Measuring Creative Destruction[†]

Abstract

We examine the link between a firm's future performance and innovations made by other firms using text-based measures of innovation displacement—how relevant one firm's innovations are to another's operations. Our findings indicate that when other major innovators' recent innovations are similar to the focal firm's technologies, the focal firm's profit growth over the next 7 years is expected to decline, with the association exacerbating annually, especially for non-innovative firms. This displacement effect persists across various firm types and model specifications. Moreover, firms exposed to higher displacement have higher risk-adjusted stock returns in the following year.

Keywords: Innovation, Displacement, Patents, Generative AI, Machine learning.
JEL Classification: C1, G10, G11, O3.

[†]This Draft: August 2024.

1 Introduction

Firms grow by successfully innovating, that is, by improving the quality or reducing the cost, of the products and services that they offer to their consumers. However, innovation often has some losers: firms that fail to innovate can be displaced by those that do. This process of creative destruction has been the lynchpin of a growing agenda in both macroeconomics and finance.[1] In this paper we provide a new direct measure of firms' exposure to innovation displacement.

We use a flexible underlying model: any innovation by any firm may have an association with any other firm's growth. The strength of the link between each pair of firms depends on three factors. First, take two firms where A is innovating and B is displaced. The size of the displacement depends on the relevance of A's recent innovations to B's daily operations and research. For example, an innovation in semiconductors is likely associated with a shock in the growth of a GPU manufacturer. However, this same innovation is less likely associated with changes in the performance of a restaurant chain. Secondly, it depends on the current aggregate innovation value of firm A. For example, innovations by a major innovator such as Apple may have a more widespread effect on other firms. On the other hand, innovations by a lower-ranked innovator may struggle to be recognized by consumers and other firms. Lastly, it depends on the characteristics of firm B. Our main analysis focuses on one characteristic: whether or not a firm is an active innovator. A firm that frequently produces innovations may be more tolerant of innovation displacement exposure (IDE) because it can quickly adjust its own focus to adopt new technologies after observing large displacements.

The first building block of our study is a measure of similarity between A's recent innovations and the current tools that B uses in its operations and research. We consider each issued patent as an innovation. For non-innovative firms, we define their assets in the innovation economy by the technologies they utilize in their day-to-day operations. For innovative firms, we include an additional aspect: their recent innovations, represented by the patents they have recently obtained. We construct two text-based measures of similarities. First, given the recent innovations made by firm A, we define our main displacement measure based on the similarity between these innovations and B's technologies. In addition, for the set of innovative firms, we compute a secondary similarity measure: the similarity between innovations of A and innovations of B.

To incorporate the second component of IDE – the current aggregate market value

[1] Aghion and Howitt (1992); Klette and Kortum (2004); Kogan et al. (2017); Kogan, Papanikolaou and Stoffman (2020).

of A's innovations – into our measure, we weigh these similarity measures using the aggregate firm-level innovation value in each year as computed by Kogan et al. (2017). The final weighted measures are our main and secondary measures of IDE. The details of these two measures are discussed in section 2 and 5.

The third factor that contributes to the size of the displacement exposure – heterogeneity across firm types– is not directly built into our measures. Instead, we conduct subsample analyses for a variety of characteristics including innovation activity level, firm size, profitability, intangible capital size, and time. For our main analysis in section 4, we focus on the difference in displacement exposure between innovative versus non-innovative firms. In the robustness section 6, we explore the other dimensions and verify that our main finding exists within different groups of firms.

Using these measures, we study the association between IDE and firms' profit growth. We have three main findings in our analysis: first, when we apply our main displacement measure to all firms, we find that IDE is significantly negatively associated with firms' profit growth, and this gap in profit growth is widening over time. In addition, we find that this displacement measure is correlated with many traditional firm characteristics such as employment and R&D expenditures. Nonetheless, a list of 9 innovation-related firm characteristics and over 1000 fixed effects can only explain about 21% of the variation in our displacement measure (19% after R^2 adjustment). Moreover, the residuals of the IDE measure (after regressing on firm characteristics) are still significantly negatively associated with profit growth. This shows that the main displacement measure captures information that is associated with profit growth and cannot be fully explained by these other characteristics.

Secondly, we show that the association between the main displacement measure and profit growth is different between innovative and non-innovative firms. In particular, we document that while both types of firms have a similar level of exposure to displacement in the near future, the magnitudes increase at different rates: over a longer horizon, the IDE-associated negative change in profit growth is lower among innovative firms. Furthermore, we compare the purely patent-based secondary displacement measure with our main measure: we find that the two measures are highly correlated at an inter-industry and inter-temporal level; however, once standardized within each year and industry category, the two measures have significant but low correlation. In addition, when we apply both measures in the subsample of innovative firms, we find that both measures capture distinct displacement effects that are negatively significantly correlated with profit growth.

To further validate this negative association between IDE and profit growth, we conduct a set of subsample analyses and use an alternative model specification to test the robustness of our main result. We divide the dataset by each firm's intangible capital level, size, and profitability and also by the time when the data was collected. Across all subsamples, we find a significantly negative association between our main displacement measure and firms' profit growth. Moreover, we use a non-linear model coupled with the debiased machine learning (DML) technique to rerun our analysis over the entire sample. This approach allows us to control for high-dimensional confounders. Using this method, we also find that the displacement measure is significantly negatively associated with profit growth and this association is increasing in magnitude over time.

In addition to developing our main displacement measure, we demonstrate two practical applications of the IDE measure. First, we construct an equal-weight long-short portfolio based on this measure, revealing that the long side outperforms the S&P 500 index by approximately 5% annually, while the short side underperforms by roughly the same amount. More importantly, we benchmark our long-short portfolio against traditional asset pricing models such as CAPM, Fama-French 3, Fama-French 5, and Fama-French 5 plus momentum. The results indicate that our portfolio achieves a significantly positive alpha across all models, with unexplained returns ranging from 3% to over 6% annually. This finding underscores that the displacement measure, in addition to being indicative of firms' future profit growth, can also enhance asset pricing exercises by providing additional gains. Moreover, the long and short sides of the IDE-based portfolio show a high correlation, largely driven by intertemporal variations in risk factors such as Fama-French 5 and momentum. We also find that the average firm characteristics at the portfolio level are significantly associated with the current returns. Specifically, for firms on the short side (low IDE values in the previous year), profitability and asset growth show significant associations: higher profitability correlates with higher monthly returns, whereas higher asset growth corresponds to lower returns. For firms on the long side, R&D expenses are significantly associated with current returns, with higher expenditures linked to lower returns.

Finally, as a second application, we train and test over 20 machine learning models to conduct out-of-sample forecasts of firm growth using IDE and traditional firm characteristics. We use the data up to 2010 for training and validation, and data after 2010 as the out-of-sample testing set to evaluate our models. We find that the best machine learning models consistently significantly outperform the majority voting baseline in predicting the direction of firms' profit changes. In addition, our IDE measure significantly contributes to this additional predictive power in longer horizons (3-year and 5-year profit changes).

Our paper contributes to multiple literature strands, particularly at the intersection of innovation, firm performance, stock market outcomes, and machine learning. It aligns closely with macroeconomic research on technological innovation displacement. Traditional methods for identifying technology shocks include measuring technological change through Solow residuals and imposing long-run restrictions on vector auto-regressions (VARs). These approaches, however, are indirect and sensitive to specific assumptions. Our approach diverges by constructing direct measures of technological innovation and displacement using unstructured text data.

Previous studies, such as Shea (1998), measured technological innovation through patents and R&D spending but found weak links between these measures and total factor productivity (TFP). This weakness likely arises from assuming all patents have equal value, a notion disproved by Kortum and Lerner (1998). Additionally, fluctuations in patent counts often reflect regulatory changes rather than genuine innovation. R&D spending is still indirect and subject to efficiency variations over time (Kortum, 1993). Alexopoulos (2011) introduced an aggregate-level measure based on technology-related books, but it lacks firm-level precision. In contrast, our work introduces a novel, text-based measure of innovation displacement, assessing how innovations by other firms impact a focal firm's future performance. This firm-level approach allows for granular analysis of reallocation and growth dynamics, revealing new insights into how external innovations have displacement effects on other firms' growth, especially firms that are not active innovators.

Additionally, our work extends the literature on the broader effects of technological innovation, particularly in balancing knowledge spillovers and business stealing. We show that similar innovations by major innovators can lead to significant displacement effects of the focal firm's current technologies, especially for firms less engaged in innovation (Kogan, Papanikolaou and Stoffman, 2020).

Our research is rooted in the tradition of endogenous growth and creative destruction (Acemoglu et al., 2018), emphasizing the role of technological displacement in shaping firm dynamics and growth. We also contribute to the literature on intangible capital and finance (e.g., Lev and Radhakrishnan (2005), Eisfeldt and Papanikolaou (2013), Peters and Taylor (2017), Crouzet et al. (2022), Chen et al. (2023)) by using firm-level measures of intangible capital and investment.

Finally, our paper makes an important methodological contribution to the economics literature. A rapidly growing branch of the big data literature in economics and finance uses natural language processing to quantify text (see e.g. Gentzkow, Kelly and Taddy

(2019)).[2] Our paper introduces a new method to estimate the similarities between financial texts based on recent advances in computer science, allowing the use of text embeddings to represent the contextual meaning of texts.

The remainder of the paper is organized as follows. Section 2 presents our main text-based measure. In Section 3, we study the association between innovation-induced displacement and firms' profit growth. Section 4 considers innovative versus non-innovative firms. In Section 5, we introduce an alternative text-based measure of displacement effect that solely relies on patents. Section 6 considers the robustness of our results by looking at different subsets of the data and different model specifications. Section 7 presents an application of our measure in asset pricing. Section 8 presents evidence that our IDE measure can be used to make effective forecasts of firms' profit growth in an out-of-sample setting. Section 9 concludes.

2 Data processing

To measure a firm's innovation displacement exposure, we construct a dataset of firm technologies and innovations spanning each year from 1980 to 2015.[3] We define a firm's innovations in year t by the patents it received from $t - 4$ to t, and we define its current technologies as those employed by the firm over the past five years in its operations and research. Specifically, we first obtain yearly technology summaries using GPT4o and then aggregate the summaries from $t - 4$ to t to build a comprehensive technology stack for the firm.

2.1 Constructing representations of firm innovations

First, we generate a numerical representation of each firm's innovations in each year with a 5-year lookback window. We start by producing an innovations summary for each firm each year based on the patents issued to it in year t. The primary challenge arises from the fact that some firms receive over a hundred patents annually, making it computationally impractical to generate a numerical vector that captures all the information from each patent. To address this issue, we employ a procedure designed to select a representa-

[2] A partial list of papers in this vein includes the work of Hansen, McMahon and Prat (2018), Chen, Wu and Yang (2019), Gentzkow, Kelly and Taddy (2019), Athey and Imbens (2019), Kelly et al. (2021), Erel et al. (2021), Fedyk et al. (2024).

[3] Our patent data begins in the year 1976; however, our displacement measure in each year is constructed with a look-back window of 5 years. Therefore, the earliest year we can compute displacement measures for is 1980.

tive sample of patents, while accounting for the variability in innovation capacity across different firms.

- If a firm i has at most 10 patents in year t, we take the abstracts of all of its patents issued in t.
- If a firm i has more than 10 patents but at most 100, we take a random subset of 10 patents issued in t.
- If a firm i has more than 100 patents, we take a random 10% subset of all of the patents issued in t.

After selecting the set of patents, we use GPT4o to summarize firm i's patents in year t in at most 500 tokens following the prompt:

> *You are an economist studying firms' innovations. Summarize the following patent abstracts. First, give a summary of the common topics covered by many abstracts, then focus on individual patents.*
>
> *{Selected patent texts.}*

Next, we generate innovation summaries for firm i in each year $t \in \{t-4, \cdots, t-1\}$ and concatenate the summaries in these 5 years chronologically from the earliest to the latest. Finally, we extract a numerical embedding based on each concatenated innovation summary using a text embedding model.[4]

2.2 Constructing representations of firm technologies

The second key component of our measures involves creating firm-level numerical representations of technologies. First, we generate a summary for each firm annually. For a given firm i in year t, we apply the following prompt to obtain a summary of i's technologies in t:

> *Instructions: You are an economist analyzing companies' technologies. In three paragraphs, describe the technological stack of name of i in the year t. Focus on technologies that the company was using in its day-to-day operations and research. Be as specific as possible and give a lot of details. For each technology, mention whether or not it was considered legacy in the field in t and whether or not there were disruptive technologies that could replace it. If there was a risk of disruption, mention companies and*

[4] We use text-embedding-3-large to generate 3072-dimensional embeddings.

technologies that were threatening to the given technology. Do not use information that became public after the year t.

Similar to constructing innovation embeddings, we obtain a 500-token summary for firm i for each year from $t-4$ through t, concatenate these summaries, and then utilize the same text-embedding model to create the technology embedding for firm i in year t.

2.3 Merging different sources to construct the main dataset

Lastly, we combine data from different sources to form our main dataset. We acquire the patent and aggregate innovation value data based on Kogan et al. (2017) and firm characteristics from COMPUSTAT, including company profits, R&D expenses, capital expenses, employment, etc. In addition, we merge stock return data from CRSP and risk factor data from the Fama-French data repository. These datasets are merged based on a unique firm identifier PERMNO and the year variable in each data set. Then, we merge this combined firm data with innovation and technology embeddings by the same identifiers.

2.4 Computing the main displacement measure

We define Innovation Displacement Exposure (IDE) as the cumulative impact of innovations from other firms on the focal firm, capturing the aggregated displacement effect of external innovations. More specifically, we compute the value-weighted similarity between other firms' innovations compared to i's technologies used in daily operations and research.

$$\text{IDE}_{i,t} = \frac{\sum_j^{n_{t,\text{Innov}}} \cos(\mathbf{Innov}_{j,t}, \mathbf{Tech}_{i,t}) \cdot A^f_{j,t}}{\sum_j^{n_{t,\text{innov}}} \text{Mk Cap}_{j,t}}, \tag{1}$$

where $n_{t,\text{Innov}}$ is the number of innovative firms in year t, $A^f_{j,t}$ is the aggregate innovation value of firm j in year t computed by Kogan et al. (2017), $\mathbf{Innov}_{j,t}$ is the innovation embedding of innovative firm j in year t, $\mathbf{Tech}_{i,t}$ is the technology embedding of firm i in year t, and $\text{Mk Cap}_{j,t}$ is the market capitalization of innovative firm j in year t. $\text{IDE}_{i,t}$ is large if other major innovators make advancements that are close to technologies used by the focal firm. We have a secondary measure of innovation displacement exposure that only uses the text of patents. Since most of the paper focuses on the main measure, we leave the details of the secondary measure in section 5.[5]

[5]Both sets of IDE values are standardized within each year and industry category.

3 Displacement and profit growth

We first study the association between innovation-induced displacement and firms' profit growth. In particular, when other firms develop innovations similar to the technologies used by the focal firm, the focal firm's operations are likely displaced. It either loses its competitive advantage or has to adopt these new technologies. In both cases, its future profit will be hindered. Therefore, we conjecture that a firm with a higher IDE value is expected to have lower profit growth. We test this hypothesis using the following 7 regressions ($k \in \{1, 2, 3, 4, 5, 6, 7\}$):

$$\Pi_{f,t+k} - \Pi_{f,t} = \alpha_1^k IDE_{f,t} + \alpha_2^k \log(AG)_{f,t} + \alpha_3^k P_{f,t} + \beta^k Z_{f,t} + \delta_{f,t}^k + \epsilon_{f,t}^k, \tag{2}$$

where $\Pi_{f,t+k}$ and $\Pi_{f,t}$ are firm f's log profits in year $t+k$ and t, $IDE_{f,t}$ is the text-based displacement measure of firm f in year t, $\log(AG)_{f,t} = \log\left(\frac{AT_{f,t}}{AT_{f,t-1}}\right)$ is the log of asset growth of the firm f at year t, $P_{f,t}$ is the profitability of firm f in year t defined by $\frac{sale_{f,t} - cogs_{f,t}}{at_{f,t}}$, $Z_{f,t}$ are the log profit, employment, and capital stock in year t of firm f, and $\delta_{f,t}^k$ are the fixed effect (year, industry code, and industry category interacted with year). All standard errors are clustered at the firm-year level. All independent variables are standardized at the industry category and year level.

Insert table A1 here.

As shown in table A1, the coefficients on profitability are significantly negative which shows a mean-reversal trend: firms that are currently highly profitable have lower future profit growth. In addition, the positively significant coefficients on asset growth align with Fama and French (2015)'s finding that asset growth captures future growth opportunities. More importantly, IDE (the main displacement measure) is significantly negatively associated with profit growth in all of the next 7 years. This means that firms that are exposed to a higher level of innovation displacement are expected to have a lower level of profit growth that sustains over at least the next 7 years. Moreover, as shown in figure 1, the magnitude increases from 1.5% in the next year to 6.7% 7 years in the future.[6] This suggests that the displacement effect exacerbates each year.

To formally test this observation of worsening displacement effect over time, we construct a dataset where we concatenate all profit growth in the next 1 through 7 years into one column, concatenate all of the control variables used in equation 2, and include an additional variable corresponds to the number of years across which the profit growth is

[6] Strictly speaking, the percent change is the change in log profit.

measured. In particular, we run the regression:

$$\begin{aligned}\Pi_{f,t+k} - \Pi_{f,t} =& \alpha_1 IDE_{f,t} + \alpha_2 \log(AG)_{f,t} + \alpha_3 P_{f,t} \\ &+ \alpha_4 IDE_{f,t} k + \alpha_5 \log(AG)_{f,t} k + \alpha_6 P_{f,t} k \\ &+ \beta_1 Z_{f,t} + \beta_2 Z_{f,t} k + \beta_3 k + \delta_{f,t,k} + \epsilon_{f,t,k},\end{aligned} \tag{3}$$

where α_1, α_2, and α_3 estimate the baseline association between displacement, asset growth, and profitability and profit growth. α_4, α_5, and α_6 estimate the change in magnitude of these associations as the forward-looking window increases by 1 year. The fixed effects include year, industry code, the interaction between industry category and year, as well as the interactions between the window size indicators (1, 2, $\cdots$, 7) and each of these fixed effects. We standardize all independent variables and cluster standard errors at the firm-year-window size level.

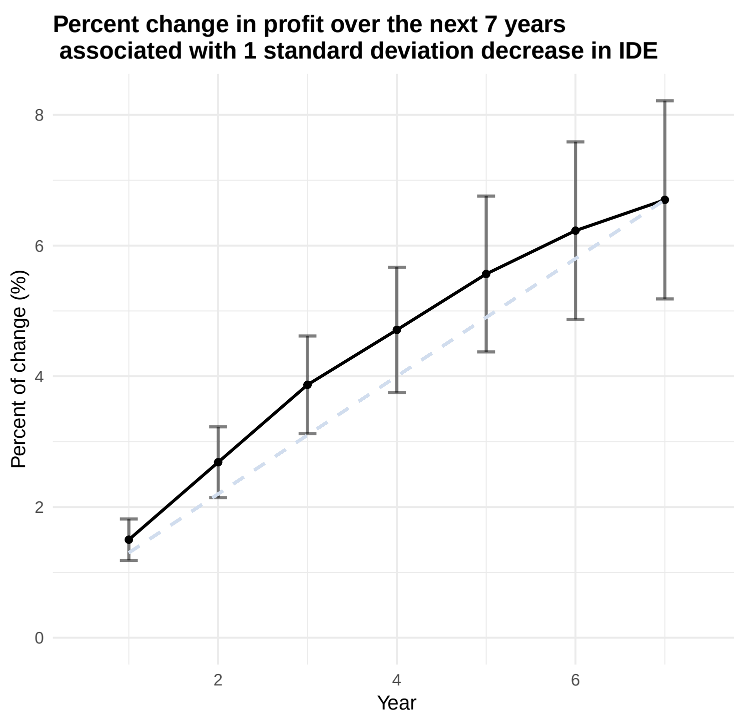

Figure 1: This figure shows the percent change of profit associated with one standard deviation **decrease** in displacement. Note that displacement is negatively associated with profit growth. The blue line represents the widening of the gap estimated using equation 3.

Insert table *A2* here.

As shown in table A2, each additional standard error increase of IDE is associated with a 0.9% additional decrease in profit growth each year. As a sanity check, the association between IDE and profit growth in the first year is the baseline plus the additional association: $0.4\% + 0.9\% = 1.3\%$ which is close to the results in table A1, and the magnitude in the subsequent years also match in the two tables. This 0.9% additional association is negatively significant. This confirms that the association between IDE and profit growth is increasing in magnitude.

Insert table *A3* here.

The natural question to test next is whether the new displacement measure is a combination of traditional innovation-related firm characteristics and whether it captures additional information indicative of firms' profit dynamics. Moreover, once we remove the portion of IDE explained by these variables, are the residuals still significantly negatively associated with profit growth?

First, we analyze which firm characteristics are associated with our displacement measure using the following regression:

$$IDE_{f,t} = \alpha^{\text{IDE}} X_{f,t} + \delta_{f,t}^{\text{IDE}} + \epsilon_{f,t}^{\text{IDE}}, \tag{4}$$

where $X_{f,t}$ are vectors of firm characteristics including the firm's log capital stock, log employment, log profit, log asset growth, profitability, total asset, market cap, capital expenditures, and R&D expenditures.[7] $\delta_{f,t}^{\text{IDE}}$ are the fixed effects corresponding to each IDE value (year, industry code, and industry category interacted with year).

We find that firms with high IDE (exposed to large innovation displacements) tend to have higher capital stocks, lower employment, lower profit, higher profitability, lower total assets, lower capital expenditure, and higher research and development expenditures. Nonetheless, these variables and the fixed effects can only explain about 19% of the variation in the displacement measure. This shows that our measure captures information that is not associated with traditional firm characteristics.

Insert table *A4* here.

More importantly, when we reestimate the regression specified by equation 2 using the residual of IDE in equation 4, we still see a significantly negative association between the residuals and profit growth as shown in table A4. Each standard deviation increase in the residual of IDE is associated with a 1.4% decrease in profit by year 1 and 5.1% by

[7] We apply an asinh transformation to R&D and capital expenditures to stabilize the variance in these variables while allowing for negative values.

year 7. This shows that the unexplained information captured by IDE is associated with negative profit growth. In other words, our main displacement measure (IDE) captures information indicative of firms' profit growth that is not included in any of the traditional innovation-related firm characteristics.

In the previous analysis, we have shown that IDE is significantly negatively associated with future profit growth even when only examining its residuals after regressing on firm characteristics; therefore, IDE is of clear interest to firm managers and investors prior to the realization of these displacement events. Hence, we study whether there are characteristics that indicate certain firms may have higher IDE values in the future. Since our IDE measure is computed based on innovation and technology descriptions with a 5-year lookback window, we start examining firm characteristics 5 years before the construction year of IDE and also consider longer horizons like 7 and 10 years ago. More specifically, we consider the following variables: total assets, R&D expenditures, profitability, and capital expenditures (standardized in each year and industry category). One issue of this analysis is that characteristics such as capital expenditures and total assets are strongly correlated. Thus, we take the following steps to remove the collinearity of these variables: first, we run a Principal Component Analysis (PCA) on the 4 independent variables; then we remove the common variations in these variables using the first 2 principal components; lastly, we standardize the residuals and run a regression where the dependent variables are the IDE values T years in the future ($T \in \{5, 7, 10\}$) and the independent variables are the residual total assets, R&D expenditures, profitability, and capital expenditures.[8]

Insert table *A5* here.

As shown in Table A5, the two variables that consistently predict variations in future IDE are total assets and profitability. Firms with more assets are more likely to be exposed to higher displacement, while firms that are more profitable tend to be exposed to less. Intuitively, firms with more assets are at risk of a larger variety of displacements by different innovations; on the other hand, firms that exhibit high profitability relative to their current assets are likely to possess unique assets that are difficult to replicate or have a stronger ability to deter competition and avoid displacement.

[8]The first 2 principal components explain $81 - 82\%$ of variation in the data.

4 Innovative versus non-innovative firms

Given the main result – IDE is negatively associated with profit growth – we study whether this association varies between two types of firms in the economy: innovative and non-innovative firms. We define a firm's innovation activity by counting the number of patents received by the firm within our time range (1976 to 2015). A firm is considered innovative if it receives at least 10 patents during this interval.

IDE is fundamentally caused by other firms' innovations outcompeting legacy technologies used by the focal firm; therefore, we conjecture that this effect should be smaller in magnitude when the focal firm is an active innovator. This result aligns with the finding in Chen, Wu and Yang (2019) which focuses on the FinTech industry. One plausible explanation is a firm that produces a large number of patents likely has the infrastructure to pivot its research and development to either adopt new or invent other competitive technologies. Take Apple as an example of an active innovator in the mobile phone industry. Its competitors like Samsung may release a new iteration of products that are temporarily the most desirable phones in the market; however, Apple can quickly overcome these challenges by incorporating these technologies and improving upon them in its own next iteration.

Insert tables *A6* and *A7* here.

We begin our analysis by running the regressions specified in equation 2 on the two subsamples separately: one for innovative firms and one for non-innovative firms. As shown in tables A6 and A7, in both subsamples, displacement is significantly negatively associated with firms' profit growth in the next 7 years. This shows our main result is robust for both innovative and non-innovative firms. In addition, as shown in figure 2, the difference in magnitude between the two subsamples is widening over time. In particular, one standard deviation increase in displacement corresponds to around 1.5% (1.3% vs 1.7%) decrease in profit growth in the first year. However, 7 years in the future, the same increase in displacement is associated with 4.1% decrease in profit growth among innovative firms, while the magnitude for non-innovative firms is at a larger 7.7%.

We statistically show that this difference is significant in the longer horizon using a permutation test. The null hypothesis is that in terms of the size of displacement effect, the innovative firms are randomly drawn from the population of firms. The alternative hypothesis is that this sample is purposely chosen to exhibit a larger coefficient on IDE over time. More specifically, we first, randomly create 1000 samples within each we randomly assign innovative or non-innovative labels to the firms (respecting the proportions in the original sample). Then, we compute the coefficients on our IDE measure (displace-

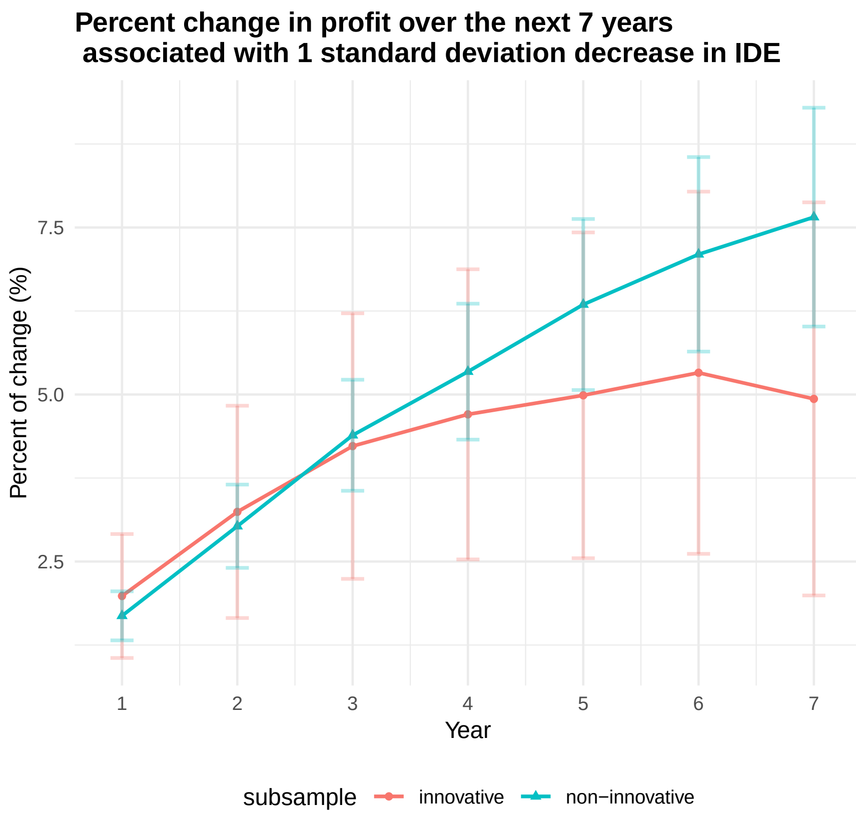

Figure 2: This figure shows the difference between innovative and non-innovative firms in terms of the percent change of profit associated with one standard deviation decrease in displacement. Note that displacement is negatively associated with profit growth.

ment) interacted with the number of years across which the profit growth is computed (analogous to equation 3) for both randomly assigned subsamples and compute the differences between each pair of coefficients over the next 7 years – one for the innovative subsample and another for the non-innovative subsample. Lastly, the p-value of this hypothesis is computed as the probability that the true difference we observe based on the true assignment is larger than the differences computed based on random assignments. As shown in figure 3, we observe a larger difference based on the original assignment of innovative versus non-innovative firms than over 90% of the randomly assigned subsamples in the permutation test. This shows that, in the true sample, there exists a significant difference between innovative and non-innovative firms. More specifically, the association between displacement and profit growth increases faster in magnitude among non-innovative firms. Economically, non-innovative firms suffer increasingly more from innovation displacement exposure.

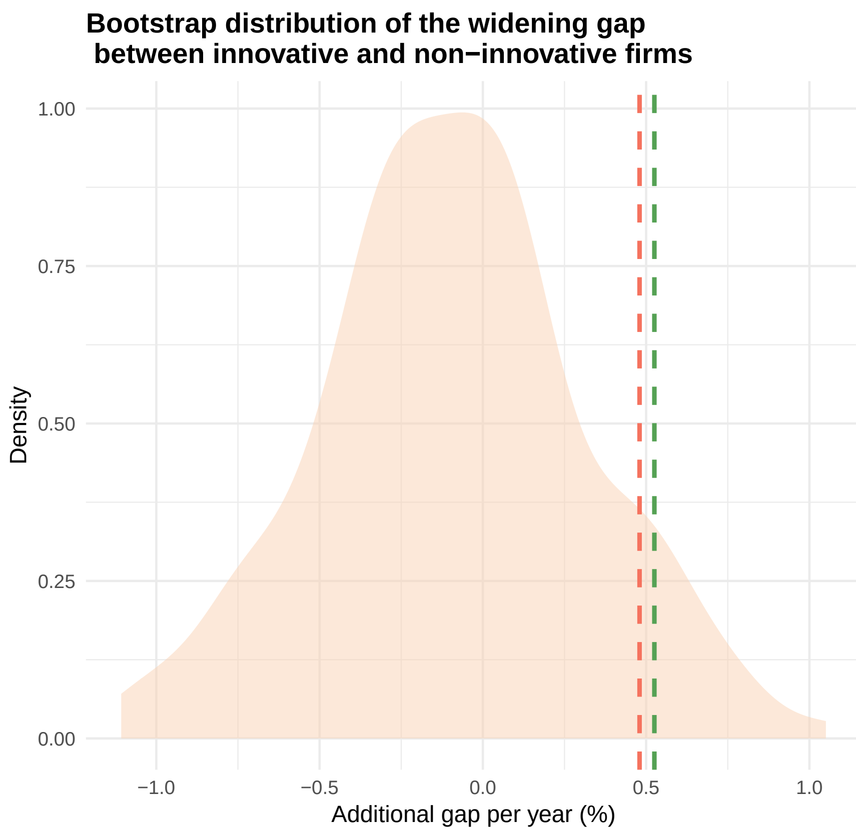

Figure 3: The distribution of the difference between the coefficients in the randomly assigned innovative and non-innovative subsamples. The coefficients of interest are the ones on the interaction between the number of years forward and IDE. The red line represents the 90th percentile of the distribution, and the green line represents the difference we compute based on the original assignment of innovative versus non-innovative firms.

5 Patent-based displacement measure

In this section, we introduce an alternative text-based measure of displacement exposure that solely relies on patents. One potential drawback of our main measure introduced in section 2 is the existence of a look-ahead bias when constructing technology descriptions – the generative model is trained on data after the specified year. By relying only on patent texts, we can eliminate this concern. More importantly, this measure captures an additional form of innovation-induced asset displacement that innovative firms experience: the displacement of their recent innovations. Specifically, when a major innovator introduces a new innovation, the recent innovations of the focal firm, along with its current technologies, may lose their competitive edge as they no longer represent the state-of-the-art in their respective fields. Therefore, our main and alternative displacement measures complement each other when studying the profit growth of innovative firms.

The alternative IDE measure is defined as

$$\mathrm{IDE}_{i,t} = \frac{\sum_j^{n_{t,\mathrm{innov}}} \cos(\mathbf{Innov}_{j,t}, \mathbf{Innov}_{i,t}) \cdot A_{j,t}^{f}}{\sum_j^{n_{t,\mathrm{innov}}} \mathrm{Mk\ Cap}_{j,t}}, \tag{5}$$

where $\mathbf{Innov}_{j,t}$ and $\mathbf{Innov}_{i,t}$ are both innovation embeddings computed based on summaries of firms' recent patents.

We study the association between this secondary displacement measure and profit growth among firms with at least 10 patents – innovative firms.[9]

First, we study the correlation between the two measures. We show that the alternative IDE measure is positively correlated with the original measure, but the correlation is mostly due to heterogeneity across year and industry categories.

Insert table *A8* here.

More specifically, we compare the two measures using two types of correlations: Pearson correlation computes the correlation of the two continuous measures, and Kendall-Tau computes the correlation between the rank orderings. As shown in table A8, the raw numbers of the main and alternative displacement measures are highly correlated at over 0.8; however, the correlation drops to a lower $6 - 11\%$ when these measures are standardized within each year and industry category. This implies that within each industry category and year, the two measures are directionally aligned but capture different

[9]Since this alternatively measure defined by equation 5 only depends on patents, we focus on the subsample of innovative firms.

information.

Next, we run a modified version of the main regression in equation 2 over the sample of innovative firms:

$$\Pi_{f,t+k} - \Pi_{f,t} = \alpha_1^k IDE_{f,t}^o + \alpha_2^k IDE_{f,t}^a + \alpha_3^k \log(AG)_{f,t} + \alpha_4^k P_{f,t} + \beta^k Z_{f,t} + \delta_{f,t}^k + \epsilon_{f,t}^k, \quad (6)$$

where $IDE_{f,t}^o$ is the standardized main displacement measure and $IDE_{f,t}^a$ is the standardized patent-based alternative measure.

Insert table A9 here.

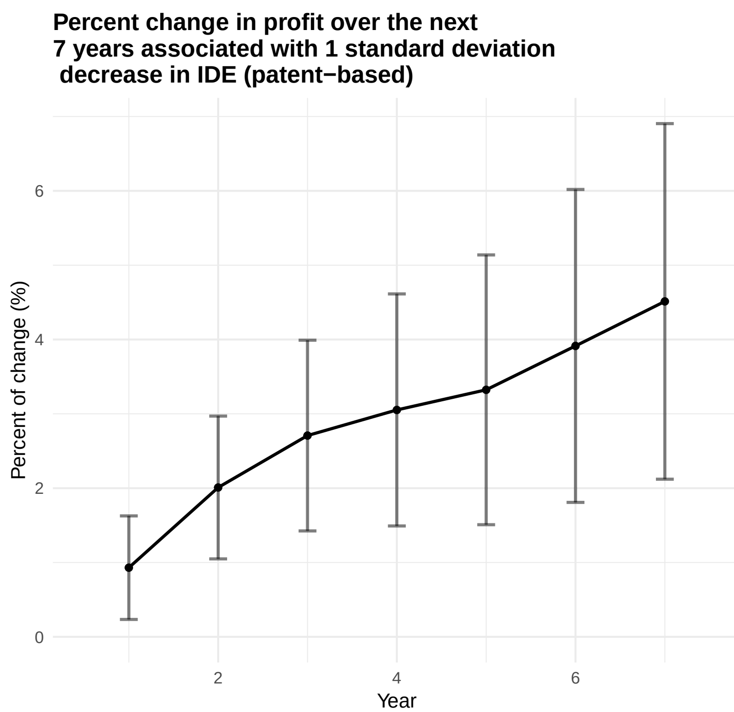

Figure 4: This figure shows the percent change of profit associated with one standard deviation **decrease** in patent-based alternative displacement measure. Note that displacement is negatively associated with profit growth.

As shown in Table A9, the alternative IDE measure identifies significant and distinct displacement effects among innovative firms. Specifically, each standard deviation increase in this patent-based displacement measure corresponds to an additional 0.9% decrease in profit growth in the following year. Consistent with the findings in Section 3, this effect intensifies over time, reaching a 4.5% reduction by year 7 (see figure 4). Together, these results demonstrate that innovation-induced displacement exposure is associated with profit growth among innovative firms through two channels: the displacement of existing technologies and the displacement of recent innovations (patents).

6 Robustness

In this section, we check the robustness of our results by looking at different subsets of the data and considering a different model specification. We start with subsample analyses: first, we divide the data based on the fraction of intangible capital of each firm relative to its book equity (Eisfeldt, Kim and Papanikolaou, 2020). Next, we divide the data based on the firms' market capitalization in each year and industry. Next, we divide the data set along the time dimension – 20th century versus 21st century. Then, we separately consider highly profitable firms and ones with lower profitability. Our last subsample analysis looks at firms in the United States with at least one patent. We show that the negative association between IDE and profit growth is significant in all subsamples, and we provide some suggestive evidence that the magnitude of firms' exposure to IDE varies across different firm types: conditioning on the same IDE value, some types of firms' (such as large firms') profit growths are hindered more heavily.

In addition to the subsample analysis, we redo our main analysis in section 3 using an alternative specification of confounding. In particular, we incorporate a nonlinear model, XGBoost, with the debiased machine learning technique to control for the high dimensional confounders. We find that our main result remains robust.

6.1 High versus low intangible capital

We divide the dataset into two based on each firm's relative intangible asset in each year within its industry category. Based on (Eisfeldt, Kim and Papanikolaou, 2020), we define a firm's relative intangible capital as this firm's intangible capital divided by its book equity in the year t:

$$G_{f,t} = \frac{Int_{f,t}}{BE_{f,t}}.$$

Based on this definition, we divide the data into two: in each year and industry category, we take firms with at or above median $G_{f,t}$ as the set of firms with high intangible capital; at the same time, we define the other firms as the low intangible capital set. We run the regression defined by equation 2 separately for both subsamples.

Insert tables A10 and A11 here.

As shown in tables A10 and A11, displacement is negatively associated with profit growth in both subsamples. Among firms with high intangible capital stock, one standard deviation increase in IDE corresponds to a 1.4% decrease in profit growth over the first year and 5.3% over the next 7 years. Focusing on firms with low relative intangible capital, one standard deviation increase in IDE corresponds to a 1.6% decrease in profit growth over the first year and 7.8% over the next 7 years. As a more direct demonstration, figure 5 shows that firms with low intangible capital are more exposed to displacement and the difference is widening: losing an additional 0.2% in profit growth by year 1 and 2.5% by year 7.

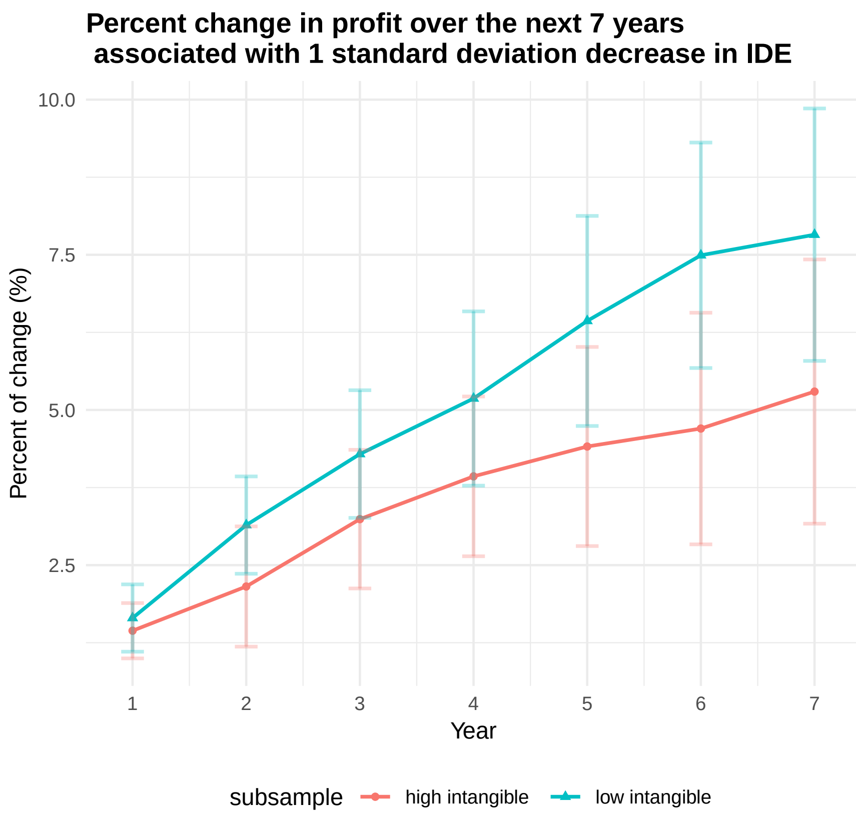

Figure 5: This figure illustrates the difference in percent change in profit growth between firms with high and low relative intangible capital, corresponding to a one standard deviation decrease in displacement. Note that displacement is negatively associated with profit growth.

6.2 Large versus small firms

Next, we split the dataset based on each firm's market capitalization within its industry category each year. Firms with market capitalization at or above the median in their respective year and industry category are considered large firms, while those below the median are small firms. Following the approach outlined in the previous subsection, we run the regression specified in equation 2 separately for each subsample.

Insert tables *A*12 and *A*13 here.

As shown in tables A12 and A13, displacement is negatively associated with profit growth in both subsamples. Among small firms, one standard deviation increase in IDE corresponds to a 1.1% decrease in profit growth over the first year and 5.6% over the next 7 years. Among larger firms, one standard deviation increase in IDE corresponds to a 1.7% decrease in profit growth over the first year and 7.3% over the next 7 years. Visually, figure 6 directly demonstrates that larger firms are more exposed to displacement than smaller firms and the difference is widening: losing an additional 0.6% in profit growth by year 1 and 1.7% by year 7.

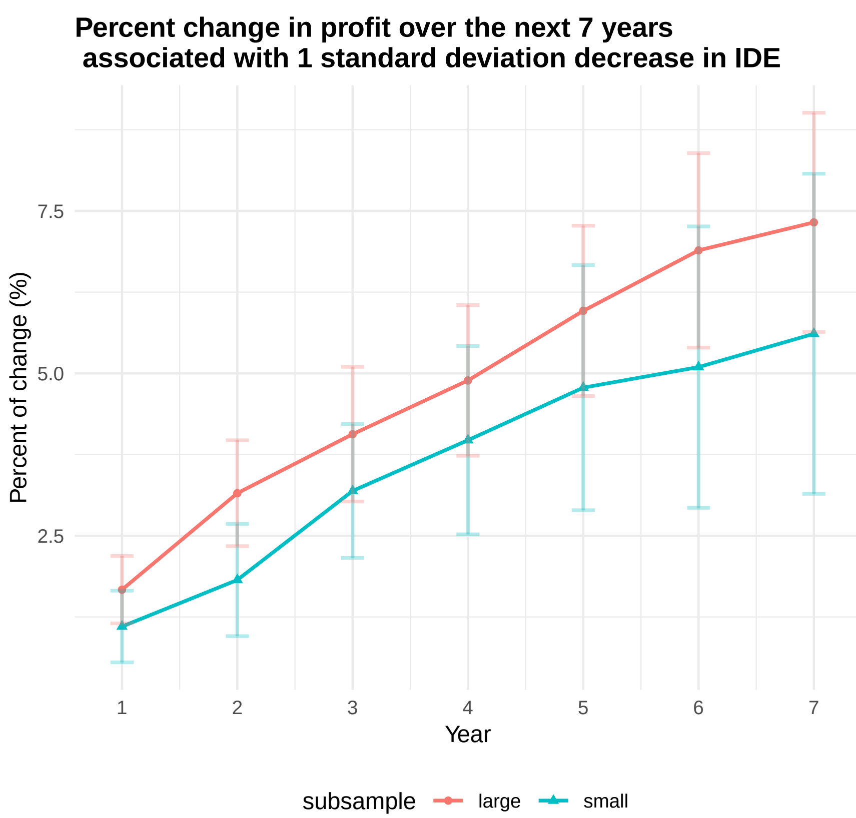

Figure 6: This figure illustrates the difference in percent change in profit growth between firms with high and low market capitalization, corresponding to a one standard deviation decrease in displacement. Note that displacement is negatively associated with profit growth.

6.3 Pre-2000 versus since 2000

Furthermore, we divide the data into two samples based on whether the year is before or after (including) 2000. Similar to the previous subsection, we run the regression defined by equation 2 for both subsamples.

Insert tables *A*14 and *A*15 here.

As shown in Tables A14 and A15, displacement is negatively associated with profit growth in both subsamples. In the 20th century data, a one standard deviation increase in IDE is associated with a 1.4% decline in profit growth in the first year and a 6.2% decline over the next 7 years. For the 21st century, the corresponding association is slightly larger, with a 1.6% decrease in the first year and 7.0% over 7 years. Figure 7 illustrates that firms in the 21st century face a greater decline in profit growth due to displacement, with the gap widening: an additional 0.4% loss by year 1 and 0.8% by year 7. Notably, this difference between recent and older data is the smallest observed among our subsample comparisons, both in absolute and relative terms.

6.4 High profitability versus low profitability

In this subsection, we partition the data into two subsamples based on each firm's profitability within each year and industry category, measured as gross profit divided by total assets: $\frac{sale_{f,t} - cogs_{f,t}}{at_{f,t}}$. Firms below the median profitability in their respective year and industry category form one subsample, while the remaining firms form the other. As in the previous subsection, we apply the regression specified in equation 2 to both subsamples.

Insert Tables *A*16 and *A*17 here.

Tables A16 and A17 show that displacement is negatively associated with profit growth in both groups. For firms with lower profitability, a one standard deviation increase in IDE corresponds to a 1.3% reduction in profit growth in the first year and 6.2% over 7 years. Among highly profitable firms, the same increase in IDE is associated with a 0.8% decrease in the first year and 4.8% over 7 years. Figure 8 further illustrates that less profitable firms are more vulnerable to displacement, with the gap widening over time: an additional 0.5% loss in profit growth by year 1 and 1.4% by year 7.

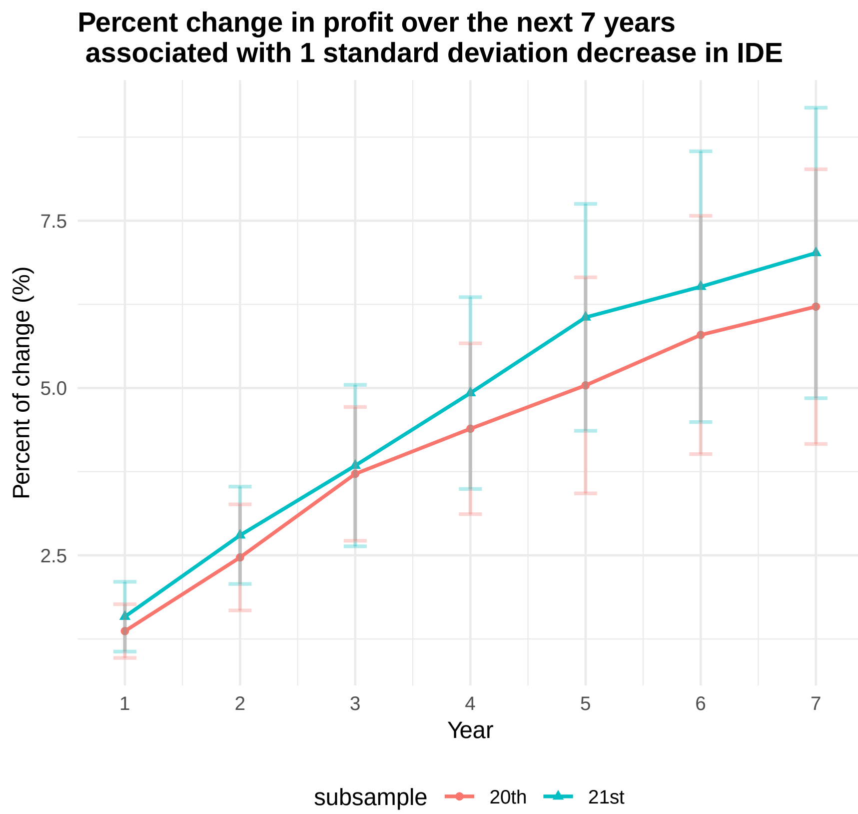

Figure 7: This figure shows the difference between data in the 20th and 21st centuries in terms of the percent change of profit associated with one standard deviation decrease in displacement. Note that displacement is negatively associated with profit growth.

6.5 Firms in the United States

In this subsection, we examine the subsample of firms in the United States that have at least one patent. First, we use a Welch t-test to compare the summary statistics between the two samples: U.S. firms with at least one patent vs. the rest. All variables are standardized each year before being split into two subsamples for interpretability. For example, without standardization, the market capitalizations of two firms in different years are not comparable. As shown by table A18, firms in the United States with at least one patent tend to be large in terms of market capitalization, more profitable, have higher current profits, and be exposed to more displacement as measured by our IDE measure defined in equation 1.

Insert tables *A*18 and *A*19 here.

Nonetheless, as shown in table A19, the negative association between IDE and profit

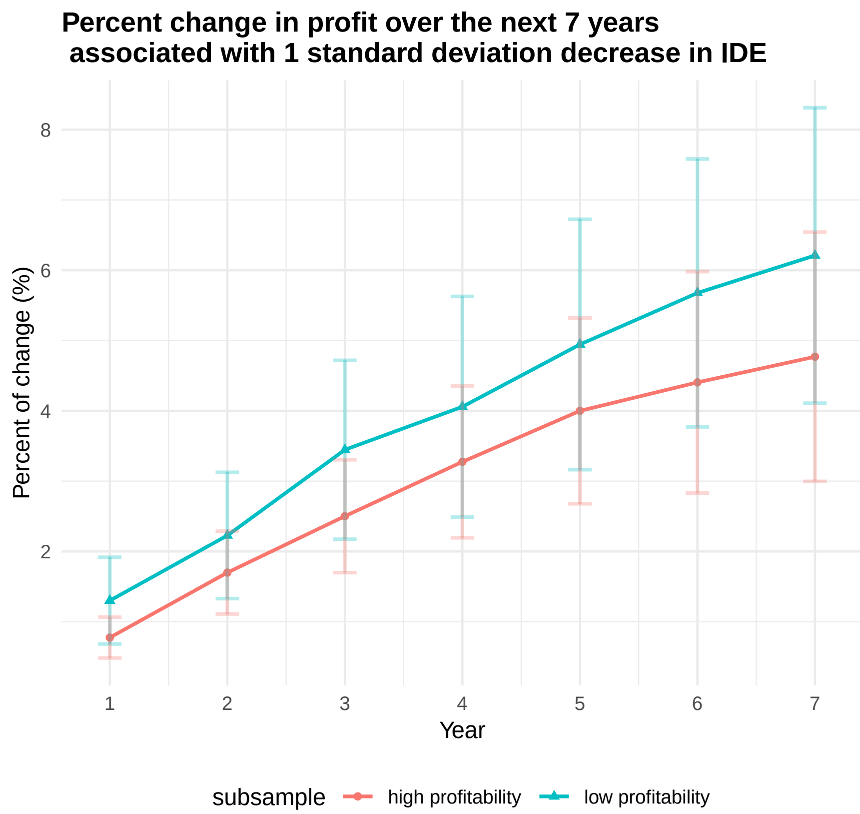

Figure 8: This figure shows the difference between high and low profitability firms in terms of the percent change of profit associated with one standard deviation decrease in displacement. Note that displacement is negatively associated with profit growth.

growth remains significant in the subsample of United States firms with at least one patent. Each standard deviation increase in IDE is associated with a 1.5% decrease in profit growth of U.S. firms in the first year, and this grows to 5.8% by the end of year 7. The direction and magnitudes are all similar to the main result across all samples shown in table A1.

Combining the results in subsections 6.1, 6.2, 6.3, 6.4, and 6.5, we see that the result that displacement is significantly negatively associated with profit growth is robust in a variety of subsamples. This shows that our results are generalizable. However, even though we observe that displacement is associated with different rates of changes in profit growth, the exact mechanisms driving these differences are still to be determined and left for future studies.

6.6 Non-linear confounders and debiased machine learning

Another potential concern for our results is that the estimates are based on a linear confounding specification, which may not accurately represent the underlying relationship between displacement and profit growth, confounded by firm characteristics and year. In this subsection, we analyze the entire sample using a non-linear model of the confounding relationship leveraging the debiased machine learning technique developed in Chernozhukov et al. (2018). Debiased machine learning allows us to estimate the parameters of interest (association between IDE and profit growth) while controlling for high-dimensional confounders.

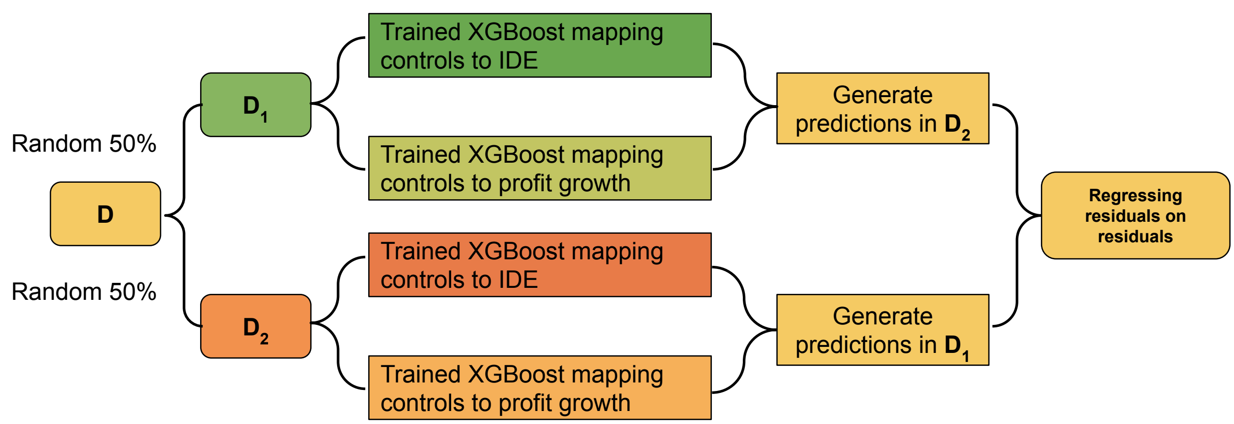

Figure 9: The diagram of debiased machine learning applied to identify the effect of displacement (IDE) on profit growth

As shown in figure 9, we utilize debiased machine learning in the following steps:

1. Randomly split the dataset into 2 equal-sized subsets D_1 and D_2.
2. Train an XGBoost model $g_T^{D_1}$ using D_1 including all of the control variables and fixed effects in equation 2 to predict IDE (our main displacement measure).
3. Train another XGBoost model $g_O^{D_1}$ using D_1 including all of the control variables and fixed effects in equation 2 to predict the profit growth (we do this 7 times because we consider the profit growth over each of the next 7 years).
4. Generate predictions of IDE and profit growth in D_2.
5. Repeat steps 2-4. Use D_2 as the training data and generate predictions in D_1. The new models are $g_T^{D_2}$ and $g_O^{D_2}$.
6. Combine all of the out-of-sample predictions to form a new dataset.

7. Regress the residuals of IDE on the residuals of profit growth:

$$\left(\Pi_{f,t+k} - \Pi_{f,t+k}\right) - \left(\widehat{\Pi_{f,t+k} - \Pi_{f,t}}\right) = \alpha_k + \tau_k^{IDE}(IDE_{f,t} - \widehat{IDE_{f,t}}) + \epsilon_{f,t+k}, \quad (7)$$

where τ_k^{IDE} is the estimated effect of displacement on profit growth.

Insert tables *A*20 and *A*21 here.

As shown in table A20, when controlling for all of the confounders (including the continuous variables and fixed effects in equation 2), the negative effect of displacement on profit growth is significant over all of the next 7 years. In particular, as shown by figure 10, the difference is widening each year from 0.4% in the first year to 3.4% in the 7th year. This increase in the effect size per year is a significant 0.5% as shown by table A21.[10] This demonstrates that our main results discussed in section 3 are unlikely a coincidence due to model misspecification.

7 Application in asset pricing

In this section, we rank firms by their IDE values to construct an equal-weight long-short portfolio. We begin by forming an equal-weight portfolio that goes long on the top 10% of firms with the highest IDE values and shorts the bottom 10%. Next, we examine the extent to which the correlation between the returns of the long and short sides of the portfolio is driven by intertemporal variations in risk factors, such as Fama-French factors and momentum. Finally, we investigate the firm characteristics associated with the current returns of firms with high versus low IDE values.

7.1 IDE-based long-short portfolio

First, we analyze the equal-weight long-short portfolio based on the IDE measure. Since the IDE measures are calculated annually, we create a portfolio at the end of each year after computing this year's IDE values. We long the top 10% of firms with the highest IDE values and short the bottom 10% (equal-weight). This portfolio is rebalanced monthly, but the firms on the long and short sides are selected annually. Figure 11 shows the P&L of this portfolio and the long and short sides separately. We observe that the long-short

[10]This growth is shown using the following regerssion analogous to equation 3:

$$\left(\Pi_{f,t+k} - \Pi_{f,t+k}\right) - \left(\widehat{\Pi_{f,t+k} - \Pi_{f,t}}\right) = \alpha + \beta_1 k + \beta_2(IDE_{f,t} - \widehat{IDE_{f,t}}) + \tau^{IDE} k(IDE_{f,t} - \widehat{IDE_{f,t}}) + \epsilon_{f,t,k}$$

.

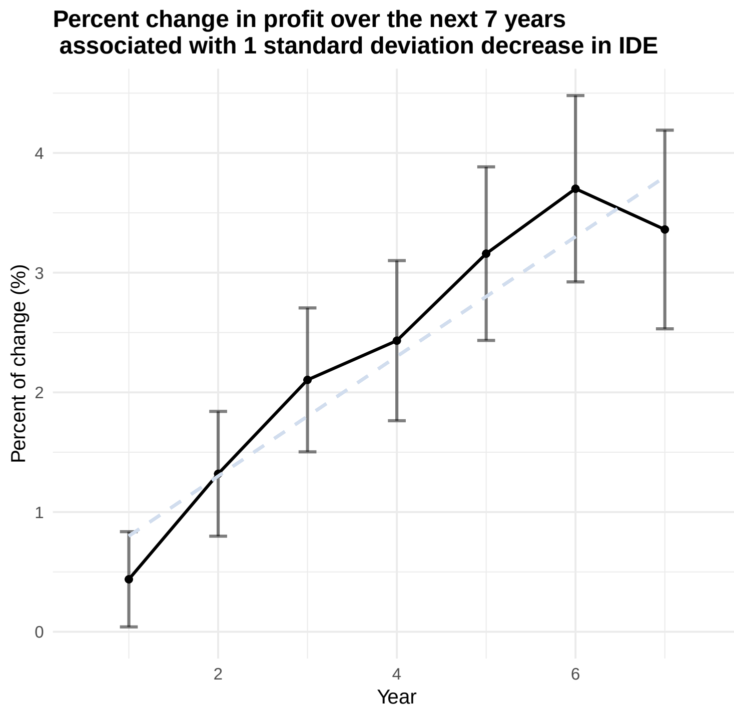

Figure 10: This figure shows the percent change of profit associated with one standard deviation **decrease** in displacement. Note that displacement is negatively associated with profit growth. The grey line shows the exacerbating negative association between IDE and profit growth estimated in table A21.

portfolio has large growth spurs from 2001 to 2003 (the dot.com bubble) and 2008 to 2009 (the Global Financial Crisis). This shows that the IDE-based portfolio gains profits during recessions. More specifically, examining the long and short sides independently, this profit is mostly due to the short side correctly picking heavily depreciating assets. This demonstrates that firms with low IDE have very low returns during recessions. Economically, a firm whose main technologies are not close to major innovators devalues quickly when the overall economy is weak.

Overall, as shown by figure 12. The short side of the IDE portfolio has an annualized relative return of $\frac{1.76}{36} = 4.9\%$ (beyond the S&P 500 index). Similarly, the annualized relative return of the long side is $\frac{1.71}{36} = 4.8\%$. This shows that roughly an equal amount of profit is made from the long and short sides of the portfolio.

Next, we compare the IDE portfolio return with well-studied factor models such as CAPM, Fama-French 3, Fama-French 5, and Fama-French 5 with momentum. We compare the monthly portfolio returns against these models.[11] We compute the α of the IDE

[11] If we use annual factor data, we only have 36 data points in our regression. Therefore, we analyze monthly data to have a meaningful number of observations.

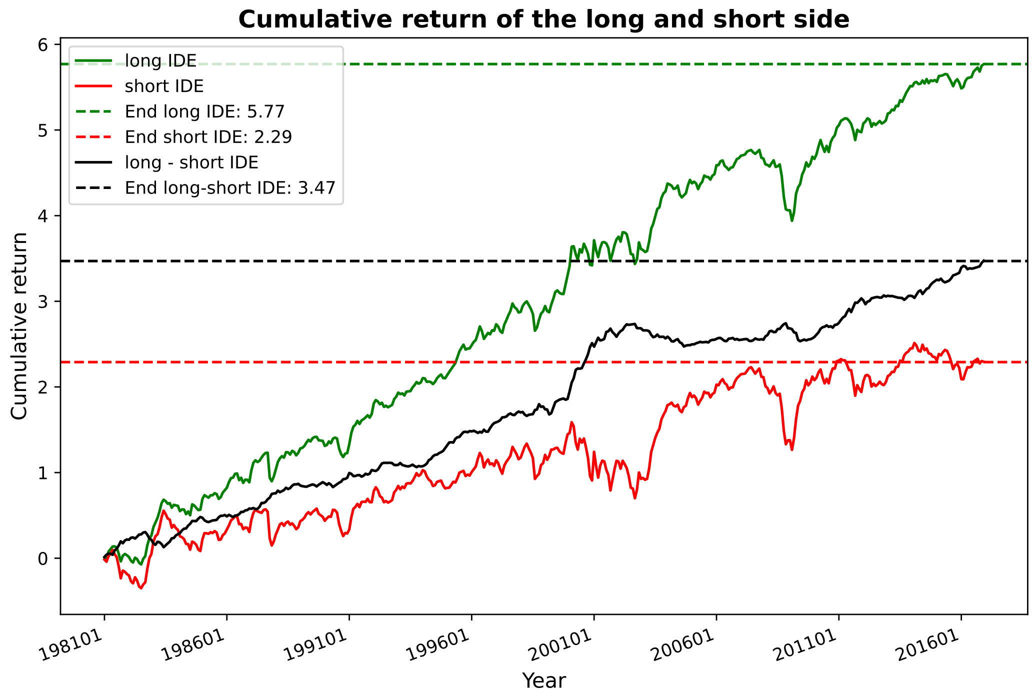

Figure 11: This figure shows the cumulative (monthly) return of the long and short sides and the overall long-short IDE portfolio. The long-short portfolio is equal-weight and rebalanced monthly.

portfolio using the following regression

$$r_t - rf_t = \alpha + \beta \cdot X_t + \epsilon_t, \tag{8}$$

where r_t is the return of the IDE portfolio in month t, rf_t is the monthly risk-free rate, and X_t are the factors used in each model in month t.

Insert table A22 here.

As shown in Table A22, the monthly α values vary across different models: CAPM yields a monthly α of 0.54% (annualized to 6.48%), FF3 yields 0.46% (annualized to 5.52%), FF5 yields 0.30% (annualized to 3.66%), and FF5 with momentum yields 0.26% (annualized to 3.12%). This shows that the IDE-based long-short portfolio generates positively significant returns that cannot be explained by these asset pricing models.

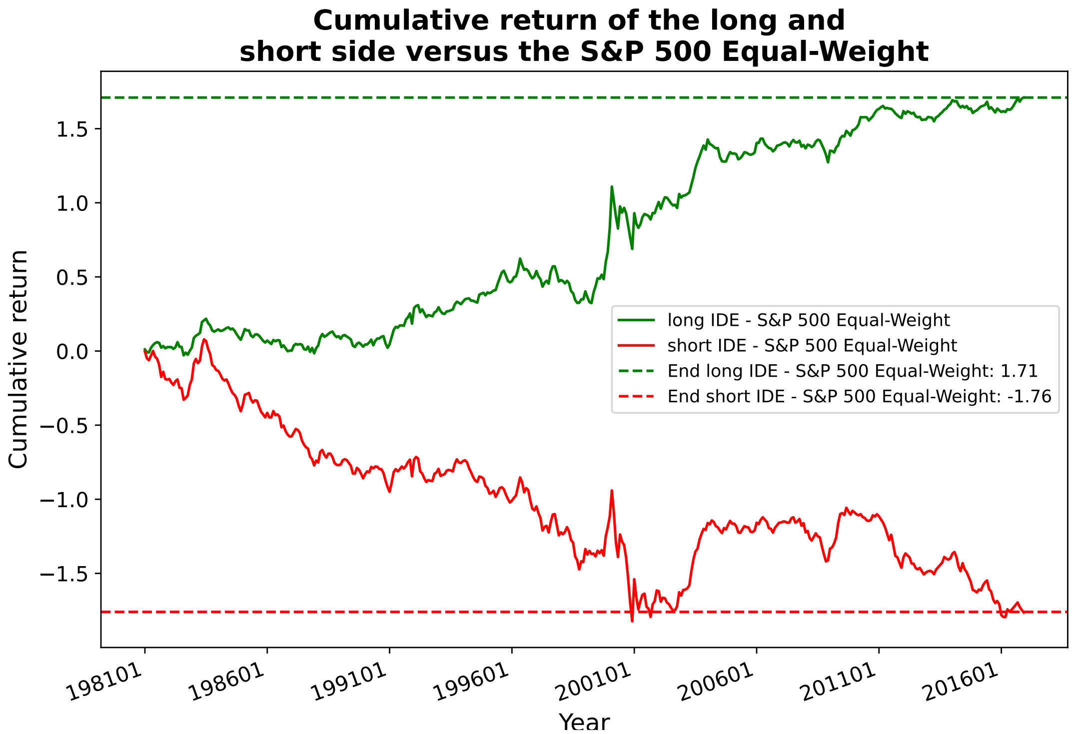

Figure 12: This figure shows the cumulative (monthly) return of the long and short sides of the IDE equal-weight portfolio relative to the equal-weight S&P500 index (without dividends). The long-short portfolios are rebalanced monthly.

7.2 Understanding the correlation between the long-side and short-side

Figure 11 shows a strong correlation between the long and short returns. In this subsection, we decompose this high correlation using traditional risk factors. To compute the amount of correlation due to the inter-temporal variation of these factors, we first establish a baseline correlation between raw returns in the long and short sides at 0.94. Then, for each set of factors, we compute their contribution to the baseline correlation in the following way:

$$\Delta\rho = 0.94 - \text{corr}(r^{\text{adj}}_{\text{long}}, r^{\text{adj}}_{\text{short}}), \tag{9}$$

where $r^{\text{adj}}_{\text{long}}$ and $r^{\text{adj}}_{\text{short}}$ are the residual of the IDE portfolio returns after regressing on the set of factors similar to equation 8.

We compute $\Delta\rho$ for CAPM, FF3, FF5, and FF5 with momentum. CAPM (the expected market return factor) contributes to roughly 0.11 in correlation. FF3 (small minus big and value minus growth) explains another 0.15 in correlation. FF5 (robust minus weak and conservative minus aggressive) explains 0.02 in correlation. Lastly, momentum contributes to another 0.08 in correlation.

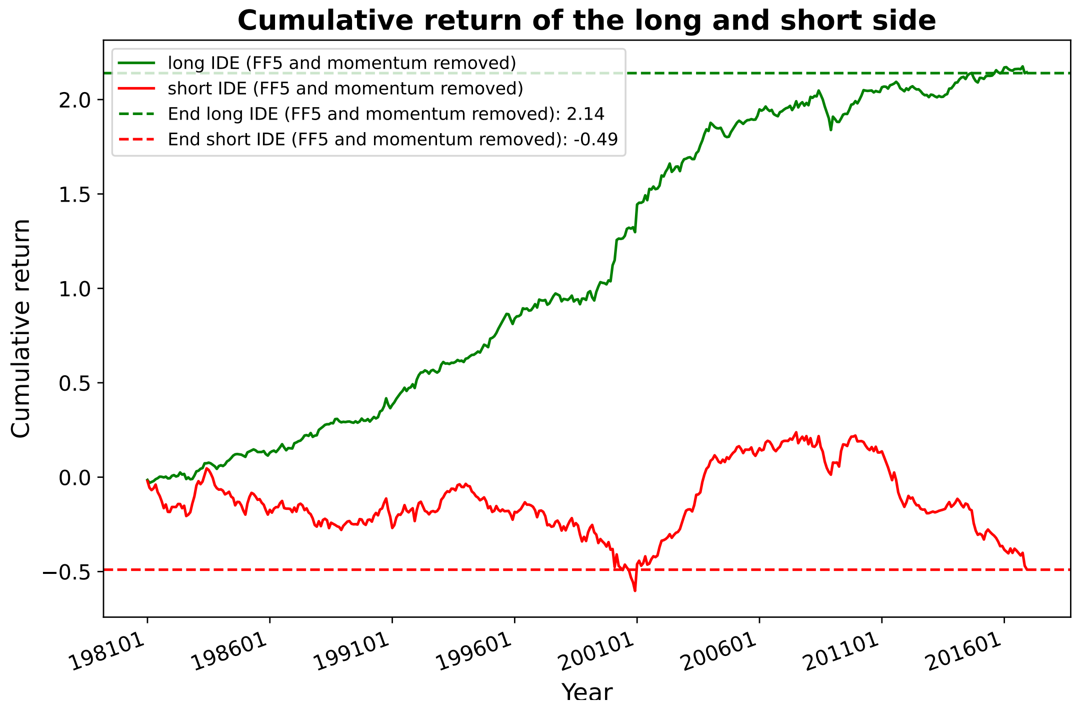

Figure 13: This figure shows the cumulative (monthly) return of the long and short sides of the IDE portfolio when Fama-French 5 factors and momentum are removed. The long-short portfolios are equal-weight and rebalanced monthly.

Overall, the biggest decrease in correlation happens when including the two factors: small minus big and value minus growth, and expected market return and momentum also contribute to large decreases in correlation. After all five Fama French factors and momentum are removed from the long and short sides, the correlation drops from 0.94 to 0.58. Figure 13 shows the P&L of the long and short sides of the IDE-based portfolio after Fama-French 5 and momentum are removed from the returns.

7.3 Firm characteristics and return in the IDE portfolio

In this section, we study how firm characteristics correlate with returns on the long and short sides of the portfolio. In particular, we run the following regression for the long and short-side portfolio returns separately:

$$r_t^s = \alpha + \beta FF_t + \gamma Z_t^s + \epsilon_t^s, \tag{10}$$

where $s \in \{\text{long}, \text{short}\}$, r_t^s is the return of the long (short) side of the portfolio in month t, FF_t are the Fama-French 5 factors in month t, and Z_t^s are the average characteristics among firms selected to be on the long (short) side in month t based on their IDE rankings.

Insert tables *A*23 and *A*24 here.

As shown in table A23, the two factors that have significant associations with the current returns of firms with the lowest IDE values are profitability and asset growth. In particular, in months when these firms are more profitable, the portfolio return of that month is expected to be high. On the other hand, in months when the firms have high asset growth, the corresponding portfolio returns are lower. In other words, the returns of firms with low displacement are positively driven by current profit but negatively driven by potential future growth opportunities. By definition, these low IDE firms are ones whose technologies are different from recent high-value innovations. Thus, it is reasonable to expect investors to value these firms based on their current ability to generate profit and against their potential to capture future opportunities. In addition, as shown in table A24, the firm characteristic significantly associated with the current return of the highest IDE firms is R&D expense. A higher R&D expenditure corresponds to a lower current return. Note that as shown in table A3, these firms with high IDE values tend to have high R&D expenses. Therefore, one potential explanation is that an even higher R&D expenditure indicates inefficiencies in the firms' current research and development infrastructure, making investors value these firms lower than more efficient firms. The exact rationalization for this phenomenon is left for future research.

8 IDE for out-of-sample prediction

In the previous sections, we have shown that IDE is significantly associated with firms' profit growth, and this significance is robust to subsetting by firm types and changing the underlying model specification. Now, we show that IDE has a significant out-of-sample predictive power of firms' profit growth. More specifically, the question we address in this section is whether our IDE measure can significantly improve the performance of predictive machine learning models.

To assess this, we partition our dataset based on a timestamp. Specifically, we train and validate more than 20 predictive machine learning models using data up to 2010, while testing their performance on data in years after. The objective of these models is to predict whether a firm's profit will increase or decrease in the next T years. Our modeling approach includes a diverse set of algorithm classes: K-nearest neighbors, tree-based methods, boosting, neural networks, and weighted ensembles of these models.

The K-nearest neighbors (KNN) model predicts outcomes in the test set by identifying the labels of the K closest points in the training set. In our analysis, the task is to predict the direction of profit growth—whether a firm's profit will increase or decrease over the next several years. Since this is a binary classification problem, the KNN model predicts the majority class among the K nearest neighbors for each test observation. Essentially, we use the training set to find the firms that have characteristics that are closest to a given firm in the test set, and we compute the label of this new data point via majority voting of the observed labels of the firms that are similar to it.

Tree-based methods construct individual decision trees as weak learners and aggregate them to form the final prediction. If we think of each weak learner as a human labeler, this tree-based approach is similar to recruiting multiple independent labelers and using the wisdom of the crowd to make the final predictions. Gradient boosting also uses multiple learners, but these learners are built sequentially to gradually improve the performance. This process is analogous to using multiple human labelers but the next labeler observes the mistakes made by the previous ones and can try to address them. For a discussion on gradient boosting techniques, see Appendix F.

Neural networks address this problem more directly by using only one model that predicts labels based on firm characteristics. These models fit nonlinear functions using the training sample, freeze the parameters, and apply the same functions in the test set to make predictions. For details on neural networks and their application, please refer to Appendix D.

The weighted ensemble approach combines predictions from multiple models (e.g., boosting, neural networks) using a stacking technique across N levels. This method involves training a meta-model that uses both the predictions from the base models and the raw input features to produce a final prediction. The stacking process is iteratively applied for N levels. For example, when $N = 1$, let $X_{n,d}$ be the $n \times d$ matrix of raw inputs and $Y_{n,M}$ be the $n \times M$ matrix of predictions from the ensemble's M base models. The ensemble then maps the concatenated matrix of $X_{n,d}$ and $Y_{n,M}$ to a set of n final predictions.

We predict the direction of profit growth over the next 1, 3, and 5 years to examine the predictive power of IDE on short- and long-term profit growth. In particular, given a forward-looking interval T, and a model m, we optimize the following objective function (summed over all firms and years):

$$\mathcal{L}(\mathbb{1}\left(\Pi_{f,t+T} > \Pi_{f,t}\right), g_m^T(X_{f,t})) = -\,\mathbb{1}\left(\Pi_{f,t+T} > \Pi_{f,t}\right)\log(g_m^T(X_{f,t})) - \left(1 - \mathbb{1}\left(\Pi_{f,t+T} > \Pi_{f,t}\right)\right)\log(1 - g_m^T(X_{f,t})), \quad (11)$$

where g_m^T is the predictive model (that outputs a number between 0 and 1, inclusive) and $X_{f,t}$ are the IDE values, capital stock, employment, current profit, profitability, asset growth, and the fixed effects year, industry code, industry category, and firm ID (PERMNO) (we treat these fixed effects as categorical inputs). All continuous independent variables are standardized within the year and industry category.[12]

Insert tables *A25* and *A26* and *A27* here.

Tables A25, A26, and A27 show that unsurprisingly weighted-ensemble with three levels of stacking performs the best in the validation set because the learned weights are multiplied to the predictions of individual models, so it is always expected to performance at least similar to the best individual model in the validation set (without time-based splitting). In addition, we observe that the performances on the validation set and test set are positively correlated. This shows that our models are likely not overfitted. A common baseline to evaluate the performance of these models is majority voting: predicting that all data points in the test set are in the majority class. In our study, the majority class is always "increasing" because firm profits naturally tend to increase over time. Our models significantly outperform the majority voting strategies in all three cases.[13] This shows our best models have learned nuanced information beyond a naive guess.

The important question is to examine how much our IDE measure contributes to this learning and whether this increment is statistically significant. We measure the importance of each variable in the best out-of-sample (test set) predictor of each forecasting horizon – neural network with bagging and fast AI implementation (1 level of stacking) for 1-year predictions, ExtraTrees with bagging based on entropy splitting rule (2 levels of stacking) for 3-year forecasting, and ExtraTrees with bagging based on Gini impurity

[12]The tree-based models based on Gini impurity use a splitting rule that minimizes the Gini Impurity of the subsamples defined by

$$\text{Gini} = 2p_1(1 - p_1) = 2p_0p_1$$

where p_0 is the proportion of samples in the subsample of the training set in class 0 (decreasing future profit) and p_1 is the proportion of samples in class 1 (increasing future profit).

[13]In the sample of firms with recorded 1-year future profits, the majority class makes up 59% of the test data, and they make up 61% among firms with recorded 3-year (5-year) future profits.

splitting rule (2 levels of stacking) for 5-year forecasting. The importance of each variable is measured using a permutation-based ablation test. In particular, we take a 5000-data point random sample from the testing data, and for each independent variable, we randomly shuffle the order of its values in this subset. Then, we compute the drop in the prediction accuracy when the perturbed data is fed into the trained model. We repeat this process 5 times with different subsamples. We compute the accuracies of these 5 subsamples and compute the means and standard errors of these perturbed accuracies. The mean drop in accuracy compared to the original data is used as the raw importance values. We then scale them so they sum up to 1, which are the values shown in figure 14. Then, we perform a t-test on the means and standard errors of the full test data and the perturbed samples to compute the statistical significance of including each variable in the best predictive model. Intuitively, this method allows us to measure the performance loss when one variable is lost (completely noisy) and evaluate the statistical significance using the accuracies before and after random shuffling.

As shown in figure 14, asset growth is the strongest predictor of profit changes in the next year. This may be due to profits continuing the asset growth momentum from the previous year. In addition, Fama and French (2015) documents asset growth as an indicator of potential growth opportunities. Having more growth opportunities in the near future may also facilitate profit growth. In 3- and 5-year predictions, firm ID and current profit are consistently the most important variables since the profit growth of firms likely depends on their identities and their current profit level. Moreover, we observe that our IDE measure is significantly important in longer-term predictions (3 and 5 years) and is ranked in the middle of all variables. This shows that displacement events caused by other firms' innovations have a significant role in the long-term growth of the focal firm.[14]

In addition, similar to section 6, we repeat this fully predictive exercise on a few subsamples. We confirm that this result – IDE is statistically significantly predictive of profit growth – is robust to sample selection. In particular, we run this analysis on large firms, firms with low profitability, and firms with below median intangible asset ratio in each year and industry category. The importance values are shown in figures F1, F2, and F3 in Appendix G. In sum, we find that IDE is predictive of profit growth in an out-of-sample setting and for a variety of firm types.

[14]It is expected that the year variable is not important in forecasting profit growth because the years in the testing sample (2011 to 2015) never appeared in the training set. Therefore, the trained models cannot use a categorical variable that takes unseen values to make predictions.

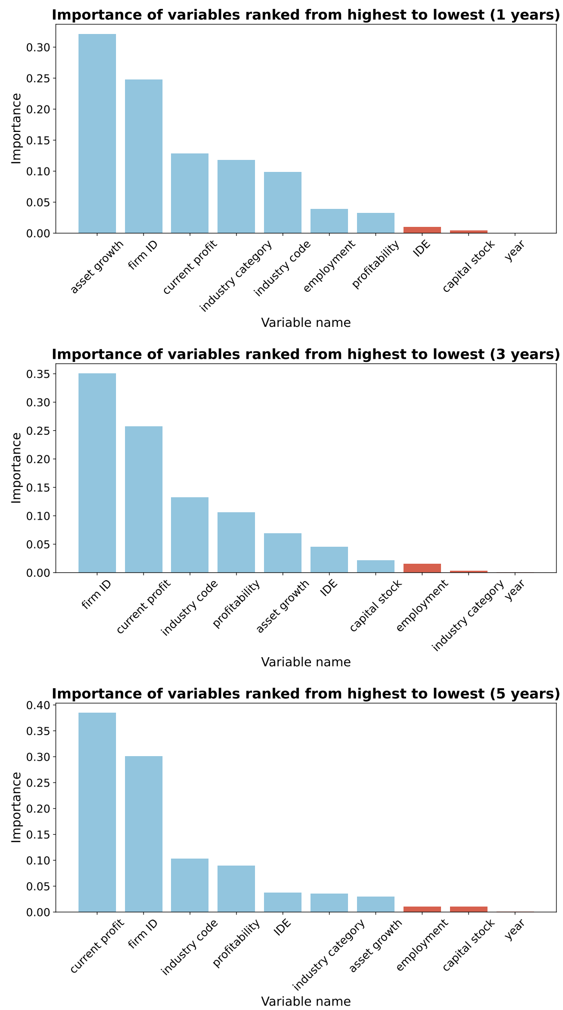

Figure 14: These figures show the importance of different variables in the best predictive model of profit direction changes over the next 1, 3, and 5 years. The importance values are scaled to sum to 1, and the ones colored blue are significant at the 5% level while the ones in red are not.

9 Conclusion

We develop two types of text-based measures of Innovation Displacement Exposure (IDE). Our main measure is based on patents and GPT4o-generated technology descriptions and can be applied to all firms. Our secondary displacement measure is purely based on patents; therefore, this measure can only be computed for innovative firms with many patents. We use these measures to study how IDE is associated with firms' profit growth. We find that the focal firm's IDE is significantly negatively associated with its profit growth in the next 7 years. This negative association exacerbates each year, and it is particularly severe among non-innovative firms. Moreover, we demonstrate that this negative innovation displacement effect (the negative association between IDE and profit growth) is prevalent among different types of firms and under different types of model specifications. In addition, we apply our main measure in two practical extensions. First, we investigate the potential of our main displacement measure in stock trading and find that stocks exposed to higher displacement are expected to have higher returns in the next year, after adjusting for exposures to factor models. This is consistent with Kogan, Papanikolaou and Stoffman (2020). Finally, we build over 20 fully predictive forecasting models that take current firm characteristics including IDE, and predict the direction of each firm's profit growth in the future. We find that IDE can significantly improve the performance of the best predictive models and is ranked around the middle of all variables in terms of predictive importance.

Our study relies on a few assumptions. First, we assume that text embedding models can effectively extract meaningful numerical representations from the information embedded within patent texts and technology descriptions. The high significance of our results indicates that the embedding model used (OpenAI's text-embedding-3-large) successfully captures information that reflects both the similarities and differences between various technologies and innovations. Moreover, as the frontier of Large Language Models advances, we believe these embeddings will become increasingly more powerful. Another limitation of our method is the potential of look-ahead biases that exist in GPT4o-generated technology descriptions. We use four approaches to alleviate this concern. First, we develop an alternative definition of innovation-based displacement that only uses patents instead of relying on the knowledge of the generative model. Our main results hold using this alternative measure. Secondly, we use SEC 10K files to show that GPT4o is aware of the temporal ordering of technologies used by these firms. In addition, we use the technology descriptions generated by GPT4o to show that it is aware of the ordering of technologies that are mentioned in its own descriptions. Lastly, we define alternative technology descriptions based on firms' 10K filings and patents. We show that the resulting IDE measure is highly positively correlated with our main IDE measure, both with and without industry-year standardization. We believe similar approaches can be used to evaluate the extent of look-ahead bias in other applications where Large Language Models are used to generate augmentative information based on their knowledge.

Our study begins an agenda to investigate firms' displacement exposures due to other firms' innovations. One potential direction is to examine the source of heterogeneity in the magnitude of the association between innovation displacement exposure and profit growth. In addition, further research can aim to develop an ex-ante measure that predicts future displacement events each firm will experience. Moreover, more efforts are needed to elicit the asset pricing implications of the IDE measure such as using it to construct a risk factor. Furthermore, even though we provide some guidelines on evaluating the look-ahead bias in LLM-based information augmentation, further research is needed to expand this analysis to different applications and develop other systematic evaluation methods.

References

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. 2018. "Innovation, reallocation, and growth." *American Economic Review*, 108(11): 3450–3491.

Aghion, Philippe, and Peter Howitt. 1992. "Growth Through Creative Destruction." *Econometrica*, 60(2): 323–51.

Alexopoulos, Michelle. 2011. "Read all about it!! What happens following a technology shock?" *American Economic Review*, 101(4): 1144–1179.

Athey, Susan, and Guido W Imbens. 2019. "Machine Learning Methods that Economists Should Know About." *Annual Review of Economics*, 11: 685–725.

Chen, Hui, Ali Kakhbod, Maziar Kazemi, and Hao Xing. 2023. *Process Intangibles and Agency Conflicts.* SSRN.

Chen, Mark A, Qinxi Wu, and Baozhong Yang. 2019. "How valuable is FinTech innovation?" *The Review of Financial Studies*, 32(5): 2062–2106.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/debiased machine learning for treatment and structural parameters." *The Econometrics Journal*, 21(1): C1–C68.

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

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

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

Erel, Isil, Léa H Stern, Chenhao Tan, and Michael S Weisbach. 2021. "Selecting directors using machine learning." *Review of Financial Studies*, 34(7): 3226–3264.

Erickson, Nick, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. 2020. "Autogluon-tabular: Robust and accurate automl for structured data." *arXiv preprint arXiv:2003.06505*.

Fama, Eugene F, and Kenneth R French. 2015. "A five-factor asset pricing model." *Journal of Financial Economics*, 116(1): 1–22.

Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier. 2024. "ChatGPT and Perception Biases in Investments: An Experimental Study." *Available at SSRN 4787249*.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as data." *Journal of Economic Literature*, 57(3): 535–74.

Hansen, Stephen, Michael McMahon, and Andrea Prat. 2018. "Transparency and deliberation within the FOMC: a computational linguistics approach." *Quarterly Journal of Economics*, 133(2): 801–870.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring technological innovation over the long run." *American Economic Review: Insights*, 3(3): 303–20.

Klette, Tor Jakob, and Samuel Kortum. 2004. "Innovating firms and aggregate innovation." *Journal of Political Economy*, 112(5): 986–1018.

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

Kogan, Leonid, Dimitris Papanikolaou, and Noah Stoffman. 2020. "Left behind: Creative destruction, inequality, and the stock market." *Journal of Political Economy*, 128(3): 855–906.

Kortum, Samuel. 1993. "Equilibrium r&d and the patent–r&d ratio: Us evidence." *The American Economic Review*, 83(2): 450–457.

Kortum, Samuel, and Josh Lerner. 1998. "Stronger protection or technological revolution: what is behind the recent surge in patenting?" Vol. 48, 247–304, Elsevier.

Lev, Baruch, and Suresh Radhakrishnan. 2005. "The valuation of organization capital." In *Measuring capital in the new economy*. 73–110. University of Chicago Press.

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

Shea, John. 1998. "What do technology shocks do?" *NBER macroeconomics annual*, 13: 275–310.

Online Appendix

1. Appendix A includes all of the tables used in this paper.
2. Appendix B includes details about validity checks against look-ahead bias in GPT4o.
3. Appendix C introduces a modified IDE measure where similarities scale nonlinearly with respect to cosine.
4. Appendix D discusses the basic architecture of neural networks and the GPT family of models.
5. Appendix E discusses the details of debiased machine learning.
6. Appendix F introduces gradient boosting and the XGBoost model.

A Tables

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.015***	-0.027***	-0.039***	-0.047***	-0.056***	-0.063***	-0.067***
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)	(0.007)	(0.008)
profitability t	-0.012***	-0.028***	-0.040***	-0.049***	-0.058***	-0.060***	-0.066***
	(0.003)	(0.005)	(0.006)	(0.007)	(0.007)	(0.008)	(0.009)
log asset growth	0.074***	0.087***	0.097***	0.098***	0.101***	0.104***	0.104***
	(0.003)	(0.005)	(0.006)	(0.007)	(0.008)	(0.008)	(0.008)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year× *IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	118,467	108,556	99,765	91,965	85,075	78,950	73,406
R^2	0.063	0.065	0.069	0.069	0.073	0.077	0.080
Adjusted R^2	0.053	0.054	0.057	0.057	0.060	0.062	0.064
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A1: This table shows the association between IDE and profit growth. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:	
	Baseline	Additional percent change in profit per year
	(1)	(2)
IDE	-0.004	-0.009***
	(0.002)	(0.001)
Year	✓	
Ind Code	✓	
Year×*IndCategory*	✓	
Observations	656,184	
R^2	0.132	
Adjusted R^2	0.119	
Note:		*p<0.1; **p<0.05; ***p<0.01

TABLE A2: This table shows the association between IDE and profit growth over time (stacking the data from years 1 to 7). We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

Variable	Estimate	Cluster s.e.	t value	Pr(>|t|)
Log capital stock	0.6024	0.0351	17.141	< 2e-16 ***
Log employment	-0.0809	0.0283	-2.855	0.00728 ***
Log profit	-0.1118	0.0299	-3.733	0.00069 ***
Profitability	0.0578	0.0111	5.189	9.77e-06 ***
Log asset growth	-0.0105	0.0074	-1.421	0.16445
Total asset	-0.0363	0.0133	-2.721	0.01019 **
Capital Expenditures	-0.3966	0.0259	-15.291	< 2e-16 ***
Research and development expenditures	0.0626	0.0177	3.537	0.00119 **
Market Cap	0.0144	0.0114	1.254	0.21824
Year		✓		
Ind Code		✓		
Year×*IndCategory*		✓		
Observations		74,397		
R^2		0.205		
Adjusted R^2		0.191		

TABLE A3: This table shows the association between firm characteristics and our displacement measure. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level. An asinh transformation is applied to capital and R&D expenditures to stable the variance and allow for non-positive values.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.014***	-0.022***	-0.027***	-0.030***	-0.036***	-0.046***	-0.051***
	(0.002)	(0.003)	(0.005)	(0.006)	(0.008)	(0.008)	(0.009)
profitability t	-0.012***	-0.029***	-0.042***	-0.050***	-0.058***	-0.058***	-0.065***
	(0.004)	(0.005)	(0.007)	(0.008)	(0.009)	(0.010)	(0.012)
log asset growth	0.071***	0.086***	0.099***	0.101***	0.107***	0.112***	0.113***
	(0.004)	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	68,628	63,110	58,210	53,836	49,921	46,454	43,346
R^2	0.102	0.118	0.126	0.129	0.138	0.142	0.145
Adjusted R^2	0.086	0.100	0.107	0.109	0.116	0.119	0.121
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A4: This table shows the association between IDE and profit growth after we remove the part of IDE that can be explained by traditional firm characteristics. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:		
	IDE		
	$t+5$	$t+7$	$t+10$
capital expenditure	-0.043	-0.050	-0.070*
	(0.029)	(0.030)	(0.036)
R&D expenditure	-0.163**	-0.107*	-0.085
	(0.064)	(0.055)	(0.051)
Asset total	0.147***	0.193***	0.259**
	(0.049)	(0.066)	(0.104)
Profitability	-0.337***	-0.335***	-0.389***
	(0.091)	(0.100)	(0.130)
Year	✓	✓	✓
Ind Code	✓	✓	✓
Year$\times$*IndCategory*	✓	✓	✓
Observations	50,325	40,980	30,086
R^2	0.257	0.267	0.281
Adjusted R^2	0.240	0.248	0.258
Note:	*p<0.1; **p<0.05; ***p<0.01		

TABLE A5: This table shows the association between future IDE and firm characteristics after we remove the first two principal components from the firm characteristics. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.020***	-0.033***	-0.043***	-0.048***	-0.051***	-0.054***	-0.050***
	(0.005)	(0.008)	(0.010)	(0.011)	(0.012)	(0.014)	(0.015)
profitability t	-0.007	-0.027***	-0.035***	-0.033***	-0.037**	-0.041**	-0.049***
	(0.008)	(0.009)	(0.010)	(0.012)	(0.014)	(0.015)	(0.017)
log asset growth	0.066***	0.072***	0.092***	0.098***	0.103***	0.105***	0.107***
	(0.007)	(0.009)	(0.010)	(0.011)	(0.011)	(0.012)	(0.013)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	20,127	20,127	20,127	20,127	20,127	20,127	20,127
R^2	0.151	0.170	0.181	0.185	0.189	0.189	0.185
Adjusted R^2	0.102	0.122	0.134	0.138	0.142	0.141	0.137
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A6: This table shows the association between IDE and profit growth among innovative firms (those who received at least 10 patents over our time interval). We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.017***	-0.030***	-0.044***	-0.054***	-0.064***	-0.071***	-0.077***
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)	(0.007)	(0.008)
profitability *t*	-0.012***	-0.028***	-0.041***	-0.053***	-0.063***	-0.066***	-0.072***
	(0.003)	(0.005)	(0.006)	(0.008)	(0.009)	(0.010)	(0.011)
log asset growth	0.075***	0.088***	0.097***	0.096***	0.098***	0.102***	0.101***
	(0.003)	(0.005)	(0.006)	(0.007)	(0.009)	(0.009)	(0.009)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year$\times$*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	98,340	88,429	79,638	71,838	64,948	58,823	53,279
R^2	0.103	0.123	0.134	0.141	0.152	0.158	0.165
Adjusted R^2	0.092	0.111	0.121	0.126	0.136	0.140	0.145
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A7: This table shows the association between IDE and profit growth among non-innovative firms (those who received fewer than 10 patents over our time interval). We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	raw measures	standardized measures
Pearson Correlation	0.96***	0.11***
Kendall-Tau correlation	0.83***	0.06***
Note:		*p<0.1; **p<0.05; ***p<0.01

TABLE A8: This table shows the correlation between the two measures: one based on patents and technologies, and an alternative purely based on patents. The left column shows the Pearson and Kendall-Tau correlation between the two measures as their raw values, and the right column shows the correlation between the two measures once standardized in each year and industry category. The sample is for innovative firms where the alternative patent-based measure is well-defined.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Original IDE	-0.017***	-0.029***	-0.038***	-0.042***	-0.045***	-0.047***	-0.043***
	(0.004)	(0.007)	(0.009)	(0.010)	(0.011)	(0.013)	(0.014)
Alternative IDE	-0.009**	-0.020***	-0.027***	-0.031***	-0.033***	-0.039***	-0.045***
	(0.003)	(0.005)	(0.006)	(0.008)	(0.009)	(0.011)	(0.012)
profitability t	-0.010	-0.030***	-0.039***	-0.040***	-0.044***	-0.047***	-0.054***
	(0.007)	(0.008)	(0.009)	(0.011)	(0.013)	(0.014)	(0.016)
log asset growth	0.065***	0.072***	0.091***	0.099***	0.106***	0.108***	0.110***
	(0.006)	(0.008)	(0.009)	(0.010)	(0.011)	(0.012)	(0.013)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year$\times$*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	25,529	25,529	25,529	25,529	25,529	25,529	25,529
R^2	0.156	0.175	0.186	0.191	0.194	0.191	0.185
Adjusted R^2	0.116	0.136	0.148	0.153	0.156	0.153	0.146
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A9: This table shows the association between the original and alternative IDE measures and profit growth. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015, and the set only covers innovative firms. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1}-\Pi_t$	$\Pi_{t+2}-\Pi_t$	$\Pi_{t+3}-\Pi_t$	$\Pi_{t+4}-\Pi_t$	$\Pi_{t+5}-\Pi_t$	$\Pi_{t+6}-\Pi_t$	$\Pi_{t+7}-\Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.014***	-0.022***	-0.032***	-0.039***	-0.044***	-0.047***	-0.053***
	(0.002)	(0.005)	(0.006)	(0.006)	(0.008)	(0.009)	(0.011)
profitability t	-0.004	-0.017***	-0.029***	-0.036***	-0.045***	-0.045***	-0.048***
	(0.003)	(0.005)	(0.007)	(0.008)	(0.010)	(0.011)	(0.013)
log asset growth	0.072***	0.084***	0.094***	0.091***	0.099***	0.105***	0.110***
	(0.003)	(0.006)	(0.006)	(0.007)	(0.008)	(0.009)	(0.010)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	52,490	47,427	43,039	39,240	35,938	33,080	30,541
R^2	0.104	0.119	0.132	0.141	0.156	0.163	0.176
Adjusted R^2	0.083	0.096	0.107	0.113	0.126	0.131	0.142
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A10: This table shows the association between IDE and profit growth among firms with high intangible capital. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.016***	-0.031***	-0.043***	-0.052***	-0.064***	-0.075***	-0.078***
	(0.003)	(0.004)	(0.005)	(0.007)	(0.008)	(0.009)	(0.010)
profitability t	-0.017***	-0.035***	-0.044***	-0.046***	-0.053***	-0.057***	-0.062***
	(0.006)	(0.008)	(0.011)	(0.012)	(0.014)	(0.015)	(0.017)
log asset growth	0.074***	0.081***	0.089***	0.090***	0.087***	0.089***	0.087***
	(0.004)	(0.007)	(0.007)	(0.010)	(0.010)	(0.010)	(0.011)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	52,445	48,630	45,140	41,951	39,090	36,491	34,115
R^2	0.121	0.145	0.160	0.168	0.180	0.187	0.189
Adjusted R^2	0.100	0.123	0.136	0.143	0.153	0.159	0.159
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A11: This table shows the association between IDE and profit growth among firms with low intangible capital. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.011***	-0.018***	-0.032***	-0.040***	-0.048***	-0.051***	-0.056***
	(0.003)	(0.004)	(0.005)	(0.007)	(0.009)	(0.011)	(0.012)
profitability *t*	-0.001	-0.016***	-0.033***	-0.044***	-0.053***	-0.056***	-0.065***
	(0.004)	(0.006)	(0.007)	(0.008)	(0.009)	(0.011)	(0.013)
log asset growth	0.063***	0.068***	0.072***	0.063***	0.062***	0.064***	0.064***
	(0.004)	(0.007)	(0.008)	(0.009)	(0.010)	(0.009)	(0.010)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	55,923	49,629	44,263	39,675	35,730	32,375	29,412
R^2	0.111	0.131	0.146	0.154	0.167	0.175	0.184
Adjusted R^2	0.091	0.109	0.122	0.127	0.137	0.143	0.149

Note: *p<0.1; **p<0.05; ***p<0.01

TABLE A12: This table shows the association between IDE and profit growth among firms with below median market capitalization. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.017***	-0.032***	-0.041***	-0.049***	-0.060***	-0.069***	-0.073***
	(0.003)	(0.004)	(0.005)	(0.006)	(0.007)	(0.007)	(0.008)
profitability t	-0.005	-0.013**	-0.019**	-0.024***	-0.030***	-0.032***	-0.036***
	(0.004)	(0.006)	(0.007)	(0.008)	(0.008)	(0.009)	(0.011)
log asset growth	0.073***	0.086***	0.100***	0.108***	0.110***	0.112***	0.110***
	(0.004)	(0.005)	(0.006)	(0.008)	(0.008)	(0.009)	(0.010)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	61,416	57,857	54,501	51,367	48,503	45,811	43,297
R^2	0.151	0.181	0.191	0.195	0.204	0.206	0.207
Adjusted R^2	0.134	0.163	0.172	0.175	0.183	0.184	0.184

Note: *p<0.1; **p<0.05; ***p<0.01

TABLE A13: This table shows the association between IDE and profit growth among firms with at or above median market capitalization. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.014***	-0.025***	-0.037***	-0.044***	-0.050***	-0.058***	-0.062***
	(0.002)	(0.004)	(0.005)	(0.006)	(0.008)	(0.009)	(0.010)
profitability t	-0.006	-0.022***	-0.033***	-0.041***	-0.051***	-0.054***	-0.059***
	(0.004)	(0.006)	(0.008)	(0.009)	(0.010)	(0.011)	(0.013)
log asset growth	0.079***	0.091***	0.102***	0.099***	0.103***	0.102***	0.099***
	(0.005)	(0.007)	(0.008)	(0.010)	(0.012)	(0.012)	(0.012)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	67,972	61,997	56,674	51,969	47,869	44,220	40,956
R^2	0.095	0.115	0.131	0.140	0.152	0.159	0.167
Adjusted R^2	0.084	0.103	0.119	0.127	0.137	0.144	0.150
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A14: This table shows the association between IDE and profit growth using data in the 20th century. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 1999. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1}-\Pi_t$	$\Pi_{t+2}-\Pi_t$	$\Pi_{t+3}-\Pi_t$	$\Pi_{t+4}-\Pi_t$	$\Pi_{t+5}-\Pi_t$	$\Pi_{t+6}-\Pi_t$	$\Pi_{t+7}-\Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.016***	-0.028***	-0.038***	-0.049***	-0.061***	-0.065***	-0.070***
	(0.003)	(0.004)	(0.006)	(0.007)	(0.008)	(0.010)	(0.011)
profitability t	-0.017***	-0.030***	-0.042***	-0.051***	-0.057***	-0.059***	-0.064***
	(0.005)	(0.006)	(0.008)	(0.009)	(0.010)	(0.012)	(0.013)
log asset growth	0.067***	0.080***	0.090***	0.094***	0.095***	0.102***	0.104***
	(0.004)	(0.006)	(0.008)	(0.009)	(0.009)	(0.010)	(0.011)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	50,495	46,559	43,091	39,996	37,206	34,730	32,450
R^2	0.118	0.136	0.140	0.141	0.148	0.148	0.150
Adjusted R^2	0.106	0.123	0.126	0.126	0.133	0.131	0.132
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A15: This table shows the association between IDE and profit growth using data in the 21st century. We control for the log asset growth and profitability among other variables. The data spans from 2000 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.013***	-0.022***	-0.034***	-0.041***	-0.049***	-0.057***	-0.062***
	(0.003)	(0.004)	(0.006)	(0.008)	(0.009)	(0.010)	(0.011)
profitability *t*	-0.140***	-0.216***	-0.289***	-0.330***	-0.361***	-0.390***	-0.405***
	(0.012)	(0.014)	(0.018)	(0.021)	(0.022)	(0.027)	(0.031)
log asset growth 0.076***	0.063***	0.070***	0.078***	0.083***	0.078***	0.076***	0.078***
	(0.004)	(0.006)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	57,406	52,290	47,901	43,946	40,484	37,526	34,835
R^2	0.129	0.163	0.181	0.189	0.200	0.212	0.220
Adjusted R^2	0.110	0.142	0.159	0.166	0.175	0.185	0.191
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A16: This table shows the association between IDE and profit growth among firms with below median profitability. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.008***	-0.017***	-0.025***	-0.033***	-0.040***	-0.044***	-0.048***
	(0.001)	(0.003)	(0.004)	(0.005)	(0.007)	(0.008)	(0.009)
profitability t	-0.007**	-0.012***	-0.014**	-0.018**	-0.020**	-0.017	-0.021*
	(0.003)	(0.004)	(0.006)	(0.008)	(0.010)	(0.010)	(0.011)
log asset growth	0.072***	0.088***	0.104***	0.112***	0.119***	0.121***	0.125***
	(0.003)	(0.005)	(0.006)	(0.006)	(0.009)	(0.009)	(0.010)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	61,061	56,266	51,864	48,019	44,591	41,424	38,571
R^2	0.117	0.115	0.121	0.128	0.140	0.139	0.145
Adjusted R^2	0.099	0.095	0.100	0.105	0.116	0.113	0.117
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A17: This table shows the association between IDE and profit growth among firms with at or above median profitability. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

<table>
<tr><th>Variable</th><th>Mean (the rest)</th><th>Mean (U.S.)</th><th>t value</th><th>Pr(>|t|)</th></tr>
<tr><td>Market cap</td><td>-0.17</td><td>0.15</td><td>-76.40</td><td>< 2e-16 ***</td></tr>
<tr><td>Profitability</td><td>-0.06</td><td>0.05</td><td>-24.63</td><td>< 2e-16 ***</td></tr>
<tr><td>Current profit (log)</td><td>-0.27</td><td>0.24</td><td>-114.96</td><td>< 2e-16 ***</td></tr>
<tr><td>IDE</td><td>-0.24</td><td>0.21</td><td>-89.35</td><td>< 2e-16 ***</td></tr>
</table>

TABLE A18: This table shows the summary statistics of the two subsamples: one is consisted of firms in the United States with at least 1 patent, and the other one contains the rest of the data. All values are standardized each year in the t-test.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.015***	-0.024***	-0.034***	-0.041***	-0.046***	-0.054***	-0.058***
	(0.002)	(0.004)	(0.006)	(0.007)	(0.008)	(0.009)	(0.009)
profitability t	-0.011**	-0.027***	-0.037***	-0.046***	-0.055***	-0.055***	-0.062***
	(0.005)	(0.006)	(0.007)	(0.009)	(0.009)	(0.010)	(0.011)
log asset growth	0.067***	0.079***	0.092***	0.094***	0.096***	0.096***	0.099***
	(0.003)	(0.005)	(0.006)	(0.007)	(0.008)	(0.009)	(0.010)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year$\times$*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	69,338	65,126	61,158	57,435	54,007	50,858	47,923
R^2	0.107	0.126	0.136	0.143	0.149	0.153	0.160
Adjusted R^2	0.091	0.109	0.118	0.124	0.130	0.132	0.138
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A19: This table shows the association between IDE and profit growth among firms in the United States with at least 1 patent. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.004**	-0.013***	-0.021***	-0.024***	-0.032***	-0.037***	-0.034***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	73,406	73,406	73,406	73,406	73,406	73,406	73,406
R^2	0.0001	0.0003	0.001	0.001	0.001	0.001	0.001
Adjusted R^2	0.0001	0.0003	0.001	0.001	0.001	0.001	0.001
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE A20: This table shows the effect of IDE on profit growth identified by an XGBoost-based debiased machine learning algorithm. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year. All fixed effects are included in the XGBoost model.

	Dependent variable:	
	Baseline	Additional percent change in profit per year
	(1)	(2)
IDE	-0.003	-0.005***
	(0.003)	(0.001)
Observations	513,842	
R^2	0.001	
Adjusted R^2	0.001	
Note:		*p<0.1; **p<0.05; ***p<0.01

TABLE A21: This table shows the growth of the effect of IDE on profit growth over time using the debiased machine learning approach and XGBoost. We control for the log asset growth and profitability among other variables. The data spans from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.

Variable	Estimate	s.e.	t value	Pr(>|t|)
CAPM	0.5396	0.111	4.862	< 2e-16 ***
Fama-French 3	0.4597	0.110	4.179	< 2e-16 ***
Fama-French 5	0.3049	0.111	2.755	0.006 ***
Fama-French 5 and momentum	0.2629	0.111	2.372	0.018***
number of observations		432		

TABLE A22: This table shows the $alpha$ of the IDE long-short portfolio after controlling for traditional asset pricing factors. The return data is from 1981 to 2016.

Variable	Estimate	s.e.	t value	Pr(>|t|)
capital stock (log)	4.8148	6.196	0.777	0.438
current employment (log)	0.2260	5.327	0.042	0.966
current profit (log)	-7.0916	8.941	-0.793	0.428
profitability	10.7642	3.785	2.844	0.005 ***
asset growth (log)	-3.3637	1.478	-2.276	0.023 **
total asset	7.4112	8.277	0.895	0.371
capital expenditures (asinh)	-3.0508	4.400	-0.693	0.488
R&D expenditures (asinh)	3.1984	3.290	0.972	0.332
market capitalization	0.9682	6.126	0.158	0.875
number of observations		420		

TABLE A23: This table shows the association of average firm characteristics and the current return of the short side of the IDE-based portfolio. The return data is from 1981 to 2015.

Variable	Estimate	s.e.	t value	Pr(>|t|)
capital stock (log)	-1.0194	7.937	-0.128	0.898
current employment (log)	-4.3701	6.134	-0.712	0.477
current profit (log)	6.3656	9.445	0.674	0.501
profitability	1.1867	4.113	0.289	0.773
asset growth (log)	-0.0996	2.653	-0.038	0.970
total asset	-5.8590	4.529	-1.294	0.196
capital expenditures (asinh)	1.1003	4.309	0.255	0.799
R&D expenditures (asinh)	-4.9746	2.687	-1.851	0.065*
market capitalization	4.3129	4.446	0.970	0.333
number of observations		420		

TABLE A24: This table shows the association of average firm characteristics and the current return of the long side of the IDE-based portfolio. The return data is from 1981 to 2015.

Model $T = 1$	**Score Test**	**Score Val**
NeuralNetFastAI_BAG_L1	0.622330	0.668022
NeuralNetTorch_BAG_L1	0.616111	0.668098
XGBoost_BAG_L1	0.613623	0.665359
LightGBMLarge_BAG_L1	0.612452	0.664844
LightGBMXT_BAG_L1	0.610770	0.671361
LightGBM_BAG_L1	0.609453	0.661991
ExtraTreesEntr_BAG_L2	0.607916	0.688947
ExtraTreesGini_BAG_L2	0.603234	0.688776
CatBoost_BAG_L1	0.598844	0.690646
WeightedEnsemble_L2	0.598844	0.690646
LightGBM_BAG_L2	0.594601	0.688012
ExtraTreesEntr_BAG_L1	0.594527	0.652459
ExtraTreesGini_BAG_L1	0.591528	0.654701
CatBoost_BAG_L2	0.591381	0.702335
WeightedEnsemble_L3	0.591381	0.702335
RandomForestEntr_BAG_L2	0.587211	0.690045
RandomForestGini_BAG_L2	0.586992	0.690875
XGBoost_BAG_L2	0.580187	0.692592
LightGBMLarge_BAG_L2	0.578870	0.688384
LightGBMXT_BAG_L2	0.575285	0.691142
KNeighborsDist_BAG_L1	0.567457	0.582172
KNeighborsUnif_BAG_L1	0.567457	0.581256
NeuralNetFastAI_BAG_L2	0.567237	0.690999
RandomForestEntr_BAG_L1	0.554946	0.655407
NeuralNetTorch_BAG_L2	0.552678	0.691419
RandomForestGini_BAG_L1	0.550995	0.652936

TABLE A25: This table shows the predictive performance of all machine learning models we consider in a binary classification task where we predict whether the profit of each firm will increase in the next 1 year. The models are ranked from best to worst based on their test accuracy.

Model $T = 3$	**Score Test**	**Score Val**
ExtraTreesEntr_BAG_L2	0.633051	0.785706
ExtraTreesGini_BAG_L2	0.632967	0.786139
WeightedEnsemble_L2	0.632127	0.777522
ExtraTreesEntr_BAG_L1	0.627257	0.713099
ExtraTreesGini_BAG_L1	0.626165	0.716035
LightGBMLarge_BAG_L1	0.626081	0.747826
CatBoost_BAG_L2	0.625829	0.795040
WeightedEnsemble_L3	0.625829	0.795040
RandomForestEntr_BAG_L2	0.625745	0.788870
LightGBM_BAG_L2	0.623814	0.786628
LightGBMXT_BAG_L2	0.623226	0.789132
XGBoost_BAG_L2	0.622890	0.789519
LightGBMLarge_BAG_L2	0.622806	0.787823
LightGBM_BAG_L1	0.622302	0.741031
RandomForestGini_BAG_L2	0.621799	0.786992
LightGBMXT_BAG_L1	0.620455	0.700123
CatBoost_BAG_L1	0.620119	0.762077
RandomForestEntr_BAG_L1	0.619783	0.715056
RandomForestGini_BAG_L1	0.617852	0.715159
XGBoost_BAG_L1	0.617096	0.685349
NeuralNetTorch_BAG_L2	0.613738	0.788688
NeuralNetTorch_BAG_L1	0.613654	0.678997
NeuralNetFastAI_BAG_L2	0.612562	0.770909
NeuralNetFastAI_BAG_L1	0.609455	0.740667
KNeighborsUnif_BAG_L1	0.571417	0.609520
KNeighborsDist_BAG_L1	0.570829	0.611649

TABLE A26: This table shows the predictive performance of all machine learning models we consider in a binary classification task where we predict whether the profit of each firm will increase in the next 3 years. The models are ranked from best to worst based on their test accuracy.

Model $T = 5$	**Score Test**	**Score Val**
ExtraTreesGini_BAG_L2	0.643922	0.838938
ExtraTreesEntr_BAG_L2	0.643253	0.839541
NeuralNetFastAI_BAG_L2	0.643157	0.830936
RandomForestGini_BAG_L2	0.639430	0.840801
NeuralNetTorch_BAG_L2	0.639048	0.840332
LightGBMXT_BAG_L2	0.638284	0.839957
WeightedEnsemble_L2	0.637806	0.834448
CatBoost_BAG_L2	0.637041	0.851818
WeightedEnsemble_L3	0.637041	0.851818
RandomForestEntr_BAG_L2	0.636755	0.840466
LightGBMLarge_BAG_L2	0.635990	0.839876
XGBoost_BAG_L2	0.635799	0.841779
LightGBM_BAG_L2	0.635321	0.838791
LightGBMLarge_BAG_L1	0.632263	0.797979
ExtraTreesEntr_BAG_L1	0.627867	0.757743
ExtraTreesGini_BAG_L1	0.624904	0.760759
NeuralNetFastAI_BAG_L1	0.624618	0.814907
LightGBM_BAG_L1	0.624427	0.794266
CatBoost_BAG_L1	0.624235	0.807187
RandomForestEntr_BAG_L1	0.623089	0.747356
RandomForestGini_BAG_L1	0.621464	0.747236
XGBoost_BAG_L1	0.620413	0.708783
LightGBMXT_BAG_L1	0.616399	0.736621
NeuralNetTorch_BAG_L1	0.606843	0.693544
KNeighborsUnif_BAG_L1	0.582378	0.634303
KNeighborsDist_BAG_L1	0.582187	0.636917

TABLE A27: This table shows the predictive performance of all machine learning models we consider in a binary classification task where we predict whether the profit of each firm will increase in the next 5 years. The models are ranked from best to worst based on their test accuracy.

B Validity checks against the potential look-ahead bias of GPT4o

One of the components of our first displacement measure is a set of descriptions of the technologies used by each firm each year. These technologies are generated by GPT4o following a prompt that specifies the firm and year (the exact prompt is shown in section 2). A potential concern of this approach is the look-ahead bias in the GPT4o-generated summaries. In our main analysis, we developed a purely patent-based alternative IDE measure to alleviate this concern, as this measure does not ask GPT4o to retrieve hidden information from its knowledge base. In this subsection, we first present two other approaches we take that show GPT4o is aware of the temporal ordering of technologies. Then, we develop another set of technology descriptions based on 10K filings and patents and show that the resulting IDE values are highly correlated with the IDE values computed using GPT4o-generated technology descriptions.

B.1 10K-based ordering of technologies

We use the SEC 10K filings as a data source of technology descriptions with a definitive time stamp. We first randomly select 500 firm-year pairs in our dataset. Then, we extract the 10K filing for each firm in the corresponding year. Some firms are foreign issuers and do not file for 10K. Therefore, after removing these firms, we have a set of 304 firm-year pairs. For each pair, we also download the 10K filing of the firm from 3 years ago. Then, we pass each 10K file to GPT4o and ask it to summarize the technologies based on each document:

> *This file is a 10K form summarizing the performance of a firm in a year. Please read this website and summarize the technologies used by this firm this year in its day to day operations and research. Use no more than 500 tokens. Do not include any information about the year. Be as specific as possible and give a lot of details. For each technology, mention whether or not it was considered legacy in the field in t and whether or not there were disruptive technologies that could replace it. If there was a risk of disruption, mention companies and technologies that were threatening to the given technology: [Downloaded file]*

Then, each pair of summaries from the same firm 3 years apart is passed to GPT4o, and we ask GPT4o to determine which one is from an earlier year. The order of the two documents is randomized: half of the time, the summary of the earlier document appears first in the prompt, and the other times, the later one appears first. We use the prompt:

Given these two summaries above, tell me which one is from an earlier year: [newline] First summary: [first summary] [newline] Second summary: [second summary].

Lastly, we compare the GPT4o ordering and the correct ordering.

As an example, take the firm Apple and the year 2010, we first download the 10K filings of Apple in 2010 and 2007. Then, we ask GPT4o to create a short summary for each 10K document focusing on the technologies described in the corresponding filing. Next, we pass these two summaries to GPT4o in a randomized order. For example, if we randomly pick the summary based on the 2010 filing as the first one in the prompt. Our query to GPT4o would be

Given these two summaries above, tell me which one is from an earlier year: [newline] First summary: [Apple 2010 technology summary] [newline] Second summary: [Apple 2007 technology summary].

Then, we check if the response is that the second summary is from an earlier year.

We apply this process to the entire sample of firm-year pairs, and we observe that GPT4o gives the correct time ordering of the two technology summaries over 90% of the time. More specifically, the accuracy is balanced regardless of whether or not the earlier summary appears first in the prompt. When the earlier document appears first, GPT4o's accuracy is 91%, and when the later one appears first, GPT4o's accuracy is 92%. The two numbers are not statistically different.

B.2 GPT4o description-based ordering of technologies

Next, we test whether GPT4o can correctly order technologies based on the summaries we use to construct our displacement measure (discussed in section 2).

We conduct this test at two levels of varying difficulty. First, we select technology summaries from the same firm 5 years apart. Then, to increase the difficulty, we choose summaries that are 3 years apart. In each test, we randomly select 500 firm-year pairs first and then find the corresponding earlier documents. We drop the ones that do not have a match (the firms were not included 5 or 3 years ago).

In particular, given two summaries, we use the following prompt:

You are an economist reading descriptions of technologies used by firm in daily operations and research. You will read two summaries and tell me which one is from an earlier year. Respond with 1 if the earlier one is the first summary and 0 if the earlier one is the second summary. Do not include any words or symbols, just use an integer:

[newline] First summary: <summary 1>

[newline] Second summary:<summary 2>

Similar to the exercise we did with 10K files, the order of the first and second summaries is randomized: half of the time, the earlier summaries appear first, and in the other cases, the later summaries appear first. Overall, we observe that when the 2 documents are 5 years apart, GPT4o correctly orders them 77% of the time. In particular, when the earlier summary appears first, GPT4o correctly identifies it as the earlier one 99% of the time. When the separation is only 3 years, GPT4o achieves a 67% accuracy and 95% when the earlier summary appears first. Note that in both the 5-year and 3-year tests, the technology summaries are similar, so this task is challenging even for humans.

These exercises show that GPT4o understands the time ordering of technologies at a high accuracy.

B.3 Technology descriptions based on 10K filings and patents

In this section, we generate technology descriptions by combining summaries of 10K filings and patents. First, we take a random 20% of all firms and extract their 10K filings from the SEC Edgar website. Then, for this subset of data, we concatenate the patent summaries with the 10K summaries (tailored toward daily technologies) to form technology descriptions. Given a 10K filing in year t, we use the following prompt to ask GPT4o to summarize the document:

> *You are an economist studying firms' innovation. Summarize the following 10-K file focusing on technologies used by this firm in its day-to-day operations and research. First, give an overview and then discuss each technology. Describe whether or not each technology is legacy in year t. If it is describe which technologies in that year could be disruptive of this technology. The summary should be no more than 2 paragraphs or 500 tokens. The 10k file is the following: +[10K]*

Then, similar to the main IDE measure discussed in section 2, we concatenate technology summaries in years $t, t-1, \cdots, t-4$ to for the firm's technology description in year t and compute its embedding.

	raw measures	standardized measures
Pearson Correlation	0.89***	0.42***
Kendall-Tau correlation	0.75***	0.29***
Note:		*p<0.1; **p<0.05; ***p<0.01

TABLE B28: This table shows the correlation between the two measures: one based on patents and technologies, and an alternative based on 10Ks and patents. The left column shows the Pearson and Kendall-Tau correlation between the two measures as their raw values, and the right column shows the correlation between the two measures once standardized in each year and industry category.

Similar to section 5, we compute the Pearson and Kendall-Tau correlations between this alternative IDE and our main measure. As shown in table B28. These two measures are highly positively correlated both in their raw form or when standardized each year and industry category. The raw values have a Pearson correlation of over 0.8 and a Kendall-Tau (rank) correlation of over 0.7. These correlations remain high when we standardize the two versions of IDE in each year and industry. The Pearson correlation becomes 0.4 and Kendall-Tau is around 0.3.

C Nonlinear IDE

Our definition of IDE in equation 1 assumes that the magnitude of exposure scales linearly with the cosine similarity between firm j's innovations and firm i's technologies. In this section, we relax this assumption and develop a measure where the cosine similarity enters the function nonlinearly. While potentially less economically interpretable, this measure is more strongly predictive of profit growth. To construct this measure, we evaluate over 20 machine learning models and select the one with the best performance on testing data. These models use modified IDE values with varying degrees of nonlinearity as inputs, and predict the next year's profit growth as targets.[15] The resulting predictions form our new IDE measure, which scales nonlinearly with respect to the similarities.

[15] These modified IDEs mostly follow the definition in equation 1 except the cosine similarity term is replaced with cos, $\cos^2$, $\cos^3$, $\cos^{\frac{1}{3}}$, and $\cos^{\frac{1}{5}}$.

We randomly split 80% of the data for training and validation, and the other 20% is used for testing. As shown by table C29, the neural network implementation using FastAI achieves the highest testing performance in terms of mean-squared-error. Therefore, we use the predictions of this model as our nonlinear IDE measure.[16]

We evaluate the informativeness of this measure using the 20% testing data. More specifically, we run the regression in our main study specified by equation 2 where the negative of the nonlinear IDEs (model predictions) are used to replace the original IDEs in the regression. We use the negative of the predictions because these model predictions are meant to be positively correlated with profit growth; however, our IDE measure is intended to have a negative association with profit growth. As a sanity check, we observe that our main IDE measure is significantly negatively associated with the model predictions. Comparing the results with the new nonlinear measure shown in table C30 and the original main result in table A1, we observe that the nonlinear measure is significantly negatively associated with profit growth similar to our main measure: each standard deviation increase in the nonlinear IDE measure is associated with a 1.1% decrease in profit growth in year 1 and 5.1% by year 7.

[16] Similar to section 8, we compute the importance of each input in this best model. We see that all versions of IDE (cosine, squared cosine, cubic cosine, cubic root of cosine, fifth root of cosine) are significantly important, and cubic of cosine is the most important one.

Model	Score Test	Score Val
NeuralNetFastAI_BAG_L2	-0.477185	-0.488348
CatBoost_BAG_L2	-0.477199	-0.488263
NeuralNetTorch_BAG_L1	-0.477202	-0.488219
WeightedEnsemble_L2	-0.477208	-0.488205
WeightedEnsemble_L3	-0.477225	-0.487909
NeuralNetFastAI_BAG_L1	-0.477226	-0.488225
NeuralNetTorch_BAG_L2	-0.477230	-0.488194
LightGBMXT_BAG_L2	-0.477232	-0.488264
CatBoost_BAG_L1	-0.477256	-0.488300
LightGBMXT_BAG_L1	-0.477265	-0.488280
LightGBM_BAG_L2	-0.477283	-0.488369
LightGBM_BAG_L1	-0.477285	-0.488415
XGBoost_BAG_L1	-0.477291	-0.488462
XGBoost_BAG_L2	-0.477354	-0.488425
LightGBMLarge_BAG_L1	-0.477371	-0.488535
LightGBMLarge_BAG_L2	-0.477409	-0.488465
ExtraTreesMSE_BAG_L2	-0.480714	-0.497547
RandomForestMSE_BAG_L2	-0.480758	-0.495065
ExtraTreesMSE_BAG_L1	-0.495232	-0.505290
RandomForestMSE_BAG_L1	-0.498903	-0.509182
KNeighborsUnif_BAG_L1	-0.525793	-0.534826
KNeighborsDist_BAG_L1	-0.532000	-0.539883

TABLE C29: This table shows the performance of all machine learning models when we use non-linear similarities to construct an alternative IDE measure that predicts next year's profit growth.

D Neural Networks and GPT

D.1 Introduction to Neural Networks

We begin by exploring a basic neural network: a linear neural network with a single hidden layer. This network is characterized by three dimensions: input dimension $\dim_{in}$, hidden dimension $\dim_h$, and output dimension $\dim_{out}$.

This one-hidden-layer network involves two mappings. The first mapping translates the input space into the hidden space. Formally, if the input data is denoted as X, then the hidden space H is given by

$$H = f_1(X) = XW_1 + B_1,$$

where W_1 and B_1 are trainable matrices of parameters.

The second mapping converts the hidden space H into the output space Y:

$$Y = f_2(H) = f_2(f_1(X)) = (XW_1 + B_1)W_2 + B_2,$$

$$= XW_1W_2 + B_1W_2 + B_2,$$

where W_2 and B_2 are also trainable matrices of parameters.

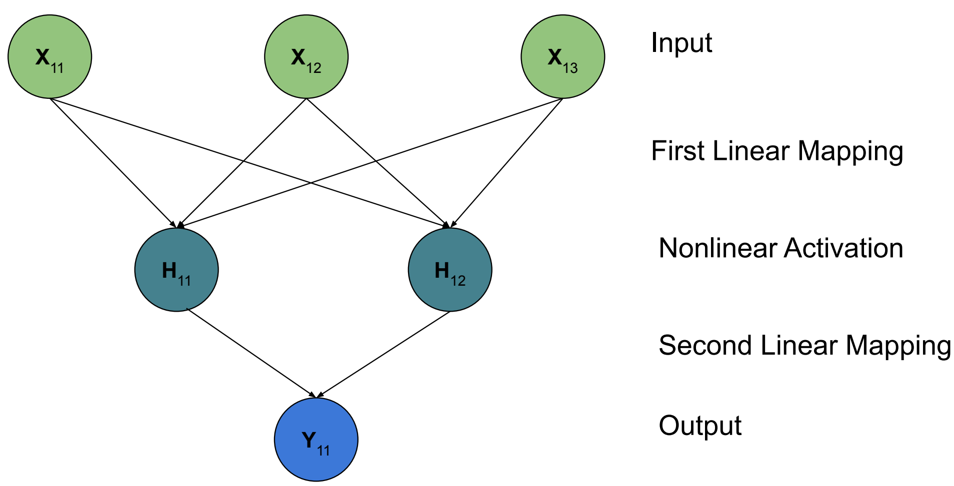

Figure C1: Diagram of a one-hidden-layer neural network with a nonlinear activation function. The top green circles represent the inputs, which are three-dimensional vectors in this figure. The turquoise circles in the middle represent the nodes in the hidden layer, and the blue circle at the bottom represents the output.

As illustrated in Figure C1, this simple linear network can be extended by introducing a nonlinear activation function $g(\cdot)$. When the activation function is applied to the first mapping, the output of the first mapping (the hidden space) becomes

$$H = g(f_1(X)) = g(XW_1 + B_1).$$

In a similar manner, applying a nonlinear activation function after any mapping adds nonlinearity to the linear transformation, enabling neural networks to more effectively model complex relationships between the input X and output Y.

D.2 Introduction to GPT-4

In this section, we delve into the key components of GPT-4. We begin by discussing the decoder architecture, which forms the foundation of models such as those in the GPT family. Following this, we explain the self-attention mechanism that empowers GPT-4 to consider the context when generating new text. We then explore the pre-training process for GPT-4's base model, and conclude by highlighting the advancements that GPT-4 (and 3.5) brings over its predecessors.

D.2.1 Decoder

The original Transformer architecture was conceived for tasks like machine translation, utilizing both an encoder and a decoder. In a typical transformer-based translation model, the encoder generates a numerical representation of the input up to token $t+1$, where the $(t+1)$th token is the next to be translated. The decoder then uses this encoder output along with a numerical representation of the t translated words to predict the translation of the $(t+1)$th word. However, GPT employs a modified version of the Transformer architecture that includes only the decoder component. This design centers exclusively on generating a numerical representation of the already generated text and predicting the next token, which is ideal for tasks such as text completion and generation.

The decoder in GPT models comprises four essential components: positional encoding, self-attention layers, position-wise feedforward networks, and layer normalization with residual connections. We will briefly discuss each of these components before focusing on the self-attention layer, which is the core mechanism driving the model's capabilities.

In the Generative Pre-trained Transformer (GPT) model architecture, positional encoding is vital for providing information about the sequence position of tokens. The Transformer architecture does not inherently account for the order of tokens, so positional encoding ensures that the model can distinguish tokens based on their positions within a sequence.

Positional encoding involves adding fixed-length vectors to the input token embeddings before feeding them into the model. These positional embeddings encode positional information relative to other tokens in the sequence. GPT uses sinusoidal functions to generate these embeddings:

$$\begin{aligned}\mathrm{PE}_{(pos,2i)} &= \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right)\\ \mathrm{PE}_{(pos,2i+1)} &= \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right),\end{aligned}$$

where $\mathrm{PE}_{(pos,2i)}$ are the positional embeddings for tokens at even positions, $\mathrm{PE}_{(pos,2i+1)}$ are the positional embeddings for tokens at odd positions, and d_{model} is the dimensionality of the embeddings (1,536 dimensions for GPT-4).

These positional encoding vectors are combined with the input embeddings of tokens, injecting positional information into the model's representation of the input. By incorporating positional encoding, GPT ensures that the model can effectively capture the sequence structure and relationships between tokens based on their positions.

Additionally, the decoder is composed of multiple layers of self-attention mechanisms. Each layer processes the input sequence independently and captures dependencies within the sequence. The self-attention mechanism allows the model to assign different importance to each token based on its relevance to other tokens in the sequence, enabling the model to understand context and generate coherent and contextually appropriate text.

After the self-attention layers, each position in the sequence passes through a position-wise feedforward neural network. This network consists of fully connected layers with non-linear activation functions, enabling the model to capture complex patterns and interactions in the data. The position-wise feedforward networks refine the token representations, incorporating both local and global context information.

To further stabilize training and improve the flow of gradients, GPT incorporates layer normalization and residual connections after each self-attention layer and position-wise feedforward network. Layer normalization normalizes the activations within each layer, reducing internal covariate shifts and enhancing training stability. Residual connections allow gradients to bypass certain layers, mitigating the vanishing or exploding gradient problem that can arise in deep neural networks.

D.2.2 Self-attention

The purpose of the self-attention mechanism is to generate a numerical embedding for each piece of text while accounting for the contextual relationships among all tokens in the text. Specifically, the raw input to the attention mechanism is a piece of text T. This text is divided into sub-word tokens through a process called tokenization. The tokenization process relies on a predefined set of tokens that can be combined to represent a vast number of unique words. For example, the prefix "un" is a token in many models because it negates the meaning of many other sub-word tokens, such as “happy.” Other tokens may represent short and common words like “and.”

After tokenization, each token is assigned an initial embedding that combines the token’s semantic meaning and its position within the text. This results in a set of initial embeddings:

$$\mathrm{EMB}_0 = [[\mathrm{BOS}], t_1, \ldots, t_N, [\mathrm{EOS}]],$$

where t_i represents the embedding for a token i, “[BOS]” (beginning of sentence) denotes the start of the sentence, and “[EOS]” (end of sentence) signifies the sentence’s conclusion.

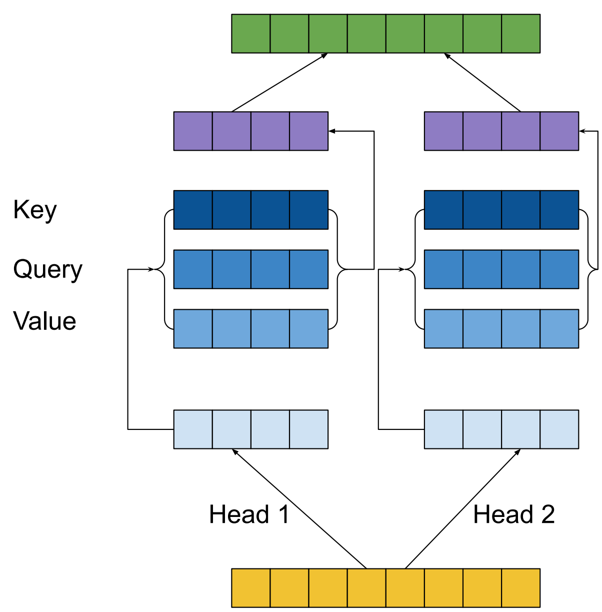

Figure C2: Diagram of a multi-head attention layer

As shown in Figure C2, a multi-head self-attention layer processes an embedding as input and produces a refined embedding as output. The input embedding is passed through three parallel linear mappings to form three matrices: the key matrix, the query matrix, and the value matrix.

$$Q = \text{EMB}_0 W^Q,$$
$$K = \text{EMB}_0 W^K,$$
$$V = \text{EMB}_0 W^V,$$

where Q, K, and V are trainable parameter matrices. For each query, a cosine similarity score is calculated between the query and all keys, including the query itself. The value of the token is then represented as a weighted combination of all the token values in the text, with the weights determined by the cosine similarities between queries and keys. Mathematically, this is represented as:

$$\text{Attention}(\text{EMB}_0) = \text{softmax}(QK^T)V.$$

In the GPT model, the attention mechanism is typically computed as:

$$\text{Attention}(\text{EMB}_0) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

where d_k is a scaling factor equal to the number of columns in K. To enhance the model's capacity to capture various aspects of the input, the embeddings EMB_0 are often divided into multiple sub-vectors of equal size. Attention is computed for each sub-vector independently, and the results are concatenated to produce a multi-head attention output for the input embedding. This attention mechanism is typically repeated multiple times, with the output of the $(i-1)$th attention layer being normalized and combined with its input to serve as the input for the ith attention layer. In GPT-4, each embedding has 1,536 dimensions.

D.2.3 Pre-training

GPT is trained using the autoregressive language modeling task, which involves predicting the next token in a sequence based on its preceding context. This task is mathematically formulated as maximizing the log-likelihood of observing the next token x_{i+1} given the preceding tokens $x_1, x_2, \ldots, x_i$ and the model parameters θ. The objective can be expressed as:

$$\mathcal{L}_{\text{pretrain}}(\theta) = \sum_{i=1}^{n-1} \log P(x_{i+1}|x_1, x_2, \ldots, x_i; \theta),$$

where $\mathcal{L}_{\text{pretrain}}(\theta)$ represents the training objective, and θ denotes the model's parameters.

The goal of this pre-training task is to enable the model to learn the intricate patterns and dependencies present in natural language. By predicting the next token based on its context, GPT captures both syntactic and semantic structures, allowing it to generate text that adheres to grammatical rules and maintains coherence. The autoregressive nature of the training process further encourages the model to understand long-range dependencies, enabling it to consider information across a wide span of tokens.

Through backpropagation and gradient descent, the model gradually adjusts its parameters to minimize the negative log-likelihood of the next token, improving its ability to capture nuanced linguistic patterns and generate coherent, contextually appropriate text. The following list includes some of the data sources used to pretrain the base model of GPT-4:

1. **Common Crawl**: A vast dataset consisting of web pages collected from across the Internet, offering a broad variety of text data.
2. **Wikipedia**: Articles from Wikipedia in various languages and domains, providing structured and comprehensive information on a wide range of topics.
3. **BooksCorpus**: A collection of books spanning different genres and authors, allowing the model to learn from literary works and fictional narratives.

D.2.4 Reinforcement Learning with Human Feedback (RLHF)

GPT-4 significantly enhances its text generation capabilities through a process known as reinforcement learning with human feedback (RLHF). This method involves training the model not only through traditional supervised learning but also by incorporating direct feedback from human evaluators to fine-tune its behavior.

In the RLHF framework, GPT-4 generates text samples based on its current model, and these samples are then evaluated by human judges or annotators. The human feedback provided acts as a reward signal, guiding the model towards generating higher-quality text over time.

Formally, consider a set S representing all possible text samples that GPT-4 can generate. The model produces text samples according to its current policy π_θ, which is parameterized by θ. Each generated sample $s \in S$ is then evaluated by human judges, resulting in a feedback signal $r(s)$. This feedback signal $r(s)$ reflects the quality and desirability of the generated text, as determined by human evaluators.

The objective of GPT-4 in this framework is to learn an optimal policy π_θ that maximizes the expected cumulative reward across the distribution of text samples. Mathematically, this can be represented as the following optimization problem:

$$\max_\theta \mathbb{E}_{s\sim\pi_\theta}[r(s)],$$

where $\mathbb{E}_{s\sim\pi_\theta}[r(s)]$ denotes the expected reward across the distribution of text samples generated by the model.

To achieve this, GPT-4 employs policy gradient methods, a class of algorithms in reinforcement learning that adjust the model's parameters θ based on the gradient of the expected reward. The model iteratively updates its policy to increase the likelihood of generating text that receives positive feedback from human judges.

The RLHF process involves multiple iterations of text generation, evaluation, and feedback, which gradually refine the model's behavior. By directly incorporating human judgments into the training loop, GPT-4 learns to align its text generation more closely with human preferences, producing text that is not only syntactically correct but also contextually relevant, coherent, and aligned with desired outcomes.

Moreover, RLHF enables GPT-4 to address complex generation tasks where the quality of output cannot be easily quantified by automated metrics alone. By leveraging human expertise, the model can better understand and adapt to nuanced requirements, leading to more effective and reliable text generation.

E Debiased machine learning

In this section, we discuss the debiased machine learning in more detail. First, we explain the overview and then we discuss more details about the cross-fitting technique and Neymann orthogonality condition.

E.1 Overview

The debiased machine learning method proposed in Chernozhukov et al. (2018) is designed to address the challenges posed by high-dimensional nuisance parameters in the estimation of low-dimensional parameters of interest. In our study, we control for continuous confounders such as current profit and current employment and we also control for over 1000 discrete variables covering inter-industrial and inter-temporal variations in the data. Traditional approaches often suffer from biases introduced by regularization techniques necessary to handle high-dimensional data, leading to inconsistent estimates. To mitigate this issue, the method leverages cross-fitting. Cross-fitting is a technique that further reduces bias by splitting the data into multiple folds. The nuisance parameters are estimated in one part of the data and then used to estimate the parameter of interest in another part. This process is repeated across different folds, and the results are averaged to obtain the final estimate. In our study, we do cross-fitting in two folds as discussed in section 6.6.

In our application, we aim to estimate the association between IDE and profit growth values over each of the next 7 years while controlling for a large set of confounders X (including profit, employment, year, industry, etc). In this debiased machine learning approach, we model IDE and profit growth as non-parametric functions of the confounders.

$$
\begin{aligned}
IDE_{f,t} &= g_T(x_{f,t}) + \epsilon^T_{f,t} && (12)\\
\Pi_{f,t+1} - \Pi_{f,t} &= g^1_O(x_{f,t}) + \epsilon^{1,O}_{f,t} && (13)\\
&\vdots \\
\Pi_{f,t+7} - \Pi_{f,t} &= g^7_O(x_{f,t}) + \epsilon^{7,O}_{f,t} && (14)
\end{aligned}
$$

There are 7 parameters of interest denoted by τ^{IDE}_k:

$$
\begin{aligned}
\epsilon^{1,O}_{f,t} &= \tau^{\text{IDE}}_1 \epsilon^T_{f,t} + \epsilon_{1,f,t} && (15)\\
&\vdots \\
\epsilon^{7,O}_{f,t} &= \tau^{\text{IDE}}_7 \epsilon^T_{f,t} + \epsilon_{7,f,t} && (16)
\end{aligned}
$$

where $\epsilon^{k,O}_{f,t}$ and $\epsilon^T_{f,t}$ are the residuals estimated in system 12, and $\epsilon_{k,f,t}$ are the unexplained variations in 15. The intuition of this debiased machine learning method can be explained in two steps. First, we remove any information in the treatment and outcome variables that are associated with the set of high-dimensional confounders. Then, the residuals capture variations in treatment and outcome that do not depend on any of the confounders. Therefore, in the second stage, when we run a regression of the residuals of the outcome variables on the residuals of the treatment, the resulting coefficient only captures changes in the outcome that are associated with changes in the treatment variable.

E.2 Implementing cross-fitting

Cross-fitting is a technique employed in debiased machine learning to reduce bias in the estimation of parameters. The core idea is that when we use machine learning models (g^T and g_O^k in system 12) to fit the entire sample, these highly complex models may lead to overfitting of the data instead of estimating the true underlying relationships.

1. **Data Splitting:** The dataset is randomly divided into two disjoint subsets: *Training Set 1* and *Training Set 2*. Each subset is used for different purposes in the estimation process to avoid overfitting and to mitigate the influence of bias introduced by the machine learning methods used to estimate the nuisance parameters.

2. **Nuisance Parameter Estimation:**

   - *Training Set 1:* The nuisance parameters, denoted by η, are estimated using the observations in Training Set 1. These nuisance parameter estimates, $\hat{\eta}_1$, are then applied to the observations in Training Set 2.
   - *Training Set 2:* Similarly, the nuisance parameters are estimated using Training Set 2, resulting in $\hat{\eta}_2$, which are then applied to the observations in Training Set 1.
   - **Treatment parameter Estimation:** Pool together the predictions on the two sets. Compute the residuals like the system of equations specified in 12, and run another set of regressions with the residuals like system 15.

E.3 Neymann Orthogonality

The Neymann orthogonality condition ensures that the moment conditions used for estimating the parameter of interest are locally insensitive to the estimation errors of the nuisance parameters. Let θ_0 be the low-dimensional parameter of interest and η_0 represent the high-dimensional nuisance parameters. The orthogonality condition requires that the score function $\psi(W;\theta,\eta)$ satisfies:

$$\mathbb{E}[\psi(W;\theta_0,\eta_0)] = 0,$$
$$\partial_\eta \mathbb{E}[\psi(W;\theta_0,\eta_0)] = 0,$$

where W denotes the observed data, and ∂_η is the derivative with respect to η. This condition implies that small deviations in the estimation of the nuisance parameter η do not affect the score function's expectation, making the estimator robust to errors in η.

In the context of debiased machine learning (DML), the orthogonality condition is used to construct a score function that reduces sensitivity to the regularization bias from machine learning methods. Suppose the moment equation is defined as:

$$\psi(W;\theta,\eta) = (Y - D\theta - g(X))(D - m(X)),$$

where $g(X) = \mathbb{E}[Y|X]$ and $m(X) = \mathbb{E}[D|X]$ are functions estimated using machine learning techniques. The DML estimator leverages the orthogonality condition by approximately removing the effect of regularization bias from the nuisance parameters, ensuring that:

$$\mathbb{E}[(D - m(X))(g(X) - \hat{g}(X))] \approx 0,$$

where $\hat{g}(X)$ is an estimator of $g(X)$.

As shown in Chernozhukov et al. (2018), the bias term in this score function can be written as

$$\mathbb{E}\left(V^2\right)^{-1}\frac{1}{n}\sum_{i\in I}(\hat{m}_0(X_i) - m_0(X_i))(\hat{g}_0(X_i) - g_0(X_i))$$

where $V = D - \hat{m}_0(X)$, n is the number of testing samples, and I is the set of testing samples.

Therefore, we need the machine learning estimators to have unbiased errors which can be satisfied by many machine learning models.[17]

F Intoduction of Gradient Boosting

F.1 Gradient Boosted trees

Gradient Boosted Trees (GBT) is an ensemble learning technique that combines the predictions of multiple weak learners, decision trees, to form a strong predictive model. The idea behind GBT is to iteratively add trees to the ensemble, each tree correcting the errors made by the previous ones. The model starts with an initial tree, and subsequent trees are trained on the residual errors (the difference between the actual target values and the

[17] Near orthogonality is satisfied in our application: the estimation errors in m and g are weakly correlated (always lower than 4%), and the estimation errors are all insignificant from 0 except for the profit growth over the next year. Nonetheless, the average error for next year's profit growth is small in magnitude at 0.4%.

current model predictions). This process is known as gradient descent in function space because each tree is added in a direction that minimizes the loss function, typically a mean squared error for regression tasks or a log-loss for classification tasks.

Mathematically, we can express the model's prediction after m trees as:

$$\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + \eta f_m(\mathbf{x}_i) \tag{17}$$

where $\hat{y}_i^{(m)}$ is the prediction for instance i after m trees, $\hat{y}_i^{(m-1)}$ is the prediction after $m-1$ trees, η is the learning rate, and $f_m(\mathbf{x}_i)$ is the prediction of the m-th tree. The objective is to minimize the loss function $L(y_i, \hat{y}_i^{(m)})$, which can be approximated by using a first-order Taylor expansion around the current prediction:

$$L(y_i, \hat{y}_i^{(m)}) \approx L(y_i, \hat{y}_i^{(m-1)}) + \frac{\partial L(y_i, \hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)}} \cdot f_m(\mathbf{x}_i) \tag{18}$$

Here, $\frac{\partial L(y_i, \hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)}}$ represents the gradient of the loss function with respect to the previous prediction, guiding the new tree to correct the errors of the ensemble model.

F.2 XGBoost: Advanced Gradient Boosted Trees

F.2.1 Intuition of XGBoost

XGBoost (eXtreme Gradient Boosting) is a highly efficient and scalable implementation of Gradient Boosted Trees, widely used in machine learning competitions and industry applications due to its performance and speed. XGBoost builds upon the fundamental concept of Gradient Boosted Trees (GBT), where multiple decision trees are sequentially added to correct the residual errors of previous models. However, XGBoost introduces various optimizations and enhancements, making it faster and more effective than standard GBT.

Similar to a standard Gradient Boosted Tree, the intuition behind XGBoost is that each new tree added to the ensemble helps to correct the mistakes of the existing trees by focusing on the most challenging instances. XGBoost incorporates both first-order and second-order (Hessian) information, which allows it to perform more precise updates in minimizing the loss function. This approach enables XGBoost to converge faster and achieve better performance with fewer trees.

F.2.2 Mathematical Formulation

As discussed in equation 17, the model's prediction after adding the m-th tree can be mathematically expressed as:

$$\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + \eta f_m(\mathbf{x}_i)$$

XGBoost optimizes an objective function that includes a loss term $L(y_i, \hat{y}_i^{(m)})$ and a regularization term $\Omega(f_m)$, which controls the complexity of the model to prevent overfitting:

$$\text{Objective} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(m)}) + \sum_{m=1}^{M} \Omega(f_m) \tag{19}$$

The loss function can be approximated using a second-order Taylor expansion around the current predictions:

$$L(y_i, \hat{y}_i^{(m)}) \approx L(y_i, \hat{y}_i^{(m-1)}) + g_i f_m(\mathbf{x}_i) + \frac{1}{2} h_i f_m(\mathbf{x}_i)^2 \tag{20}$$

where:

- $g_i = \frac{\partial L(y_i, \hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)}}$ is the gradient of the loss function.
- $h_i = \frac{\partial^2 L(y_i, \hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)2}}$ is the Hessian (second derivative), which provides information on the curvature of the loss function.[18]

F.2.3 Controlling Overfitting with Early Stopping

We use a technique known as early stopping to prevent overfitting when training the XGBoost models, especially when training models with a large number of trees. Overfitting occurs when the model learns noise and patterns specific to the training data rather than general trends, resulting in poor performance on unseen data.

Early stopping monitors the model's performance on a validation set during training and stops the training process if the performance does not improve for a specified

[18] For a more detailed discussion for the implementation of these models see Erickson et al. (2020).

number of consecutive iterations (trees). This is done by specifying a patience parameter, commonly referred to as ‘early_stopping_rounds‘. If the evaluation metric does not improve after the given number of rounds, training halts, and the model reverts to the best iteration observed. In our analysis, we set this hyperparameter as 5 rounds.

Mathematically, let $M(t)$ be the evaluation metric on the validation set at iteration t. Early stopping stops training at iteration t^* if:

$$M(t) \geq M(t^*) \quad \text{for all} \quad t^* \leq t < t^* + \text{early\_stopping\_rounds} \tag{21}$$

where t^* is the best iteration so far. This process helps to prevent adding unnecessary complexity to the model and ensures that the training process stops when the validation performance starts to degrade, effectively controlling overfitting.

By using early stopping, XGBoost finds the optimal number of boosting rounds automatically, enhancing generalization and improving model performance on new data.

G Variable importance figures of subsample cuts

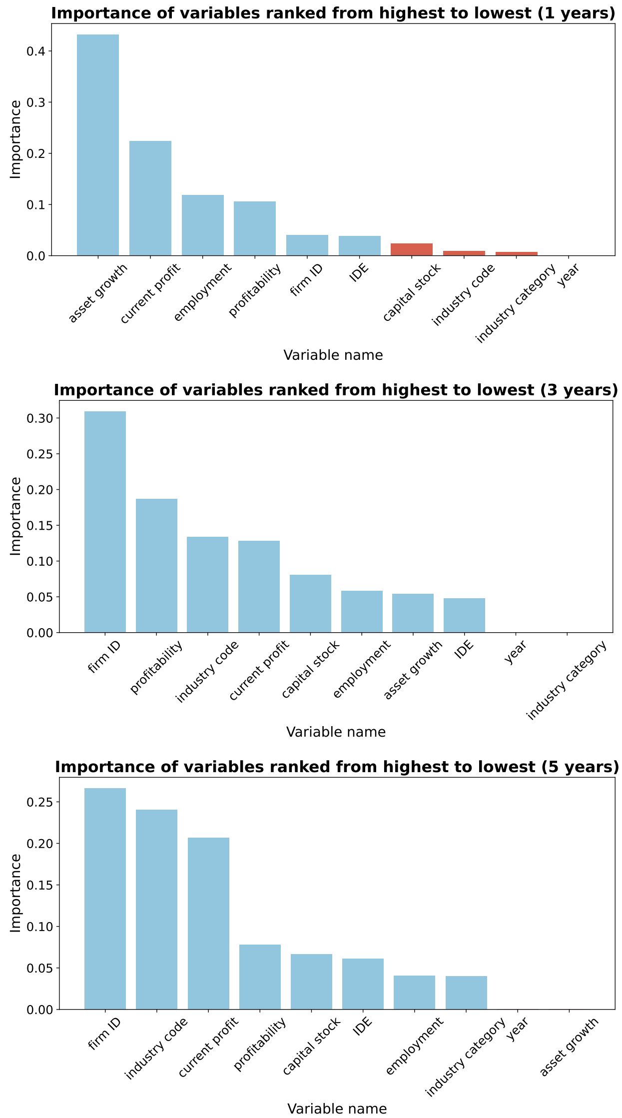

Figure F1: These figures show the importance of different variables in the best predictive model of profit direction changes over the next 1,3, and 5 years among firms with above median market capitalization in each year and industry category. The importance values are scaled to sum to 1, and the ones colored blue are significant at the 5% level while the ones in red are not. When the variables are not significant, it is possible that some of them may have negative importance values by chance (statistical zeros); in these cases, as long as the negative values are not significant, we treat the importance as 0.

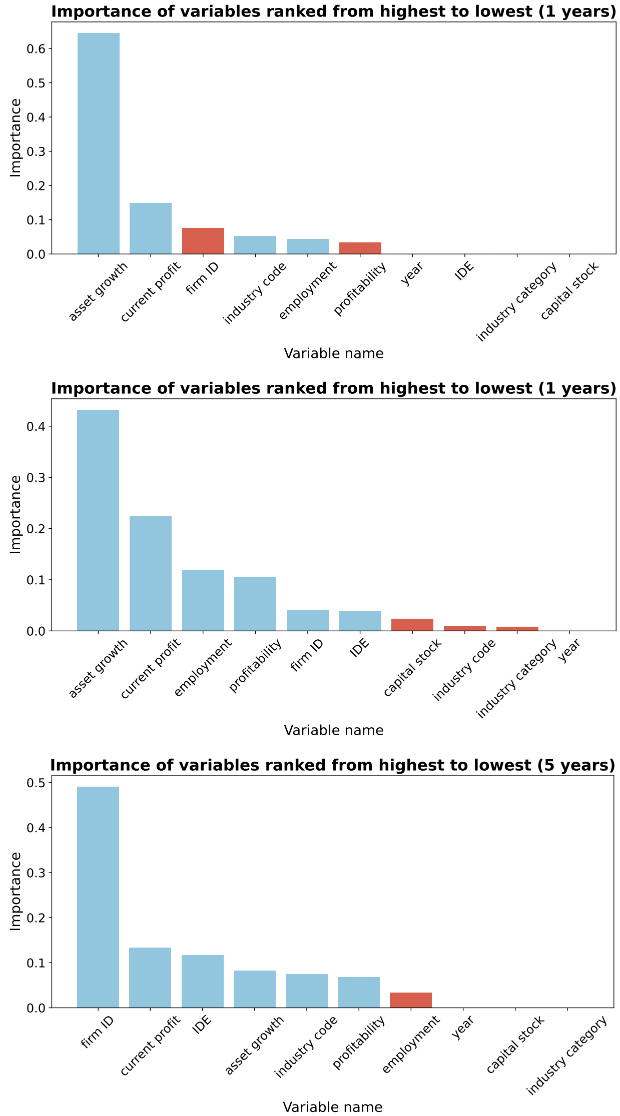

Figure F2: These figures show the importance of different variables in the best predictive model of profit direction changes over the next 1,3, and 5 years among firms with below median profitability in each year and industry category. The importance values are scaled to sum to 1, and the ones colored blue are significant at the 5% level while the ones in red are not.

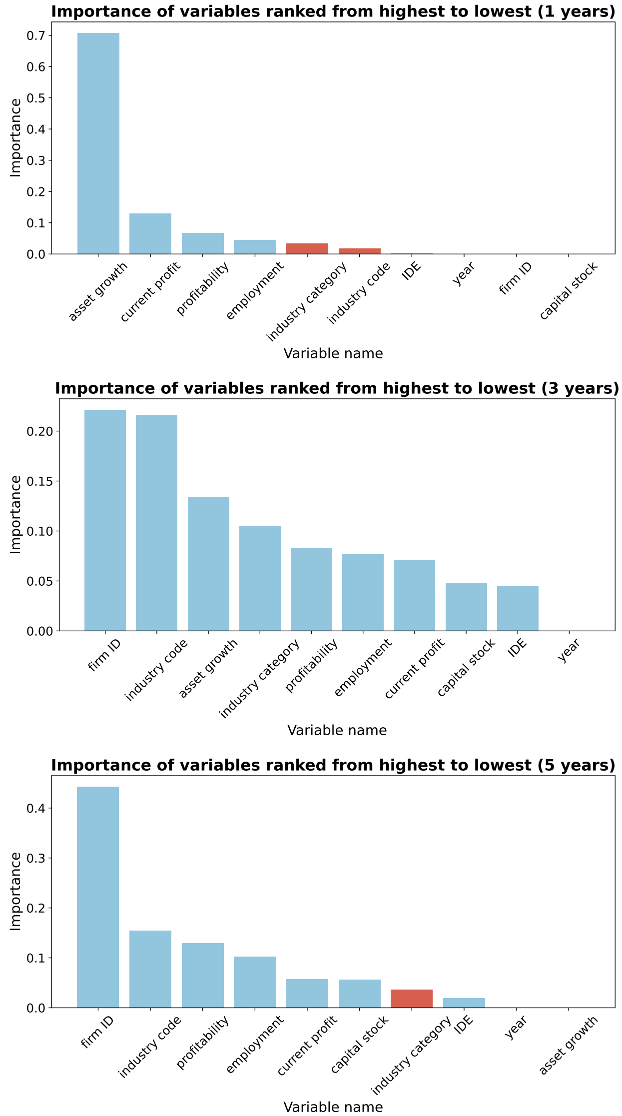

Figure F3: These figures show the importance of different variables in the best predictive model of profit direction changes over the next 1,3, and 5 years among firms with below median intangible asset ratio (defined in the description of equation 2) in each year and industry category. The importance values are scaled to sum to 1, and the ones colored blue are significant at the 5% level while the ones in red are not.

	Dependent variable:						
	$\Pi_{t+1} - \Pi_t$	$\Pi_{t+2} - \Pi_t$	$\Pi_{t+3} - \Pi_t$	$\Pi_{t+4} - \Pi_t$	$\Pi_{t+5} - \Pi_t$	$\Pi_{t+6} - \Pi_t$	$\Pi_{t+7} - \Pi_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IDE	-0.011***	-0.018***	-0.030***	-0.035***	-0.040***	-0.046***	-0.051***
	(0.004)	(0.004)	(0.006)	(0.008)	(0.009)	(0.010)	(0.009)
profitability t	-0.007	-0.021**	-0.034***	-0.042***	-0.058***	-0.058***	-0.066***
	(0.006)	(0.008)	(0.009)	(0.010)	(0.011)	(0.012)	(0.013)
log asset growth	0.076***	0.092***	0.096***	0.100***	0.099***	0.099***	0.102***
	(0.006)	(0.008)	(0.010)	(0.011)	(0.013)	(0.014)	(0.012)
Year	✓	✓	✓	✓	✓	✓	✓
Ind Code	✓	✓	✓	✓	✓	✓	✓
Year×*IndCategory*	✓	✓	✓	✓	✓	✓	✓
Observations	23,702	21,695	19,984	18,354	16,976	15,750	14,658
R^2	0.128	0.153	0.160	0.169	0.180	0.185	0.192
Adjusted R^2	0.082	0.104	0.107	0.112	0.118	0.118	0.121
Note:							*p<0.1; **p<0.05; ***p<0.01

TABLE C30: This table shows the association between nonlinear IDE and profit growth. We control for the log asset growth and profitability among other variables. The data is the 20% testing data from 1980 to 2015. All independent variables are standardized within each industry category and year, and all standard errors are clustered at the firm-year level.