT}}/_{K^{PH}}$ distribution have a value of $^{K^{IT}}/_{K^{PH}}$ of 0.06 and Q^{PH} of 1.15. For firms in these industries, the Hayashi (1982) framework fits well. For the top quintile industries, $^{K^{IT}}/_{K^{PH}}$ is 5.27 and Q^{PH} is 7.61.

Column 1 of Table 5 presents the coefficient estimate of β from equation 5 that corresponds to the bin-scatter. There is strong explanatory power; $^{K^{IT}}/_{K^{PH}}$ alone explains almost 80% of the variation in Q^{PH}. Column 2 of Table 5 presents a placebo test in which residual SG&A after removing sales and marketing expense is capitalized into its own "intangible" value. The inclusion of the placebo intangible value is statistically insignificant, and it adds no additional explanatory power.

Column 3 explores the effect of industry-median long-term revenue growth forecasts, risk, and industry-median age on Q^{PH}.[22] As equation 5 shows, the cost of capital and growth are two impor-

[21] A caveat to this framework is that the Crouzet and Eberly (2023) model treats markups as exogenous, and as such the study itself notes that it is not the best model to investigate intangible investment that is designed to boost markups over time. With customer capital, the intangible capital value is inextricably linked to markups.

[22] Age of a firm is measured as years since IPO, long-term revenue growth forecasts of a firm is measured as the average analyst forecast recorded in the I/B/E/S data set, and the asset beta of the industry is based on estimates by Aswath Damodaran of NYU Stern.

tant variables that should also explain Q^{PH}. Both median firm age and long-term revenue growth forecasts are included to control for differences across industries in growth, and the estimated asset beta of the industry is included to control for differences across industries in the cost of capital. As shown in column 3, the coefficient estimate on the long-term revenue growth forecast is indeed positive and significant, consistent with the theory. As shown in column 4, inclusion of these three control variables does not increase the statistical power of the specification materially, and there is only a minimal effect on the coefficient estimate for K^{IT}/K^{PH}. Taken together, the results in Tables 4 and 5 are consistent with the view that sales and marketing expenses are investments that translate into firm value.[23]

4.2 Evidence from transactions

Sales and marketing expense and R&D expense also predict value paid for intangible capital in firm acquisitions, a result shown using the PPA data described above in Section 2.3. These data are used to conduct both an industry-level and firm-level analysis. For the industry-level analysis, PPA data for all firms acquired (both public and private) are used to construct an industry-level average of the ratio of intangible asset value to revenue. The data set used to estimate this industry-level variable includes 6,761 transactions. Only industries for which there are at least 5 transactions are included in the regression analysis. The total value of intangible assets is then split between customer and non-customer related intangible value.[24]

With these dependent variables in hand, Table 6 reports estimates of regressions similar to the specification outlined in equation 6. The PPA data do not have information on R&D expense or sales and marketing expense for the target firms, and so the regression analysis uses the same industry-level medians calculated from Compustat and Capital IQ. The central question is: in industries where firms from Compustat/Capital IQ spend more on R&D and sales and marketing, are target firms (both public and private) paid more for intangible assets when they are acquired?

As Panel A of Table 6 shows, the answer is "yes." Column 1 shows the correlation of the industry-level total intangible asset to revenue ratio from the PPA transactions with the industry-level median I^{SM}/Rev and I^{RD}/Rev from Compustat/Capital IQ. Both variables are positively correlated. Columns 3 and 5 split total intangible asset value into the value associated with customer

[23] Table A5 in the Appendix shows robustness of these results when using a flat 15% or 30% for assumed depreciation rate of customer capital in the capitalization. The level of K^{IT}/K^{PH} is lower using a 30% depreciation rate for customer capital, which boosts the slope coefficient of the regression. But the statistical power is unchanged.

[24] For the PPA data, the industry-level average ratio of intangible asset value to revenue is calculated as a weighted average where the weights are the revenue of the target firms, as opposed to using the industry-level median ratio. The reason for this change is that the PPA data contain a large number of transactions for firms that are quite small relative to the public firms in Compustat/Capital IQ, and so the revenue-weighted averages are closer to the type of firms for which the right hand side variables are measured. All results are qualitatively similar if industry-level medians as opposed to weighted averages are calculated from the underlying PPA data.

and non-customer related assets. $^{I^{SM}}/_{Rev}$ predicts customer-related value and $^{I^{RD}}/_{Rev}$ predicts non-customer-related value. In terms of magnitudes, a one dollar increase in $^{I^{SM}}/_{Rev}$ translates into a 3 to 4 dollar increase in the value paid for intangible assets, and this comes completely from intangible assets associated with customers. A similar magnitude applies to R&D expense, except the value comes from non-customer related assets.

In columns 2, 4, and 6, the residual SG&A to revenue ratio at the industry-level is added. It has no additional explanatory power. The coefficients are insignificant, and the R^2 does not change meaningfully. As with the analysis using Q^{PH}, this supports the view that residual SG&A as a whole after removing sales and marketing expense does not represent an investment in intangible capital.

A firm-level analysis is also possible given that some of the firms acquired in the PPA data are publicly traded and therefore are included in the Compustat/Capital IQ sample prior to being acquired. For these targets, it is possible to relate the value paid for a specific target's intangible assets to the investment in intangible assets made by the target just prior to the acquisition.[25] There are 448 such firms in the PPA data. The specific estimated equation relates the intangible asset value to revenue ratio (e.g., a revenue-based transaction multiple for the intangible assets) to the target-level $^{I^{SM}}/_{Rev}$ and $^{I^{RD}}/_{Rev}$. Revenue, R&D, and sales and marketing are measured at the target-level from Compustat/Capital IQ for the latest time period for which the data are available, which is usually the year prior to the acquisition. The regressions are weighted by the revenue of the acquired firm.[26]

Panel B of Table 6 shows the results. The target-level results are qualitatively similar to those at the industry-level. Firms that spend more on sales and marketing just prior to the acquisition have higher value paid for intangible assets associated with customers, and firms that spend more on R&D have higher value paid for other intangible assets such as those associated with research or technology. Residual SG&A after removing sales and marketing expense does not have predictive power.[27]

[25] This analysis is closely related to the analysis in Ewens et al. (2024) who also match PPA data for publicly traded targets to the Compustat data. The main differences are (1) the use of sales and marketing expense as opposed to overall SG&A, and (2) the emphasis on customer versus non-customer related intangibles.

[26] The firm-level regressions are weighted by revenue given that the standard deviation of the intangible asset to revenue ratio is significantly larger for targets with lower revenue, suggesting heteroskedasticity across revenue in the error term. Following the test suggested in Solon, Haider and Wooldridge (2015), we find that the squared predicted residuals from unweighted specifications are significantly larger for targets with lower revenue, thereby justifying a weighted specification.

[27] As pointed out in Ewens et al. (2024), there is an underlying structural relationship between intangible investment and the book value of intangible capital. Such a structural relationship, however, must be used cautiously in this setting. The value in the PPA data is not a book value but a market value, and this is the market value of a selected set of firms that were acquired and therefore were likely to have been more successful than firms that failed, for example. Furthermore, investment in intangible capital is likely complementary to the value paid in an acquisition for the tangible assets, and so the spending on these inputs affects value beyond just the price paid for intangibles.

4.3 Evidence from 10-K filing text and the entrepreneurship literature

How do firms describe their sales and marketing strategy?

As mentioned in Section 2.2, firms extensively discuss their sales and marketing strategies in their 10-K filings. To uncover whether firms describe their strategies as an investment or as a cost, Gemini is provided the following prompt:

> Firms frequently describe some of their operations as 'sales and marketing'. We are economists trying to determine whether firms conduct sales and marketing activities as a cost of doing business, where they have to spend on sales and marketing to make each sale, or as an investment in long-lived customer relationships that prove valuable to the company over time. Please read the following document and tell us whether the firm describes its sales and marketing expenses as costs of making each sale or as an investment in building and maintaining a customer base, which retains some of its value over time like a type of asset. Please provide an answer that is only a single word, either 'cost' or 'investment'. Here is the document:

Gemini's answer to this question is "investment" for 78% of the firm-year observations in the sample. Furthermore, as shown in the left panel of Figure 8, firms in industries with a higher median $^{I^{SM}}/_{Rev}$ use language more consistent with their sales and marketing strategy being an investment. For the industries in the top quintile of the median $^{I^{SM}}/_{Rev}$ distribution, over 90% of firms use language more consistent with the sales and marketing strategy being an investment.

One concern is that firms engage in "cheap talk": investors view these expenses more positively if they are described as an investment rather than a cost, and firms with high sales and marketing expense use language that caters to this view. The right panel shows that industry-median Q^{PH} is strongly positively correlated with the fraction of firms in the industry that describe sales and marketing expenses as investments. The language firms use appears to match reality: firms that describe their sales and marketing strategy as building a valuable customer base are firms that have higher market value attached to their intangible capital.

We also ask Gemini to explain its reasoning for its classification, and its answers elucidate how sales and marketing strategies are a long-term investment in the customer base. For example, after reading the 2014 10-K filing of Consolidated Communications Holdings Inc, a communications services company, Gemini says:

> Investment. Consolidated Communications explicitly states that their sales and marketing activities aim to increase customer loyalty and retention (page 10). The company also highlights their efforts to "increase revenue per customer" through service bundles

> and value-added services (page 17), indicating a focus on building long-term relationships rather than simply covering the cost of each individual sale. The description of their marketing strategy as a key component of their overall business strategy further supports the interpretation of sales and marketing as an investment in customer relationships.

For the 2019 10-K filing of Dean Foods Company, a food and beverage company focused on dairy products, Gemini points to the importance of branding:

> Investment. Dean Foods describes their sales and marketing as integral to their long-term strategy. The company states their strategy is "to invest and grow our portfolio of brands while strengthening our operations and capabilities to achieve a more profitable core business" (Item 1. Business, Current Business Strategy). This explicitly frames sales and marketing not as a per-unit cost, but as an investment in brand building and growth, implying a long-term return on these expenditures. Further, the description of their sales force operating on a local and regional basis, with some national accounts managed centrally, suggests a focus on building and maintaining customer relationships rather than simply making individual sales.

In contrast, Gemini views the sales and marketing efforts of HKN, Inc, an energy-based company that owns oil and gas assets in the Gulf Coast region, as more consistent with a variable input cost. After reading HKN Inc's 2010 10-K filing, Gemini says:

> Cost. The provided 10-K filing focuses heavily on the firm's oil and gas operations, describing sales and marketing activities implicitly rather than explicitly. The document details production costs, revenue generation tied directly to production volume and commodity prices, and expenses related to maintaining and improving existing assets (e.g., upgrades to Main Pass 35). There is no mention of marketing campaigns, advertising, customer relationship management systems, or other activities typically associated with building long-term customer relationships. The emphasis is entirely on the cost of extraction and sale of oil and gas, with revenue directly tied to production. Therefore, sales and marketing are treated as a cost of doing business, inextricably linked to the immediate act of production and sale.

Consistent with the broader pattern in Figure 8, HKN, Inc does not break out sales and marketing expense as a sub-component of SG&A, and Gemini finds "minimal" spending on sales and marketing based on its reading of the text.

Evidence from the entrepreneurship literature

A common framework used to value companies in the entrepreneurship literature and the venture capital industry is to compare the lifetime value of the customer (LTV) to the customer acquisition cost (CAC). It is difficult to pinpoint the original source of this framework, but it is commonly taught in entrepreneurship classes and discussed extensively in industry literature, especially for the Software as a Service (Saas) sector.[28] This framework corresponds closely to the NPV framework taught widely in finance classes to evaluate whether an investment should be undertaken.

More specifically, the CAC includes both variable and new fixed costs to acquire a new customer. To calculate CAC, the framework advises to use total sales and marketing expense per customer and then subtract the part of the sales and marketing expense that is directed toward existing customers. This is the "investment cost" per new customer. The LTV then includes the present value of the future marginal profits per customer per year. It is crucial in this step to use marginal revenue minus marginal cost per customer per year, which corresponds closely to the idea of a markup. The flow of future profits earned per customer is then discounted using standard finance tools. The ratio of the LTV to the CAC then represents the net present value of an investment in customer capital, where a ratio above 1 would be considered a positive NPV project.[29]

The professionals who developed this framework clearly interpret the expenses to acquire new customers as an investment. For example, David Skok writes:

> SaaS businesses face significant losses in the early years (and often an associated cash flow problem). This is because they have to invest heavily upfront to acquire the customer, but recover the profits from that investment over a long period of time. The faster the business decides to grow, the worse the losses become. Many investors/board members have a problem understanding this, and want to hit the brakes at precisely the moment when they should be hitting the accelerator.

The SaaS Metrics Framework by Updata Partners concludes by describing the rCAC (Return on Total Customer Acquisition Cost) as:

> rCAC is the multiple of acquisition cost provided by a customer's lifetime gross profit (rCAC = LTV/tCAC). rCAC provides a churn-adjusted view of unit economics by combining GMPP [Gross Margin Payback Period] with expected customer lifetime. In conventional terms, this is the ROI [Return on Investment] on the spend to acquire a customer, perhaps the most important thing to know when analyzing a business model.

[28] Two prominent examples are the SaaS Metrics Framework by Updata Partners, and the SaaS Metrics 2.0 framework by David Skok. The description of this framework in this section follows from the teaching slides of Steve Kaplan at Chicago Booth.

[29] Professionals usually state that a good business will have an LTV to CAC ratio higher than 3. This is likely because they are considering the average of the projects undertaken by the firm, not the marginal project.

Recent developments in venture capital suggest that VCs are willing to finance such investment in customer capital directly. For example, the venture capital firm General Catalyst has a current product that provides non-dilutive financing that is directly tied to a company's sales and marketing budget. The specific product provides up to 80% of the sales and marketing budget of the company, where the payment of the company back to General Catalyst depends on the growth in the customer base.[30] As General Catalyst notes:

> GC created the Customer Value strategy to solve the issue of how to fund S&M/CAC [sales and marketing/customer acquisition cost] ... We did this by treating S&M/CAC as though it's an asset. With this strategy, GC pre-funds a company's S&M budget. In return, GC is entitled only to the customer value created by that spend, and GC's entitlement is capped at a fixed amount. After GC reaches that fixed amount, the remaining lifetime value of the customers is the company's to keep forever.

5 Explaining industry-level variation

Sales and marketing expense leads to valuable intangible capital. What is the precise mechanism? This section explores the determinants of the large and persistent variation across industries in customer capital investment in order to shed light on which mechanism appears to be most important.

5.1 Explanatory variables

There are four factors that vary across industries that are explored as determinants of the level of investment in customer capital, $^{I^{SM}}/_{Rev}$. These four factors are motivated by theory, by past empirical work, and by a careful reading of the 10-K filings for the random sample of 150 firms.

Primary customers are households: Household demand may be more elastic with respect to investment in advertising and branding relative to business demand or government demand. We expect, therefore, industries in which the primary customers are households to have higher equilibrium investment in customer capital. For every firm-year observation, Gemini is asked to read Item 1 of the 10-K SEC filing and answer the following question: "does the firm primarily market its products to households, businesses, or the government?" The precise measure is the fraction of firms in an industry for which Gemini answers this question with the word "households".[31]

Products are sold online: A large body of research in marketing suggests that the ability to target consumers online is a major advancement in marketing technology. See, for example, Goldfarb (2014) for a summary of the evidence. As such, we expect that firms in industries in which

[30] Details on this financing arrangement can be found here and here. We are grateful to Sean Higgins for informing us of this product.

[31] The exact text of all prompts given to Gemini are listed in Appendix Table A1.

products are sold online have a larger impact on consumer demand through sales and marketing efforts. For every firm-year observation, Gemini is asked to read Item 1 of the 10-K SEC filing and answer the following question: "Does the firm generate revenue by selling to its customers through online or digital avenues?" The precise measure is the fraction of firms in an industry for which Gemini answers this question with the word "yes".

A platform business model: The key characteristic of platforms is network effects in demand, or the idea that a given consumer's utility of a product is higher if other consumers also buy the product. As shown in the seminal work by Katz and Shapiro (1985), markets characterized by network effects in demand feature multiple equilibria, and as such firms can earn significant profits by successfully convincing consumers that other consumers will also use the product (see also Farrell (2007)). This is closely related to the idea in Rochet and Tirole (2003) that a crucial part of the business model based on a two-sided platform is to "get both sides of the market on board," as succinctly put in the original article.

In such industries, where the value of the platform good depends on having a large user base, the return on sales and marketing efforts may be particularly large. Sales and marketing efforts in this context both increase the user base directly, by persuading potential customers to join the network, and indirectly, by influencing potential customers' expectations about the total size of the user base. Furthermore, for every consumer a firm wins (or loses), there is an amplification effect of other consumers moving in the same direction as the consumer.[32]

For every firm-year observation, Gemini is asked to read Item 1 of the 10-K SEC filing and answer the following question: "We are economists conducting research on the underlying business models used by firms. One business model involves building a platform on which individuals or other entities interact. A platform business model involves profiting from a platform that allows two or more groups of users to interact. Your task is to read the following document and answer the following question: Is such a platform part of the business model of the firm?" The precise measure is the fraction of firms in an industry for which Gemini answers this question with the word "yes".

High-tech industries: The motivation for examining this characteristic comes from the manual reading of 10-K SEC filings. A number of firms in high-tech industries explicitly note that the technical nature of their products requires a highly skilled sales force to help acquire and maintain customers. Here is one such example from the 2012 10-K filing of Kopin Corp, a company that sells high-resolution microdisplays and optics:

> We believe that the technical nature of our products and markets demands a commit-

[32] While Katz and Shapiro (1985) do not explicitly model sales and marketing in the presence of network effects in demand, the conclusion notes that "given the possibilities of multiple equilibria ... firm's reputations may play a major role in determining which equilibrium actually obtains ... It would also be useful to consider firm's expenditures to influence consumers' expectations, such as precommitments to a given level of software". This idea is explored in more detail in Jullien and Pavan (2019).

> ment to close relationships with our customers. Our sales and marketing staff, assisted by our technical staff and senior management, visit prospective and existing customers worldwide on a regular basis. We believe these contacts are vital to the development of a close, long-term working relationship with our customers, and in obtaining regular forecasts, market updates and information regarding technical and market trends. We also participate in industry specific trade shows and conferences.
>
> Our design and engineering staff is actively involved with a customer during all phases of prototype design and production by providing engineering data, up-to-date product application notes, regular follow-up and technical assistance. In most cases, our technical staff works with each customer in the development stage to identify potential improvements to the design of the customer's product in parallel with the customer's effort.

This type of language is common for firms that produce technical products; more examples of such text from 10-K filings are in Appendix Section B.2. To measure whether an industry is high tech, the Revelio Labs data is used to construct the ratio of salaries paid to engineers to revenue. This approach follows Heckler (2005), which uses a similar classification to measure high-tech industries.[33]

5.2 Results

Figures 9 and 10 show industry-level bin-scatter plots relating industry-median $^{I^{SM}}/_{Rev}$ to each of the four explanatory variables described above. As the left panel of Figure 9 shows, there is no obvious relation between industry-median $^{I^{SM}}/_{Rev}$ and the fraction of firms in an industry with households as the primary customer base. While there are some industries that both sell primarily to households and have high amounts of sales and marketing expense (e.g., beverage and tobacco manufacturing (312), non-store retail (454), and personal services (812)), many industries with high levels of sales and marketing expense are primarily business-to-business oriented (e.g., data processing and hosting (518), medical equipment manufacturing (339), and professional services (541)).

In contrast, the three other variables are positively correlated with $^{I^{SM}}/_{Rev}$. The most robust correlation with $^{I^{SM}}/_{Rev}$ is the fraction of firms in an industry that have a platform-business model. To test the relationship formally, Table 7 presents results from industry-level regressions of the following form:

$$\left(^{I^{SM}}/_{Rev}\right)_j = \alpha + \beta X_j + \epsilon_j \tag{6}$$

[33]The industry-level summary statistics for these covariates are in Appendix Table A4.

where $\left({}^{I^{SM}}/_{Rev}\right)_j$ is the median sales and marketing expense to revenue ratio in industry j, and X_j is a 4 by j matrix of the demand characteristics of industries explained above. The estimated coefficients β are reported in Table 7.

Consistent with the figures, the most powerful determinant in the univariate specifications is the fraction of firms in the industry that operate a platform business model. This single variable explains almost 60% of the variation in the data in the univariate specification. A one standard deviation increase in the share of firms in an industry operating a platform business model is associated with a 4.4 percentage point higher ${}^{I^{SM}}/_{Rev}$.

To shed more light on this result, Table 8 shows the largest 25 firms by revenue in 2022 that Gemini classifies as having a platform business model for every year they are in the sample. The sales and marketing expense to revenue ratio for 2022 is also reported for these firms, as is the industry in which they are classified. Many of these firms are in the broader information sector (two-digit NAICS code of 51), but there are also other industries represented. The average sales and marketing expense to revenue ratio for these 25 firms is 18%.

The fraction of firms in an industry that sell their product online also has strong predictive power, as does the salaries paid to engineers.[34] When all four covariates are included in the estimation, the platform and high tech measures appear to be the strongest. The online and platform covariates are highly correlated (0.56), and so it is not surprising that the inclusion of both leads to a change in the coefficient estimates.

One concern with the correlations in Table 7 is that they may be spuriously driven by age, growth, or risk of an industry. For example, perhaps 2007 to 2022 was an especially strong time period for growth for platform-based business models, and any industry that is growing requires large expenditures on a sales force. Column 6 of Table 7 shows the correlation between ${}^{I^{SM}}/_{Rev}$ and the median age of firms in an industry, the median forecasted long-term revenue growth of firms in an industry, and the estimated asset beta from Aswath Damodaran. The median industry-level revenue growth forecast is positively correlated with ${}^{I^{SM}}/_{Rev}$, and the statistical power is strong. However, as shown in column 7, including these variables as controls does not meaningfully change the fit of the overall model nor the coefficients on the other explanatory variables.

Table 9 presents estimates where the left hand side variable is ${}^{(WL)^{SM}}/_{Rev}$. The numerator is measured using the Revelio Labs data on salaries paid to workers in sales and marketing functions. The stability of the estimates both qualitatively and quantitatively is notable given that these two variables are constructed from completely different underlying data sets. The platform covariate is the strongest predictor, but both online and high-tech are powerful predictors as well.

[34]The sample of industries shrinks when the regression includes Revelio Labs data because we require that an industry have at least five firms with all underlying data to be included in the sample.

5.3 Sales and marketing strategies

The specific sales and marketing strategies undertaken by firms are also systematically related to the underlying characteristics of the industry. To show these patterns, the data generated from Gemini's analysis of the specific sales and marketing strategies is used. Specifically, a regression specification similar to equation 6 is estimated, with the main difference being that the left hand side variable is changed to the fraction of firms in an industry that Gemini determines have a specific sales and marketing strategy such as advertising or a sales force.

Figure 11 summarizes the results of these 20 univariate regressions (5 strategies by 4 covariates). Each bar in Figure 11 is generated from a separate univariate regression of the strategy in question on the underlying covariate in question. The height of the bar represents how a one standard deviation change in the covariate in question affects the propensity of the firms in the industry to undertake the strategy in question. So, for example, the first bar in the figure starting from the left has a height of 0.11; this indicates that a one standard deviation increase in the fraction of firms that have households as their primary customers leads to a 0.11 increase in the fraction of firms that emphasize advertising as a key component of their sales and marketing strategy.

Industries that have households as their primary customers are significantly more likely to emphasize advertising and brand value as part of their sales and marketing strategy. They put no differential emphasis on the use of a sales force. In contrast, high tech industries put the most emphasis on a sales force and some emphasis on the use of customer data. They put almost no differential emphasis on advertising or brand value.

Firms in industries that sell their products online emphasize all five strategies prominently in their business descriptions. Firms in industries with platform business models also emphasize all five strategies. Perhaps the most notable pattern for these two covariates is the strong emphasis on the acquisition and use of customer data. A one standard deviation increase in either the fraction of firms selling online in an industry or the fraction of firms operating a platform business model leads to a 12-13% increase in the acquisition and use of customer data as a sales and marketing strategy. Data seems to be a critical element to the business strategy for firms in these industries (e.g., Begenau, Farboodi and Veldkamp (2018)).

The patterns in Figure 11 offer insight into the prototypical examples of the firms that invest heavily in customer capital. One such prototype is the firm selling consumer goods to households that spend heavily on advertising and brand value. Another prototype is the high tech manufacturing firm that must have a highly trained and specialized sales force to earn and to keep customers. Finally, firms engaged in selling products online and firms that operate platform business models engage in a large number of strategies to boost customer capital, with the acquisition and use of

customer data being a central component.[35]

6 Conclusion and future directions

This study provides evidence that firm spending on sales and marketing should be treated as investment in customer capital. Furthermore, this investment, along with spending on R&D and externally acquired intangible assets, explains a substantial amount of the variation across industries in measures of the value of intangible capital. Using acquisition data, it is possible to link the investment directly to the value of assets created by the investment. The across-industry variation in customer capital investment is large, with industries focused on platform business models, online sales, and high tech manufactured goods having the highest amount.

The data set compiled in this study is constructed from sources that are available to all researchers, and our plan is to make all of the code to build and analyze the data publicly available. As such, a primary goal of this study is to inspire more empirical and theoretical research on customer capital. There are many avenues for future research in this area. We emphasize four. First, is there a rise in investment in customer capital over time? If so, why? And what are the broader implications? Kaplan and Zoch (2020) and Bronnenberg et al. (2022) use different data sets to find that sales and marketing efforts appear to be increasing over time in the United States; Traina (2018) shows that SG&A has been growing as a share of operating expenditures since the 1950s. This study shows that industries that invest heavily in customer capital are the fastest growing in terms of revenue and enterprise value. More research on these questions is needed.

Second, how does investment in customer capital affect within-industry competition? This study focuses exclusively on the across-industry variation, but the within-industry variation may help answer fundamental questions on business dynamism, market power, and market concentration. Some of the largest firms in the economy have platform business models and invest heavily in customer capital. These include household names such as Amazon, Facebook, and Google. What is the role of customer capital investment in firm growth? How does the customer capital of incumbents affect entry into markets? Are trends in markups related to trends in customer capital investment?

Third, what are the implications for finance? Is investment in customer capital riskier than investment in other types of capital? Is it easier to finance such investment with debt, equity, or some combination of the two? Anecdotally, venture capital investment disproportionately goes to firms in business services and the software industry, which this study shows are two industries that have high rates of investment in customer capital. Is this a coincidence or is it illustrative of a deeper

[35] In Appendix Table A6, measures of investment in customer capital at the industry level are regressed on measures of the industry-level technical returns to scale from production. The estimates on the returns to scale are from McAdam, Meinen, Papageorgiou and Schulte (2024) and Lenzu, Rivers and Tielens (2022). In general, there is no strong positive or negative correlation between measures of investment in customer capital and technical returns to scale.

mechanism?

Finally, what are the normative implications of customer capital investment? There is a large body of research focused on whether investment in sales and marketing is socially useful.[36] When firms invest more in customer capital, are customers better off? Is such investment complementary to investment in improving the efficiency of the production process? Can better measurement of the actual efforts undertaken by firms help resolve this debate? We look forward to more research on these questions.

References

Alexander, Lewis and Janice Eberly, "Investment hollowing out," *IMF Economic Review*, 2018, *66*, 5–30.

Argente, David, Doireann Fitzgerald, Sara Moreira, and Anthony Priolo, "How do entrants build market share? The role of demand frictions," *American Economic Review: Insights*, 2021.

Bagwell, Kyle, "The economic analysis of advertising," *Handbook of industrial organization*, 2007, *3*, 1701–1844.

Baker, Scott R, Brian Baugh, and Marco Sammon, "Customer churn and intangible capital," *Journal of Political Economy Macroeconomics*, 2023, *1* (3), 447–505.

Beaumont, Paul and Clémence Lenoir, "Building a customer base under liquidity constraints," *Available at SSRN 3232638*, 2019.

Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp, "Big data in finance and the growth of large firms," *Journal of Monetary Economics*, 2018, *97*, 71–87.

Belo, Frederico, Vito D Gala, Juliana Salomao, and Maria Ana Vitorino, "Decomposing firm value," *Journal of Financial Economics*, 2022, *143* (2), 619–639.

_ , Xiaoji Lin, and Maria Ana Vitorino, "Brand capital and firm value," *Review of Economic Dynamics*, 2014, *17* (1), 150–169.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch, "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data," *Journal of Monetary Economics*, 2021, *121*, 1–14.

Bronnenberg, Bart J, Jean-Pierre Dubé, and Chad Syverson, "Marketing investment and intangible brand capital," *Journal of Economic Perspectives*, 2022, *36* (3), 53–74.

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

[36] This historical debate is summarized in the introduction of Bronnenberg et al. (2022).

Cabral, Luis, "Dynamic pricing in customer markets with switching costs," *Review of Economic Dynamics*, 2016, *20*, 43–62.

Cai, Wei, Yue Chen, Shiva Rajgopal, and Li Azinovic-Yang, "Diversity targets," *Review of Accounting Studies*, 2024, *29* (3), 2157–2208.

Corrado, Carol, Charles Hulten, and Daniel Sichel, "Intangible capital and US economic growth," *Review of Income and Wealth*, 2009, *55* (3), 661–685.

_ , Jonathan Haskel, Cecilia Jona-Lasinio, and Massimiliano Iommi, "Intangible capital and modern economies," *Journal of Economic Perspectives*, 2022, *36* (3), 3–28.

Covarrubias, Matias, Germán Gutiérrez, and Thomas Philippon, "From good to bad concentration? US industries over the past 30 years," *NBER Macroeconomics Annual*, 2020, *34* (1), 1–46.

Crouzet, Nicolas and Janice C Eberly, "Understanding weak capital investment: The role of market concentration and intangibles," Technical Report, National Bureau of Economic Research 2019.

_ and Janice Eberly, "Intangibles, markups, and the measurement of productivity growth," *Journal of Monetary Economics*, 2021, *124*, S92–S109.

_ and _ , "Rents and intangible capital: A q+ framework," *The Journal of Finance*, 2023, *78* (4), 1873–1916.

_ , Janice C Eberly, Andrea L Eisfeldt, and Dimitris Papanikolaou, "The economics of intangible capital," *Journal of Economic Perspectives*, 2022, *36* (3), 29–52.

David, Joel and François Gourio, "The rise of intangible investment and the transmission of monetary policy," *Chicago Fed Letter*, 2023, (482).

Dell'Ariccia, Giovanni, Dalida Kadyrzhanova, Camelia Minoiu, and Lev Ratnovski, "Bank lending in the knowledge economy," *The Review of Financial Studies*, 2021, *34* (10), 5036–5076.

Dixit, Avinash K and Joseph E Stiglitz, "Monopolistic competition and optimum product diversity," *The American economic review*, 1977, *67* (3), 297–308.

Dou, Winston Wei, Yan Ji, David Reibstein, and Wei Wu, "Inalienable customer capital, corporate liquidity, and stock returns," *The Journal of Finance*, 2021, *76* (1), 211–265.

Einav, Liran, Benjamin Klopack, and Neale Mahoney, "Selling Subscriptions," Technical Report, National Bureau of Economic Research 2023.

_ , Peter J Klenow, Jonathan D Levin, and Raviv Murciano-Goroff, "Customers and retail growth," Technical Report, National Bureau of Economic Research 2021.

Eisfeldt, Andrea L and Dimitris Papanikolaou, "Organization capital and the cross-section of expected returns," *The Journal of Finance*, 2013, *68* (4), 1365–1406.

_ **and** _ , "The value and ownership of intangible capital," *American Economic Review*, 2014, *104* (5), 189–194.

_ **and Gregor Schubert**, "AI and Finance," Technical Report, National Bureau of Economic Research 2024.

Eslava, Marcela, John Haltiwanger, and Nicolas Urdaneta, "The size and life-cycle growth of plants: The role of productivity, demand, and wedges," *Review of Economic Studies*, 2024, *91* (1), 259–300.

Ewens, Michael, Ryan H Peters, and Sean Wang, "Measuring intangible capital with market prices," *Management Science*, 2024.

Falato, Antonio, Dalida Kadyrzhanova, Jae Sim, and Roberto Steri, "Rising intangible capital, shrinking debt capacity, and the US corporate savings glut," *The Journal of Finance*, 2022, *77* (5), 2799–2852.

Farrell, J, "Coordination and Lock-In: Competition with Switching Costs and Network Effects," *Handbook of Industrial Organization*, 2007, *3*.

Feng, Hui, Neil A Morgan, and Lopo L Rego, "Marketing department power and firm performance," *Journal of Marketing*, 2015, *79* (5), 1–20.

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi, "How exporters grow," *Review of Economic Studies*, 2024, *91* (4), 2276–2306.

Fort, Teresa C, "The changing firm and country boundaries of US manufacturers in global value chains," *Journal of Economic Perspectives*, 2023, *37* (3), 31–58.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

_ **,** _ **, and** _ , "The slow growth of new plants: Learning about demand?," *Economica*, 2016, *83* (329), 91–129.

Goldfarb, Avi, "What is different about online advertising?," *Review of Industrial Organization*, 2014, *44*, 115–129.

Gourio, Francois and Leena Rudanko, "Customer capital," *Review of Economic Studies*, 2014, *81* (3), 1102–1136.

Gutiérrez, Germán and Thomas Philippon, "Investmentless growth: An empirical investigation," *Brookings Papers on Economic Activity*, 2017, (Fall), 89–169.

Hastings, Justine, Ali Hortaçsu, and Chad Syverson, "Sales force and competition in financial product markets: the case of Mexico's social security privatization," *Econometrica*, 2017, *85* (6), 1723–1761.

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

He, Bianca, "Unveiling intangibles," *Working Paper Available at SSRN 4321762*, 2022.

Heckler, Daniel E, "High-technology employment: a NAICS-based update," *Monthly Lab. Rev.*, 2005, *128*, 57.

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

Jullien, Bruno, Alessandro Pavan, and Marc Rysman, "Two-sided markets, pricing, and network effects," in "Handbook of industrial organization," Vol. 4, Elsevier, 2021, pp. 485–592.

_ **and** _ , "Information management and pricing in platform markets," *The Review of Economic Studies*, 2019, *86* (4), 1666–1703.

Kaplan, Greg and Piotr Zoch, "Markups, labor market inequality and the nature of work," Technical Report, National Bureau of Economic Research 2020.

Katz, Michael L and Carl Shapiro, "Network externalities, competition, and compatibility," *The American Economic Review*, 1985, *75* (3), 424–440.

Kepler, John D, Charles McClure, and Christopher R Stewart, "Competition Enforcement and Accounting for Intangible Capital," *Fama-Miller Working Paper, Chicago Booth Research Paper*, 2024, (24-20).

_ **, Vic Naiker, and Christopher R Stewart**, "Stealth acquisitions and product market competition," *The Journal of Finance*, 2023, *78* (5), 2837–2900.

Khandelwal, Amit, "The long and short (of) quality ladders," *The Review of Economic Studies*, 2010, *77* (4), 1450–1476.

Larkin, Yelena, "Brand perception, cash flow stability, and financial policy," *Journal of Financial Economics*, 2013, *110* (1), 232–253.

Lenzu, Simone, David Rivers, and Joris Tielens, "Financial shocks, productivity, and prices," *Working Paper*, 2022.

Lev, B, "The Valuation of Organization Capital," 2005.

Loukas, Lefteris, Manos Fergadiotis, Ion Androutsopoulos, and Prodromos Malakasiotis, "EDGAR-CORPUS: Billions of Tokens Make The World Go Round," in "Proceedings of the Third Workshop on Economics and Natural Language Processing" Association for Computational Linguistics Punta Cana, Dominican Republic November 2021, pp. 13–18.

Markovitch, Dmitri G, Dongling Huang, and Pengfei Ye, "Marketing intensity and firm performance: Contrasting the insights based on actual marketing expenditure and its SG&A proxy," *Journal of Business Research*, 2020, *118*, 223–239.

McAdam, Peter, Philipp Meinen, Chris Papageorgiou, and Patrick Schulte, "Returns to scale: New evidence from administrative firm-level data," Technical Report, Deutsche Bundesbank Discussion Paper 2024.

McGrattan, Ellen R, "Intangible capital and measured productivity," *Review of Economic Dynamics*, 2020, *37*, S147–S166.

Menzio, Guido, "Markups: A search-theoretic perspective," Technical Report, National Bureau of Economic Research 2024.

Morlacco, Monica and David Zeke, "Monetary policy, customer capital, and market power," *Journal of Monetary Economics*, 2021, *121*, 116–134.

Nath, Pravin and Neeraj Bharadwaj, "Chief marketing officer presence and firm performance: Assessing conditions under which the presence of other C-level functional executives matters," *Journal of the Academy of Marketing Science*, 2020, *48* (4), 670–694.

Peters, Ryan H and Lucian A Taylor, "Intangible capital and the investment-q relation," *Journal of Financial Economics*, 2017, *123* (2), 251–272.

Ptok, Annette, Rupinder P Jindal, and Werner J Reinartz, "Selling, general, and administrative expense (SGA)-based metrics in marketing: Conceptual and measurement challenges," *Journal of the Academy of Marketing Science*, 2018, *46*, 987–1011.

Rochet, Jean-Charles and Jean Tirole, "Platform competition in two-sided markets," *Journal of the European Economic Association*, 2003, *1* (4), 990–1029.

Roussanov, Nikolai, Hongxun Ruan, and Yanhao Wei, "Marketing mutual funds," *The Review of Financial Studies*, 2021, *34* (6), 3045–3094.

Solon, Gary, Steven J Haider, and Jeffrey M Wooldridge, "What are we weighting for?," *Journal of Human Resources*, 2015, *50* (2), 301–316.

Sun, Qi and Mindy Z Xiaolan, "Financing intangible capital," *Journal of Financial Economics*, 2019, *133* (3), 564–588.

Traina, James, "Is aggregate market power increasing? production trends using financial statements," *Production Trends Using Financial Statements (February 8, 2018)*, 2018.

Vigna, Stefano Della and Ulrike Malmendier, "Paying not to go to the gym," *american economic Review*, 2006, *96* (3), 694–719.

Table 1: Impact of Filtering Conditions on Sample Size and Characteristics

		N	Avg Rev	Median Rev
1	Baseline Sample	55,104	3,599	333
2	With Capital IQ data	55,064	3,597	333
3	Capital IQ Sales & Marketing expense available	30,107	3,626	295
4	Compustat R&D expense available	33,609	3,895	240
5	Sales & Marketing expense available, after Gemini results	36,347	3,392	246
6	R&D expense available, after Gemini results	45,305	3,615	306
7	CAPX available	55,000	3,605	334
8	Capitalization possible	34,778	3,837	317
9	Revelio data available	28,954	5,584	833

The baseline sample starts with annual data for all non-financial U.S. publicly traded firms from 2007 to 2022.

Table 2: Investment Facts

	N	Wgt Avg	p10	p25	Median	p75	p90
Panel A							
Sales and Marketing expense to revenue $(I^{SM}/_{Rev})$	36,347	0.041	0.000	0.002	0.029	0.148	0.359
R&D expense to revenue $(I^{RD}/_{Rev})$	45,305	0.034	0.000	0.000	0.017	0.151	0.738
Capital expenditure to revenue $(I^{CX}/_{Rev})$	55,000	0.065	0.004	0.013	0.031	0.077	0.253
Residual SGA to revenue $(^{RSGA}/_{Rev})$	35,869	0.140	0.061	0.110	0.199	0.377	1.618
Panel B. All investment variables available							
Sales and Marketing expense to revenue $(I^{SM}/_{Rev})$	30,928	0.042	0.000	0.003	0.032	0.162	0.378
R&D expense to revenue $(I^{RD}/_{Rev})$	30,928	0.038	0.000	0.000	0.019	0.163	0.721
Capital expenditure to revenue $(I^{CX}/_{Rev})$	30,928	0.056	0.004	0.012	0.029	0.065	0.200
Residual SGA to revenue $(^{RSGA}/_{Rev})$	30,928	0.137	0.062	0.111	0.198	0.379	1.700
Panel C. Comparison with Revelio when both are available							
Sales and Marketing expense to revenue $(I^{SM}/_{Rev})$	18,627	0.041	0.000	0.003	0.027	0.125	0.328
SM salaries to revenue $(^{(WL)^{SM}}/_{Rev})$	18,627	0.030	0.008	0.022	0.053	0.121	0.274
Panel D. Components of total capital							
Physical capital $(K^{PH}/_{K})$	34,730	0.559	0.055	0.125	0.325	0.731	0.961
Intanglible capital $(K^{IT}/_{K})$	34,730	0.441	0.039	0.269	0.675	0.875	0.945
Capitalized Sales and Marketing expense $(K^{SM}/_{K})$	34,730	0.150	0.000	0.016	0.142	0.432	0.654
Capitalized R&D expense $(K^{RD}/_{K})$	34,730	0.065	0.000	0.000	0.017	0.202	0.509
Externally acquired intangible capital $(K^{EI}/_{K})$	34,730	0.226	0.000	0.003	0.080	0.282	0.534

Table 3: Comparing Industry-Level Measures of Investment in Customer Capital

	(1) $I^{SM}/_{Rev}$, wgt avg	(2) $I^{SM}/_{Rev}$, avg	(3) $(WL)^{SM}/_{Rev}$, median	(4) $I^{SM}/_{Rev}$, median ('18 - '22)	(5) $I^{SM}/_{K}$, median
$I^{SM}/_{Rev}$, median	0.740**	1.235**	0.523**		0.463**
	(0.072)	(0.086)	(0.055)		(0.057)
$I^{SM}/_{Rev}$, median ('07 - '11)				1.121**	
				(0.086)	
Constant	0.013**	0.032**	0.023**	0.004	0.013**
	(0.003)	(0.006)	(0.004)	(0.005)	(0.002)
Observations	67	67	62	66	64
R^2	0.769	0.733	0.495	0.727	0.697

This table presents industry-level regressions of various measures of investment in customer capital on the industry-level median ratio of sales and marketing expense to revenue. Robust standard errors are in parentheses. $^{*}p < 0.05$,$^{**}p < 0.01$.

Table 4: Explaining Q^{PH} with ${}^{I^{SM}}/_{Rev}$

	Dependent variable: Q^{PH}, median				
	(1)	(2)	(3)	(4)	(5)
${}^{I^{SM}}/_{Rev}$, median	38.39**		32.74**	32.49**	34.46**
	(5.71)		(4.94)	(5.15)	(5.12)
${}^{(WL)^{SM}}/_{Rev}$, median		47.08**			
		(14.77)			
${}^{I^{RD}}/_{Rev}$, median			16.73**	14.59**	16.98**
			(2.44)	(4.07)	(2.50)
${}^{RSGA}/_{Rev}$, median				2.49	
				(2.59)	
${}^{COGS}/_{Rev}$, median					1.51
					(1.19)
Constant	1.64**	1.12*	1.55**	1.18**	0.47
	(0.25)	(0.48)	(0.20)	(0.32)	(0.87)
Observations	67	62	67	67	67
R^2	0.576	0.460	0.703	0.707	0.706

This table presents estimates of industry-level regressions of the enterprise value to book physical capital value ratio (Q^{PH}) on measures of intangible capital investment. $RSGA$ is residual SG&A after subtracting sales and marketing expense. Robust standard errors are in parentheses. $^{*}p < 0.05,^{**} p < 0.01$.

Table 5: Explaining Q^{PH} with ${}^{K^{IT}}/_{K^{PH}}$

	Dependent variable: Q^{PH}, median			
	(1)	(2)	(3)	(4)
${}^{K^{IT}}/_{K^{PH}}$, median	1.17**	1.16**		1.09**
	(0.14)	(0.17)		(0.11)
Placebo ${}^{K^{IT}}/_{K^{PH}}$, median		0.01		
		(0.09)		
Rev g forecast, median			43.39**	2.45
			(14.14)	(9.94)
β^A, Damodoran			0.66	0.38
			(1.14)	(0.67)
Firm age, median			-2.16	-1.67*
			(1.37)	(0.69)
Constant	1.29**	1.27**	0.38	1.55
	(0.15)	(0.13)	(1.81)	(0.96)
Observations	67	67	67	67
R^2	0.789	0.789	0.381	0.803

This table presents estimates of industry-level regressions of the enterprise value to book physical capital value ratio (Q^{PH}) on measures of intangible capital and characteristics of firms in the industry. Rev g forecast is the analyst long-term growth forecast from I/B/E/S and β^A is the asset beta of the industry as measured by Aswath Damodaran. Robust standard errors are in parentheses. $^*p < 0.05,^{**} p < 0.01$.

Table 6: Explaining Value Paid for Intangible Assets in Acquisitions

	V^{IT}/Rev, median		V^{IT}_{cust}/Rev, median		V^{IT}_{other}/Rev, median	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Industry level						
I^{SM}/Rev, median	3.307**	3.250**	3.497**	3.466**	-0.190	-0.216
	(1.199)	(1.207)	(1.126)	(1.126)	(0.321)	(0.323)
I^{RD}/Rev, median	3.897**	3.413**	0.214	-0.052	3.683**	3.465**
	(0.541)	(0.876)	(0.481)	(0.815)	(0.134)	(0.255)
$RSGA/Rev$, median		0.565		0.310		0.254
		(0.594)		(0.529)		(0.219)
Constant	0.278**	0.194*	0.156**	0.110	0.122**	0.085
	(0.053)	(0.097)	(0.034)	(0.080)	(0.032)	(0.045)
Observations	67	67	67	67	67	67
R^2	0.439	0.445	0.289	0.293	0.584	0.587
	V^{IT}/Rev		V^{IT}_{cust}/Rev		V^{IT}_{other}/Rev	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B. Target level						
I^{SM}/Rev	1.399*	1.385*	1.478**	1.525**	-0.114	-0.181
	(0.587)	(0.645)	(0.507)	(0.503)	(0.544)	(0.598)
I^{RD}/Rev	5.129**	5.171**	0.115	-0.022	4.994**	5.188**
	(0.416)	(0.445)	(0.265)	(0.249)	(0.369)	(0.500)
$RSGA/Rev$		-0.105		0.344		-0.485
		(0.836)		(0.271)		(0.713)
Constant	0.526**	0.540*	0.224**	0.178*	0.275*	0.340
	(0.151)	(0.229)	(0.055)	(0.069)	(0.135)	(0.207)
Observations	448	448	448	448	448	448
R^2	0.245	0.245	0.0821	0.0874	0.296	0.297

Panel A presents estimates of industry-level regressions of the median value paid for intangible assets in acquisitions on measures of intangible capital investment from publicly traded firms. Panel B presents estimates from target-level regressions of the value paid for intangible assets on measures of intangible capital investment just prior to the acquisition. $RSGA$ is residual SG&A after subtracting sales and marketing expense. Robust standard errors are in parentheses. $^{*}p < 0.05$,$^{**}p < 0.01$.

Table 7: Explaining Variation in $^{I^{SM}}/_{Rev}$ Across Industries

	Dependent variable: $^{I^{SM}}/_{Rev}$, median						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Households	0.001				0.027		0.026
	(0.018)				(0.014)		(0.014)
Online		0.098**			0.030		0.034
		(0.027)			(0.017)		(0.025)
Platform			0.276**		0.141**		0.119*
			(0.038)		(0.046)		(0.055)
$^{(WL)^{EG}}/_{Rev}$, median				0.720**	0.568**		0.526**
				(0.172)	(0.141)		(0.149)
Rev g forecast, median						1.239**	0.371
						(0.420)	(0.216)
β^A, Damodoran						-0.010	-0.015
						(0.024)	(0.023)
Firm age, median						0.027	0.024
						(0.032)	(0.030)
Constant	0.039**	-0.000	0.002	0.003	-0.033**	-0.045	-0.050
	(0.013)	(0.007)	(0.005)	(0.008)	(0.008)	(0.043)	(0.035)
Observations	67	67	67	62	62	62	62
R^2	0.000	0.231	0.581	0.441	0.704	0.317	0.723

This table presents estimates of industry-level regressions of sales and marketing expense as a share of revenue ($^{I^{SM}}/_{Rev}$) on characteristics of firms in the industry. The first four covariates are the share of firms in each industry that have one of the four characteristics described in Section 5. $^{(WL)^{EG}}/_{Rev}$ is total salaries of engineers scaled by revenue. Rev g forecast is the analyst long-term growth forecast from I/B/E/S and β^A is the asset beta of the industry as measured by Aswath Damodaran. Robust standard errors are in parentheses. $^*p \leq 0.05,^{**} p \leq 0.01$.

Table 8: Examples of Companies with Platform Business Models

	$I^{SM}/_{Rev}$	NAICS
AMAZON COM INC	0.082	454
ALPHABET INC.	0.094	518
WALGREENS BOOTS ALLIANCE, INC.	0.006	456
META PLATFORMS, INC.	0.125	519
CISCO SYSTEMS, INC.	0.176	334
UBER TECHNOLOGIES, INC	0.149	485
SALESFORCE, INC.	0.431	513
PAYPAL HOLDINGS, INC.	0.082	518
COUPANG, INC.	0.029	455
BLOCK, INC.	0.117	518
LIVE NATION ENTERTAINMENT, INC.	0.002	711
CARVANA CO.	0.053	441
EBAY INC	0.218	518
AIRBNB, INC.	0.180	721
ACTIVISION BLIZZARD, INC.	0.162	513
ELECTRONIC ARTS INC.	0.132	513
SERVICENOW, INC.	0.388	513
DOORDASH, INC.	0.256	519
PETCO HEALTH & WELLNESS COMPANY, INC.	0.034	459
PALO ALTO NETWORKS INC	0.391	518
IAC INC.	0.365	519
SNAP INC	0.236	519
ZOOM VIDEO COMMUNICATIONS, INC.	0.386	519
LYFT, INC.	0.129	485
TWILIO INC	0.326	518

This table lists the largest 25 companies by revenue as of 2022 that have a platform business model, along with $I^{SM}/_{Rev}$ for 2022.

Table 9: Explaining Variation in $^{(WL)^{SM}}/_{Rev}$ Across Industries

	Dependent variable: $^{(WL)^{SM}}/_{Rev}$, median						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Households	-0.004				0.002		-0.005
	(0.014)				(0.008)		(0.011)
Online		0.081**			0.038**		0.052*
		(0.020)			(0.013)		(0.021)
Platform			0.215**		0.118**		0.109**
			(0.025)		(0.039)		(0.038)
$^{(WL)^{EG}}/_{Rev}$, median				0.489**	0.287**		0.342**
				(0.087)	(0.089)		(0.075)
Rev g forecast, median						0.588	-0.183
						(0.312)	(0.179)
β^A, Damodoran						-0.019	-0.030
						(0.025)	(0.020)
Firm age, median						-0.006	-0.011
						(0.033)	(0.021)
Constant	0.045**	0.010	0.015**	0.018**	-0.004	0.024	0.037
	(0.010)	(0.005)	(0.003)	(0.005)	(0.003)	(0.045)	(0.034)
Observations	62	62	62	62	62	62	62
R^2	0.001	0.299	0.594	0.367	0.682	0.152	0.719

This table presents estimates of industry-level regressions of firms' salaries paid to workers in sales and marketing as a share of revenue ($^{WL^{SM}}/_{Rev}$) on characteristics of firms in the industry. The first four covariates are the share of firms in each industry that have one of the four characteristics described in Section 5. $^{(WL)^{EG}}/_{Rev}$ is total salaries of engineers scaled by revenue. Rev g forecast is the analyst long-term growth forecast from I/B/E/S and β^A is the asset beta of the industry as measured by Aswath Damodaran. Robust standard errors are in parentheses. $^*p \leq 0.05$,** $p \leq 0.01$.

Figure 1: Kernel Densities for Investment Variables

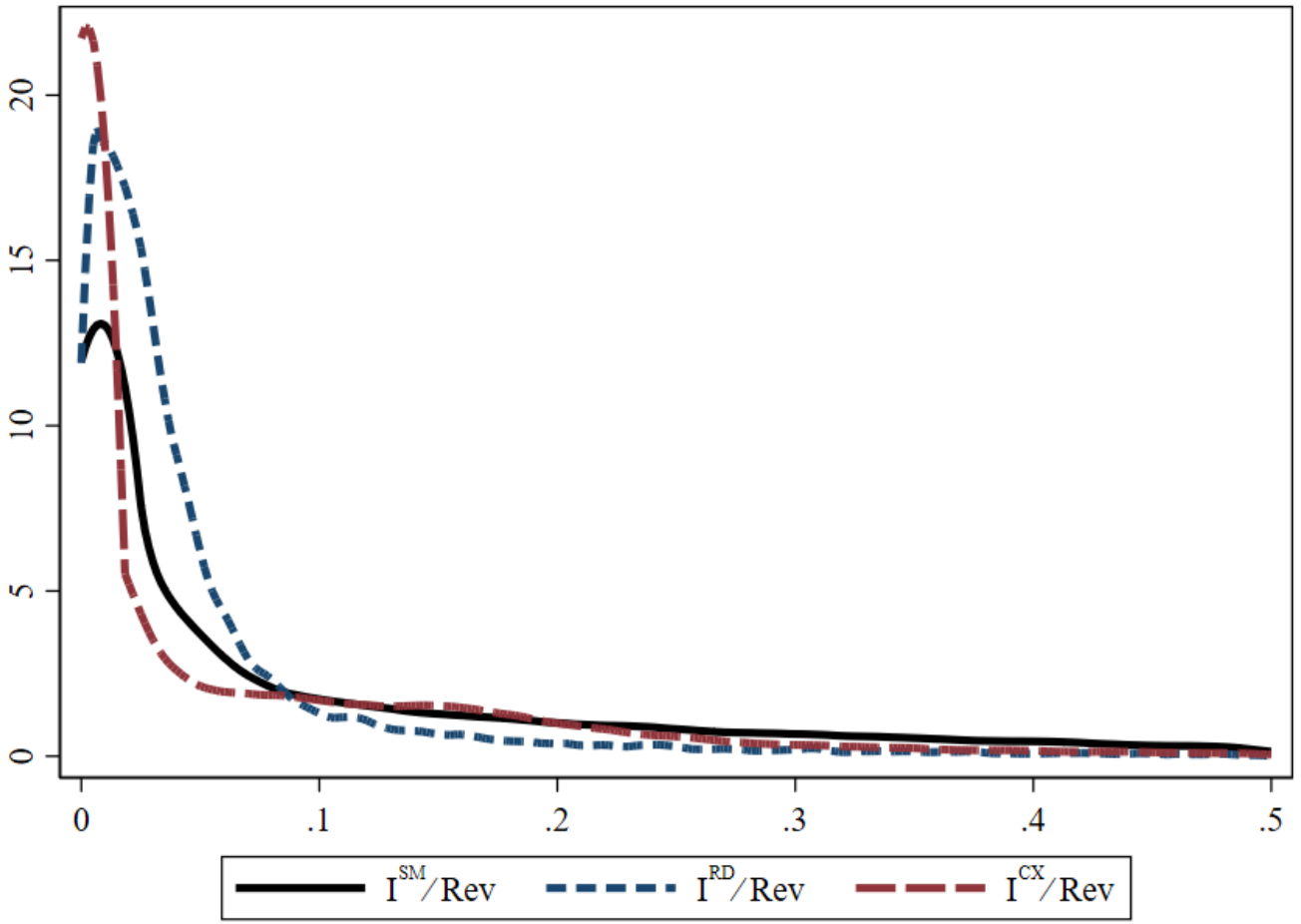


Figure 2: Sales and Marketing and Advertising Across the Distribution

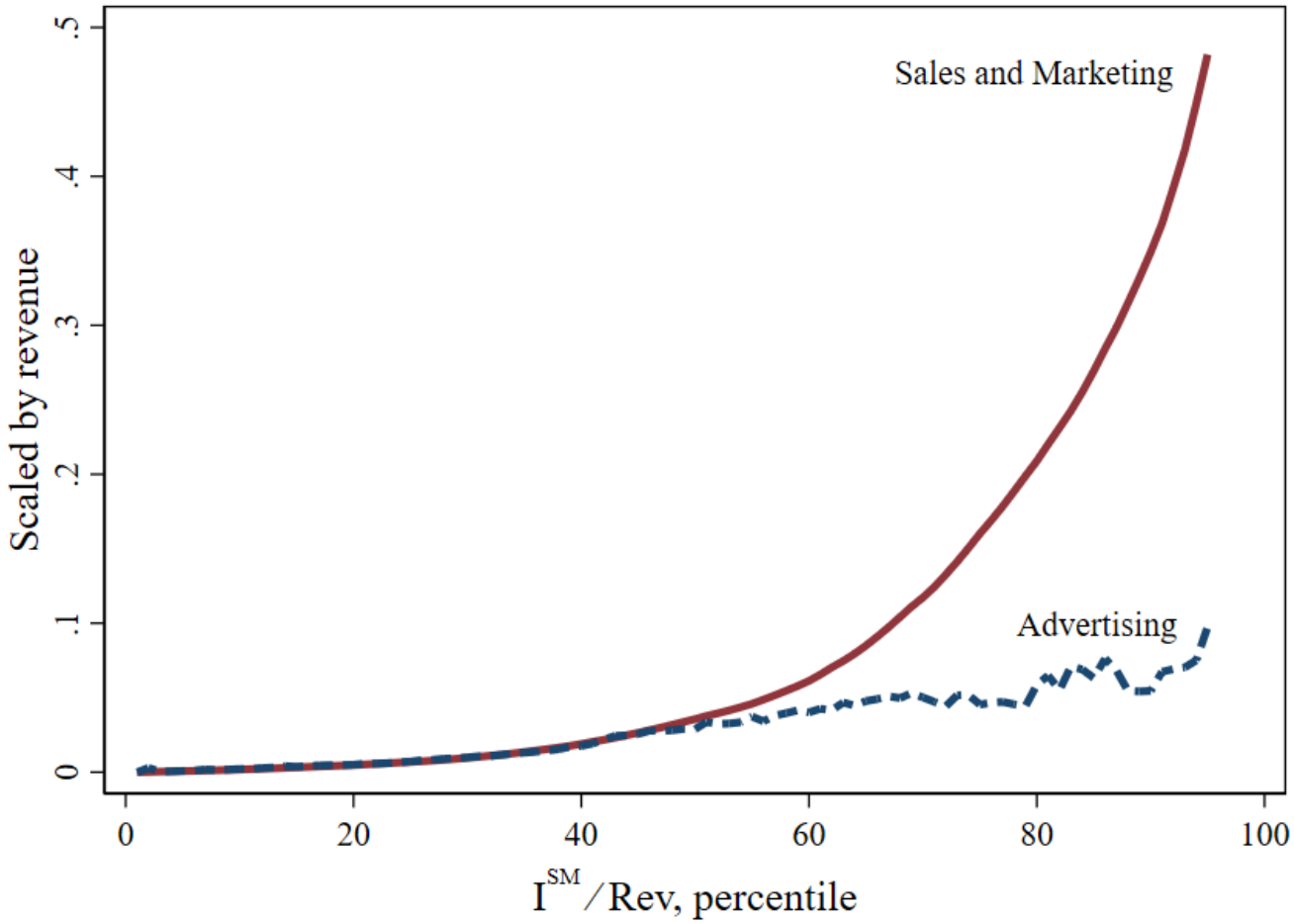
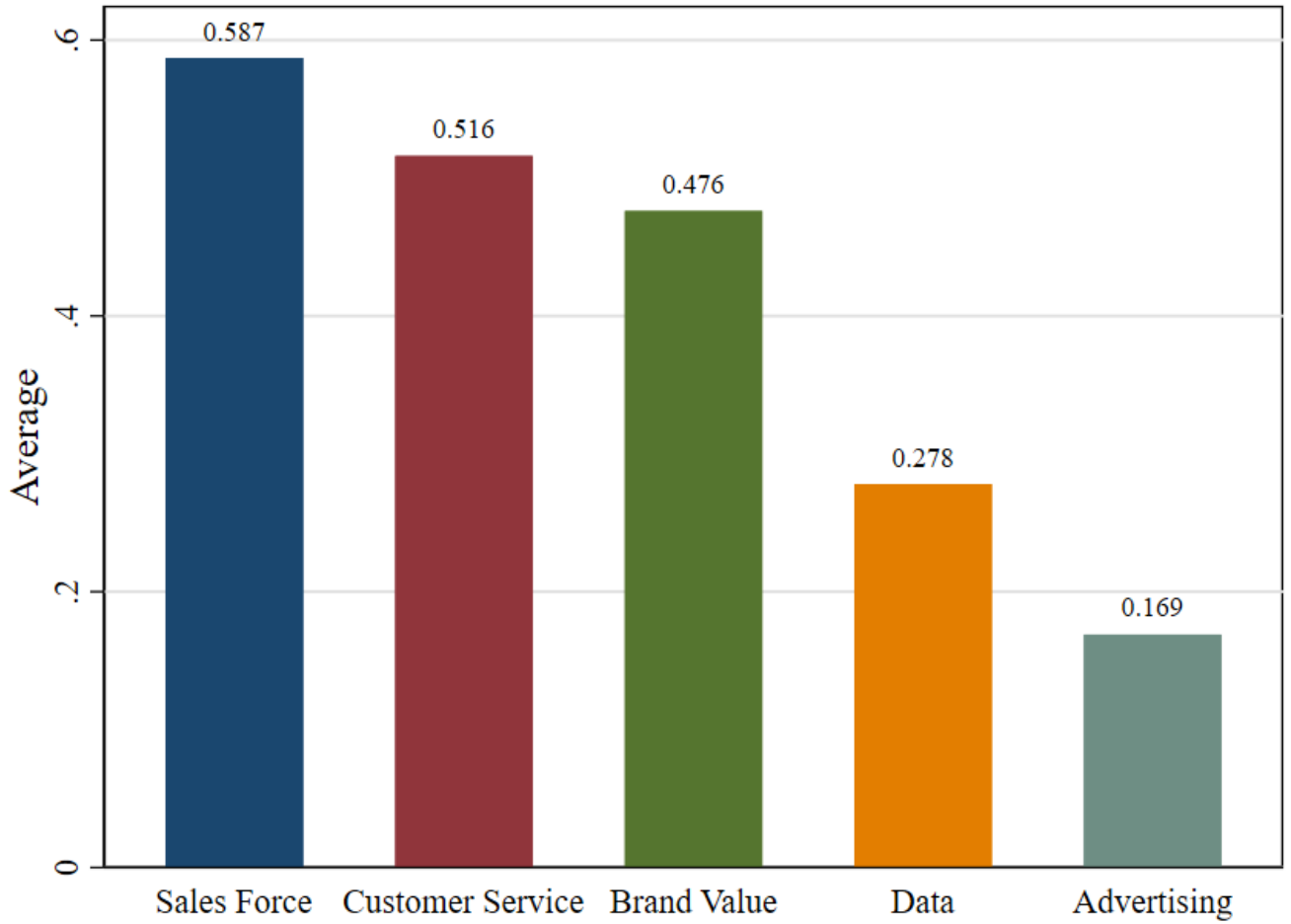
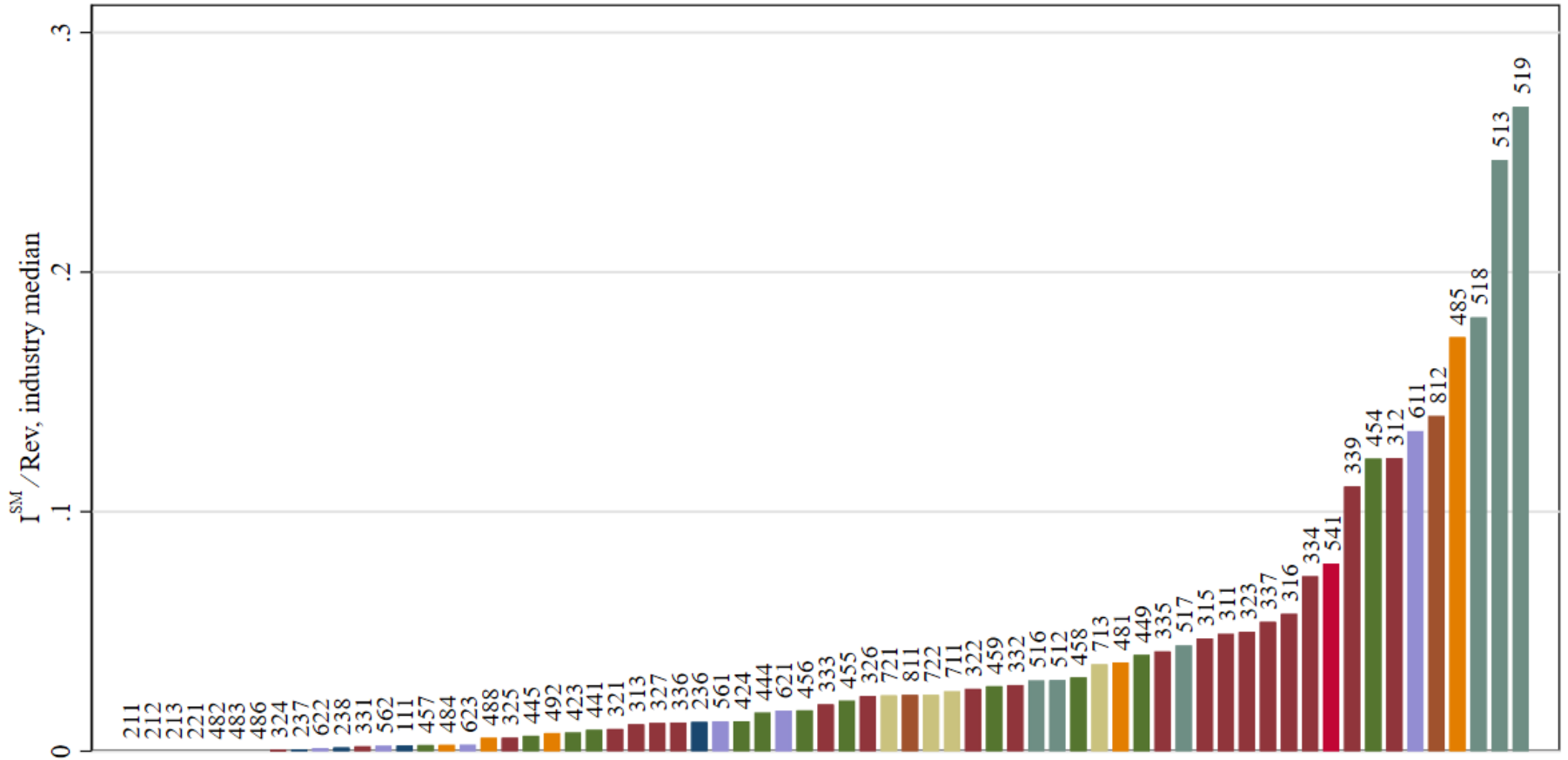


Figure 3: Fraction of firms citing each strategy



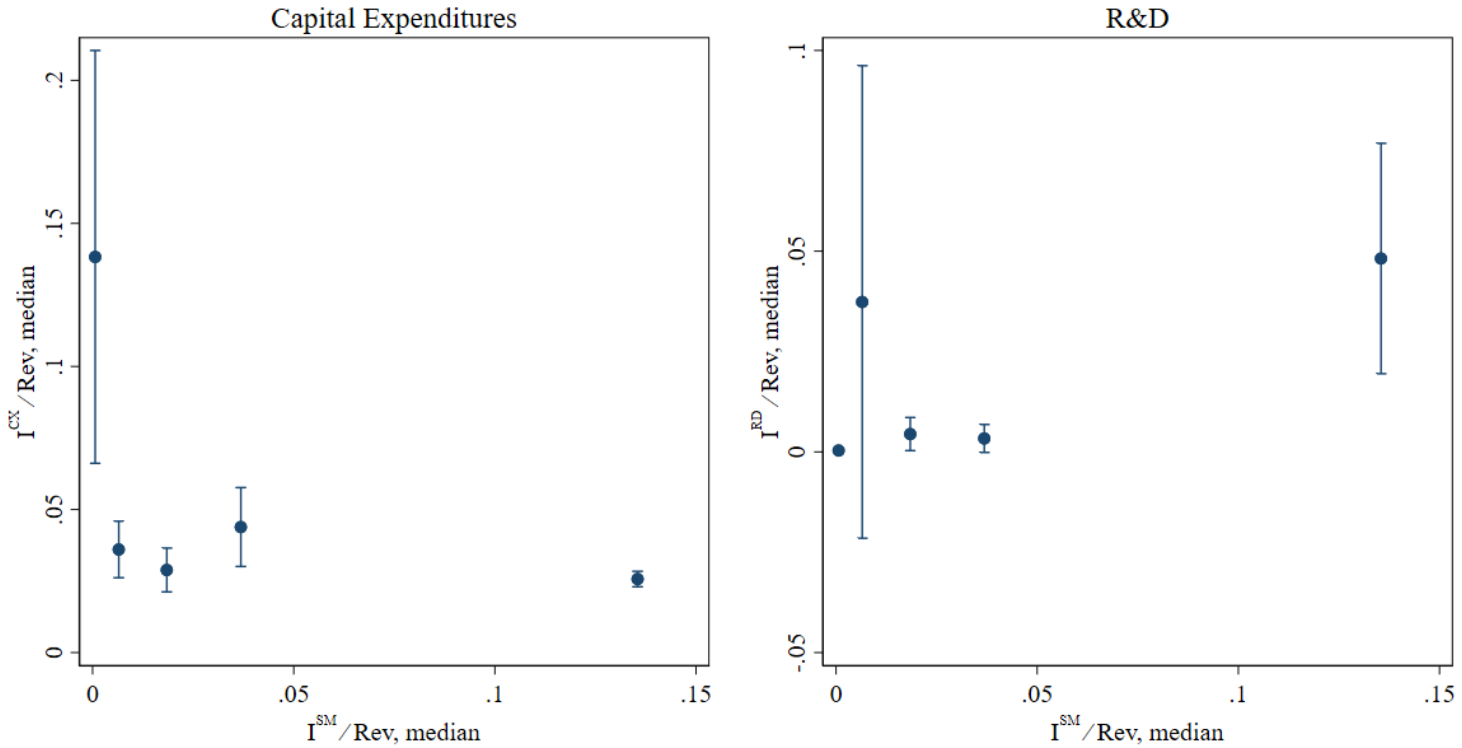
Each bar represents the fraction of firms in the sample that mention the sales and marketing strategy in question in their 10-Ks, as identified by Gemini. See Appendix Table A1 for the prompts used.

Figure 4: Median Ratio of Sales and Marketing Expense to Revenue Across Industries



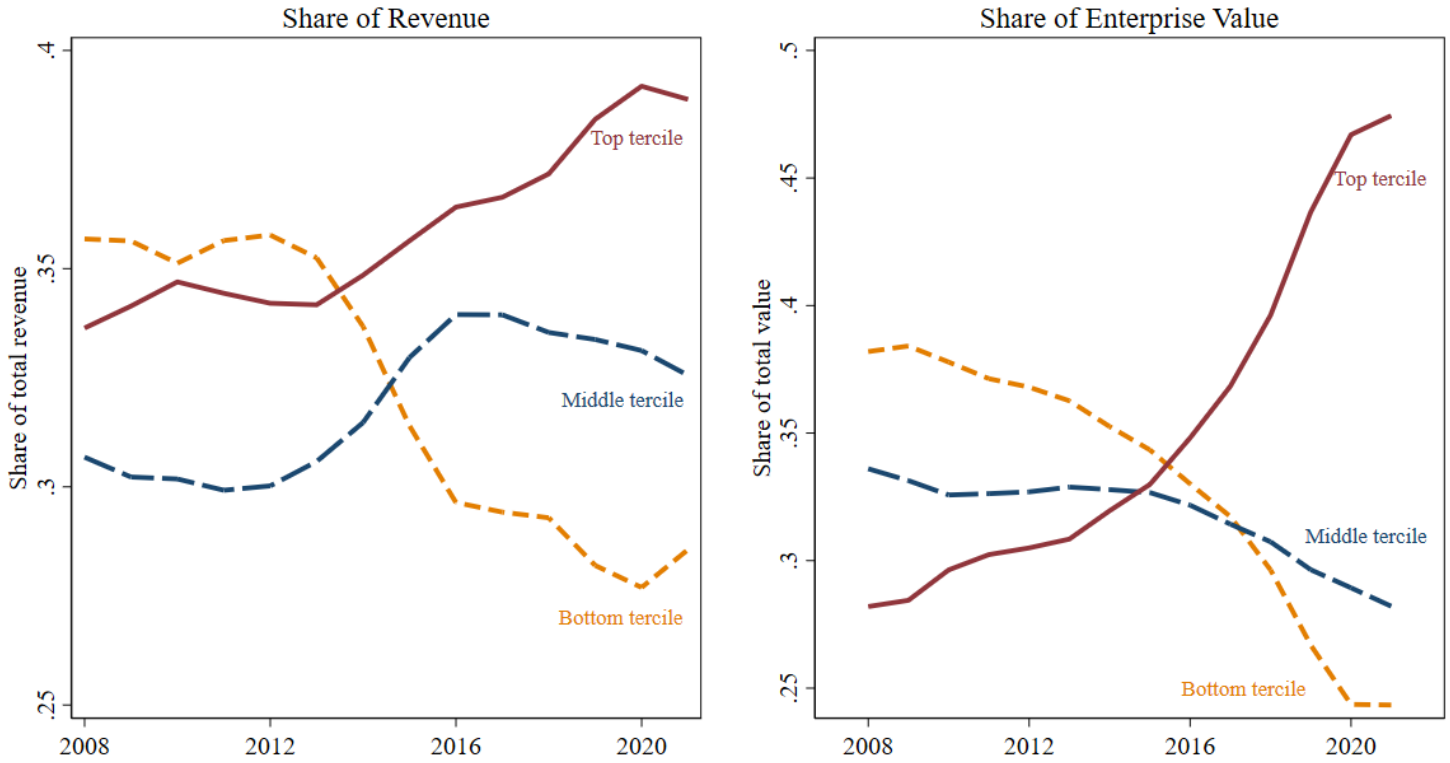
The color of the bar represents the broad category of the industry. See Appendix Table A2 for the names of each industry.

Figure 5: Comparison of I^{SM} with Capital Expenditures and R&D



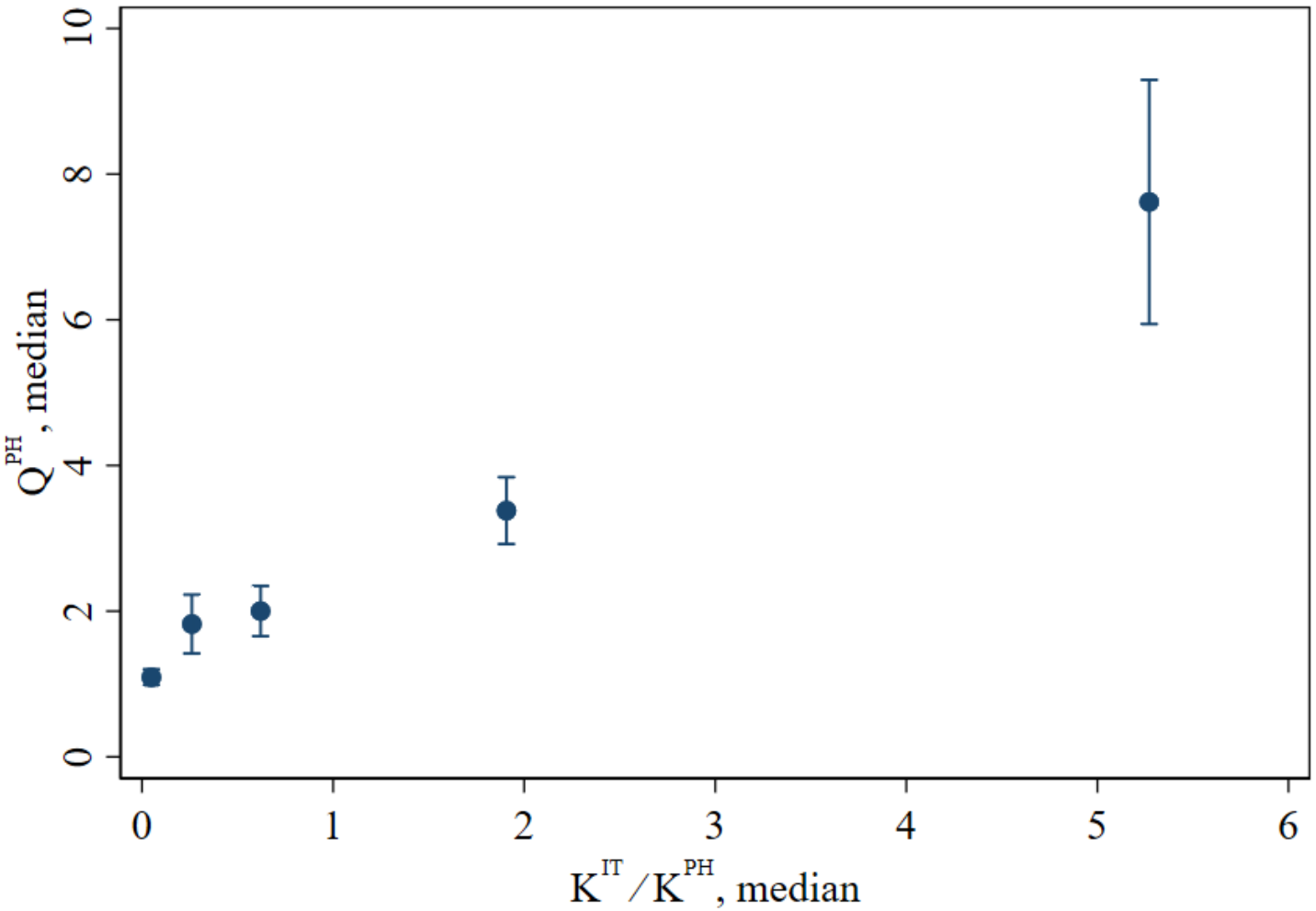
Bin-scatter plots at the industry level for capital expenditures (I^{CX}/Rev, left panel) and R&D (I^{RD}/Rev, right panel) against sales and marketing expense (I^{SM}/Rev), all scaled by revenue. Confidence bars are at the 90% level. See Appendix Figure A4 for full industry scatter plots.

Figure 6: Shares of Revenue and Enterprise Value, by Industry-level $I^{SM}/_{Rev}$



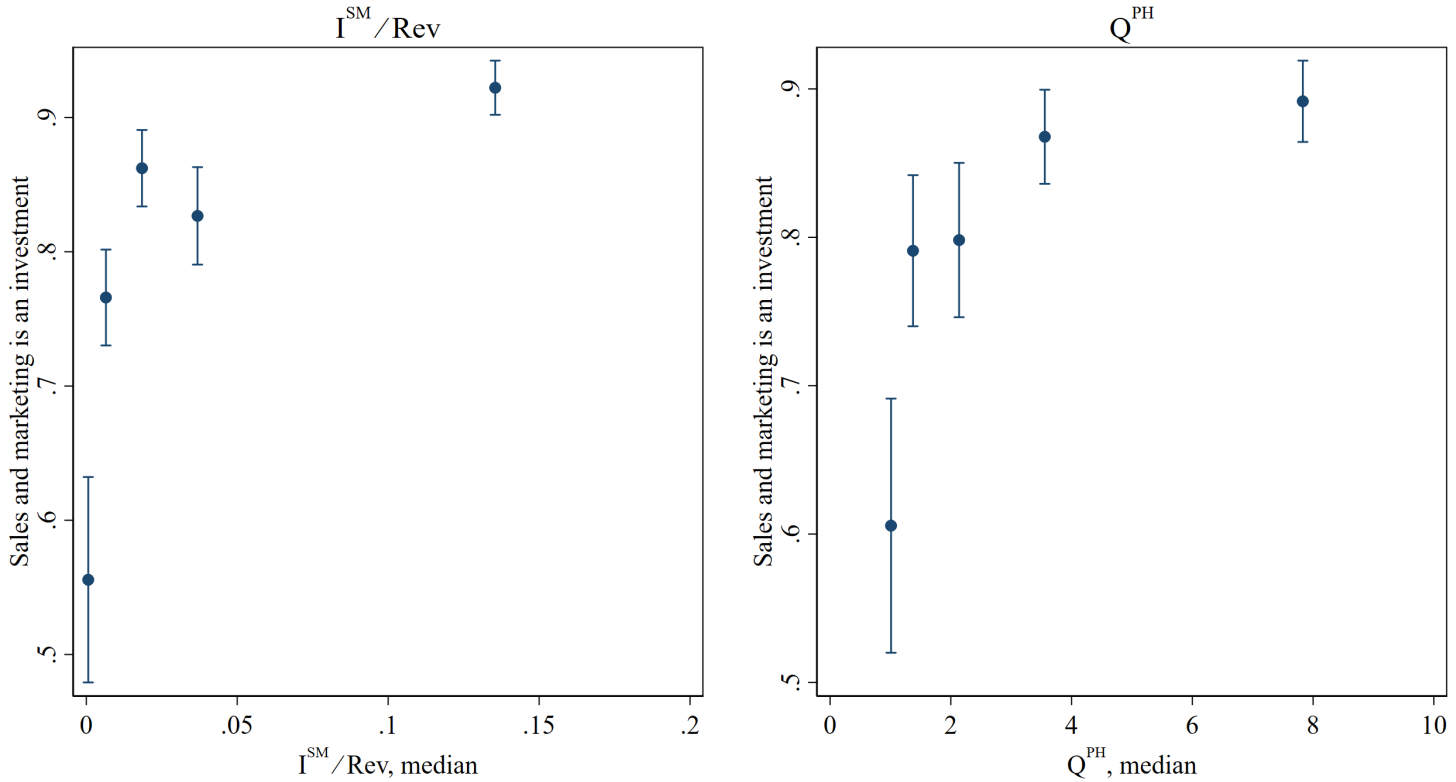
All industries are sorted into terciles based on the industry-level median ratio of sales and marketing expense to revenue. The share of total revenue and total enterprise value for each tercile is shown on the left and right, respectively.

Figure 7: Explaining Q^{PH}



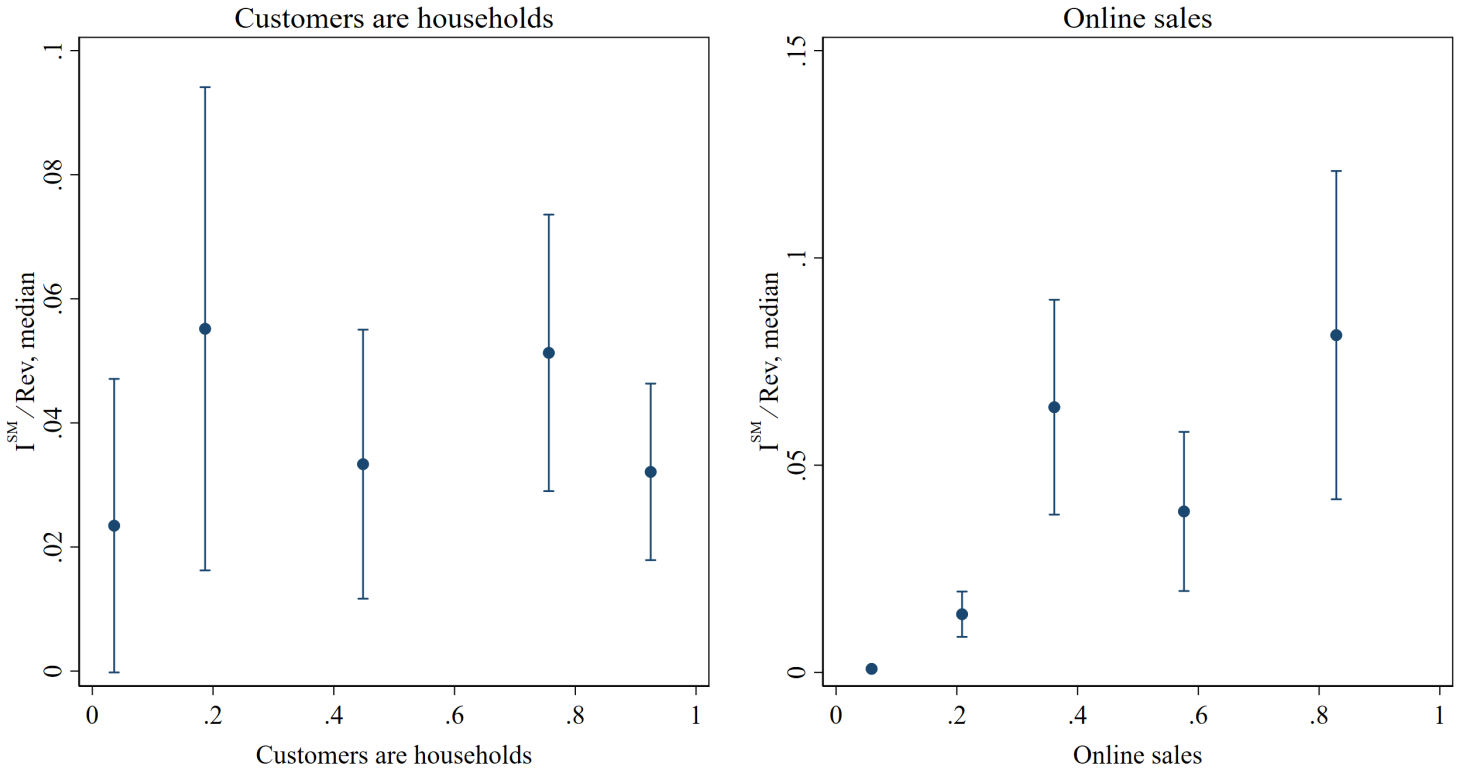
Bin-scatter plot at the industry level for the ratio of enterprise value to physical capital (Q^{PH}) against the ratio of intangible capital to physical capital (K^{IT}/K^{PH}). Confidence bars are at the 90% level. See Appendix Figure A5 for the full industry scatter plot.

Figure 8: Is Sales and Marketing an Investment or Cost?



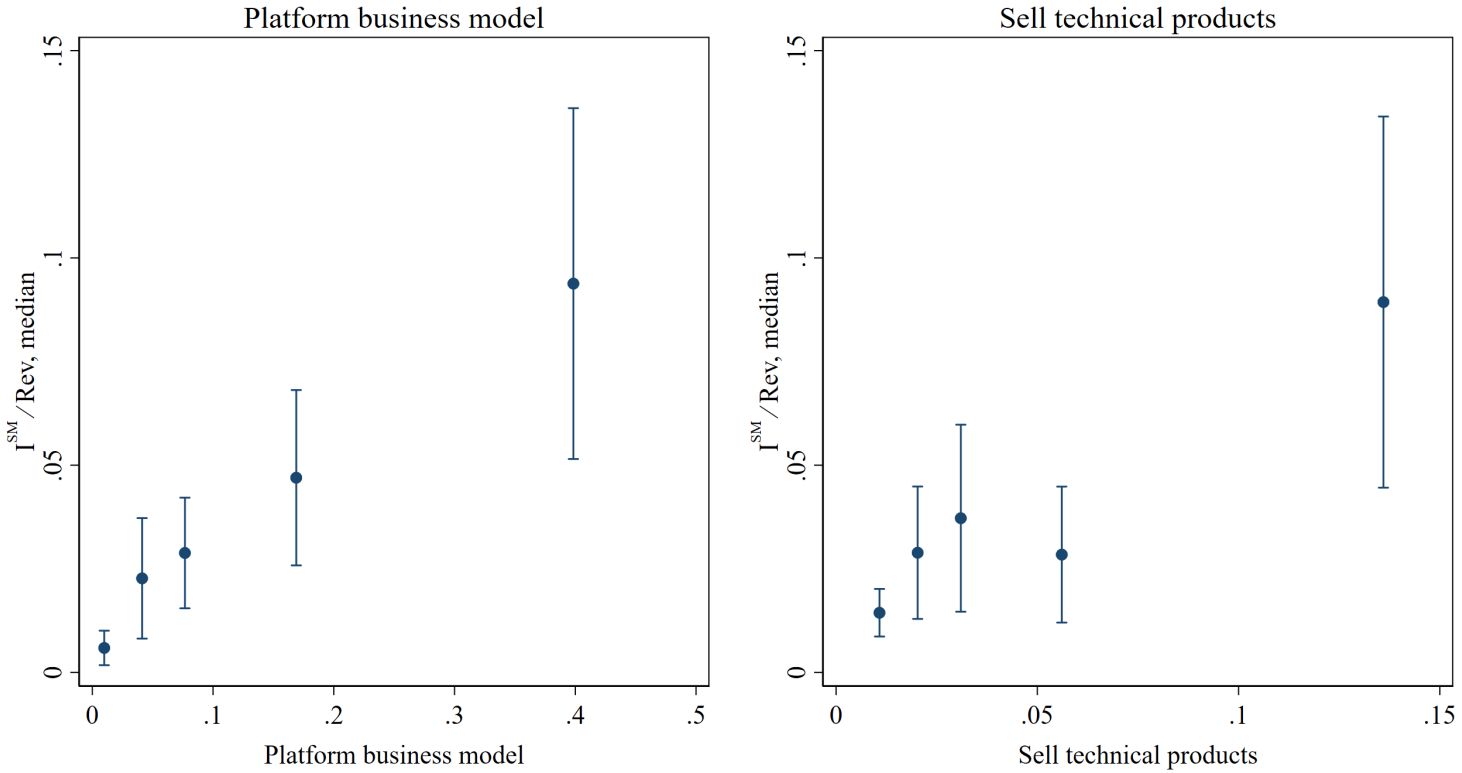
Bin-scatter plots at the industry level for the share of firms in that industry that describe their sales and marketing expense as being an investment, as identified by Gemini through textual analysis of firms' 10-Ks, against sales and marketing expense as a share of revenue ($I^{SM}/_{Rev}$, left panel) and enterprise value to physical capital ratio (Q^{PH}, right panel). Confidence bars are at the 90% level. See Appendix Figure A6 for full industry scatter plots.

Figure 9: Explaining I^{SM}/Rev



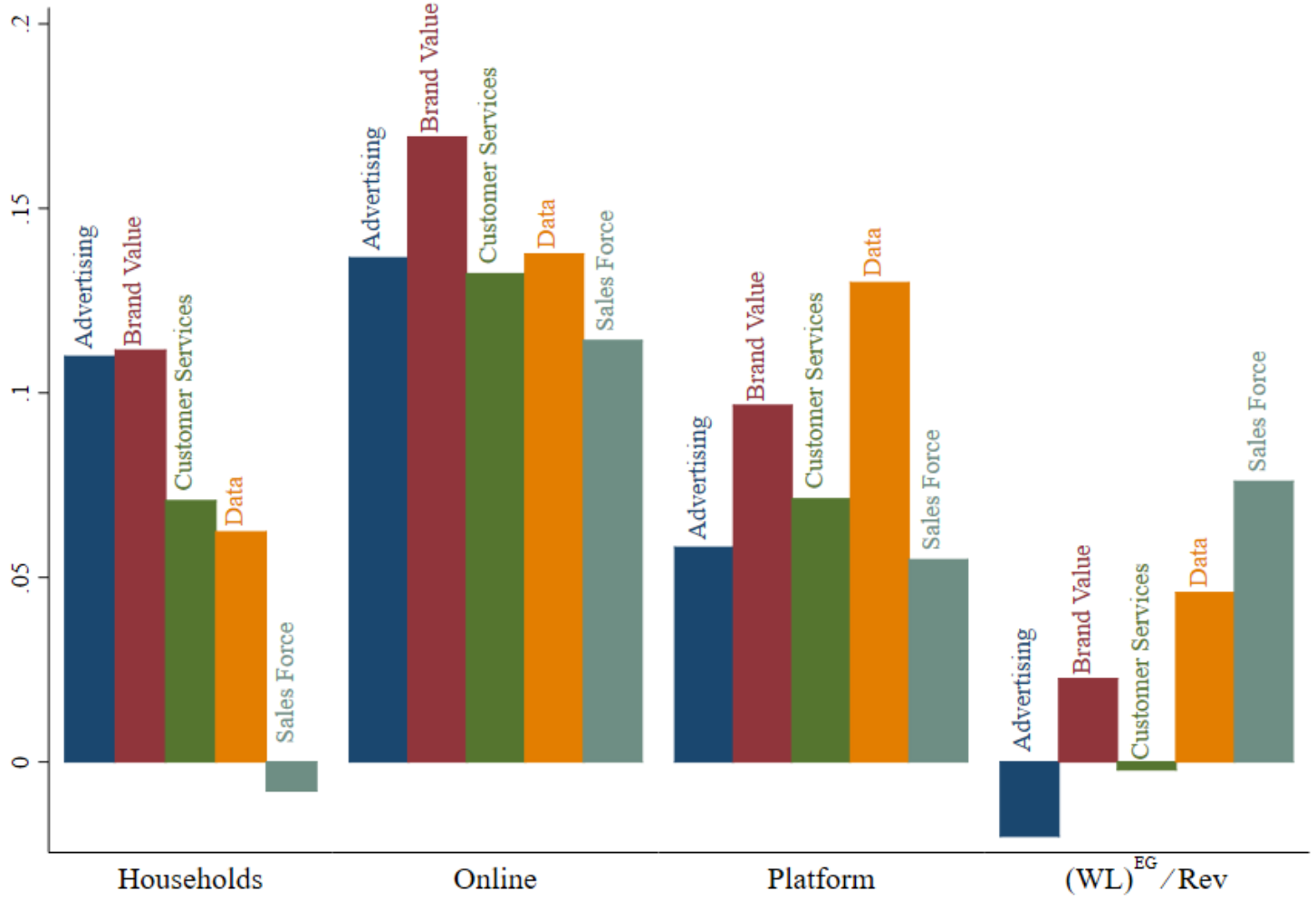
Bin-scatter plots at the industry level for the ratio of sales and marketing expense to revenue (I^{SM}/Rev) against the fraction of companies in each industry that primarily market their products to households (left panel) and that sell predominantly online (right panel), as identified by Gemini through textual analysis of firms' 10-Ks (see Appendix Table A1 for the prompts used). Confidence bars are at the 90% level. See Appendix Figure A7 for full industry scatter plots.

Figure 10: Explaining I^{SM}/Rev (continued)



Bin-scatter plots at the industry level for the ratio of sales and marketing expense to revenue (I^{SM}/Rev). The left panel plots this against the fraction of companies using a platform business model in each industry, as identified by Gemini through textual analysis of firms' 10-Ks (see Appendix Table A1 for the prompts used). The right panel shows the relationship with salaries paid to engineers at the firm as a share of revenue ($(WL)^{SM}/Rev$). Confidence bars are at the 90% level. See Appendix Figure A8 for full industry scatter plots.

Figure 11: Explaining Sales and Marketing Strategies



Each bar in this graph represents how much a one-standard deviation change in the covariate on the horizontal axis affects the sales and marketing strategy listed at the top of each bar. For example, a one standard deviation increase in the share of firms in an industry selling to households leads to an 11 percentage point increase in that industry's share of firms emphasizing advertising as a strategy.

Appendix

A Measuring SG&A

Compustat and Capital IQ measure SG&A differently. A regression of one on the other yields an R^2 of 0.89. As a general rule, the mapping from the information on the 10-K filing to the Capital IQ data is more transparent than the mapping to Compustat. For this reason, this study uses the Capital IQ measure of SG&A. Two adjustments to the Compustat measure of SG&A explain the discrepancy. As is well known, Compustat's measure of SG&A includes R&D expenses; Capital IQ does not include R&D in SG&A. In addition, if the firm separately reports a line item called "general and administrative" expenses, it appears that Compustat reclassifies these expenses into COGS. Several other researchers have noted the lack of clarity in the mapping of 10-K information to the actual variables in Compustat as well. See, for example, page 271 of the appendix of Peters and Taylor (2017) in which there is a detailed discussion on the difficulties in separating R&D expense from SG&A expense in Compustat. As a general rule, it is difficult to ascertain what exactly is included in the *xsga* variable in Compustat. Capital IQ provides a transparent definition of SG&A, and the Capital IQ interface contains hyper-links that allow one to see what information in the underlying 10-K filing generates the value recorded in Capital IQ. In particular, the definition of SG&A in Capital IQ includes: Equipment expense, salaries and other employee benefits, occupancy expense, insurance expenses, stock-based compensation (some of which is also included in sales and marketing expense based on the position of the employee being compensated), net rental expense, selling and marketing expense, and general and administrative expense.

B Text Data Collection and Processing

This section describes the process employed in obtaining quantitative data from the text of SEC 10-K filings. Section B.1 describes the manual phase of data collection, which generated a few insights that guide our systematic processing of the text using Gemini. Section B.2 outlines the tasks we pose to Gemini to collect these variables from the text, and discusses checks for performance quality.

B.1 Manual Data Collection

We begin by manually consulting the qualitative information contained in the text of firms' annual filings for a random sample of 150 firm-years dating back to 1997. We initially search the filings for mentions of "sales," "marketing," "advertising," or "promotions," which generates a few useful insights that we use to design our approach to systematically processing these texts using the Gemini LLM.

First, the company's sales and marketing strategy is most prominently discussed in the Item 1 Business Description, and to a lesser extent in Item 7 Management Discussion and Analysis. It is rarely discussed elsewhere in the filing, enabling us to restrict attention to these parts of the text when working with the full sample. Item 1 frequently contains a subsection that discusses the company's strategy for acquiring and retaining customers, often titled "Sales and Marketing" or

"Marketing Strategy." Below is an excerpt from the 2001 10-K filing for Gottschalks, Inc., a chain of department stores and specialty apparel retailers.

> **Marketing Strategy**
>
> The Company's marketing strategy is based on a multi-media approach, using newspapers, television, radio, direct mail and catalogs to highlight seasonal promotions, selected brand-name merchandise and frequent storewide sales events. Advertising efforts are focused on communicating branded merchandise offered by the Company, and the high levels of quality, value and customer service available in the Company's stores. In its efforts to improve the effectiveness of its advertising expenditures, the Company uses data captured through its proprietary credit card to develop segmented advertising and promotional events targeted at specific customers who have established purchasing patterns for certain brands, departments or store locations.
>
> The Company's sales promotion strategy also focuses on special events such as fashion shows, bridal shows and wardrobing seminars in its stores and in the communities in which they are located to convey fashion trends to its customers. The Company receives reimbursement for certain of its promotional activities from some of its vendors.

Moreover, the presence of this subsection is common across a wide variety of industries, not only for firms that produce consumer products. Below is an excerpt from the 2017 10-K filing for Iteris, Inc., a producer of sensors that markets its products to government agencies and other firms.

> **Sales and Marketing**
>
> We currently sell our Roadway Sensors products through both direct and indirect sales channels. In the territories where we sell direct, we use a combination of our own sales personnel and outside sales organizations to sell, oversee installations and set-up issues, and support our products. Our indirect sales channel is comprised of a network of independent distributors in the U.S. and select international locations, which sell integrated systems and related products to the traffic management market. In the fourth quarter of our fiscal year ended March 31, 2018 ("Fiscal 2018"), we entered into a distribution agreement to expand our northern European sales coverage in the U.K. and Ireland. Our independent distributors are trained in, and primarily responsible for, sales, installation, set-up and support of our products, maintain an inventory of demonstration traffic products from various manufacturers, and sell directly to government agencies and installation contractors. These distributors often have long-term arrangements with local government agencies in their respective territories for the supply of various products for the construction and renovation of traffic intersections, and are generally well-known suppliers of various high-quality ITS products to the traffic management market. We periodically hold technical training classes for our distributors and end users, and maintain a full-time staff of customer support technicians throughout the U.S. to provide technical assistance when needed.

Here is an excerpt from the 2022 10-K filings of Chegg, Inc., a software company focusing on an education platform.

Sales and Marketing

Students Our direct to consumer marketing strategy focuses on brand and performance marketing. Brand marketing increases awareness of the Chegg brand and its services while performance marketing drives traffic to our site. We use several major direct marketing channels to reach students. The strength of our content flywheel drives significant organic traffic to Chegg. Our lifecycle marketing focuses on increasing activation, engagement and retention. We utilize three types of email marketing campaigns: onboarding programs to drive activation and retention, personalized cross-sell campaigns to deepen engagement, and promotional campaigns to drive sales and interests.

Brands We secure contracts with brands through direct sales by our field sales organization, which sells brand advertising services to large brand advertisers seeking to reach and engage college and high school students. This team has field sales people and marketing support.

Student Advocacy We are committed to providing a high level of customer service to our students and to fulfilling our brand promise of putting students first. We trust our students, understand the critical role our products and services have in their learning journey, and strive to resolve all problems quickly and thoroughly. Our student advocacy team can be reached directly through phone, email, and online chat during business hours. We also proactively monitor social media to identify and solve problems before we are otherwise informed of their existence. We endeavor to respond to students' concerns within five minutes.

Item 7, Management Discussion and Analysis, sometimes describes the company's sales and marketing activities that are included in either the advertising expense or selling, general, and administrative expense. One such example is this excerpt from the 2005 filing for Inventure Foods, Inc., a snack food manufacturer:

Critical Accounting Policies and Estimates

Advertising, Promotional Expenses and Trade Spending. The Company expenses production costs of advertising the first time the advertising takes place, except for cooperative advertising costs which are expensed when the related sales are recognized. Costs associated with obtaining shelf space (i.e. "slotting fees") are expensed in the period in which such costs are incurred by the Company. Anytime the Company offers consideration (cash or credit) as a trade advertising or promotion allowance to a purchaser of products at any point along the distribution chain, the amount is accrued and recorded as a reduction in revenue. Any marketing programs that deal directly with the consumer are recorded in selling, general and administrative expenses.

Second, we use the excerpts collected for this random sample of filings to assemble a taxonomy of the primary types of activities that companies identify as "sales and marketing," or core to their strategy to retain and acquire customers. The most common activities discussed in the text are advertising, brand value, customer service, customer data, and employing a sales force. We use this taxonomy later in constructing the list of questions to ask Gemini about the text.

Third, manual readings reveal that a common reason firms describe their internal sales force as being so crucial to the success of the business is the technological sophistication of their products. An example of this language from the 2012 filing for Kopin Corp. is included in Section 5. Below is another example, from the 2012 10-K filing for Atmel Corp., a semiconductor manufacturer.

> In addition, new product introductions frequently depend on our development and implementation of new process technologies, and our future growth will depend in part upon the successful development and market acceptance of these process technologies. Our integrated solution products require more technically sophisticated sales and marketing personnel to market these products successfully to customers. We are developing new products with smaller feature sizes and increased functionality, the fabrication of which will be substantially more complex than fabrication of our current products.

Finally we identify a set of common false positives, discussions of activities using similar words to those in the above sales and marketing descriptions, but that are semantically different. For example, it is common in the logistics industry and for oil and gas producers to use the word "marketing" to describe the process of transporting output from the production site to the distributor, or from the distributor to the customer. Another type of false positive is particular to the pharmaceutical and biotechnology industries, where firms frequently describe the regulations they face that prohibit marketing their products before they have obtained approval to do so from the Food and Drug Administration. Third, gaming companies and other online platforms frequently earn revenues from advertising, and describe this revenue source in their business description. In our assessments of Gemini's performance at reading and understanding the filings, we ensure it does not mistakenly interpret these types of text as "sales and marketing" for our purposes.

B.2 Text Processing Using Gemini

Based on our observations from reading the random sample of filings, we assemble the text for Item 1 Business Description and Item 7 Management Discussion and Analysis from the 10-K filings for each firm and year in the sample. To do this, we use the edgar-crawler GitHub repository created by Loukas et al. (2021). These codes pull the html file for each filing from the SEC Edgar API, then clean the html text and divide it into each item using regular expressions. We pull the filings directly from the SEC's API rather than using the edgar-corpus dataset on HuggingFace for two main reasons. First, the HuggingFace dataset ends in 2020, whereas our sample extends forward to 2022. Second, the GitHub repository contains updated bug fixes, which substantially improve the accuracy of splitting the full text of the filings into the constituent items.

We begin by ensuring that the LLM reads and understands the item text we give it similarly to our manual readings of the same text. In early iterations of our prompts, we describe to the model that sales and marketing activities are those that are aimed at acquiring and retaining customers. We then give it the Item 1 text from the filing, and ask the LLM to repeat back the sentences (if any) that discuss sales and marketing. We pose this task to Google's Gemini 1.5 Flash and GPT's 4o model in order to assess how well they each understand the text and the task; Gemini performs significantly better. Gemini responds with exactly the text we previously identified manually, and sometimes found additional information from the item that correctly describes sales and marketing activities but were missed in the manual reading. GPT 4o, in contrast, frequently responded with additional

text from the item that was not relevant. Neither model exhibited "hallucinations," responding with new text that did not originate from the item text we included in the prompt. Due to Gemini's superior performance in this initial step, we proceed with prompt engineering for Gemini only. We set temperature to zero in order to minimize randomness in the model's responses, and turn off all safety features in order to ensure that no relevant content is mistakenly omitted.

We opt to proceed with prompt engineering only, rather than employing retrieval-augmented generation (RAG) or fine-tuning techniques. After only a few attempts at improving our prompts, the model performs extremely well at each task we pose. Since RAG or fine-tuning would therefore yield only modest improvements in performance at the price of replication, we forego them.

In constructing our text-based variables for the full sample, we use the following system prompt for all requests: "You are a marketing consultant, specialized in reading SEC 10-K filings and understanding how firms conduct sales, marketing, and other activities associated with retaining and acquiring customers." We then ask the LLM thirteen questions in separate requests. We ask each question in a separate request so that the model's answer to each question is not contaminated by the other questions we ask, or by its own responses to the other questions. The exact text of each task prompt, which item text we include in the prompt, and the economic concept of interest are included in Table A1. Consistent with the recommendations in Eisfeldt and Schubert (2024), we provide the model with a rubric, describing in detail what kinds of activities we consider to be "sales and marketing," and what we mean by terms such as "platforms," "subscriptions," "promotions," and others.

Our questions fall into three main categories: intangible investment intensity, types of sales and marketing activities, and expected determinants of sales and marketing investments. Our questions regarding sales and marketing intensity and R&D intensity are aimed at imputing zeroes for a subset of the firm-years that do not report these investments. While no firms report these investments to be 0, it is clear from the text of the filing that some firms that do not report this number as a separate line item still engage in sales and marketing or R&D investments, so not all missing values can sensibly be interpreted as zero. We ask the model to read item 1 and item 7 separately and classify the intensity of sales and marketing or R&D investment into one of three categories: minimal, moderate, or substantial. If the model determines that sales and marketing or R&D investment is minimal based on both item 1 and item 7 independently, we impute that line item as zero.

There are a few tests that are conducted to ensure the sensibility of Gemini's responses. One such test is to compare Gemini's responses based on whether the firm reports sales and marketing as a non-zero expense. As a fraction of the total sample, only 1% of firm-year observations report sales and marketing as a non-zero expense and have Gemini classifying their sales and marketing expense to be minimal based on its reading of the text. It is extremely rare for a firm to report non-zero sales and marketing expenses as a line item and Gemini to classify their spending as minimal. However, conditional on Gemini reporting substantial spending on sales and marketing expense based on its reading of both Item 1 and Item 7, 10% of firm-year observations have the income statement line-item for sales and marketing expense missing. This suggests that it would be a mistake to classify sales and marketing expense as zero if it is not detailed in the income statement.

We conducted another test that urges caution on imputing zeros for missing values of sales and marketing expense. This test exploits the availability of the Revelio Labs data for both firms with and without sales and marketing expense reported. In particular, the Revelio Labs data allows us to measure the salaries paid to sales and marketing professionals for firms with and without a line item reported for sales and marketing expenses. Using these data, we can measure the industry-

level median salaries paid to sales and marketing workers to revenue ratio for both firms with and without sales and marketing expense reported in the income statement. The two measures are highly correlated across industries. In other words, the firms that do not report sales and marketing expense on their income statement have similar salaries paid to sales and marketing employees as firms that do report that are in the same industry. This test also helps show the robustness of the industry-level variation in sales and marketing expense.

We also ask the model to determine which types of sales and marketing activities the firm engages in based on its reading of the item 1 text. Based on the taxonomy constructed from the manual reading, we ask the model about whether the following types of activities are core to the firm's sales and marketing strategy: advertising, building the value of the brand, acquisition and use of customer data, customer service, and an internal sales force.

It is well-known that LLMs are trained on an immense corpus of information, and therefore Gemini could be answering our questions using knowledge outside of the text we provide in the prompt, such as the reported sales and marketing expense amount listed elsewhere in the filing. We think the risk of this is minimal for a few reasons. First, Gemini occasionally responds that the text does not contain enough information to answer the question. If the model were easily able to reference outside information, it would likely be able to infer from other sources whether, for example, the firm's primary customers are households, other businesses, or the government. Since it sometimes cannot find an answer to the question, this suggests the model likely restricts attention to the text provided rather than consulting all sources it theoretically can access. Second, we do not explicitly tell Gemini what firm and year the filing text came from, so this information would have to be inferred from the text by the model, and it would then further have to reference additional information for this issue to be of practical significance. We expect the probability of it doing this to be low, relative to answering the question based on the text provided.

Third, we conduct an additional test, in which we compare Gemini's response to our question with Gemini's response when asked to explain its reasoning. When Gemini outlines its reasoning for these answers, it never cites information such as the firm's industry, additional variables from elsewhere in the filing, news sources, or other outside information. On rare occasions, the model demonstrates making inferences from the text in order to answer the question. For example, when justifying its reasoning for why sales and marketing expenses appear to be "minimal" for George Foreman Enterprises, Inc., part of Gemini's reasoning cited the following logic: "The 10-K highlights the company's financial difficulties, including a severe cash shortage and default on promissory notes. This suggests limited resources available for marketing investments." Again, these instances are rare, and it is more likely that the model makes these inferences when asked to explain its reasoning, relative to answering the question without explaining.

It should be noted that LLMs sometimes deliver different answers when asked a question outright and when asked to explain its reasoning. In order to minimize the occurrence of such disagreements, we asked the model to explain its reasoning *after* providing the answer; this leaves little room for the model to "think" too hard about the question, and potentially generate the unwanted logical inferences described above. This strategy appears to have been effective, as Gemini's answers when asked to explain its reasoning deviated from the original answer only in rare cases. We cannot fully rule out the possibility that our text-based variables are mechanically correlated with other measures included in the regressions, but we believe the risk of this contamination is low.

C Customer Capital from Purchase Price Allocation

The purchase price allocation dataset comes from Business Valuation Resources' (BVR) DealStats database, which tracks M&A transaction records. BVR collects information on transactions related to public firms from SEC filings, including 10-K, 10-Q, 8-K(A), S-1, and S-4(A), and private firm transactions from various national and regional brokerage associations. BVR employs a team of financial analysts to verify the database's accuracy. For more information on the database and purchase price allocation, refer to He (2022). This section focuses on the extraction process for customer relationship intangible valuations, the accounting methods used for their valuation, and how estimates of useful life can affect depreciation rate calculations.

In PPA, the types of assets and their corresponding valuations for the target firm are reported.[37] For example, in the 8-K/A filing of Men's Wearhouse, Inc. on September 2, 2014, detailing its acquisition of Jos. A. Bank Clothiers, Inc., the trade name of Jos. A. Bank Clothiers was valued at $539.1 million, and customer relationships were valued at $53 million. We extract textual information using regular expressions—"tradename: $539.1 million" and "customer relationship: $53 million." In this case, no other customer-related intangible assets were recorded, so the total customer capital for this deal amounted to $592.1 million.

We also consider other potential terms related to customer capital; for example, for customer relationships, we include valuation associated with "customer", "client", "loyalty program", "user base", "customer base", and "membership". For a detailed list of items we include in our analysis, refer to Table A7. Customer capital includes "customer relationship", "customer list", "customer contract", "trademark/trade names", "brands", "business relationship", and "domain".

The fundamental idea behind valuing these assets is to assess the firm's value with and without those assets. One method to achieve this is by projecting the future cash flows generated by the customer capital and discounting them to present value. In practice, accountants have developed several commonly used methods for this purpose.

The most common approach to valuing customer relationships is the Multi-Period Excess Earnings Method (MPEEM), which focuses on estimating future revenue and earnings specifically attributable to these relationships. This method uses a discounted cash flow analysis to estimate the present value of future cash flows attributable to each customer segment, factoring in customer attrition rates and excess earnings generated by these customers. The projected cash flows are then discounted to present value. Critical factors in valuing customer capital include customer retention rates and the churn rate (i.e., the rate customers leave).

Another valuation method for assets like trademarks and domain names is the Relief from Royalty Method (RRM). This method is ideal for assets tied to specific revenue streams, where data on royalty fees from comparable market transactions are available. The RRM calculates value based on the hypothetical royalty payments saved by owning the asset rather than licensing it. The logic behind this approach is that owning an intangible asset allows the entity to avoid paying royalties to use that asset.

The main analysis uses estimates of customer capital depreciation based on the useful life estimates of customer-related intangibles from the PPA data. An alternative method is to estimate rates of customer churn. Gourio and Rudanko (2014) cite a number of industry assumptions on the turnover in the customer base to justify a depreciation rate of 15% for customer capital. We sepa-

[37] Goodwill, including synergies, is reported separately.

rately collect data from industry reports, and we also collect rates of customer churn from Baker et al. (2023). Generally, the rates of customer churn are significantly higher, in the 30 to 35% range.

Our preference for using the PPA information to estimate customer capital depreciation is three-fold. First, a disadvantage of using the rates of customer churn from industry reports or from Baker et al. (2023) are that they are only available for a sub-set of the industries in the sample. The PPA data set is large enough to allow industry-level estimates of depreciation for all 2-digit NAICS codes int the sample. Second, the industry-level reports do not contain sufficient detail as to the inputs that justify the ultimate depreciation rates; in addition, while the Baker et al. (2023) estimates of customer churn are a major step forward in the literature, they are based on consumer-level data obtained through credit card transactions. These data cannot be used to estimate depreciation rates on customer capital for business-to-business-focused companies. Finally, the depreciation rate calculated from useful life estimates of actual customer-related assets is the closest conceptually to a true depreciation rate on customer capital, at least in our opinion.

Given the large level difference between the PPA estimates and the customer churn estimates, we conduct robustness exercise for Table 5 where we assume a depreciation rate of 30% across all industries. This is reported in Appendix Table A5. Not surprisingly, this leads to a significantly smaller amount of the book value of customer capital. However, the industry-level variation remains similar qualitatively, and the core results using the capitalized book value are robust.

D A static model of sales and marketing

Following the model in the appendix of Bond, Hashemi, Kaplan and Zoch (2021), this appendix section presents a model in which sales and marketing expense is an input into the generation of firm revenue that works through shifting demand. Suppose a firm produces output Q using a single flexible production input X^Q.

$$Q = \mathcal{F}(X^Q)$$

where $\mathcal{F} : \mathbb{R}_+ \to \mathbb{R}_+$ and is twice continuously differentiable. D is a demand shifter that the firm can influence through a demand-shifting input, X^D.

$$D = \mathcal{D}(X^D)$$

The firm's revenue function is given by

$$\mathcal{R} \equiv \mathcal{P}(Q, D)Q$$

where the firm's revenue $\mathcal{R}$ depends on the price $\mathcal{P}(Q, D)$, which is a function of both the quantity produced Q and the demand shifter D.

The input prices of X^Q and X^D are denoted W^Q and W^D, respectively, and they are taken as given. The firm faces the following profit maximization problem:

$$\Pi = \max_{Q,D} \mathcal{P}(Q, D)Q - C_Q(Q; W^Q) - C_D(D; W^D)$$

where $C_Q(\cdot)$ is the firm's cost function for producing output, defined by

$$C_Q(Q; W^Q) = \min_{X^Q} W^Q X^Q$$
$$\text{s.t. } Q \leq \mathcal{F}(X^Q)$$

and $C_D(\cdot)$ is the firm's cost function for shifting demand, defined by

$$C_D(D; W^D) = \min_{X^D} W^D X^D$$
$$\text{s.t. } D \leq \mathcal{D}(X^D)$$

Let λ denote the Lagrange multiplier on the demand-shifting constraint. By the envelope theorem, this quantity captures the marginal cost of influencing demand.

$$\lambda = \frac{\partial C_D(\cdot)}{\partial D} \tag{7}$$

Likewise let ζ denote the Lagrange multiplier on the output constraint, which captures the marginal cost of production.

$$\zeta = \frac{\partial C_Q(\cdot)}{\partial Q}$$

From the cost minimization first order conditions, we have

$$W^D = \lambda \frac{\partial \mathcal{D}(\cdot)}{\partial X^D} \tag{8}$$

$$W^Q = \zeta \frac{\partial \mathcal{F}(\cdot)}{\partial X^Q} \tag{9}$$

Profit maximization implies

$$\frac{\partial \mathcal{P}(\cdot)}{\partial D} Q = \frac{\partial C_D(\cdot)}{\partial D} \tag{10}$$

$$\frac{1}{P} \frac{\partial C_Q(\cdot)}{\partial Q} = 1 - \eta^{-1}$$

Where η is the absolute value of the price elasticity of demand ($|\frac{P}{Q}\frac{\partial Q}{\partial P}|$). Let ρ denote the elasticity of revenue with respect to the demand shifter D.

$$\rho \equiv \frac{D}{PQ} \frac{\partial \mathcal{R}(\cdot)}{\partial D}$$

Using (10), this implies

$$\rho = \frac{D}{PQ} \frac{\partial C_D(\cdot)}{\partial D} \tag{11}$$

Let θ^D denote the elasticity of D with respect to the demand-shifting variable input X^D.

$$\theta^D \equiv \frac{X^D}{D}\frac{\partial \mathcal{D}(\cdot)}{\partial X^D} \tag{12}$$

Finally, let the share of revenue paid to the demand shifting input be given by

$$\alpha^D \equiv \frac{W^D X^D}{PQ}$$

Using (7), (8), and (10), we can rewrite the above as

$$\alpha^D = \frac{X^D}{PQ}\frac{\partial \mathcal{P}(\cdot)}{\partial D}Q\frac{\partial \mathcal{D}(\cdot)}{\partial X^D}$$

Therefore, using (11) and (12) we can rewrite the share of revenue paid to the demand-shifting input as the product of two elasticities: the elasticity of revenue with respect to the demand shifter, and the elasticity of the demand shifter with respect to the demand-shifting variable input.

$$\alpha^D = \rho\theta^D$$

where ρ is the elasticity of revenue with respect to demand ($\frac{D}{PQ}\frac{\partial \mathcal{R}(\cdot)}{\partial D}$), and θ^D is the elasticity of demand with respect to the variable demand-shifting input ($\frac{X^D}{D}\frac{\partial \mathcal{D}(\cdot)}{\partial D}$). The equilibrium share of sales and marketing expense to revenue is equal to fundamental parameters of the revenue function: ρ and θ^D. It is therefore natural to use the variable ${}^{I^{SM}}/_{Rev}$ as the key measure of a firm's spending on sales and marketing.

E Appendix Tables and Figures

Table A1: Questions Used for 10-K Analysis

10-K Item	Concept of Interest	Question
Item 1, 7	Sales & marketing intensity	"We are economists conducting research on the spending done by firms on sales and marketing. Your task is to read the following document and determine the extent to which the firm spends resources on marketing, advertising, product promotion, branding, customer service, sales force, and other closely related activities. Based on your reading of the document, please use your best judgement to classify the extent of their spending on such activities into one of three categories: minimal, moderate, or substantial. Please limit your answer to one word from the following three: minimal, moderate, or substantial. Here is the document:"
Item 1	Brand value	"We are economists conducting research on the spending done by firms on sales and marketing. Your task is to read the following document and determine the specific factors that the firm spends resources on in their sales and marketing strategy. Based on your reading of the document, please use your best judgement to answer the following question: Is an emphasis on increasing brand value an important element in the firm's sales and marketing strategy? Please provide an answer that is only a single word, either yes or no. Here is the document:"
Item 1	Sales force	"We are economists conducting research on the spending done by firms on sales and marketing. Your task is to read the following document and determine the specific factors that the firm spends resources on in their sales and marketing strategy. Based on your reading of the document, please use your best judgement to answer the following question: Is an emphasis on a sales force or a sales staff an important element in the firm's sales and marketing strategy? Please provide an answer that is only a single word, either yes or no. Here is the document:"

Continued on next page

Table A1 – continued from previous page

10-K Item	Concept of Interest	Question
Item 1	Advertising	"We are economists conducting research on the spending done by firms on sales and marketing. Your task is to read the following document and determine the specific factors that the firm spends resources on in their sales and marketing strategy. Based on your reading of the document, please use your best judgement to answer the following question: Is an emphasis on advertising an important element in the firm's sales and marketing strategy? Please provide an answer that is only a single word, either yes or no. Here is the document:"
Item 1	Customer data usage	"We are economists conducting research on the spending done by firms on sales and marketing. Your task is to read the following document and determine the specific factors that the firm spends resources on in their sales and marketing strategy. Based on your reading of the document, please use your best judgement to answer the following question: Is an emphasis on obtaining and using customer data an important element in the firm's sales and marketing strategy? Please provide an answer that is only a single word, either yes or no. Here is the document:"
Item 1	Customer service	"We are economists conducting research on the spending done by firms on sales and marketing. Your task is to read the following document and determine the specific factors that the firm spends resources on in their sales and marketing strategy. Based on your reading of the document, please use your best judgement to answer the following question: Is an emphasis on customer service an important element in the firm's sales and marketing strategy? Please provide an answer that is only a single word, either yes or no. Here is the document:"

Continued on next page

Table A1 – continued from previous page

10-K Item	Concept of Interest	Question
Item 1	Platform business model	"We are economists conducting research on the underlying business models used by firms. One business model involves building a platform on which individuals or other entities interact. A platform business model involves profiting from a platform that allows two or more groups of users to interact. Your task is to read the following document and answer the following question: Is such a platform part of the business model of the firm? Please provide an answer that is only a single word, either yes or no. Here is the document:"
Item 1	Online/digital sales	"We are economists conducting research on the underlying business models used by firms. A particular issue in which we are interested is how companies reach their customers and generate sales through online or digital avenues. Your task is to read the following document and answer the following question: Does the firm generate revenue by selling to its customers through online or digital avenues? Please provide an answer that is only a single word, either yes or no. Here is the document:"
Item 1	Customers	"We are economists conducting research on the spending done by firms on sales and marketing. Your task is to read the following document and determine the extent to which the firm spends resources on marketing, advertising, product promotion, branding, customer service, sales force, and other closely related activities. Your task is to read the following document and determine the primary customers of the firm in question. Specifically, does the firm primarily market its products to households, businesses, or the government? Please provide an answer that is only a single word: households, businesses, or the government. Here is the document:"

Continued on next page

Table A1 – continued from previous page

10-K Item	Concept of Interest	Question
Item 1, 7	R&D intensity	"We are economists conducting research on the spending done by firms on research and development. Your task is to read the following document and determine the extent to which the firm spends resources on research and development and other closely related activities. Based on your reading of the document, please use your best judgement to classify the extent of their spending on such activities into one of three categories: minimal, moderate, or substantial. Please limit your answer to one word from the following three: minimal, moderate, or substantial. Here is the document:"
Items 1, 7 (combined)	Cost vs. Investment	"Firms frequently describe some of their operations as 'sales and marketing'. We are economists trying to determine whether firms conduct sales and marketing activities as a cost of doing business, where they have to spend on sales and marketing to make each sale, or as an investment in long-lived customer relationships that prove valuable to the company over time. Please read the following document and tell us whether the firm describes their sales and marketing expenses as costs of making each sale or as an investment in building and maintaining a customer base, which retains some of its value over time like a type of asset. Please provide an answer that is only a single word, either 'cost' or 'investment'. Here is the document:"

Table A2: Industries, 3-Digit NAICS, and Median $I^{SM}/_{Rev}$

		$I^{SM}/_{Rev}$, median
Agri, Mining, Const, Utilities		
111	Crop Production	0.002
211	Oil & Gas Extraction	0.000
212	Mining (except Oil & Gas)	0.000
213	Mining Support Activities	0.000
221	Utilities	0.000
236	Constr of Buildings	0.012
237	Heavy & Civil Eng Constr	0.001
238	Specialty Trade Contractors	0.002
Manufacturing		
311	Food Mfg	0.049
312	Bev & Tobacco Prod Mfg	0.122
313	Textile Mills	0.011
315	Apparel Mfg	0.047
316	Leather Product Mfg	0.057
321	Wood Product Mfg	0.009
322	Paper Mfg	0.026
323	Printing Support Activities	0.050
324	Petrol & Coal Prod Mfg	0.000
325	Chemical Mfg	0.006
326	Plastics & Rubber Prod Mfg	0.023
327	Mineral Product Mfg	0.012
331	Primary Metal Mfg	0.002
332	Fabricated Metal Prod Mfg	0.028
333	Machinery Mfg	0.020
334	Computer & Electronic Mfg	0.073
335	Electrical Equip Mfg	0.042
336	Transp Equipment Mfg	0.012
337	Furniture Mfg	0.054
339	Medical Equip Mfg	0.111
Wholesalers & Retail		
423	Durable Goods Whslrs	0.008
424	Nondurable Goods Whslrs	0.013
441	Motor Vehicle Retail	0.009
444	Building Material Retail	0.016
445	Food & Bev Retail	0.006
449	Appliances & Elec Retail	0.040
454	Nonstore Retail	0.122
455	General Merch Retail	0.021
456	HealthCare Retail	0.017
457	Gas Stations & Fuel Dealers	0.003
458	Clothing & Accs Retail	0.031
459	Misc Retail	0.027
Transportation		
481	Air Transp	0.037
482	Rail Transp	0.000
483	Water Transp	0.000
484	Truck Transp	0.003
485	Ground Passenger Transit	0.173
486	Pipeline Transp	0.000
488	Transp Support Svcs	0.006
492	Couriers & Messengers	0.008
Telecom & Info Services		
512	Film & Sound Recording	0.030
513	Publishing Industries	0.247
516	Broadcasting & Media	0.030
517	Telecom	0.044
518	Data Processing & Hosting	0.181
519	Other Information Svcs	0.269
Professional Services		
541	Professional Svcs	0.078
Admin & Healthcare		
561	Admin & Support Svcs	0.012
562	Waste Mgmt Svcs	0.002
611	Educational Svcs	0.134
621	HealthCare Svcs	0.017
622	Hospitals	0.001
623	Nursing Facilities	0.003
Performing Arts & Accomodation		
711	Performing Arts & Sports	0.025
713	Amusement & Recreation	0.036
721	Accommodation	0.023
722	Food Svcs & Drink Places	0.023
Maintenance & Personal Services		
811	Repair & Maintenance	0.023
812	Personal Svcs	0.140

Table A3: Additional Summary Statistics

	N	Wgt Avg	p10	p25	Median	p75	p90
Panel A. Scaled by capital variables							
I^{SM}/K	31,695	0.030	0.000	0.003	0.028	0.085	0.146
I^{RD}/K	29,510	0.025	0.000	0.000	0.014	0.075	0.154
I^{CX}/K	34,514	0.049	0.002	0.009	0.023	0.050	0.094
Q^{PH}	34,298	5.463	0.649	1.234	3.502	11.071	33.024
Panel B. All investment variables available							
I^{SM}/K	27,023	0.031	0.000	0.004	0.031	0.087	0.147
I^{RD}/K	27,023	0.024	0.000	0.000	0.015	0.074	0.149
I^{CX}/K	27,023	0.048	0.002	0.008	0.021	0.046	0.086
Q^{PH}	26,851	5.485	0.683	1.379	3.924	11.949	35.164

Table A4: Summary Statistics for Industry Level Covariates

	N	Average	Median	SD
Households	67	0.466	0.405	0.342
Online	67	0.404	0.371	0.281
Platform	67	0.137	0.074	0.158
$(WL)^{EG}/Rev$	67	0.053	0.029	0.057

Table A5: Explaining Variation in Q^{PH} across Industries, Depreciation Robustness

	Dependent variable: Q^{PH}, median			
	(1)	(2)	(3)	(4)
${}^{K^{IT}}/_{K^{PH}}, \delta^{SM} = .15$	1.19**		1.11**	
	(0.14)		(0.12)	
${}^{K^{IT}}/_{K^{PH}}, \delta^{SM} = .30$		1.43**		1.33**
		(0.17)		(0.15)
Rev g forecast, median			3.78	5.60
			(8.77)	(9.19)
Firm age, median			-1.41*	-1.07
			(0.66)	(0.66)
Constant	1.31**	1.34**	1.73*	1.51
	(0.15)	(0.15)	(0.74)	(0.77)
Observations	67	67	67	67
R^2	0.799	0.801	0.809	0.811

Table A6: Returns to Scale and Investment in Customer Capital

	Dependent variable: ${}^{I^{SM}}/_{Rev}$, median					
	(1)	(2)	(3)	(4)	(5)	(6)
RTS_{PC}	-0.233			-0.229		
	(0.238)			(0.147)		
RTS_{IC}		0.026			0.002	
		(0.071)			(0.040)	
RTS_L			0.057			-1.074
			(1.265)			(0.621)
Constant	0.270	0.016	-0.011	0.267	0.041	1.136
	(0.235)	(0.075)	(1.281)	(0.145)	(0.042)	(0.632)
Observations	49	47	23	45	44	22
R^2	0.019	0.005	0.000	0.046	0.000	0.039

This table presents regressions of ${}^{I^{SM}}/_{Rev}$ on measures of the returns to scale from McAdam et al. (2024) and Lenzu et al. (2022). RTS_{PC} is the measure of returns to scale assuming perfect competition and RTS_{IC} assumes imperfect competition.

Table A7: Customer Capital Categorization

Category	**Key Terms**
Customer relationship	customer, client, loyalty program, user base, customer base, membership
Customer list	customer list, phone number
Customer contract	customer contract, customer agreement
Trademark/trade name	trademark, masthead
Brand	name, brand, marketing related
Business relationship	business relationship, record, network, deposit intangibles
Domain	website, domain

This table lists the key terms included in the regular expressions used to extract the corresponding customer capital values.

Figure A1: Investment in Sales and Marketing and Residual SG&A

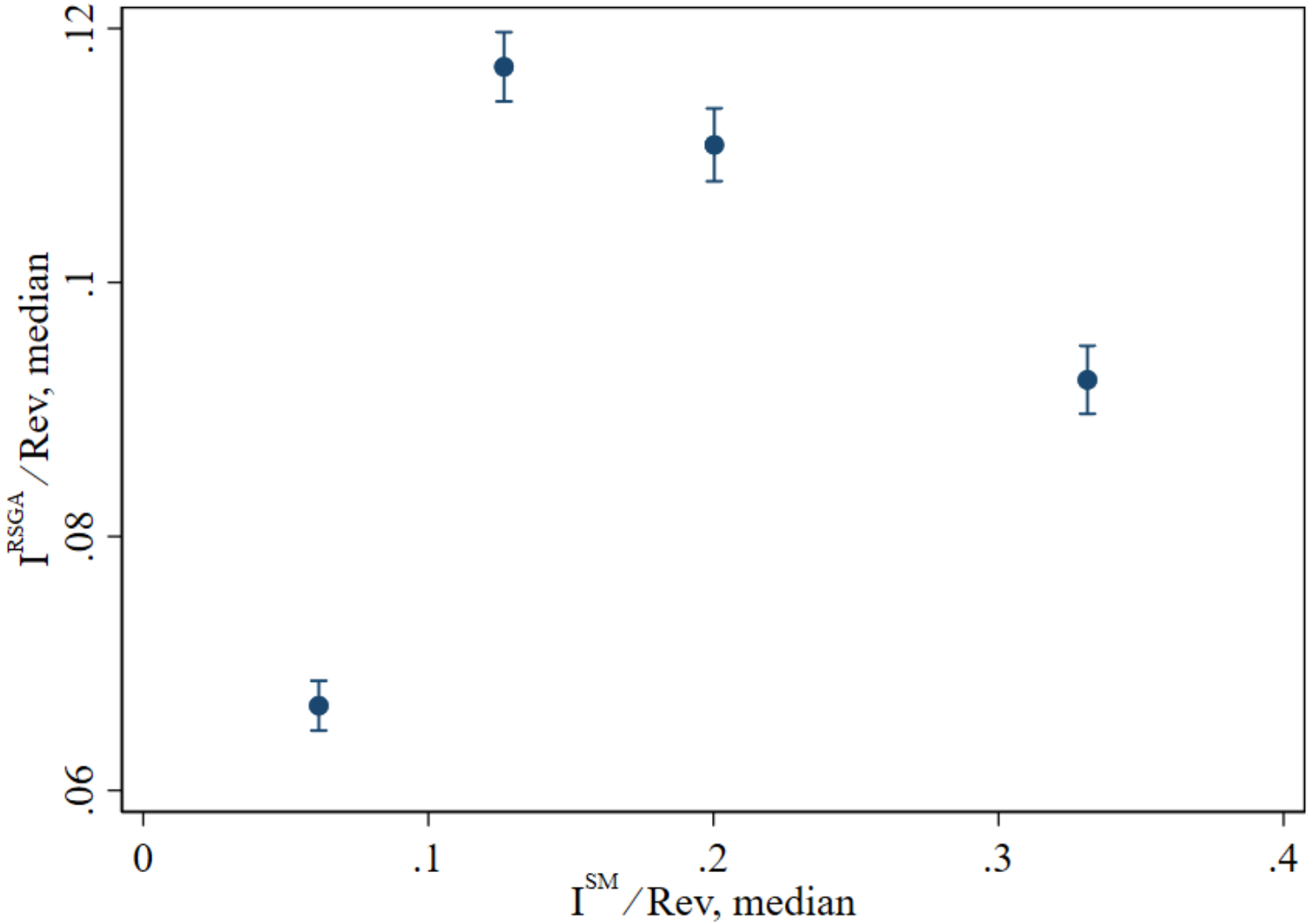


Figure A2: Sales and Marketing Investment and Salaries

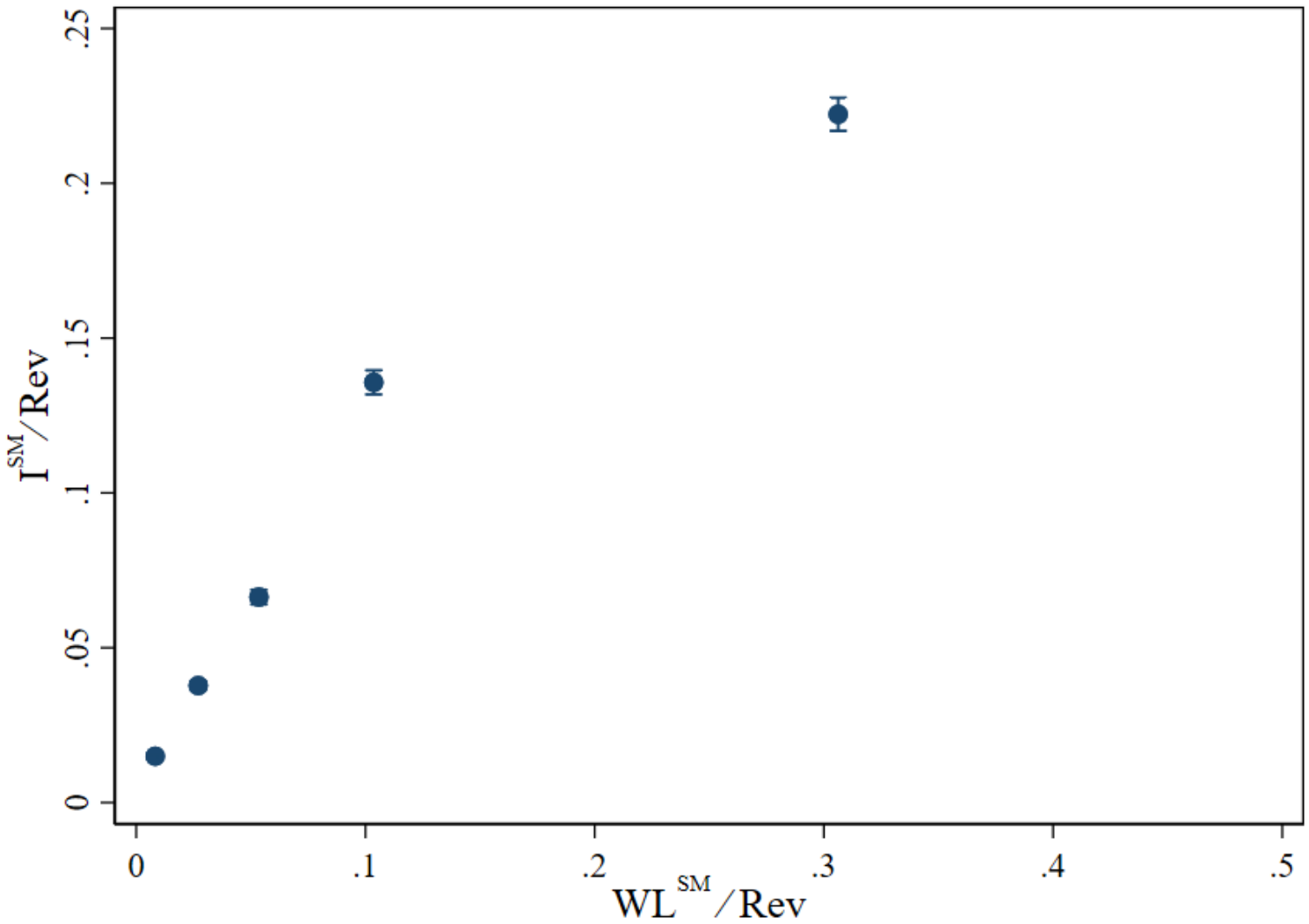


Figure A3: Sales and Marketing and Advertising at the industry level

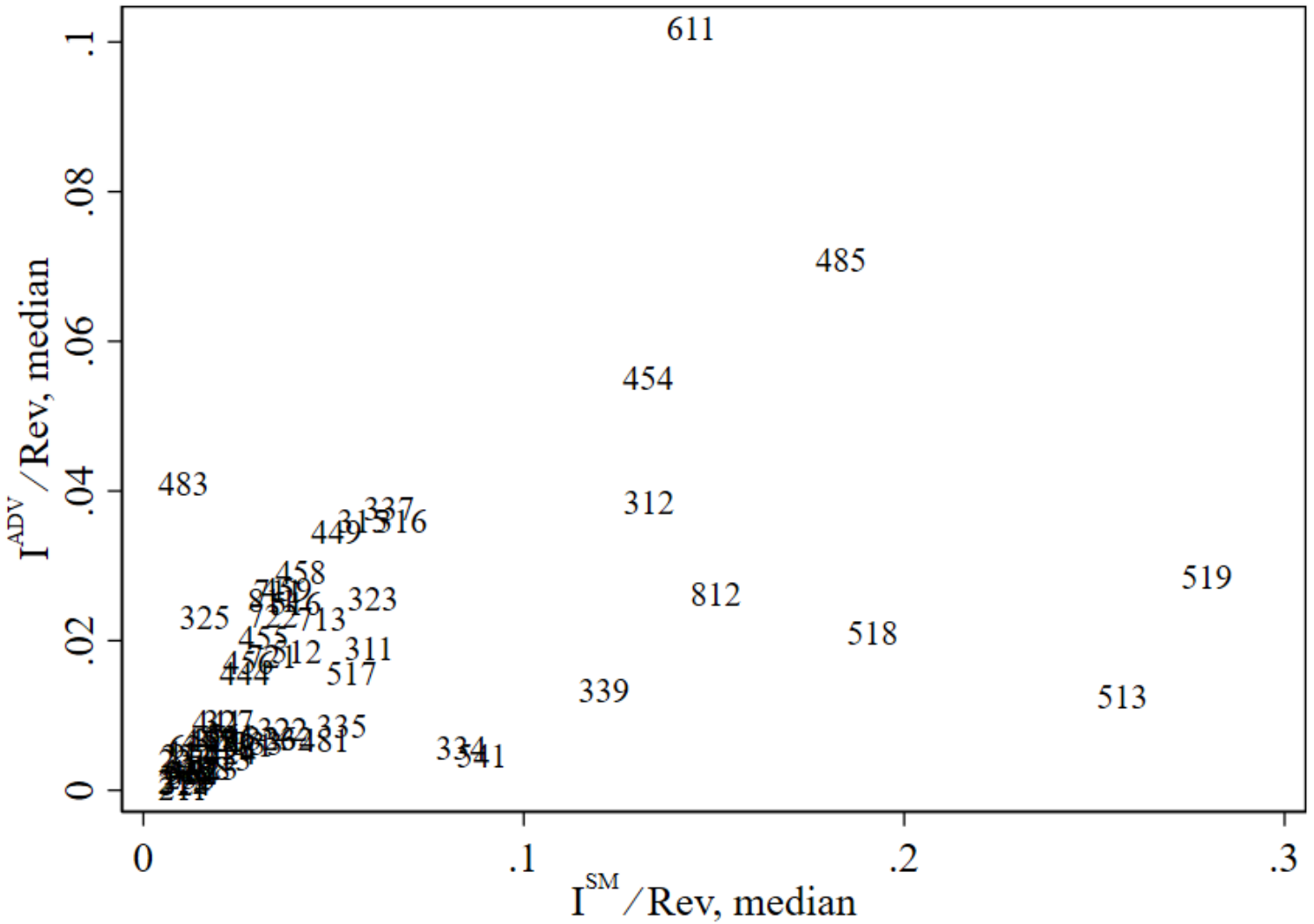


Figure A4: Comparison of I^{SM} with Capital Expenditures and R&D

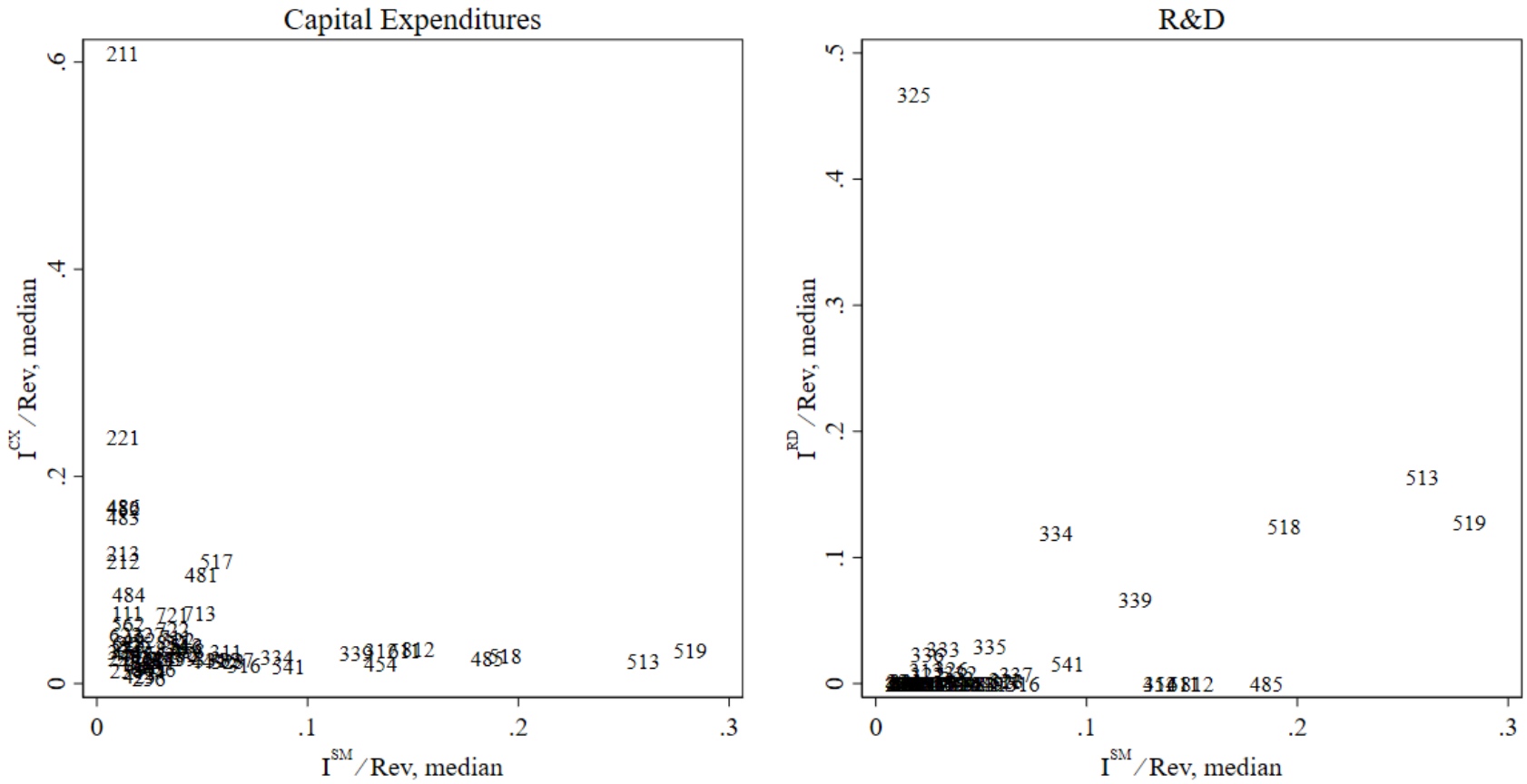


Figure A5: Explaining Q^{PH}

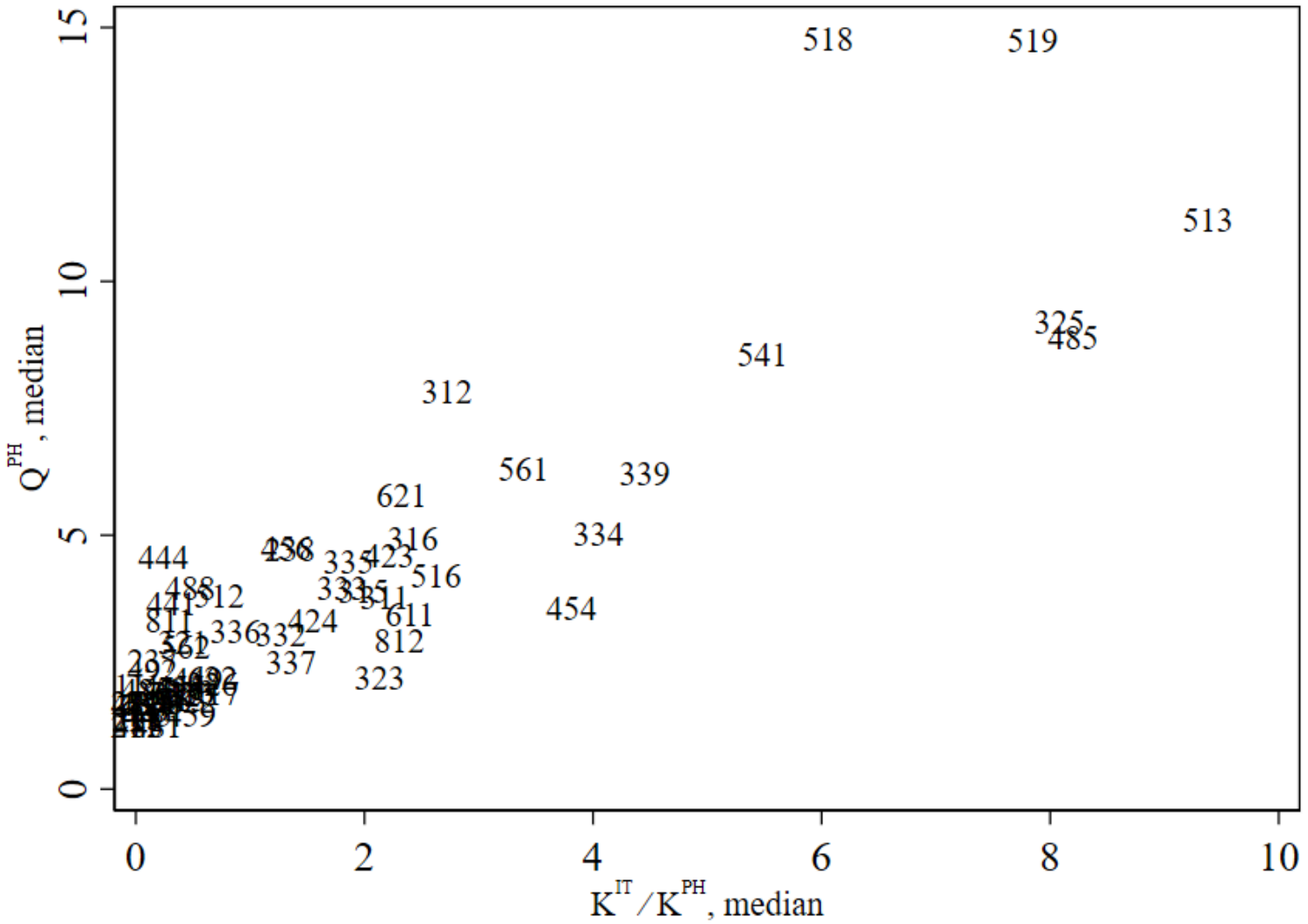


Figure A6: Is Sales and Marketing an Investment or Cost?

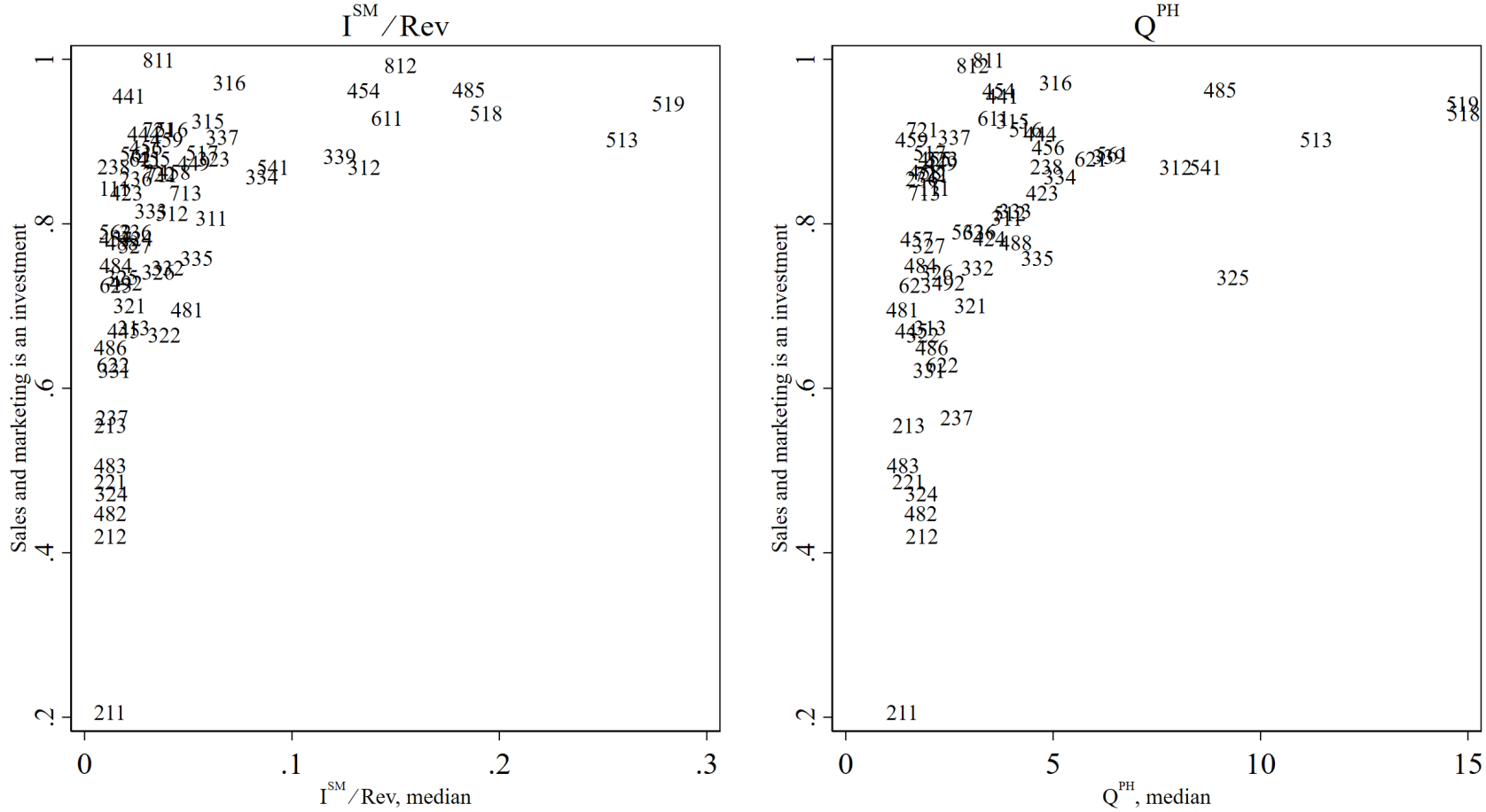


Figure A7: Explaining I^{SM}/Rev

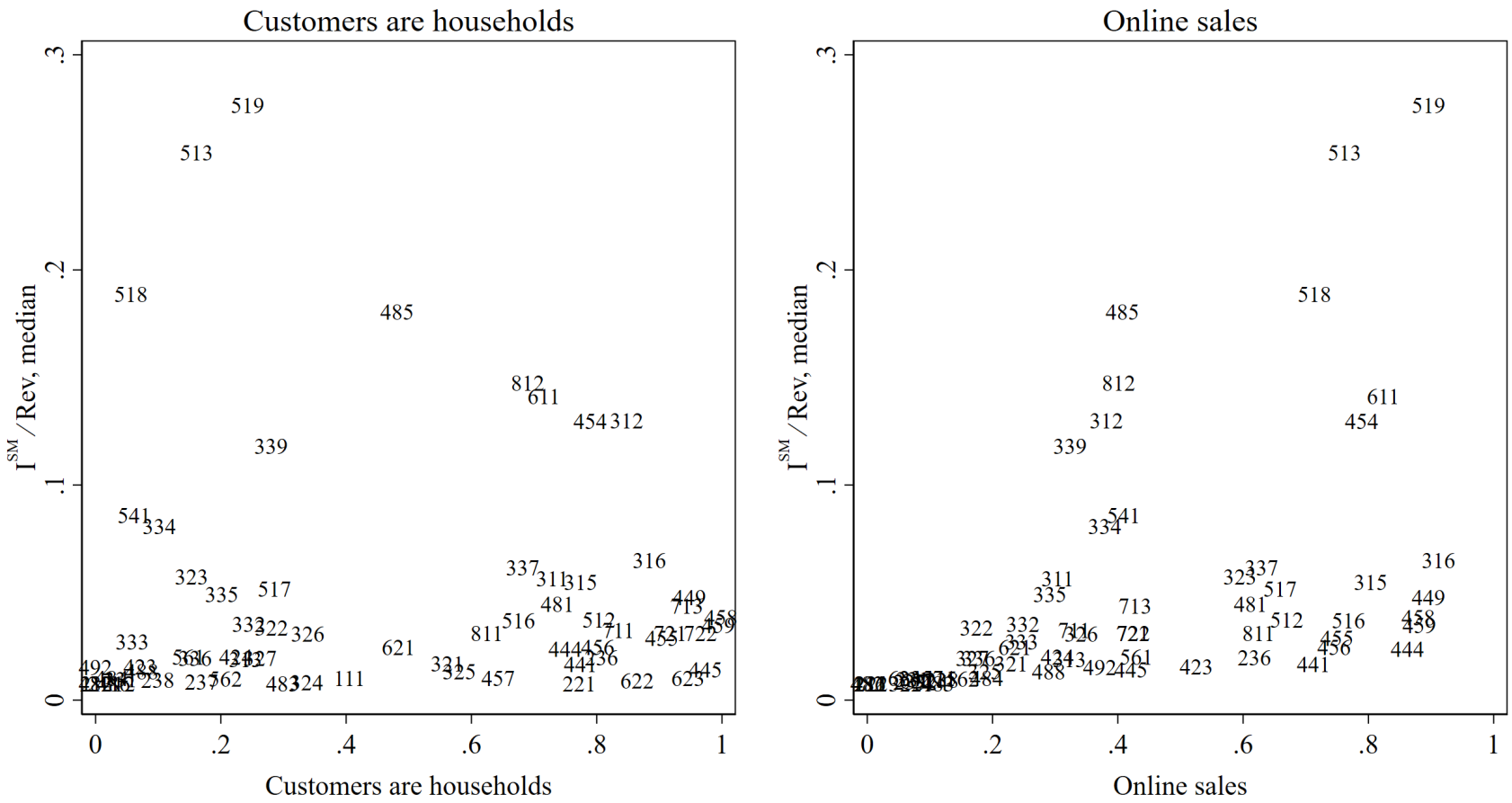


Figure A8: Explaining I^{SM}/Rev (continued)

