

Online appendix to “Intangibles, Investment and Efficiency”

Nicolas Crouzet*

Janice Eberly†

This version: January 4, 2018

A Data appendix

This section summarizes the data sources, the construction of variables, and the creation of the samples which we use in the main text of the paper.

A.1 Aggregate data

We use the following data series as deflators: the consumer price index for all urban consumers (FRED series CPIAUSCL); the gross domestic product implicit price deflator (FRED series GDPDEF); the fixed nonresidential investment implicit price deflator (FRED series A008RD3Q086SBEA). Additionally, we retrieve a measure of multi-factor productivity for the retail sector from the KLEMS database. Specifically, we use the base 2009 multifactor productivity annual time series for NAICS sectors 44-45 (line 39, table 10.2 from the KLEMS database). All these public data sources are included in the online replication files, in the folder [/input/](#).

A.2 Firm-level data

A.2.1 Data sources

Our firm-level data source is the Compustat-CRSP merged database. Specifically, we download from WRDS the Compustat fundamentals annual files (names, identifiers, and data items), the CRSP file, and the bridge between the two, merge the the fundamentals annual to the CRSP file using the bridge, and apply standard screens. The shell script [/input/rundirectory.sh](#) calls the stata scripts [/code/download_from_wrds.do](#) and [/code/merge_data.do](#) and can be used to replicate our construction of the data, provided the user has WRDS credentials; see [/code/readme.txt](#) for more details on executing the script.

A.2.2 Sample selection

We keep firms-year observations incorporated in the USA (Compustat variable `fic` is “USA”). Our main industry identifier is the 3-digit NAICS code, which we obtain by extracting the first three digits of the NAICS code associated with the firm’s `naics` code (Compustat variable `naics`), as

*Northwestern University, Kellogg School of Management, e-mail: n-crouzet@kellogg.northwestern.edu

†Northwestern University, Kellogg School of Management, e-mail: eberly@kellogg.northwestern.edu

given in the WRDS file `/nam/name`.¹ We drop firm-year observations with 3-digit NAICS codes 999 (multinationals), 22x, 52x and 53x (utilities, finance, and real estate), and with missing NAICS or less than 3-character long NAICS codes. We drop firm-year observations with missing or weakly negative sales, missing or weakly below 1m\$ book assets, and missing market value.² Finally, we keep only firm-year observations between 1989 and 2015. The resulting baseline annual sample contains 263101 observations from 24412 firms. Our time variable is the Compustat variable `fyear` and our firm identifier is the Compustat variable `gvkey`. The stata script `/code/prepare_panel.do` replicates the selection of the sample.

In order to construct **figure 1**, we add to this baseline sample the Fama-French 49 industrial classification. This classification is constructed using the Stata program `ffind`, which takes as an input the firm-year observation’s SIC code, which we take to be the Compustat variable `sic`, also retrieved from the `name` file on WRDS. In figure 1, the solid line is constructed using all firm-year observations from the baseline annual sample; the dashed line drops firm-year observations from the FF 43 group (retail); the dotted line additionally drops all observations from the FF 30 group (petroleum and natural gas). The FF 43 group contains almost exclusively firms from the NAICS 3-digit sectors 441-448 and 451-454. Section B.1 of this appendix contains more details on the dynamics of concentration in the oil & gas sector.

In order to construct **figures 2, 3 and 4**, we further select firm-year observations satisfying the following criteria: the firm-year observation is in the FF-43 group; employment is non-missing and positive; cost of goods sold is non-missing and positive; earnings before interest, taxes, depreciation and amortization is non-missing; debt currently due and due in more than one year are non-missing; current assets are non-missing; capital expenditures are non-missing; gross property, plant and equipment are non-missing. This reduces the sample to 6259 observations for 662 distinct firms. Furthermore, in order to construct **figure 3**, we require that the firm-year has information on Q and cash flows one year prior, and has positive values for Q and cash flows. We also winsorize investment rates at the top and bottom 1%, and Q and cash flow to assets at the top 1%. This leaves 4083 observations for 507 distinct firms in the sample used to construct figure 3.

A.2.3 Variable definitions and details of construction of figures

All figures are constructed by the stata scripts `/code/figure_x.do`, where `x` is the number of the figure in the paper.

Figure 1 Figure 1 only uses the Fama-French classification, as well as firm-level sales (Compustat variable `sale`). For each of the three cases, the line reported is the average HHI of sales across NAICS 3-digit sectors, weighting HHIs by their sector’s share of total (aggregate) sales.

Figure 2 In all four panels of figure 2, the solid line is the average HHI of sales across NAICS 3-digit sectors, weighting HHIs by their sector’s share of total (aggregate) sales. In the top panel, the dashed line is the ratio of revenue per employee. Revenue, or sales (Compustat variable `sale`), is deflated using the CPI for all urban consumers; the number of employees if the Compustat variable `emp`. This ratio is then winsorized at the top and bottom 1%. The dotted line is the sales-weighted average of the ratio across firms in the FF-43 group. For markups (the bottom left panel), we proceed identically, with markups defined as the ratio of sales to cost of goods sold (Compustat

¹See `/input/rundirectory.sh` for a list of the precise location of the files we download from the WRDS server.

²See below for a definition of these variable based on primitive data items.

variable `cogs`), both undeflated. For inventory needs, we also proceed in this way, with inventory needs defined as the ratio of the inventory stock (Compustat variable `invnt`) to sales, multiplied by twelve in order to express it in months.

Figure 3 In figure 3, the investment rate is defined as the ratio of capital expenditures (Compustat variable `capex`) to gross property, plant and equipment (Compustat variable `ppegt`), both undeflated. We measure Q as the ratio of (market value of equity + book value of debt – book value of current assets) to gross property, plant and equipment. We deflate the numerator by the GDP deflator and the denominator by the price deflator for non-residential investment. The market value of equity is defined as the product of the CRSP variables `prc` and `shrout`.³ The book value of debt is the sum of the Compustat variables `dlc` and `dltt`, and the book value of current assets is the Compustat variable `act`. Finally, the ratio of cash flow to assets is measured as `ebitda` (Compustat variable `ebitda`) to gross property, plant and equipment; the numerator is deflated by the GDP deflator, and the denominator is deflated by the price deflator for non-residential investment.

Figure 4 In figure 4, we use the same measure as in figure 2 for revenue per employee. The intangible share is computed as the ratio of balance sheet intangibles (Compustat variable `intan`) to total book assets (Compustat variable `at`). Section B.4 of this appendix reports results using a wider measure of intangible assets, constructed by Peters and Taylor (2017). The right panel uses the weighted average intangible share within 3-digit NAICS industry, where the weights are firm-level sales; a point on this graph is a year/NAICS 3-digit sector. The left panel of figure 4 contains the weighted average of the intangible share across 3-digit NAICS sectors.

B Additional results

This section provides additional results mentioned in the main text.

B.1 The role of the oil & gas sector

Figure 1 provides a description of the effect of changes in concentration in the oil & gas sector on aggregate concentration. We start with the same sample as in figure 1. Different from figure 1, firm-level data in the oil and gas sector has been restated to reflected three important mergers which took place in late 1990’s and early 2000’s: the Conoco/Phillips merger, the Exxon/Mobil merger and the Chevron/Texaco merger. Prior to the merger year, restating is done by adding the sales of the firm that is eventually absorbed in Compustat to the firm that survives; for instance, the sales of Conoco are added to the sales of ConocoPhillips, which exists in Compustat, under this name, prior to the merger. The oil and gas sector is defined as the firms in the Fama-French group 30; the retail firms are defined as those in the Fama-French group 43.

With the restated oil and gas sector, concentration is somewhat higher than in figure 1 of the main paper, especially in the earlier parts of the sampel (1990-1996). The dashed line shows concentration dropping the (re-stated) oil and gas sector; it shows that after re-stating these three mergers, the sector does not add to the trend increase in concentration after 1995. Finally, the dotted line drops both the (re-stated) oil and gas sector, and the retail sector, and is identical to the dotted line in figure 1. The gap between the dashed and the dotted lines in figure 1 then

³The variable is constructed by the script [/code/merge_data.do](#).

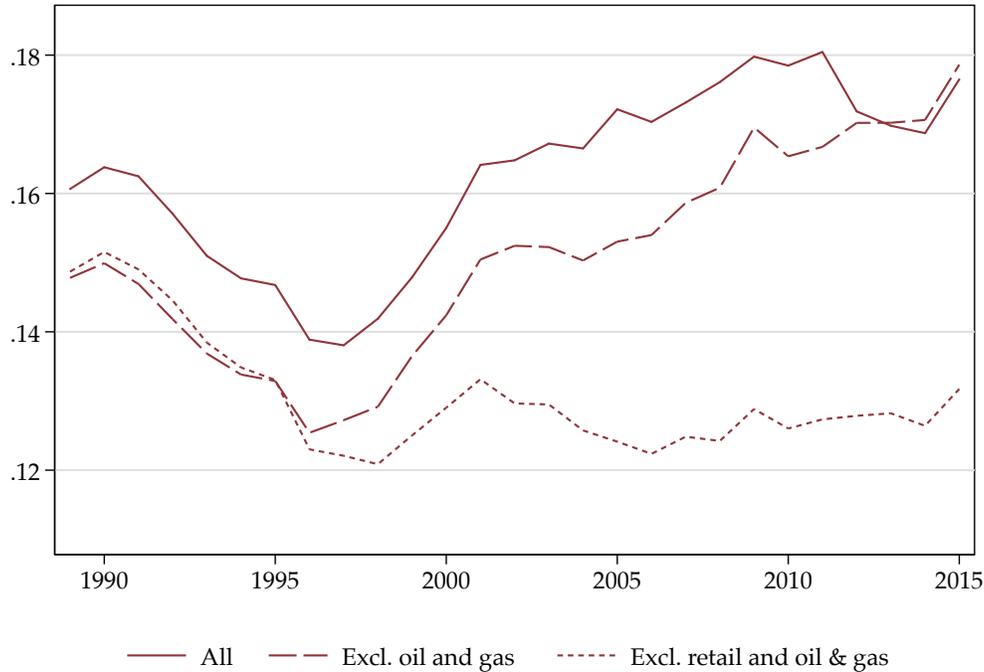


Figure 1: The role of the oil and gas sector in the evolution of aggregate concentration. Each line represents the increase in the average HHI of sales across NAICS 3-digit sectors. The average HHI is weighted by the sector’s share of total sales. The oil and gas sector has been restated to reflect important mergers. See text for details on the construction of each line.

indicates that, putting the oil and gas sector aside, retail accounts for virtually all of the increase in concentration in sales among publicly traded US firms since 1995.

B.2 The increase in concentration within NAICS 3-digit sectors in retail

Figure 2 reports the sales HHI for the different sub-sectors that make up the Fama-French retail group. We drop sectors 337, 339, 421 and 722 because there are too few observations to compute a complete HHI series (either in the time or in the panel dimension). We are left with the three-digit NAICS sectors 441-448 and 451-454. The most important one, in terms of its sales share, is the General Merchandise Stores sector (452), which contains the big-box stores. The increases in sales HHI is however widespread, with the notable exceptions of Motor Vehicle Dealers and Gasoline Stations.

B.3 Markups at the largest retail firms

Figure 3 reports the average markup, measured as the ratio of sales to cost of goods sold, for all firms (dashed blue line) and for the three largest firms, in each year, by sales (crossed blue line). The dashed blue line is identical to the markup line reported in the bottom panel of figure 2; in both cases, to maintain comparability with the main text, the averages are weighted by firm sales. The top three firms of each NAICS 3-digit group thus tend to have slightly lower markups, on

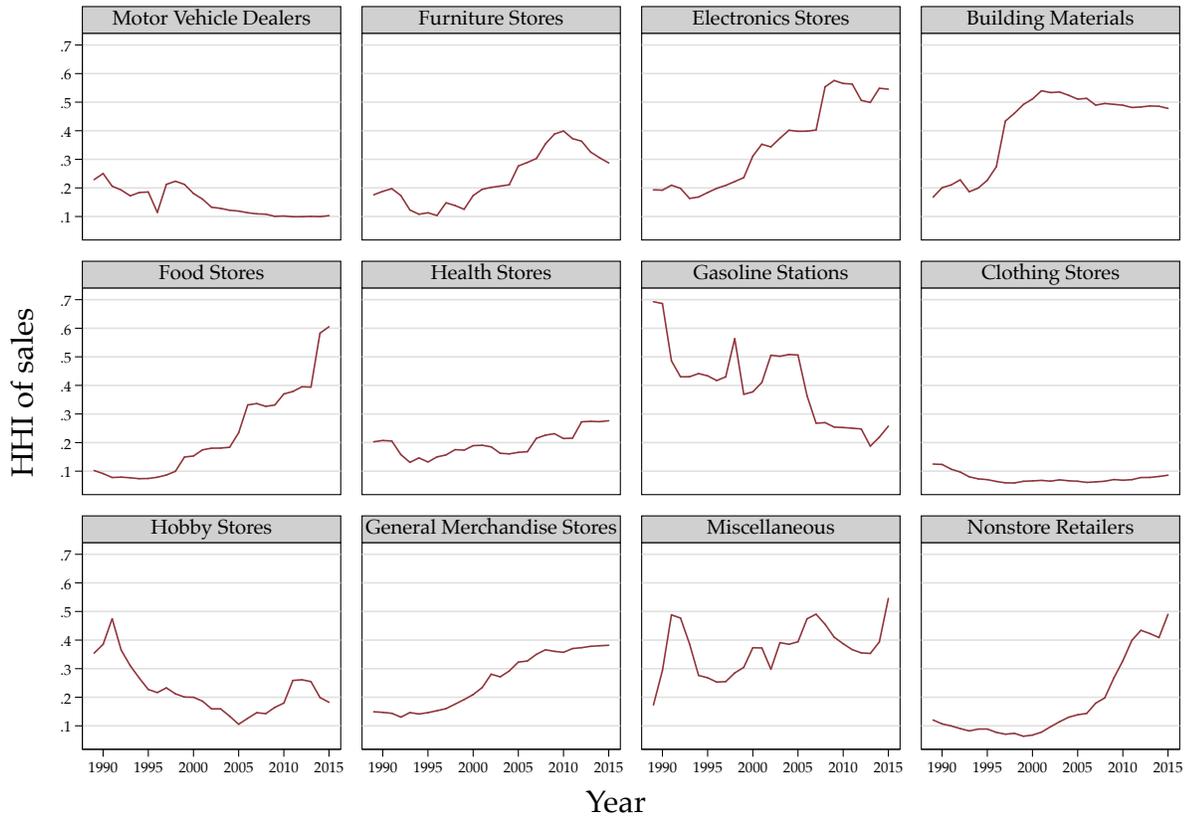


Figure 2: Sales HHI for the different sub-sectors that make up the Fama-French retail group.

average, than the rest; moreover, these markups are somewhat declining toward the end of the sample.

B.4 Analysis with the Peters-Taylor (2017) intangible stock measures

Figure 4 replicates the exercises reported in figure 4 of the paper, with different measures of the stock of intangible assets of firms. Specifically, we use the dataset constructed by Peters and Taylor (2017) and available on WRDS.⁴ This dataset contains measures for two additional potential intangible stocks: “knowledge” capital, computed as capitalized expenditures on research and development; and “organizational” capital, computed as (a constant share of) capitalized selling, general and administrative expenses. The top panel of figure 4 repeats the same exercise as figure 4 in the paper, but defining intangible assets as the sum of book intangibles (Compustat variable `intan`) and the Peters-Taylor knowledge capital variable; and defining total assets as the sum of intangible assets and gross property, plant and equipment. The results are qualitatively identical to those of the main paper, and quantitatively close.

The bottom panel of figure 4 then adds the “organizational” capital stock measure of Peters and Taylor (2017), to both the numerator and the denominator of the intangible share. The correlation between the log of the intangible share and the log of revenue per employee at the NAICS 3-digit

⁴This dataset is also downloaded when executing the shell script `/input/rundirectory.sh`.

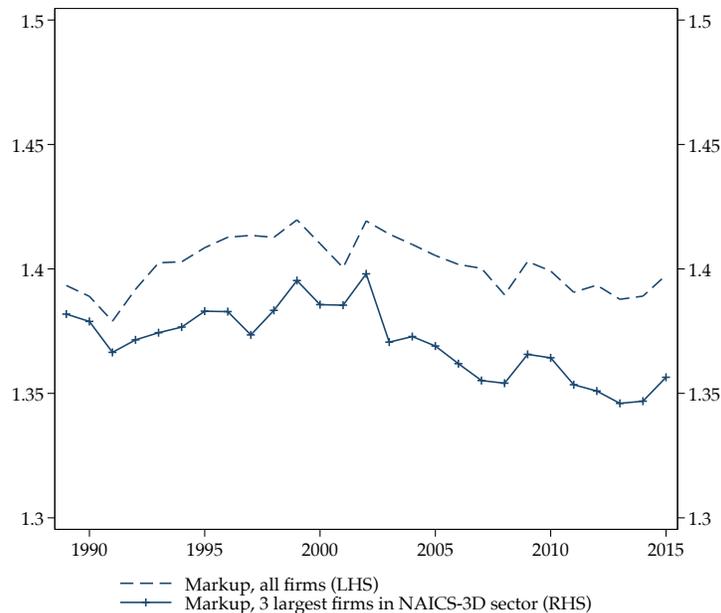


Figure 3: Average markups for firms overall and for the largest three firms in each NAICS 3-digit sector. In each case, the averages are computed by weighting by firm sales, as in figure 2. The dashed blue line is identical to the dashed blue line in the bottom left panel of figure 2.

sector level remains; although the estimate of the OLS coefficient changes in magnitude (as the typical intangible share is now substantially higher than in the previous measures), it remains significant, with a p-value of 2.9%. Moreover, after 1995, the aggregate trend in the intangible share is consistent with the previous findings. The main difference with the results presented in the paper is that there is a trend downwards in the intangible share measured in this manner, prior to 1995. This trend is driven by the fact that the ratio of capitalized SG&A to total assets is falling over time, a fact that persists throughout the sample (but overtaken by the growth in intangibles and R&D capital in the latter parts of the sample).

A potential reason for the downward trend in the SG&A capital stock, relative to assets, is that in retail, SG&A expenses seem to be inconsistently defined across firms. In particular, in some cases, they include payments to labor at retail locations. If payments to this form of labor have been declining over time, a downward trend in the SG&A stock may result. According to their 2015 10Ks, Walmart and Walgreens both allocate wages and salaries to SG&A, while Costco splits them between SG&A and cost of sales, based on whether they are paid to warehouse or store employees. The possibility that SG&A is inconsistently treated across retail firms, and may in some primarily reflect wages and salaries of non-administrative workers and in others not, as well as the fact that the results with and without the capitalized R&D measure are fairly similar, motivates our not focus on book intangibles in the main paper.

B.5 Acquisition expenditures and the intangibles share

Figure 5 provides a back-of-the envelope computation for what the path of the intangible share of assets should have been, given observed acquisition expenditures, and assuming that assets valuations remained consistent with the pre-1995 values. The idea is the following. Assume that

acquisition expenditures, $P_{i,t}$, are related to the book value of assets of the target, $A_{i,t}$, through:

$$P_{i,t} = Q_t A_{i,t},$$

where Q_t is the average value of a measure of Tobin’s Q in the sector. The amount of goodwill involved in the purchase is then $G_{i,t} = (Q_t - 1)A_{i,t} = \left(1 - \frac{1}{Q_t}\right)P_{i,t}$. This amount of goodwill corresponds to the increase in balance sheet intangibles that the acquisition will generate. For each firm-year observation, we thus estimate the amount of goodwill generated by acquisitions using this formula. For Q_t , we use the average of the ratio of book debt (the sum of Compustat variables `d1c` and `d1tt`) and the market value of equity (as defined in section A.2.3), divided by the book value of assets (Compustat variables `at`). This average is equal to 1.45 in the sample used in figures 2, 3 and 4. For acquisition expenditures, we use the Compustat variable `aqc`; we replace it by 0 when it is either missing or negative. We then take the cumulative sum of acquisition expenditures and add it to the initial value of book intangibles, computed in 1989.⁵

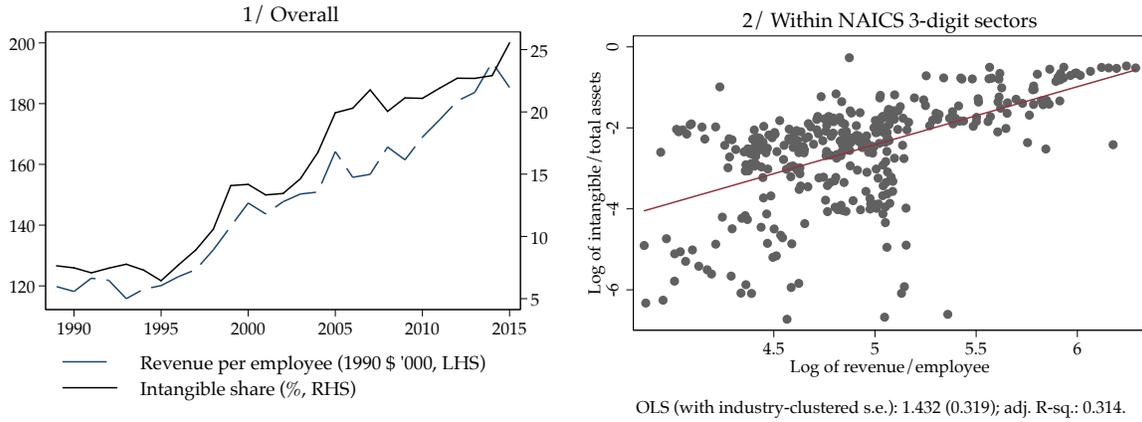
The resulting (counterfactual) change in the intangible share of total assets is the dashed line of figure 5; the solid black line is the actual change in the intangible share. (Both coincide in 1989 by construction). From 1995 onwards, the counterfactual intangible share increases only from 3.1% to 6.1%, instead of the 4.1% to 18.6% actually observed. This suggests that, relative to the pre-1995 period, an increasing fraction of acquisition expenditures among retail firms consistutes the acquisition of intangible capital, rather than traditional assets.

B.6 The investment gap with intangible investment measures

Figure 6 conducts an exercise similar to that of figure 3 in the main text, but for two distinct measures of investment: “traditional” investment (capital expenditures), as in the main text; and “total” investment, which also includes a measure of intangible investment. We construct intangible investment as the sum of net acquisitions, R&D expenditures, and expenditures on advertising. Net acquisitions are defined as the difference between Compustat variables `aqc` and `sppe` (using only `aqc` does not affect the results). R&D expenditures are Compustat variable `xrd`, and advertising expenditures are `xad`, both set to missing if they are observed to be negative. We do not include (a portion of) SG&A expenditures for the reasons outlined in section B.4. Furthermore, we deflate intangible investment by the GDP deflator, instead of the price deflator for non-residential fixed investment. We then define the total investment rate as the ratio of intangible plus physical investment, divided by total book assets (the latter deflated by the GDP deflator). Likewise, we use book assets in the numerator of our measures of Q and the ratio of cash flow to assets when we estimate the “total investment” regression. Finally, we select a more restricted sample than in figure 3, because we require our measure of intangible investment to be populated. The sample has 2180 observations for 331 distinct firms. The resulting investment gap is smaller, on average, than for traditional investment only; the cumulative gap is 3.7%, compared to 10.7%. (The point estimate of the cumulative investment gap differs slightly from what is reported in figure 3 of the main paper, because of the differences in the sample used in the estimation).

⁵Note that we adopt this procedure because we have no good data source on the book value of targets acquired. We have explored the possibility of using the SDC platinum data; while it records a large number of acquisitions and divestitures, the book value of targets’ assets is seldom reported, at least in the retail sector.

Peters-Taylor measure of intangible capital
(including only book intangibles and capitalized R&D)



Peters-Taylor measure of intangible capital
(including book intangibles, capitalized R&D and capitalized SG&A)

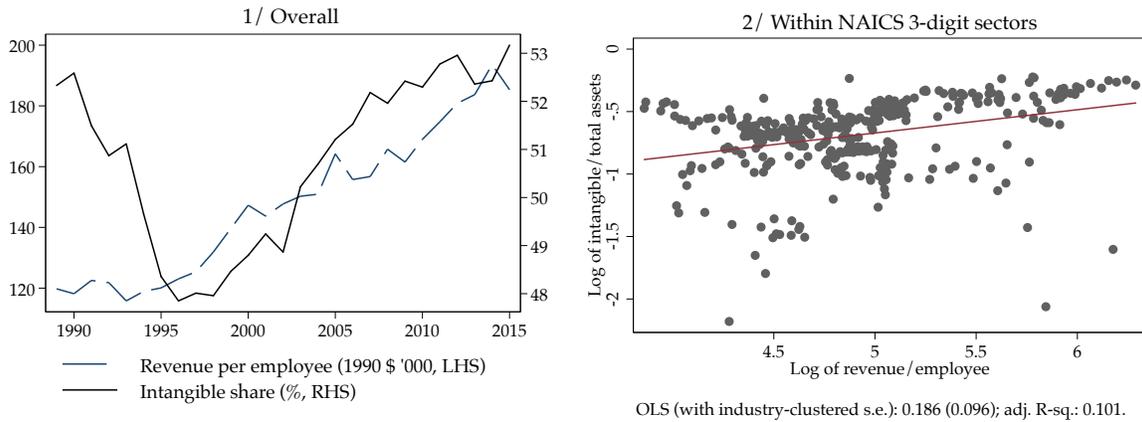


Figure 4: The relationship between intangible capital and labor revenue productivity, using the Peters and Taylor (2017) measures of intangible capital stock. The top panel contains uses only the R&D capital stock measure, while the bottom panel uses both the R&D and the SG&A capital stock measure. See text for more details.

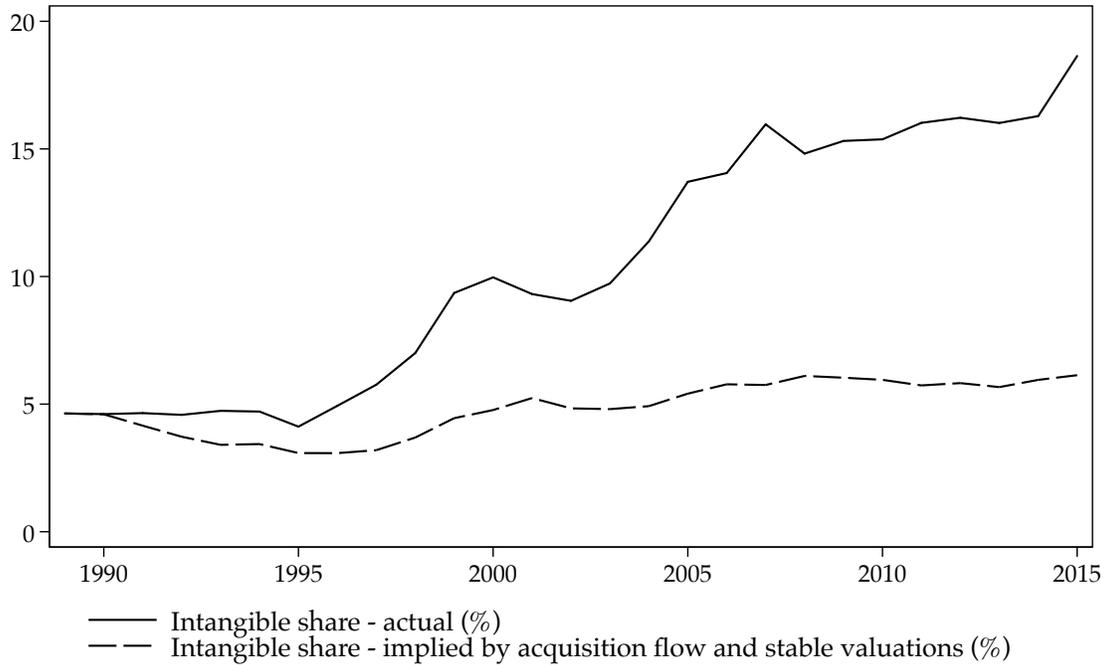


Figure 5: Actual and counterfactual path of the share of intangible assets on balance sheet. The counterfactual path is computed using observed acquisition expenditures and assuming that targets are valued using multiples of book assets consistent with valuations in 1989-1995. See text for more details.

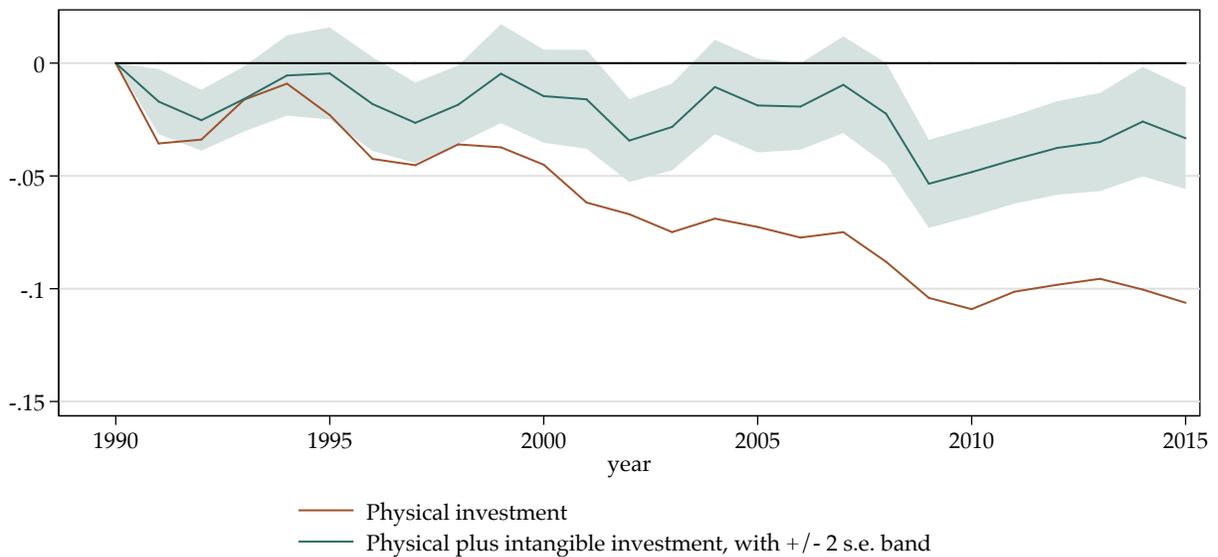


Figure 6: The investment gap using traditional investment (maroon line) and traditional plus intangible investment (blue line, with +/- 2 standard error bands). Standard errors are clustered at the firm level. See the text for the definition of intangible investment adopted in this graph.