Investing in Customer Capital*

Bianca He
Chicago Booth

Lauren Mostrom
Chicago Booth

Amir Sufi
Chicago Booth & NBER

November 2024

Abstract

Firms invest heavily in customer capital, and such investment is a main source of intangible capital value. This study measures investment in customer capital 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. Firms emphasize brand value, sales force, customer service, advertising, and the acquisition and use of customer data as sales and marketing strategies. Industries focused on platform business models, online sales, and the production of high tech manufactured goods invest most heavily in customer capital. Industry-level variation in the intensity of sales and marketing expense and R&D expense explains a large amount of the variation across industries in the value of intangible capital. Residual sales, general, and administrative expense after removing sales and marketing expense is uncorrelated with intangible capital value. Industries that invest most heavily in customer capital are growing as a share of aggregate revenue and enterprise value.

*Carlo Geat, Freddy Pinzon, and Sixun Tang provided excellent research support. We are grateful to Nicolas Crouzet, Andrea Eisfeldt, Steve Kaplan, Simone Lenzu, Quinn Maingi, Dimitris Papanikolaou, Chad Syverson, and seminar participants at 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.

Firm value is increasingly determined by investment in *intangible capital*—non-physical capital that is more difficult to quantify and therefore omitted from balance sheets—and this trend carries with it profound implications for the broader economy. Research posits that the rise of intangible capital shapes key economic outcomes such as investment policy, markups and profits, valuation, financial policy, employee compensation, and productivity growth. However, in order to understand these far-reaching implications, it is crucial to measure and delineate investment in intangible capital to the best degree possible.

The central point of this study is that investment in *customer capital* is a quantitatively large component of investment in intangible capital. Specifically, this study provides comprehensive measures of investment in customer capital, and it seeks to explain the determinants and effects of this investment. 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 current customers.[1] 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 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 administrate expense (SG&A). Compustat, the data set produced by S&P that is widely used by researchers, does not systematically collect or report these data; however, another data set produced by S&P, Capital IQ, does. 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 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 sales-weighted average annual sales and marketing expense to revenue ratio across U.S.

[1]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 3.1.

publicly traded firms is 4.1%, which is higher than research and development (R&D) expenses and two-thirds of capital expenditures. Previous research has focused 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 16% of firm-year observations discuss prominently a focus on advertising in their business descriptions; it is more common that firms discuss their efforts in maintaining good customer service (51%), building a sales force (49%), 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 no resources on sales and marketing, whereas the median firm in the information industry, which includes 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 aggregate revenue and aggregate 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 determinants and implications of this industry-level variation.[2]

Theoretical models posit a close relationship between the amount of investment in customer capital and the sensitivity of customer demand to such investment. Three variables, which are interpreted as proxies for the nature of demand, explain a large amount of the cross-sectional variation in the amount of investment across industries. The most powerful of these variables is the fraction

[2] 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).

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.

The text from business descriptions allows for an exploration of the channels of customer capital investment across industries. Industries in which firms primarily target households focus on advertising and building brand value. Industries in which firms sell more technical products place a strong emphasis on the importance of a sales force, whereas industries in which firms sell products online emphasize the acquisition and use of customer data along with brand value. Industries in which firms have a platform business model emphasize all five channels prominently: advertising, brand value, customer service, sales force, and customer data.

Investment in customer capital has important implications for enterprise value and profits. A fundamental premise of classifying sales and marketing expenses as "investment" is that such expenses build capital that is durable and valuable to the owners of the firm. The evidence supports this premise. An empirical specification motivated by the empirical work in Gutiérrez and Philippon (2017) and the theoretical framework in Crouzet and Eberly (2023) 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 to which we refer to as Q^{PH}. Industries with more R&D expenses also have higher Q^{PH}; however, industries with higher residual SG&A expenses once sales and marketing expenses are removed do not. 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.

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.

Consistent results are obtained when using data on the sources of value in acquisitions from purchase price allocations, information which has been previously used in He (2022), Kepler, Naiker and Stewart (2023), and Ewens, Peters and Wang (2024). These data are especially useful in assessing the value implications of investment on 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. As in the analysis using enterprise value of publicly traded firms, residual SG&A expenses after removing sales and marketing expenses are uncorrelated with the value paid for intangible capital, again supporting the view that these residual SG&A expenses do not reflect an investment in intangible 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.

It is not obvious whether firms in intangible capital-intensive industries should earn higher economic profits; while they are able to capture consumer surplus once they obtain a customer, they must invest heavily in building and maintaining the customer base. The results show that firms in intangible capital-intensive industries earn substantially higher profits if one ignores the capital expenses of the business. However, capital expenses are substantially higher for high intangible capital industries because of significantly higher depreciation rates. Depreciation rates on intangible capital tend to be higher, but even depreciation rates on physical capital are higher in intangible capital-intensive industries. Once depreciation is taken into account, there is some weak evidence that firms in intangible capital-intensive industries earn higher economic profits, but in general, the evidence on profits is inconclusive. The analysis suggests that an evaluation of economic profits across industries or firms that use different levels of intangible capital requires careful consideration of the differences in depreciation costs across those industries.

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 Papanikolaou (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. The former is often referred to as investment in "knowledge capital" and the latter is often referred to as investment in "organizational capital." 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. The point that overall SG&A may be a poor proxy for investment in customer capital is also made by Ptok, Jindal and Reinartz (2018).

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 et al. (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. In general, to the best of our knowledge, the data set collected for this study on investment in customer capital is among the most comprehensive in terms of the number of firms covered, the length of the time series, and the details on the type of investment spending.

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) 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.[3]

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 describes the data set. Section 2 presents summary statistics and the industry-level variation in customer capital investment. Section 3 shows evidence on the variables that explain the cross-sectional variation across industries in customer capital investment. Section 4 focuses on the value and profit implications of customer capital investment, and Section 5 concludes.

[3]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.

1 Data

1.1 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 55,101 observations, as shown in the 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.

The specific data items retrieved from Capital IQ include revenue, costs of goods sold (COGS), SG&A, R&D, depreciation and amortization, sales and marketing expense, advertising expense, marketing expense, general and administrative expense, and net rental expense. Comparing the Capital IQ data to the underlying 10-K SEC filing reveals excellent coverage if the information is in the 10-K filing.

While the revenue reported in Capital IQ and Compustat is almost identical, reported SG&A is not. A regression of one on the other yields an R^2 of 0.89. 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. As a general rule, the mapping from the information on the 10-K filing to the Capital IQ data is more transparent relative to a mapping from the information on the 10-K filing to Compustat.[4]

[4] Several researchers have described the lack of clarity in the mapping of 10-K information to the actual variables in Compustat. 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 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, there is a question of what exactly is included in the sales and marketing expense line reported by firms. 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. Interestingly, in response to the prompt, Gemini also says that a "thorough understanding requires looking beyond the single line item. SEC filings (like 10-Ks) often provide more granular detail in the footnotes of management discussion and analysis (MD&A)."[5]

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.

1.2 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 A 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.[6] We start with a manual

[5]We had already conducted the large-scale analysis of the text of 10-K filings described in the following sub-section before giving Gemini this prompt and reading its response. The exact prompt and the full response are in Appendix Section A.3.

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

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.[7] 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 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.[8]

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

[7] For this draft, 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 include problems processing the original text from the SEC and missing filings for a given firm-year observation. We are working to improve this match to be close to 100%.

[8] 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.

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 45,305, 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.[9]

The appendix reports more 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 from Capital IQ. 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 strategies. The exact prompts given to Gemini to obtain these variables are listed in Appendix Table A1.

Finally, Gemini is also used to create variables representing "primitives" of the business model of the firm. These measures are used to predict the type of firms that spend the most on investment in customer capital. 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 2.4.

1.3 Capitalization

The variables discussed above measure investment in customer capital; however, it is also necessary to measure the stock of customer capital. To do so, the capitalization methodology of Eisfeldt and Papanikolaou (2013), as implemented by subsequent work by Peters and Taylor (2017), is followed.

[9]The approach taken here is conservative in the use of Gemini's reading of the text. Gemini is used only to impute zeros for missing values; it is not used to try to quantify an exact non-zero amount for firms that reveal moderate or substantial spending in the text of their filing but no sales and marketing line item in the income statement. This is potential material for future research: using text to predict non-zero values. However, given the novelty in the use of LLMs in the literature, we choose to err on the side of caution in the use of the text to predict values of sales and marketing and R&D expense.

While the methodology of Peters and Taylor (2017) is followed closely, a crucial difference is that Peters and Taylor (2017) uses 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.[10]

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 book knowledge capital (based on R&D), the amount of book customer capital (based on sales and marketing expense), and the book 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 assumed depreciation rate for customer capital follows the evidence discussed in Gourio and Rudanko (2014). They cite a number of industry assumptions on the turnover in the customer base to justify a depreciation rate of 15% for customer capital.

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.

1.4 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

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

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 Labs Workforce Dynamics 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 shown below, despite being constructed from a distinct data source, there is a high correlation between the two measures of customer capital investment at the industry level.

2 Investing in customer capital: the facts

2.1 Sample summary statistics

Table 2 presents the summary statistics on the investment undertaken by firms in the sample. Theory offers two potential choices for how to scale such investment: static models that explore spending on demand-shifting inputs suggest revenue as the appropriate scaling variable (e.g., Bond, Hashemi, Kaplan and Zoch (2021)), whereas dynamic models treating such spending as investment suggest scaling by the book value of capital (e.g., Hayashi (1982), Whited (1992), Crouzet and Eberly (2023)). These models are described in more detail in Section 3.1. While the dynamic models are closer to the underlying economics of such spending, a challenge is that the book value of intangible capital is not reported on firm balance sheets. Measurement of the book value requires strong assumptions and extensive historical data which results in many missing observations (see the discussion above in Section 1.3). No such measurement issues arise when scaling by revenue, as it is reported on the income statement for all firm-year observations in the sample. Given these data considerations, the primary focus of the analysis is on measures of investment scaled by revenue, but investment scaled by the book value of capital is also reported as a robustness test.[11]

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. The median sales and marketing expense to revenue ratio 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.

Panel C of Table 2 presents measures of investment scaled by the book value of capital, where the book value of intangible capital is estimated using the capitalization technique discussed in Section 1.3. Scaled by the total book value of capital (both physical and intangible), the median annual investment rate for capital expenditures is 2.4%, whereas the median investment rate for sales and marketing expense is 2.8%. As with the statistics scaled by revenue, both the R&D and sales and marketing distributions have more mass in the right tail of the distribution; at the 90th percentile, investment rates for these two types of investment are 15%.

Also included are two estimates of Q (e.g., Hayashi (1982)). The first is enterprise value scaled by physical capital (Q^{PH}), which is the common measure used in the investment literature.[12] The second is enterprise value scaled by total capital (Q^{tot}), which includes the capitalized value of spending on sales and marketing and R&D and externally acquired intangible capital. The estimate of Q^{tot} has a revenue-weighted average of 1.8 and a median across the distribution of 1.1. Such estimates are closer to what one would expect from theory assuming reasonable adjustment costs. Section 4.1 elaborates on this point.

Panel D of Table 2 summarizes the Revelio Labs data for the sample where both the salary and the sales and marketing expense variables are available. The revenue-weighted average salaries paid to sales and marketing employees is 3% of revenue. It is important to note that salaries paid to sales and marketing employees may not always be accounted for within the overall sales and marketing expense line item on the income statement; it is likely that some of these expenses are accounted for in COGS. Regardless, it is helpful to have an alternative measure of customer capital

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

[12] The measurement of Q^{PH} follows Crouzet and Eberly (2023). The numerator is $(prcc_f * csho) + (dltt + dlc) - che$ and the denominator is $ppegt$. However, for one industry, 3-digit NAICS 236 "Construction of buildings" we add the book value of inventories to the denominator. In this industry, inventories are houses for sale and they are enormous. The median inventory to property, plants, and equipment ratio is 17.1 for this industry; the next highest industry (423 – merchant wholesalers, durable goods) has a ratio of 1.1. Failure to account for inventories for this industry leads to measures of Q^{PH} above 20.

investment that comes from a different data source.

The previous literature often uses advertising expense or 30% of overall SG&A to measure intangible investment in customer or organizational capital. For completeness, Table 2 includes summary statistics for residual SG&A, which is SG&A from Capital IQ less sales and marketing expense.[13] One result of the analysis below is that residual SG&A is uncorrelated with measures of value once sales and marketing expense is removed, which suggests that sales and marketing expense is the main component of SG&A that should be considered an investment in intangible capital.

Advertising is a small part of sales and marketing expense, especially for firms that invest heavily in customer capital. Figure 1 plots the sales and marketing expense to revenue ratio 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. An initial look at the data shows that firms invest in customer capital through channels beyond advertising.

Finally, Panel E of Table 2 summarizes the data from the Gemini results on the type of sales and marketing strategies employed by firms. As the weighted average shows, firms emphasize customer service, a sales force, and brand value the most. Building and using data on customers is mentioned in 28% of the observations, and advertising is mentioned in only 16% of the observations.

2.2 Industry-level variation in customer capital investment

Figure 2 shows a striking amount of variation across industries in the median ratio of sales and marketing expense to revenue ($^{I^{SM}}/_{Rev}$).[14] 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 infrastruc-

[13] SG&A from Capital IQ, as opposed to Compustat, is the preferred measure used in the analysis. As mentioned in Section 1.1, the two main differences are (1) Compustat includes R&D in SG&A while Capital IQ does not, and (2) Compustat reclassifies general and administrative expense into COGS if it is separately reported. 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.

[14] 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.

ture 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.[15] There is a high correlation between the two ratios, 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

[15]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 reported in Appendix Section A.2. For Table 3, only industries with at least five firms with Revelio data are included, which explains the smaller sample size.

2011) and the last five years of the sample (2018 to 2022). The two measures are highly correlated. Figure 3 shows a scatter plot of the early and late industry-level medians; the persistence of the level of spending on sales and marketing is evident. 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 1.3.

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, the rest of the analysis of this study focuses on this industry-level variation.

Column 6 shows that the industry-level variation between sales and marketing and residual SG&A after removing sales and marketing is positively correlated, but both the level and explanatory power is substantially weaker. The R^2 in column 6 is only 0.06. The same is true for advertising, as is shown in column 7. For both of these outcomes, the information sector industries (513, 518, and 519) are the largest outliers. These industries do not spend extensive amounts on advertising, and they do not have large SG&A budgets with the exception of their sales and marketing expense. The importance of customer capital investment for these industries is underestimated in analyses focused on broader SG&A or advertising.[16]

Industries that invest heavily in customer capital are becoming more important over time, as shown in Figure 4. For both figures, 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 4, respectively. Industries with the highest $^{I^{SM}}/_{Rev}$ experience the largest increase in the share of both revenues and enterprise value over time.[17]

[16] Appendix Figure A1 shows the industry-level scatter plot of sales and marketing expenses against advertising expenses and residual SG&A expenses.

[17] Firms are sorted into industries for Figure 4 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), which shows that the value added by manufacturing firms in the United States is often attributed to establishments that are in industries other than manufacturing.

2.3 Comparison with capital expenditures and R&D

The left panel of Figure 5 compares the across-industry distribution in sales and marketing to capital expenditures. There are a large number of industries that hug either the vertical axis or horizontal axis; many industries spend either on capital expenditures or sales and marketing, but not both. Agriculture, mining, and rail/water/pipeline/truck transportation companies have large capital expenditures and almost no sales and marketing. Of the 18 industries that have a ratio of sales and marketing expense to revenue above the revenue-weighted average of the entire sample, only one has a capital expenditure to revenue ratio 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 important and growing sectors of the economy.

In contrast, the right panel of Figure 5 shows that sales and marketing and R&D are more complementary in production for several industries. Industries that have both high R&D and high sales and marketing 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).

Another noticeable pattern in the right panel of Figure 5 is the out-sized role that one industry, chemical manufacturing (325), plays in R&D. This industry includes firms that manufacture pharmaceuticals. Sales and marketing expense is more evenly distributed across many sectors of the economy. Over 80% of industries have a higher median I^{SM}/Rev than median I^{RD}/Rev. In terms of investment in intangible capital, the right panel of Figure 5 shows that the breadth of investment in customer capital is far wider than investment in knowledge capital via R&D.

2.4 Strategies by industry

There are also important differences across industries in the underlying strategies used when investing in customer capital. For each of the five strategies for sales and marketing (brand value, advertising, sales force, customer service, and customer data acquisition and use), we calculate the fraction of the firms in an industry that engages in each given strategy according to the answers given by Gemini when reading Item 1 of the 10-K filing.

Table 4 shows the results. Three industries are among the top five for both brand value and advertising: leather manufacturing, apparel manufacturing, and beverage and tobacco manufacturing. Food manufacturing is also on the advertising list, and furniture manufacturing is on the brand value list. These industries primarily sell consumer goods to households, and so it is not surprising that these industries place emphasis on advertising and brand value. Most of the literature on advertising tends to focus on these strategies.

Two of the industries on the sales force list are high tech manufacturing industries: computer and electronic product manufacturing and medical equipment manufacturing. Companies in these industries often emphasize that a high-skilled sales force is required given the technical nature of the products. For example, Kopin Corp, a company that sells high-resolution microdisplays and optics, says in their 2012 10-K filing:

> We believe that the technical nature of our products and markets demands a commitment 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.

There are numerous examples of similar statements on the 10-K filings of high tech manufacturing firms. There appears to be a high degree of complementarity between the sale of technical products and the use of a sales force; this is explored further in Section 3.

The pattern for customer service is more mixed, with manufacturing firms, non-store retailers, and personal and laundry service firms all making the top five list. For the sales and marketing strategy of acquiring and using customer data, the list includes the same three information industries that have the highest overall sales and marketing ratio plus transit and ground transportation and non-store retailers. In the latter two categories, prominent firms include Uber Technologies Inc, Lyft Inc, and Amazon.com Inc. Many of these companies operate digital platforms that bring together providers and users of goods and services. This is explored further in the next section.

Table 5 shows the industry-level correlation among these five strategies along with the correlation with the industry-level median ${}^{I^{SM}}\!/_{Rev}$. The ratio is most highly correlated with brand value and the acquisition and use of customer data. Brand value and advertising are highly correlated with one another. The weakest correlation among the strategies is between the use of a sales force and the acquisition and use of customer data, and the use of a sales force and advertising.

3 Explaining industry-level variation

What are the primitive factors that explain the large amount of variation across industries in the level of investment in customer capital? This section provides a brief theoretical discussion, and then shows that a few variables related to the nature of demand across industries explains a large amount of the variation in such investment.

3.1 Guidance from theory

Given the paucity of empirical evidence on investment in customer capital, it should not be surprising that there is limited theory on the underlying determinants of investment in customer capital. Furthermore, it is unlikely that a single theoretical framework can capture the myriad sales and marketing strategies undertaken by firms. With these limitations in mind, this section is designed to provide some high level insights from theory that can help motivate and interpret the empirical analysis.

Investing to shift demand

To help develop intuition, this sub-section presents a model in which sales and marketing expense is an input into the generation of firm revenue that works through shifting demand. The model here is a simplified version of the model in the appendix of Bond, Hashemi, Kaplan and Zoch (2021). 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

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

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

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

The derivation is detailed in Appendix Section B. Define α^D to be the ratio of expenses on the demand-shifting input to revenue ($\frac{X^D W^D}{PQ}$). Then the firm's first order conditions imply:

$$\alpha^D = \rho \theta^D \tag{1}$$

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 ($\frac{X^D}{D}\frac{\partial \mathcal{D}(\cdot)}{\partial D}$). Equation 1 relates the equilibrium cost share ratio of sales and marketing to fundamental parameters of the revenue function: ρ and θ^D.

In terms of the across-industry variation in the ratio of sales and marketing expense to revenue, the intuition of equation 1 points to the effectiveness of such spending on shifting demand, and also the effects of a shift in demand on a firm's ability to charge higher prices for a given quantity of output. For example, if an industry is characterized by a greater ability of firms to shift demand through advertising or building brand value (high θ^D), then we should expect more spending on sales and marketing in that industry. In addition, if an industry is characterized by firms' ability to charge a significantly higher price for a given shift in demand (high ρ), then we should expect that industry to have a high equilibrium expenditure share on sales and marketing.[18]

[18] This simplified model is static and therefore there is no build-up of customer capital. However, we believe the main insights carry over in a dynamic model in which firms invest in customer capital which similarly shifts demand and depreciates at some rate. Our hypothesis is that in such a dynamic model the steady state level of investment in

Building a customer base

The model in Gourio and Rudanko (2014) highlights how firm spending on sales and marketing helps build a valuable customer base in industries characterized by search frictions. There are a few key assumptions. First, the product market has search frictions that require a buyer to meet with a sales representative in order to become a new customer of the firm. Second, firms entice customers through up-front discounts. Therefore, the firm must spend up front on a sales force and accept low prices initially to win customers. Once there is a match between a firm and a customer, the customer sticks with the firm 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 implies that the firm is able to capture all of the consumer surplus of the relationship after it is formed.

The 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.[19] The marginal benefit is a function of the future value of the customer relationships gained. This value is what Gourio and Rudanko (2014) call *customer capital*. Companies are able to capture the consumer surplus of long-term customer relationships, and so the number of long-term customer relationships is valuable to the firm.

If we assume that cost curves for the inputs are similar across industries, then the cross-sectional variation across industries in the amount of spending on such inputs is determined by the value to the firm of winning new customer relationships. Taking the model literally, the value of new customer relationships is related to how large the search frictions in the product market are for a given industry. Gourio and Rudanko (2014) conduct an empirical analysis in which they approximate cross-sectional variation in search frictions by sorting industries according to their SG&A expense, with the assumption that high SG&A expense industries have higher search frictions.

Stepping outside the specific assumptions of the model, the broader intuition is that industries that spend more on sales and marketing are industries in which consumers are willing to continue with their relationship with a firm even if the prices charged are relatively high. This makes the acquisition of a new customer especially valuable. The "stickiness" of the relationship could be for a variety of reasons: there may be poor product substitutes, there may be a lot of difficulty in identifying an alternative seller, or customers may be lazy about searching for new products. This is an open empirical question.[20]

---

customer capital is positively related to ρ and θ^D, and also positively related to the depreciation rate of customer capital. This is material we hope to include in the next draft.

[19] This is equation 2.8 on page 1108 of the Gourio and Rudanko (2014) study.

[20] The Baker et al. (2023) measure of "customer churn" is perhaps the closest empirical analog to the idea in Gourio and Rudanko (2014). As they show, some firms have a customer base that is quite stable, and such firms tend to have lower risk and volatility.

3.2 Explanatory variables

Theory suggests that primitive factors related to the ability of firms in an industry to affect demand through their sales and marketing efforts determine the cross-sectional variation across industries in investment in customer capital. In this section, four characteristics of demand across industries are used to explain such variation. They are:

- **Primary customers are households**: Household demand may be more elastic with respect to demand-shifting inputs such as advertising and branding than 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".[21]

- **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 products are sold online have a larger impact on consumer demand through sales and marketing efforts, corresponding to a high elasticity θ^D. 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 sponsor should face a high elasticity of demand to sales and marketing (θ^D). Sales and marketing efforts in this context both increase the user base directly, by persuading potential customers to join

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

the network, and indirectly by influencing potential customers' expectations about the total size of the user base. Following a similar logic, markets characterized by network effects in demand are likely to have a high value of ρ, or the elasticity of the price a firm can charge (keeping quantity sold constant) with respect to a shift in demand. For every consumer a firm wins (or loses), there is an amplification effect of other consumers moving in the same direction as the consumer.[22] 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. As mentioned in Section 2.4, 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. 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.[23]

3.3 Results

Table 6 presents results for industry-level regressions of the following form:

$$\left({}^{I^{SM}}\!/_{Rev}\right)_j = \alpha + \beta X_j + \epsilon_j \tag{2}$$

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 Panel A of Table 6.

Industries that target households as their primary customers do not have a statistically significant difference in ${}^{I^{SM}}\!/_{Rev}$. This is contrary to our prior, which was shaped by research on the importance

[22]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).

[23]The industry-level summary statistics for these covariates are in Panel A of Appendix Table A3.

of advertising and brand value. Industries that primarily target businesses as customers also engage in significant sales and marketing efforts.

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 7 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.[24] 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.

Panel B of Table 6 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.[25]

The specific sales and marketing strategies undertaken by firms are also systematically related to the underlying characteristics of demand in the industry. To show these patterns, a regression specification similar to equation 2 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 6 summarizes the results of these 20 univariate regressions. Each bar in Figure 6 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

[24]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.

[25]In Appendix Table A4, these regression specifications are estimated using the ratio of sales and marketing expense to book value of capital ($^{I^{SM}}\!/_{K}$) as the left hand side variable. The results are broadly similar, although the online and platform covariates have equivalent statistical power in the univariate specifications, and the platform result is weaker in the multivariate specification given its high correlation with the online and $^{(WL)^{EG}}\!/_{K}$ covariate.

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.

The patterns in Figure 6 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.[26]

4 Value and Profits

4.1 Valuation using Q

Spending on sales and marketing generates firm value. Two types of analysis show this result: analysis of the market value of publicly-traded firms and analysis using prices paid in acquisitions of firms. This sub-section presents results motivated by the Crouzet and Eberly (2023) framework that relates the traditional measure of Q to the stock of tangible and intangible capital at the firm.

[26] In Appendix Table A5, 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.

The Crouzet and Eberly (2023) model augments the classic Hayashi (1982) investment framework with two additional ingredients: (1) firms may earn markups 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 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 $Q^{PH} = {}^{V}\!/_{K^{PH}}$ 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^{PH} is the ratio of enterprise value to the book value of physical capital (${}^{V}\!/_{K^{PH}}$, traditionally called Q in the literature), q is the marginal effect of an additional unit of investment on enterprise value, μ is a measure of markups derived from the profit function, r is the external cost of capital (ignoring depreciation and adjustment costs), g is the steady state growth rate, and K is the value of book capital.

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}.

Before presenting the results, a few caveats are in order. First, 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. Second, the regression coefficient estimate of β in equation 4 is only identified if the other variables in question are uncorrelated with the residual, which is highly unlikely in reality. It seems reasonable to assume, for example, that the marginal effect on enterprise value from investment in physical capital (q^{PH}) varies across industries, and that this variation may be correlated with the ratio of intangible to tangible capital ratio. Finally, even if the coefficient β were identified, it would be impossible to separate the pure effect of intangible capital versus the interactive effect of markups and intangible capital on average Q. These are indeed strong caveats, and so we avoid interpreting coefficients in a structural manner. However, it is useful to see that the Crouzet and Eberly (2023) framework provides a theoretical underpinning for the reduced form regressions shown in this section.

Table 8 presents the results. The right hand side variables for the first four columns are the industry-level median flow measures of intangible investment scaled by revenue. 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 the coefficient estimate of β from equation 5. The right hand side variable is the median intangible to tangible book capital ratio ($^{K^{IT}}/_{K^{PH}}$) across firms in the industry. 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, following the methodology in Peters and Taylor (2017). There is strong explanatory power; $^{K^{IT}}/_{K^{PH}}$ alone explains over 80% of the variation in Q^{PH}. Figure 7 shows the scatter-plot version of the specification estimated in column 5, which helps to visualize the strong predictive power of $^{K^{IT}}/_{K^{PH}}$.[27]

4.2 Valuation using transactions

Sales and marketing expense and R&D expense also predict value paid for intangible capital in firm acquisitions. To establish this result, 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 pri-

[27] With the necessary caveats in mind, it is worth noting that the coefficient estimate in column 5, 1.19, carries with it a structural interpretation under strong assumptions on the error term. For example, in a world without markups, a coefficient estimate of 1.19 in the Crouzet and Eberly (2023) framework implies an estimate of q^{IT} of 1.19 (see equations 3 and 4 above). In the specific model of Crouzet and Eberly (2023), the balanced growth value of q^{IT} should be $1 + \gamma g$, where gamma is the adjustment cost parameter.

vate 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."

The PPA 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. This is essentially an industry-level revenue transaction multiple for intangible assets. The data 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.[28]

With these dependent variables in hand, Table 9 reports estimates of regressions similar to the specification outlined in equation 2. 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 9 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, and the R^2 of the regression is 0.44. Columns 3 and 5 split total intangible asset value into the value associated with customer 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,

[28]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 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.[29] 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.[30]

Panel B of Table 9 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.

In general, the results suggest that a one dollar increase in intangible investment is associated with a three to four dollar increase in market value of the intangibles. 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

[29]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.

[30]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.

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. Finally, the guidelines used by valuation professionals in PPA require an estimate of the counter-factual value of the asset if it were not owned by the target firm. This likely underestimates the true overall value of the intangible assets to the firm, given that some of the value is likely lost if the intangible asset is separated from the firm.

4.3 Intangible capital and profits

From a theoretical perspective, firms in customer capital-intensive industries capture more of the surplus associated with customer relationships. However, firms need to spend more on building and maintaining these relationships. It is not obvious, therefore, whether firms in industries more reliant on customer capital should see higher or lower economic profits.

This section presents an empirical analysis of profits, where profits are defined according to the Crouzet and Eberly (2023) framework. The key assumptions of this framework are that the production function is homogeneous of degree 1, the profit function is homogeneous of degree $\frac{1}{\mu}$, and that each type of capital has its own depreciation rate and adjustment cost. Along the balanced growth path, the equilibrium rents are defined as:

$$\mu = \frac{\pi^{inc}}{R^{PH}K^{PH} + R^{IT}K^{IT}} \tag{6}$$

where π^{inc} is the profits of the firm ignoring the cost of capital, and R^n and K^n are the per-unit internal cost of capital and book value of capital of type n, respectively. In the Crouzet and Eberly (2023) framework,

$$R^n = r + \delta^n + \gamma^n rg \tag{7}$$

where r is the external per-unit cost of capital, δ^n is the depreciation rate of capital type n and γ^n is the adjustment cost parameter for capital of type n.[31] The measure of markups in Crouzet and Eberly (2023) is intuitive: if the after-expense income generated from capital exceeds the true overall economic cost of capital, the firm is said to earn economic profits, and so $\mu > 1$.

The estimate of profits at the industry level in this study follows from equation 6, but we estimate profits as a share of revenue, and we use the difference between income and the cost of capital instead

[31] Appendix Section IA.B of Crouzet and Eberly (2023) shows that this general framework nests a number of models of profits and markups. However, the framework cannot separate profits due to decreasing returns to scale ("quasi-rents") from pure rents ("markups"). Crouzet and Eberly (2023) show that it is not possible to separate these two sources of profits without estimating or assuming the production function. The De Loecker, Eeckhout and Unger (2020) framework separates pure markups from quasi-rents by estimating a revenue elasticity at the firm level; however, the De Loecker et al. (2020) technique of using estimated revenue elasticities as a proxy for output elasticities has been the subject of a large debate (see, e.g., Bond et al. (2021), De Ridder et al. (2024).

of the ratio:

$$\frac{\pi}{Rev} = \frac{\pi^{inc} - R^{PH}K^{PH} - R^{IT}K^{IT}}{Rev}$$

Using equation 7, this equation can be re-written as:

$$\frac{\pi}{Rev} = \frac{\pi^{inc}}{Rev} - \frac{r(K^{PH} + K^{IT})}{Rev} - \frac{\delta^{PH}K^{PH} + \delta^{IT}K^{IT}}{Rev} - \frac{(\gamma^{PH}K^{PH} + \gamma^{IT}K^{IT})rg}{Rev} \tag{8}$$

The numerator of the first term on the right hand side of equation 8 is revenue less all non-investment related costs. Importantly, the numerator does not subtract capital expenditures, R&D, or sales and marketing expense, as these are investments, not expenses. This term can be thought of as the pure "income" part, π^{inc}, of overall profits, π. The second, third, and fourth terms are the overall external cost of capital and the incremental part of the cost of capital due to depreciation and adjustment costs, respectively.

For the analysis here, each separate term on the right hand side of equation 8 is first measured at the firm level, and then it is averaged at the industry-level using firm-level revenue as weights.[32] The term π^{inc} is measured using the Compustat and Capital IQ data. For the external cost of capital, we follow the literature that uses the implied cost of capital from analyst forecasts and accounting information to obtain an estimate of r at the firm-year level.[33] The book value of intangible capital is estimated as described in Section 1.3. The depreciation rates for the capitalized value of R&D and sales and marketing are the same as assumed for the capitalization methodology. The depreciation rates for physical capital are estimated at the firm level using the Compustat variable dp scaled by the book value of physical capital.[34] Following Belo et al. (2022) and Crouzet and Eberly (2023), the adjustment cost parameters γ^n are assumed to be 0.03 and 0.12 for physical and intangible capital, respectively. The growth rate g is assumed to be 2%.

The key question explored in this section is whether the cross-sectional variation in these terms across industries is related to intangible investment intensity. Regressions similar to equation 2 are estimated with the profit measures as left hand side variables. Panel A of Table 10 presents the results. Column 1 shows that industries with higher intangible investment have higher profits (π^{inc}) when ignoring the cost of capital. The relationship is strong; the two variables produce an R^2 of almost 60%.

In terms of the overall cost of capital, the external component of the cost of capital, $r(K^{PH} + K^{IT})$, without consideration of depreciation does not vary with the level of intangible investment,

[32]Revenue-weighted averages are used given that the goal is to estimate if the industry as a whole earns profits; hence larger firms should be weighted more heavily. The summary statistics for all profit variables are reported in Panel B of Appendix Table A3.

[33]These data were graciously provided to us by Niels Gormsen; see Gormsen and Huber (2024), Mohanram and Gode (2013), and Eskildsen et al. (2024) for more details on the calculation of the implied cost of capital.

[34]If the variable dp is missing at the firm level, then the depreciation rate on physical capital is assumed to be the same as the industry median.

as shown in column 2 of Panel A of Table 10. However, industries with more intangible investment have higher depreciation costs, $\delta^{PH}K^{PH}+\delta^{IT}K^{IT}$, as shown in column 3. This is true for industries with higher sales and marketing expense and for industries with higher R&D expenses. Column 4 shows that total adjustment costs are higher for industries that invest more heavily in intangible capital, but the estimates are not precise and the magnitude is small.

Putting everything together, column 5 shows the coefficients when regressing industry-level measures of π on industry-level measures of intangible capital investment. The point estimates are positive, but not precisely estimated. In terms of magnitudes, the point estimate implies that a one standard deviation increase in the sales and marketing expense of an industry leads to a 0.015 increase in the profit measure, which is almost exactly a one-eighth standard deviation of the left hand side variable. The point estimate is small, and it is not precisely estimated.

Panel B of Table 10 decomposes these results further. Recall that the overall external cost of capital is a function of the total amount of capital $K^{PH} + K^{IT}$ and the cost per unit of capital r. Columns 1 and 2 of Panel B show that the null hypothesis of zero difference cannot be rejected for either of the two components. Column 3 shows that the main difference in the overall cost of capital comes from differential depreciation rates. The left hand side variable in column 3 is total depreciation divided by total book capital, which is called δ, or the overall depreciation rate of capital. Industries with more intangible investment have higher depreciation rates. In terms of magnitudes, a one standard deviation in investment in customer capital leads to a 1.4 percentage point increase in the depreciation rate of capital, which is about 1/3 a standard deviation of the left hand side variable.

The result in column 3 is partially by assumption: the depreciation rates on the capitalized value of sales and marketing and R&D are assumed to be high. However, it is important to recognize that these assumed depreciation rates are based on research by Gourio and Rudanko (2014) and Ewens et al. (2024), who cite industry research and conduct empirical analysis to justify these high depreciation rates. In addition, in column 4, the left hand side variable is isolated to depreciation rates on physical capital, which are derived from the underlying Compustat data and therefore are explicitly estimated with data. As the estimates in column 4 show, even the depreciation rate on physical capital is higher in industries that spend more on R&D and sales and marketing.

Overall, the evidence on profits is inconclusive. While firms operating in industries with more investment in intangible capital earn more profits if the cost of capital is ignored, they also face a higher overall cost of capital driven by a higher depreciation rate. Intangible assets depreciate faster and even physical capital in intangible capital-intensive industries depreciates faster. The main lesson for future research is that differences in the cost of capital should be acknowledged in any analysis of profit patterns. For example, treating the cost of capital as constant across industries mistakenly yields results that suggests intangible capital-intensive industries earn significantly higher economic profits.

5 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, 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?

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 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.[35] 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.

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.

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.

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

— , **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.

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, 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.

Eskildsen, Marc, Markus Ibert, Theis Ingerslev Jensen, and Lasse Heje Pedersen, "In search of the true greenium," *Available at SSRN*, 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.

Gormsen, Niels Joachim and Kilian Huber, "Firms' Perceived Cost of Capital," Technical Report, National Bureau of Economic Research 2024.

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.

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

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.

Mohanram, Partha and Dan Gode, "Removing predictable analyst forecast errors to improve implied cost of equity estimates," *Review of Accounting Studies*, 2013, *18*, 443–478.

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.

Ridder, Maarten De, Basile Grassi, and Giovanni Morzenti, "The Hitchhiker's Guide to Markup Estimation: Assessing Estimates from Financial Data," *Working Paper*, 2024.

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.

Whited, Toni M, "Debt, liquidity constraints, and corporate investment: Evidence from panel data," *Journal of Finance*, 1992, *47* (4), 1425–1460.

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

		N	Avg Rev	Median Rev
1	Baseline Sample	55,101	3,595	333
2	With Capital IQ data	55,058	3,593	333
3	Capital IQ Sales & Marketing expense available	30,109	3,627	296
4	Compustat R&D expense available	33,611	3,892	240
5	Sales & Marketing expense available, after Gemini results	36,350	3,394	247
6	R&D expense available, after Gemini results	45,305	3,615	306
7	CAPX available	54,997	3,601	335
8	Capitalization possible	34,799	3,833	317
9	Revelio data available	28,957	5,569	832

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

Table 2: Summary Statistics

	N	Wgt Avg	p10	p25	Median	p75	p90
Panel A. Summary Statistics							
Sales and Marketing expense to revenue ($^{I^{SM}}/_{Rev}$)	36,350	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 ($^{I^{CX}}/_{Rev}$)	54,997	0.065	0.004	0.013	0.031	0.077	0.253
Residual SGA to revenue ($^{RSGA}/_{Rev}$)	35,865	0.140	0.061	0.110	0.199	0.377	1.619
Panel B. All investment variables available							
Sales and Marketing expense to revenue ($^{I^{SM}}/_{Rev}$)	30,925	0.042	0.000	0.003	0.032	0.162	0.378
R&D expense to revenue ($^{I^{RD}}/_{Rev}$)	30,925	0.038	0.000	0.000	0.019	0.163	0.721
Capital expenditure to revenue ($^{I^{CX}}/_{Rev}$)	30,925	0.056	0.004	0.012	0.029	0.065	0.200
Residual SGA to revenue ($^{RSGA}/_{Rev}$)	30,925	0.137	0.062	0.111	0.198	0.379	1.700
Panel C. Scaled by K							
Sales and Marketing expense to total capital ($^{I^{SM}}/_{K}$)	31,709	0.031	0.000	0.003	0.028	0.087	0.149
R&D expense to total capital ($^{I^{RD}}/_{K}$)	29,528	0.025	0.000	0.000	0.014	0.075	0.155
Capital expenditure to total capital ($^{I^{CX}}/_{K}$)	34,535	0.049	0.002	0.009	0.024	0.050	0.094
Enterprise value over physical capital (Q^{PH})	34,320	5.470	0.649	1.234	3.502	11.067	33.005
Enterprise value over total capital (Q^{tot})	34,571	1.830	0.305	0.605	1.078	2.102	4.497
Panel D. Comparison with Revelio when both are available							
Sales and Marketing expense to revenue ($^{I^{SM}}/_{Rev}$)	18,630	0.041	0.000	0.003	0.027	0.125	0.328
SM salaries to revenue ($^{(WL)^{SM}}/_{Rev}$)	18,630	0.030	0.008	0.022	0.053	0.121	0.274
Panel E. Strategies							
Brand value	49,748	0.465	0.000	0.000	0.000	1.000	1.000
Advertising	49,758	0.164	0.000	0.000	0.000	0.000	1.000
Sales force	49,759	0.491	0.000	0.000	1.000	1.000	1.000
Customer service	49,754	0.507	0.000	0.000	1.000	1.000	1.000
Data	49,765	0.282	0.000	0.000	0.000	1.000	1.000

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	I^{SM}/Rev, wgt avg	I^{SM}/Rev, avg	$(WL)^{SM}/Rev$, median	I^{SM}/Rev, median ('18 - '22)	I^{SM}/K, median	$RSGA/Rev$, median	I^{AD}/Rev, median
I^{SM}/Rev, median	0.740**	1.235**	0.522**		0.463**	0.390*	0.139*
	(0.072)	(0.086)	(0.055)		(0.059)	(0.152)	(0.059)
I^{SM}/Rev, median ('07 - '11)				1.124**			
				(0.192)			
Constant	0.013**	0.032**	0.023**	0.004	0.013**	0.153**	0.010**
	(0.003)	(0.006)	(0.004)	(0.004)	(0.002)	(0.015)	(0.002)
Observations	67	67	63	66	64	67	61
R^2	0.769	0.734	0.496	0.729	0.696	0.057	0.225

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: Sales and Marketing Strategies Cited by Industries

		Brand Value	I^{SM}/Rev, median
Panel A. Brand Value			
316	Leather Products	0.861	0.057
315	Apparel Mfg	0.828	0.047
312	Bev & Tobacco Mfg	0.768	0.122
519	Information Services	0.744	0.269
337	Furniture Mfg	0.738	0.054
		Advertising	I^{SM}/Rev, median
Panel B. Advertising			
316	Leather Products	0.606	0.057
315	Apparel Mfg	0.549	0.047
454	Nonstore Retail	0.480	0.122
312	Bev & Tobacco Mfg	0.382	0.122
311	Food Mfg	0.370	0.049
		Sales Force	I^{SM}/Rev, median
Panel C. Sales Force			
316	Leather Products	0.829	0.057
334	Computer & Electronics Prod	0.819	0.073
339	Medical Equip Mfg	0.794	0.111
513	Publishing Industries	0.788	0.247
337	Furniture Mfg	0.764	0.054
		Customer Service	I^{SM}/Rev, median
Panel D. Customer Service			
454	Nonstore Retail	0.842	0.122
812	Personal Services	0.775	0.140
337	Furniture Mfg	0.682	0.054
323	Printing Services	0.650	0.050
339	Medical Equip Mfg	0.647	0.111
		Data	I^{SM}/Rev, median
Panel E. Data			
519	Information Services	0.676	0.269
518	Data Processing & Hosting	0.623	0.181
485	Ground Passenger Transit	0.593	0.173
454	Nonstore Retail	0.559	0.122
513	Publishing Industries	0.525	0.247

Each panel lists the top five industries citing the strategy in question, along with the fraction of firms in the industry citing the strategy and the industry-level median I^{SM}/Rev.

Table 5: Correlation Across Industries for Sales and Marketing Strategies

	I^{SM}/Rev	Brand Value	Advertising	Sales Force	Customer Service	Data
I^{SM}/Rev	1					
Brand Value	0.509	1				
Advertising	0.318	0.831	1			
Sales Force	0.327	0.631	0.445	1		
Customer Service	0.277	0.704	0.640	0.656	1	
Data	0.632	0.795	0.698	0.420	0.687	1

Table 6: Explaining Variation in $I^{SM}/_{Rev}$ Across Industries

	Dependent variable: $I^{SM}/_{Rev}$, *median*				
	(1)	(2)	(3)	(4)	(5)
Panel A					
Households	0.001				0.027
	(0.018)				(0.014)
Online		0.098^{**}			0.030
		(0.027)			(0.016)
Platform			0.276^{**}		0.142^{**}
			(0.038)		(0.046)
$(WL)^{EG}/_{Rev}$, median				0.723^{**}	0.567^{**}
				(0.172)	(0.141)
Constant	0.039^{**}	-0.000	0.002	0.002	-0.033^{**}
	(0.013)	(0.007)	(0.005)	(0.008)	(0.008)
Observations	67	67	67	63	63
R^2	0.000	0.231	0.581	0.442	0.705
	Dependent variable: $(WL)^{SM}/_{Rev}$, *median*				
	(1)	(2)	(3)	(4)	(5)
Panel B					
Households	-0.004				0.003
	(0.014)				(0.008)
Online		0.080^{**}			0.037^{**}
		(0.019)			(0.013)
Platform			0.214^{**}		0.118^{**}
			(0.025)		(0.039)
$(WL)^{EG}/_{Rev}$, median				0.489^{**}	0.286^{**}
				(0.087)	(0.089)
Constant	0.045^{**}	0.010^{*}	0.015^{**}	0.018^{**}	-0.003
	(0.009)	(0.005)	(0.003)	(0.005)	(0.003)
Observations	63	63	63	63	63
R^2	0.001	0.301	0.593	0.368	0.681

Table 7: 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 8: Explaining ${}^{V}/_{K^{PH}}$, or Q^{PH}

	Dependent variable: ${}^{V}/_{K^{PH}}(Q^{PH}), median$				
	(1)	(2)	(3)	(4)	(5)
${}^{I^{SM}}/_{Rev}$, median	38.37**		32.71**	32.46**	
	(5.70)		(4.93)	(5.14)	
${}^{(WL)^{SM}}/_{Rev}$, median		47.18**			
		(14.75)			
${}^{I^{RD}}/_{Rev}$, median			16.73**	14.61**	
			(2.44)	(4.08)	
${}^{RSGA}/_{Rev}$, median				2.48	
				(2.59)	
${}^{K^{IT}}/_{K^{PH}}$, median					1.19**
					(0.14)
Constant	1.64**	1.10*	1.55**	1.18**	1.30**
	(0.25)	(0.48)	(0.20)	(0.32)	(0.15)
Observations	67	63	67	67	67
R^2	0.576	0.461	0.703	0.707	0.800

This table presents estimates of industry-level regressions of the enterprise value to book physical capital value ratio (${}^{V}/_{K^{PH}}$, or Q^{PH}) on measures if intangible capital investment and intangible capital. Robust standard errors are in parentheses. ${}^{*}p < 0.05, {}^{**}p < 0.01$.

Table 9: 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	3.277**	3.223**	3.469**	3.439**	-0.191	-0.217
	(1.200)	(1.208)	(1.126)	(1.126)	(0.322)	(0.324)
$^{I^{RD}}/_{Rev}$, median	3.904**	3.438**	0.224	-0.025	3.680**	3.464**
	(0.541)	(0.876)	(0.479)	(0.814)	(0.136)	(0.255)
$^{RSGA}/_{Rev}$, median		0.543		0.291		0.253
		(0.593)		(0.527)		(0.219)
Constant	0.282**	0.201*	0.159**	0.116	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	0.438	0.443	0.286	0.290	0.583	0.587
	V^{IT}/Rev		V^{IT}_{cust}/Rev		V^{IT}_{other}/Rev	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B. Target level						
$^{I^{SM}}/_{Rev}$	1.558*	1.568*	1.576**	1.637**	-0.129	-0.198
	(0.587)	(0.649)	(0.517)	(0.514)	(0.541)	(0.596)
$^{I^{RD}}/_{Rev}$	5.200**	5.173**	0.153	-0.021	4.991**	5.189**
	(0.420)	(0.432)	(0.279)	(0.259)	(0.370)	(0.500)
$^{RSGA}/_{Rev}$		0.068		0.437		-0.497
		(0.820)		(0.242)		(0.716)
Constant	0.474**	0.465*	0.195**	0.137*	0.279*	0.345
	(0.146)	(0.221)	(0.049)	(0.055)	(0.137)	(0.209)
Observations	448	448	448	448	448	448
R^2	0.268	0.268	0.113	0.124	0.295	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. Robust standard errors are in parentheses. $^*p < 0.05, ^{**}p < 0.01$.

Table 10: Explaining Profits

	(1)	(2)	(3)	(4)	(5)
	π^{inc}/Rev	$r(K^{PH}+K^{IT})/Rev$	$\delta^{PH}K^{PH}+\delta^{IT}K^{IT}/Rev$	$(\gamma^{PH}K^{PH}+\gamma^{IT}K^{IT})rg/Rev$	π/Rev
Panel A					
I^{SM}/Rev, median	1.014**	-0.046	0.709*	0.052	0.292
	(0.309)	(0.211)	(0.320)	(0.033)	(0.346)
I^{RD}/Rev, median	1.543**	0.267	1.095	0.065	0.111
	(0.474)	(0.218)	(0.556)	(0.052)	(0.423)
Constant	0.158**	0.108**	0.127**	0.011**	-0.087**
	(0.014)	(0.015)	(0.014)	(0.001)	(0.020)
Observations	60	60	60	60	60
R^2	0.594	0.010	0.393	0.285	0.023
	(1)	(2)	(3)	(4)	
	$K^{PH}+K^{IT}/Rev$	r	δ	δ^{PH}	
Panel B					
I^{SM}/Rev, median	0.482	-0.043	0.275*	0.157	
	(2.344)	(0.049)	(0.106)	(0.080)	
I^{RD}/Rev, median	3.533	0.000	0.312*	0.549**	
	(2.716)	(0.059)	(0.142)	(0.107)	
Constant	1.359**	0.080**	0.113**	0.079**	
	(0.170)	(0.002)	(0.007)	(0.007)	
Observations	60	60	60	60	
R^2	0.025	0.035	0.272	0.335	

This table presents estimates from industry-level regressions of sales-weighted profit measures on measures of intangible capital investment. Robust standard errors are in parentheses. $^{*}p < 0.05, ^{**}p < 0.01$.

Figure 1: Sales and Marketing and Advertising Across the Distribution

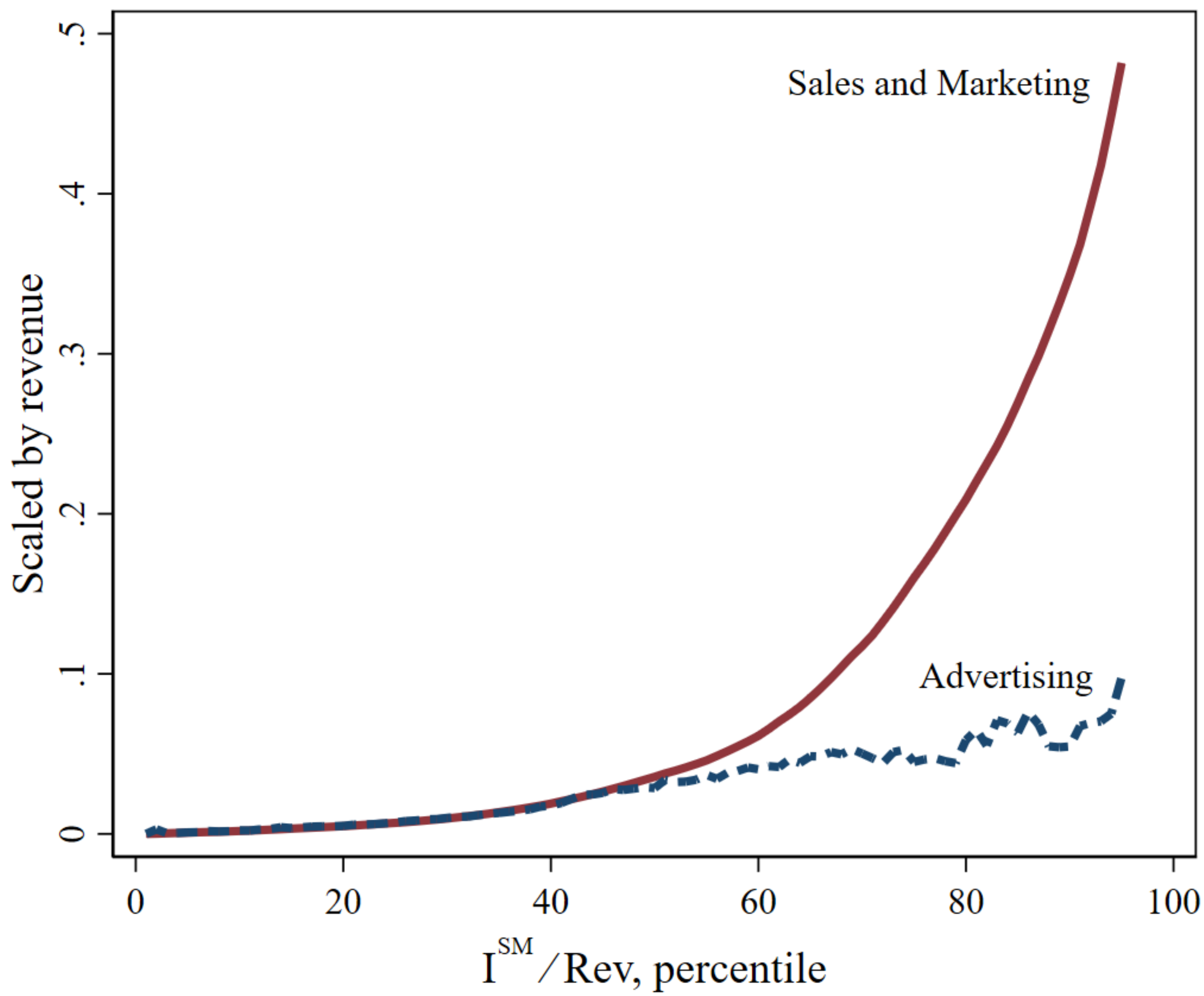
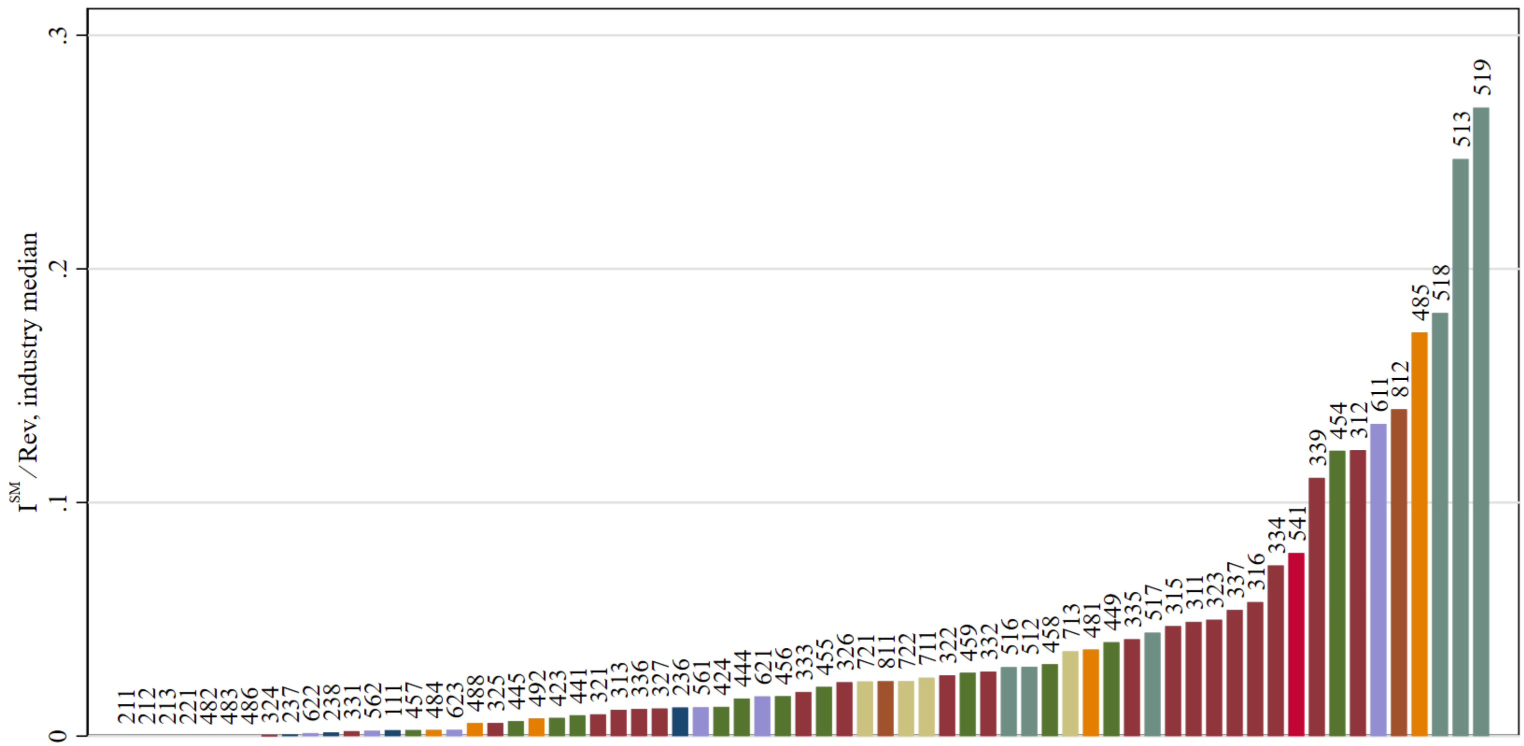


Figure 2: Median Ratio of Sales and Marketing Expense to Revenue Across Industries



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

Figure 3: The Persistence of Industry-Level ${}^{I^{SM}}\!/_{Rev}$

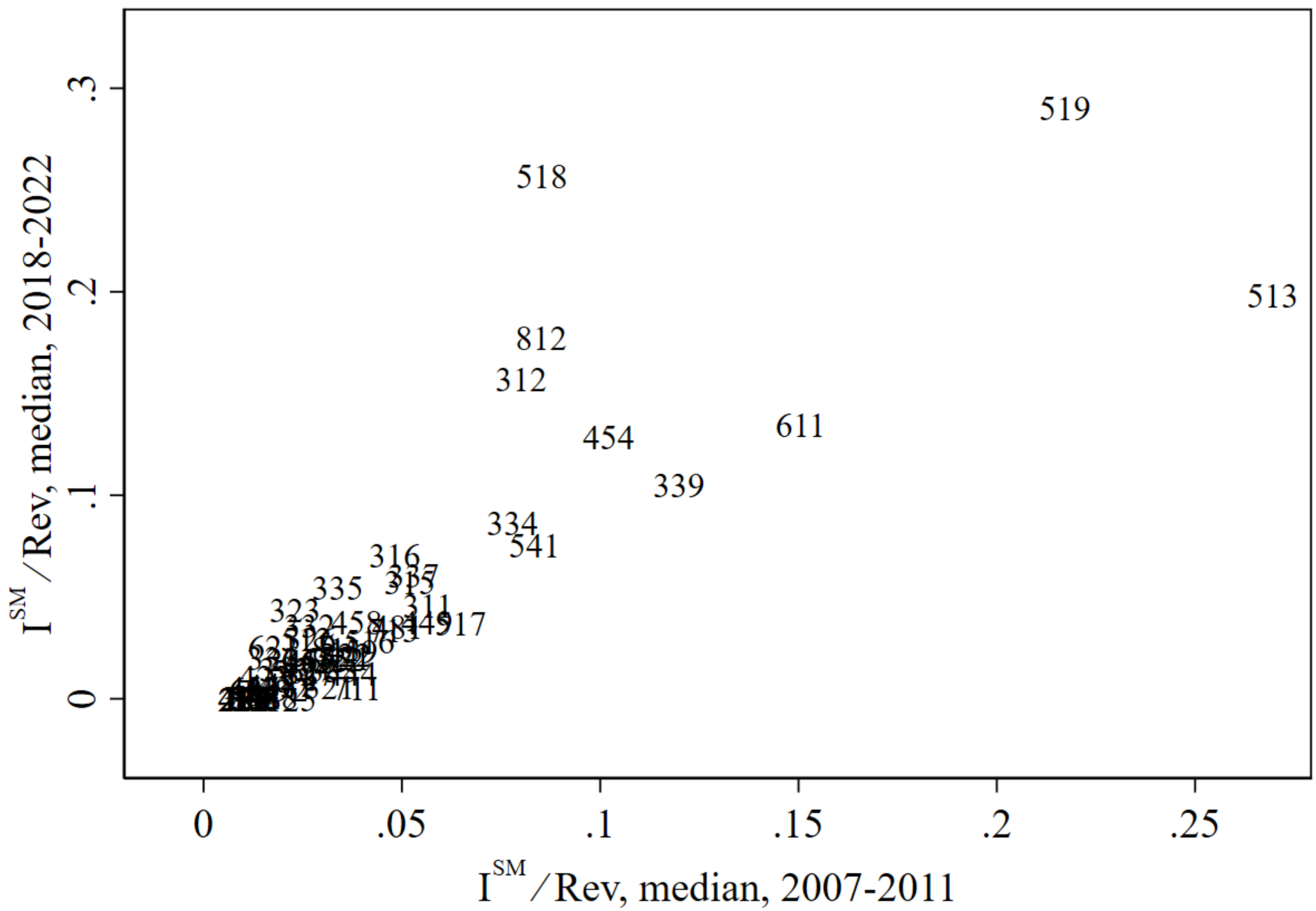
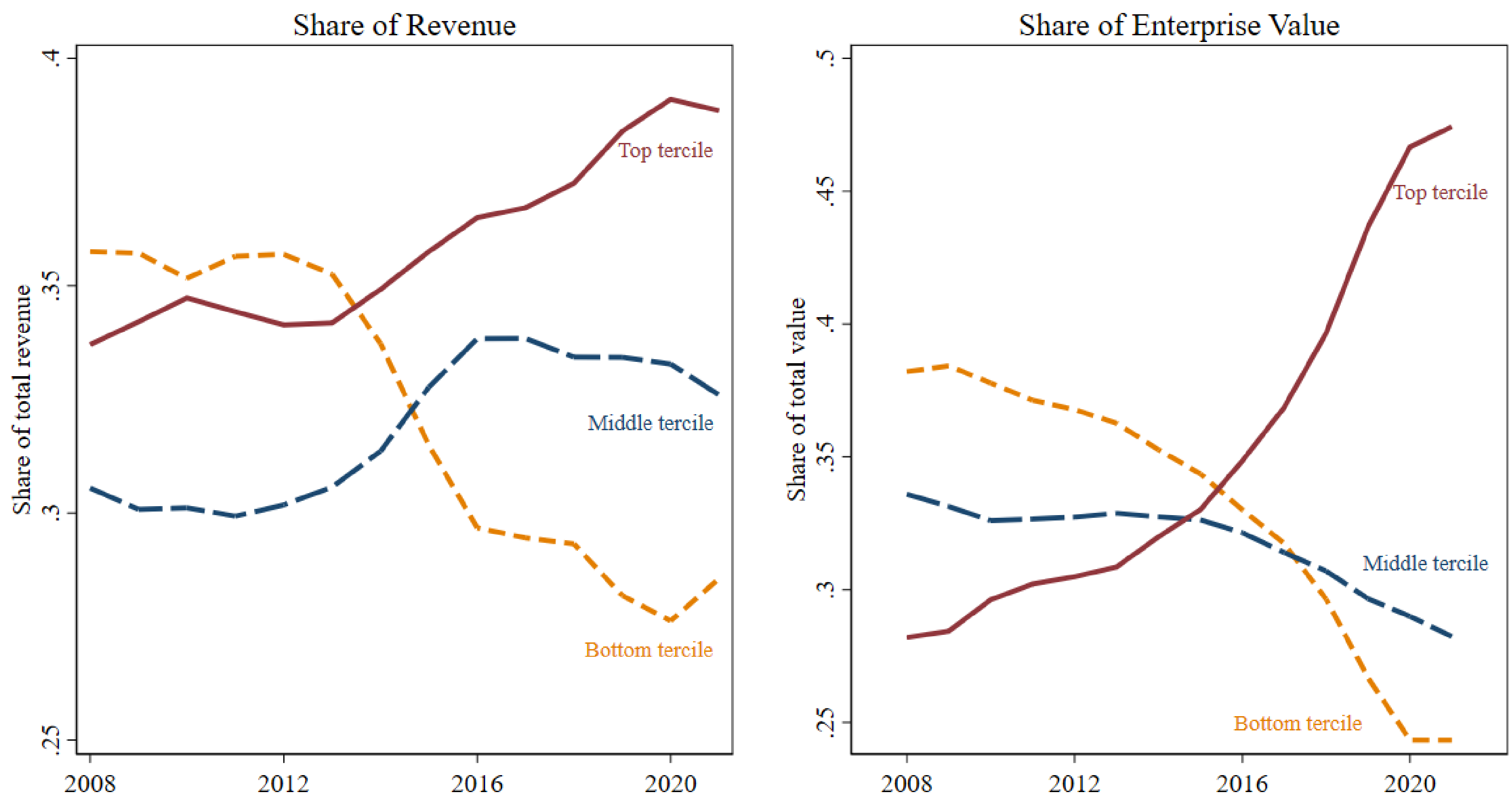


Figure 4: Shares of Revenue and Enterprise Value, by Industry-level $^{I^{SM}}/_{Rev}$



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

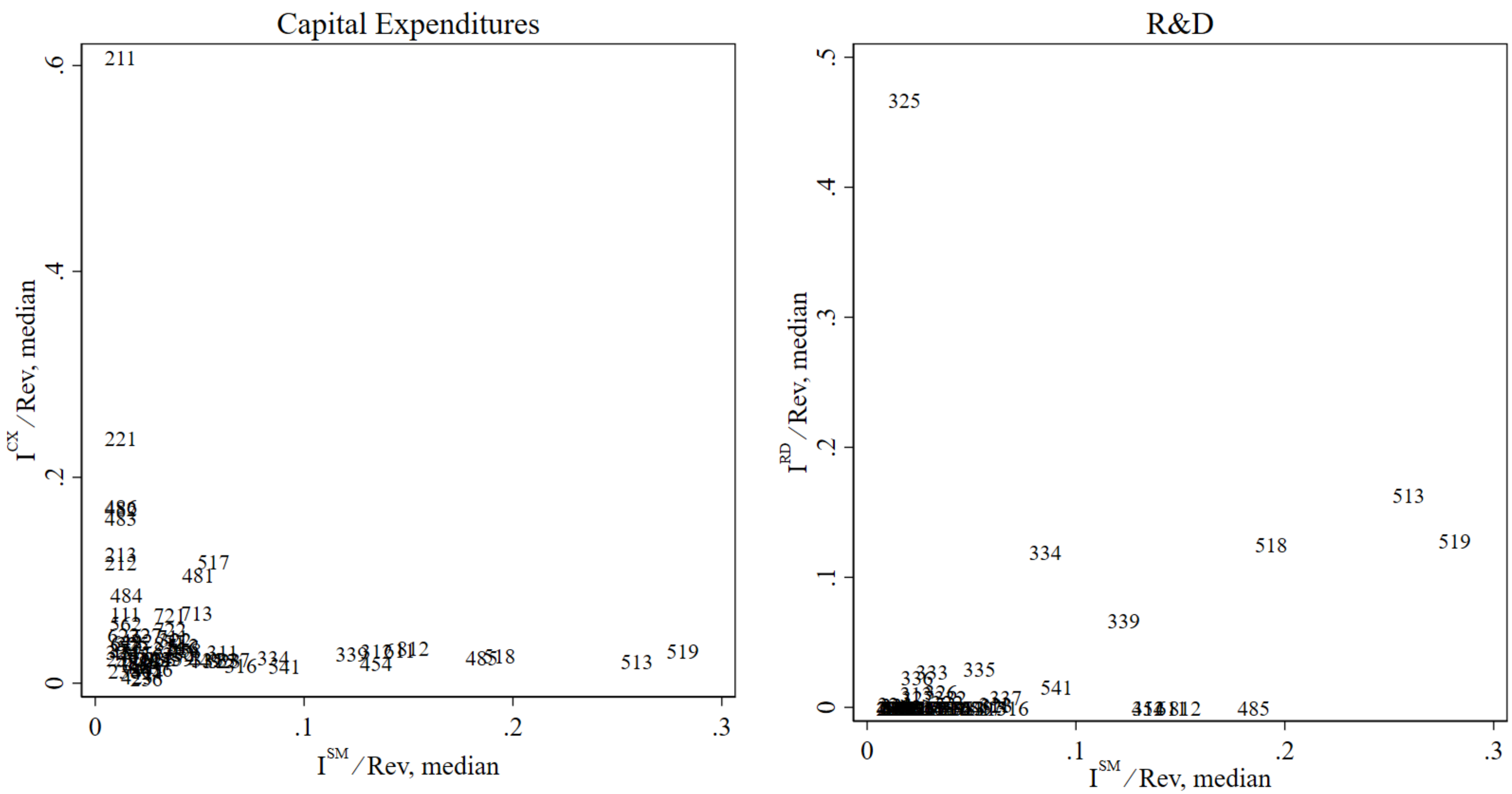
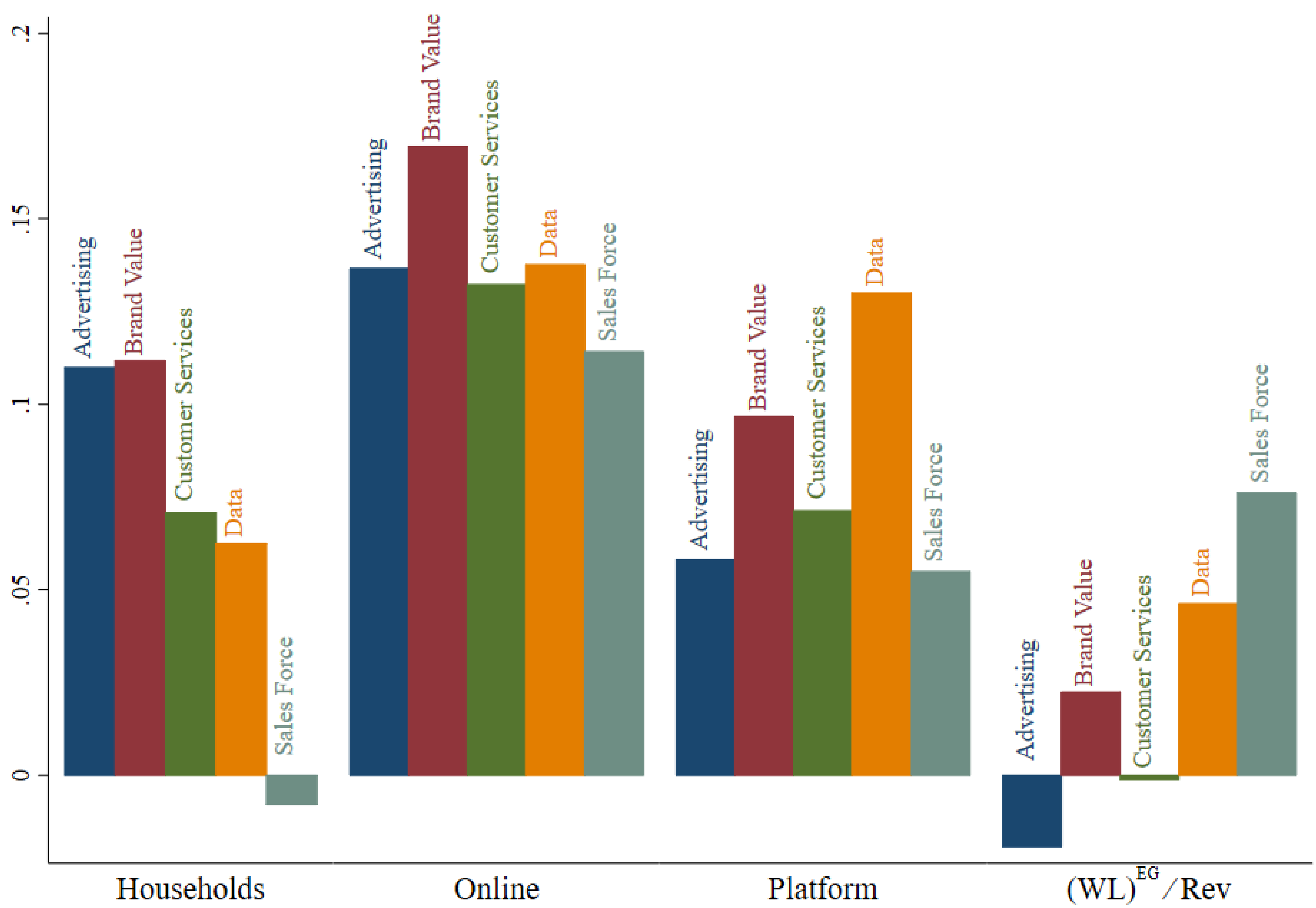
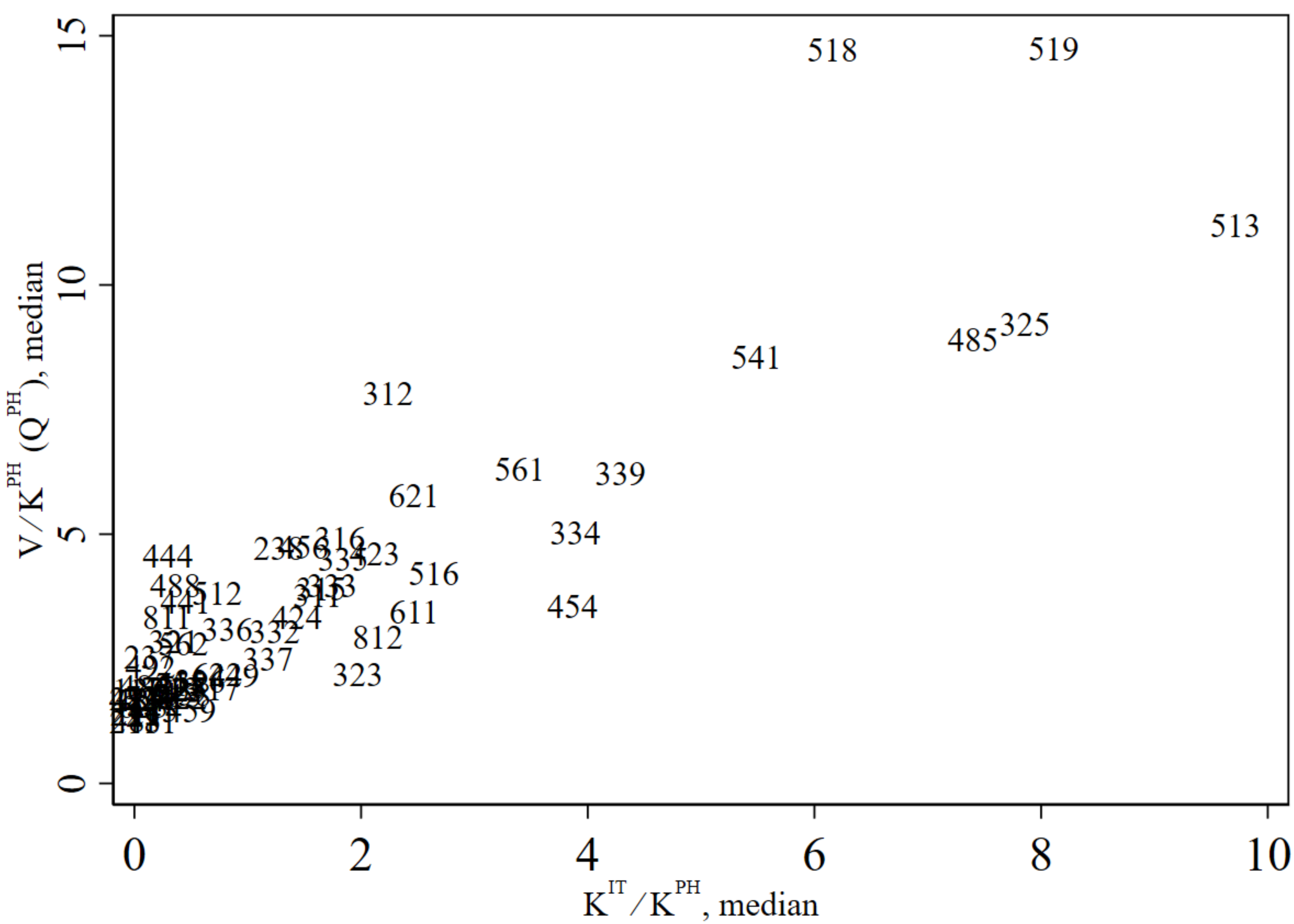


Figure 6: Explaining Sales and Marketing Strategies



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.

Figure 7: Explaining V/K^{PH}, or Q^{PH}



Appendix

A Text Data Collection and Processing

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

A.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.

Here is an excerpt from the 2019 10-K of The RealReal, Inc, a company that brings together sellers and buyers of luxury retail goods:

> A strong network effect drives the growth of our online marketplace. As we bring more consignors onto our platform, we unlock more high-quality, luxury supply, which increases our merchandise assortment and attracts more buyers. This, in turn, increases sales velocity and commissions for our consignors. In addition, a meaningful share of our consignors become buyers and vice versa, which creates a differentiated flywheel that enhances the network effect of our online marketplace.
>
> Our sales and service organization, as of December 31, 2019, included more than 240 luxury managers serving more than 40 major metropolitan markets in the United States and is responsible for obtaining exclusive supply for our online marketplace. Our sales professionals generate a robust pipeline of new consignors and build lasting relationships, which cannot be easily replicated. They consult on the consignment process and leverage data to advise consignors on pricing, expected selling time and market trends.
>
> *We deliver an end-to-end service experience.* We remove friction from the consignment process by providing multiple consignment methods: White Glove in-home consultation and pickup; drop off at one of our ten luxury consignment offices, four of which are located in our retail stores; or complimentary shipping directly to our merchandising and fulfillment facilities.
>
> *We do the work on behalf of consignor.* Once consigned items reach one of our four merchandising and fulfillment facilities, we authenticate, write the associated copy, photograph, price, sell and handle all fulfillment and returns logistics, making the consignment process seamless.
>
> *We generate high commissions for consignors.* Our scale and global reach combined with our technology-driven online marketplace and proprietary data enable consignors to realize optimal value for their pre-owned luxury goods. Our consignors earn up to 85% in commissions and achieved an average commission rate of approximately 64% in 2019.
>
> *We drive rapid monetization.* Our online marketplace efficiently matches supply with demand resulting in exceptional sales velocity. In 2019 and 2018, approximately 60% and 80% of the products on our online marketplace sold within 30 days and 90 days, respectively. In addition to sales velocity, we measure the ratio of demand versus supply in a given period, which we refer to as our online marketplace sell-through ratio. Sell-through ratio is defined as GMV in the period divided by the aggregate initial value of items added to our online marketplace in that period. In 2019 and 2018, our online marketplace sell-through ratios were 94% and 96%, respectively.

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 2.4. 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.

A.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 twelve 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."

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.

A.3 Gemini's description of what is in sales and marketing expense

The strategies undertaken by firms as part of sales and marketing were grouped based on our manual reading of the 150 firm-year observations as described above. We did not use Gemini to assist in categorizing the strategies. However, after the analysis was completed, we gave Gemini the following prompt, which is followed by its full response.

Prompt:

> "I am an economist seeking to better understand how firms spend resources to build and maintain a customer base. Firms often provide in their income statement a line that is called 'sales and marketing expense.' Can you tell me what exact types of spending fall within this category?"

Response:

> The "Sales and Marketing Expense" line item on a U.S. publicly traded firm's income statement is a broad category encompassing a variety of costs aimed at generating sales and building customer relationships. While the specific breakdown isn't mandated to be hyper-detailed in the financial statements themselves (following GAAP, Generally Accepted Accounting Principles), a thorough understanding requires looking beyond the single line item. SEC filings (like 10-Ks) often provide more granular detail in the footnotes or management discussion and analysis (MD&A).
>
> Here's a comprehensive breakdown of the types of spending that typically fall under Sales and Marketing Expense, categorized for clarity:
>
> **Advertising & Promotion:**
>
> - Traditional Media: Television, radio, print (newspapers, magazines), billboards, direct mail.
> - Digital Marketing: Search engine marketing (SEM, including PPC ads), social media marketing (paid campaigns), display advertising (banner ads), email marketing, influencer marketing.

- Public Relations (PR): Press releases, media outreach, event sponsorships, crisis management (related to reputation).
- Trade Shows and Conferences: Participation fees, booth rentals, travel expenses for attendees.
- Sales Literature & Collateral: Brochures, catalogs, website materials, presentations.
- Promotional Campaigns: Discounts, coupons, rebates, contests, giveaways, loyalty programs (the expense portion, not the liability for future rewards).

Sales Force Compensation & Operations:

- Salaries and Wages: For sales representatives, sales managers, and marketing personnel.
- Commissions: Paid to sales representatives based on sales performance.
- Bonuses: Performance-based incentives for sales teams.
- Travel and Entertainment: Expenses related to client meetings, conferences, and business trips.
- Sales Training: Costs associated with educating and developing sales personnel.
- Sales Support Staff: Salaries and wages for individuals providing administrative support to the sales team.

Market Research & Analysis:

- Customer surveys and focus groups: Gathering data on customer preferences and needs.
- Market analysis studies: Assessing market size, competition, and trends.
- Data analytics and CRM software: Tools for tracking customer behavior and managing relationships. (Note: The software purchase is often capitalized as an asset and depreciated, while maintenance and software subscriptions are expensed).

Customer Relationship Management (CRM):

- Software licenses and maintenance: Costs associated with CRM software.
- Personnel costs: Salaries of staff managing and utilizing CRM systems. Data acquisition and integration: Costs related to obtaining and integrating customer data.

Other Potential Items (depending on the nature of the business):

- Website development and maintenance: (Can be capitalized or expensed, depending on the nature of the expense and company policy).
- Branding and design: Costs associated with creating and maintaining the company's brand identity. Warranty costs (if tied to sales promotion): Costs of replacing or repairing defective products under warranty, especially if offered as a sales incentive.

Important Accounting Considerations:

- Materiality: Items are typically grouped based on materiality. Insignificant expenses might be aggregated.
- Capitalization vs. Expensing: While most sales and marketing costs are expensed in the period incurred, some long-term assets (like major website overhauls or software development) might be capitalized and amortized over their useful lives. This impacts both the income statement and balance sheet.
- Matching Principle: GAAP requires that expenses be recognized in the same period as the revenues they generate. Therefore, the timing of recognition for sales and marketing expenses is crucial.

It's crucial to examine the footnotes to the financial statements for a more comprehensive understanding of how a specific company categorizes and reports its sales and marketing expenses. This provides a level of detail not usually visible in the main income statement.

B Model derivation

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

$$\lambda = \frac{\partial C_D(\cdot)}{\partial D} \tag{9}$$

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

$$W^D = \lambda \frac{\partial \mathcal{D}(\cdot)}{\partial X^D} \tag{10}$$

$$W^Q = \zeta \frac{\partial \mathcal{F}(\cdot)}{\partial X^Q} \tag{11}$$

Profit maximization implies

$$\frac{\partial \mathcal{P}(\cdot)}{\partial D} Q = \frac{\partial C_D(\cdot)}{\partial D} \tag{12}$$

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

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 (12), this implies

$$\rho = \frac{D}{PQ}\frac{\partial C_D(\cdot)}{\partial D} \tag{13}$$

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

$$\theta^D \equiv \frac{X^D}{D}\frac{\partial \mathcal{D}(\cdot)}{\partial X^D} \tag{14}$$

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 (9), (10), and (12), 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 (13) and (14) 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.

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

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 and the accounting methods used for their valuation.

In PPA, the types of assets and their corresponding valuations for the target firm are reported.[36] 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.

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

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 A6. 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.

D 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 rollowing 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:"

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

		I^{SM}/Rev, median
Agri, Mining, Const, Utilities		
111	Crop Production	0.003
211	Oil & Gas Extraction	0.000
212	Mining (Non-Oil & Gas)	0.000
213	Mining Support	0.000
221	Utilities	0.000
236	Buildings Constr	0.012
237	Heavy & Civil Eng	0.001
238	Specialty Trade Contractors	0.002
Manufacturing		
311	Food Mfg	0.049
312	Bev & Tobacco Mfg	0.122
313	Textile Mills	0.011
315	Apparel Mfg	0.047
316	Leather Products	0.057
321	Wood Prod Mfg	0.009
322	Paper Mfg	0.026
323	Printing Services	0.050
324	Petrol & Coal Prod Mfg	0.000
325	Chemical Mfg	0.006
326	Plastics & Rubber Prod	0.023
327	Mineral Prod Mfg	0.012
331	Primary Metal Mfg	0.002
332	Fabricated Metal Mfg	0.028
333	Machinery Mfg	0.019
334	Computer & Electronics Prod	0.073
335	Electrical Equip Mfg	0.041
336	Transp Equip 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	Furniture & Elec Retail	0.040
454	Nonstore Retail	0.122
455	General Merch Retail	0.021
456	Health Care Retail	0.017
457	Fuel Dealers	0.003
458	Clothing & Accs Retail	0.031
459	Misc Retail	0.027

		I^{SM}/Rev, median
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	Information Services	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	Recreational Svcs	0.036
721	Accommodation	0.023
722	Food & Drink Svcs	0.023
Maintenance & Personal Services		
811	Repair & Maintenance	0.023
812	Personal Services	0.140

Table A3: Additional Summary Statistics

	N	Average	Median	Std Dev
Panel A: Industry-level Covariates				
Households	67	0.466	0.413	0.342
Online	67	0.404	0.371	0.281
Platform	67	0.137	0.074	0.158
$(WL)^{EG}/Rev$	63	0.050	0.029	0.052
Panel B: Measures of Profits				
π^{inc}/Rev	60	0.219	0.187	0.135
$r(K^{PH}+K^{IT})/Rev$	60	0.110	0.093	0.082
$\delta^{PH}K^{PH}+\delta^{IT}K^{IT}/Rev$	60	0.169	0.150	0.117
$(\gamma^{PH}K^{PH}+\gamma^{IT}K^{IT})rg/Rev$	60	0.014	0.013	0.009
π/Rev	60	-0.075	-0.046	0.122
$K^{PH}+K^{IT}/Rev$	60	1.430	1.183	0.973
r	60	0.078	0.078	0.012
δ	60	0.128	0.132	0.047
δ^{PH}	60	0.093	0.077	0.047

Table A4: Explaining Variation in I^{SM}/K across Industries

	Dependent variable: I^{SM}/K, median				
	(1)	(2)	(3)	(4)	(5)
Households	0.015				0.002
	(0.011)				(0.009)
Online		0.074^{***}			0.059^{***}
		(0.011)			(0.013)
Platform			0.128^{***}		0.041
			(0.019)		(0.025)
$(WL)^{EG}/K$, median				0.357^{**}	0.261^{**}
				(0.112)	(0.080)
Constant	0.025^{***}	0.002	0.014^{***}	0.018^{***}	-0.009
	(0.006)	(0.003)	(0.003)	(0.005)	(0.005)
Observations	67	67	67	63	63
R^2	0.025	0.419	0.395	0.132	0.585

Table A5: Returns to Scale and Investment in Customer Capital

	Dependent variable: $^{I^{SM}}/_{Rev}$, median					
	(1)	(2)	(3)	(4)	(5)	(6)
RTS$_{PC}$	-0.234			-0.228		
	(0.238)			(0.146)		
RTS$_{IC}$		0.026			0.003	
		(0.071)			(0.040)	
RTS$_{L}$			0.063			-1.074
			(1.266)			(0.621)
Constant	0.271	0.016	-0.018	0.266	0.041	1.137
	(0.235)	(0.075)	(1.282)	(0.144)	(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 A6: 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: Comparison of I^{SM} with Advertising and Residual SG&A