Complex Technologies and Economic Value

October 14, 2026

Abstract

This paper proposes a measure for industry-level technological complexity using a novel Network Diversity Score, which captures the connections between fundamental knowledge components. This approach outperforms other complexity metrics in key tests. The findings show that technological complexity rose steadily at both technology sector and NAICS industry levels until the Covid-19 pandemic, after which labor-intensive industries experienced a sharp decline. It is also found that the drivers of technological complexity vary with industry concentration. In more competitive industries, greater investment intensity is strongly associated with subsequent technological complexity, whereas this relationship is substantially weaker in more concentrated markets. The economic impact analysis further shows that technological complexity primarily benefits the labor market rather than firms. Greater technological complexity is associated with hiring more lower-paid workers but not with firms' higher subsequent profitability and market valuation.

1 Introduction

Technology- and innovation-driven growth is widely viewed as a fundamental source of long-term economic development because it improves productivity, expands production possibilities, and improves the efficiency with which firms and economies transform input into output. With advances in technology, novel products are created, enabling more efficient production processes and stronger organizational capabilities, thereby supporting sustained gains in competitiveness and living standards. This perspective has become especially salient in recent discussions of artificial intelligence, the productivity effects of computational power may strengthen as adoption deepens and complementary investments accumulate.[1]

At the same time, the economic effects of innovation are unlikely to be uniform across technologies. Some innovations are relatively standardized and easy to diffuse, whereas others require integrating diverse knowledge components. This heterogeneity motivates growing interest in technological complexity, which captures the extent to which a technology reflects sophisticated, interdependent, and difficult-to-replicate combinations of knowledge. In the past literature, a universally applied method of quantifying technological complexity and its impact are still lacking, while some papers studied the links between technological complexity and economic factors, eg. Nepelski and De Prato (2020) links countries' positions in the global technology space to economic development, arguing that technological complexity is an important determinant of development outcomes. This paper develops a novel approach using graph theory on U.S. patent data to quantify technological complexity and compute annual complexity scores for each industry. I then compare my approach with existing measures to highlight its advantages and investigate the key drivers of technological complexity at the industry level.

Innovation continually adds new features, components, and processes to existing technologies. Over successive waves of innovation, these additional developments, building on existing ones, transform once-simple tools and mechanisms into richly elaborate processes. Ultimately, suc-

[1] See Erik Brynjolfsson, "Productivity is 'Much Higher' Due to AI," *Wall Street Journal*. Brynjolfsson argues that AI's productivity effects are not limited to the direct use of AI tools; rather, they become substantial when firms make complementary investments in workflow redesign, organizational change, and human capital, allowing AI to function as a general-purpose technology.

cessive innovations drive technological complexity upward. This knowledge accumulation has long been recognized, as Newton famously wrote, "If I have seen further, it is by standing on the shoulders of Giants."[2] To the best of our knowledge, the only paper that measures technological complexity from an innovation standpoint using patent data is Broekel (2019). His paper links technological complexity to regional growth in Mewes and Broekel (2022). I adopt the methodology outlined in his paper and extend it to a broader application by linking it to the NAICS industry level and analyzing the determinants of technological complexity and the subsequent economic impacts.

Patents are the primary output of the inventive process and a well-established proxy for innovation, and prior research has long linked them to corporate valuation through measures such as patent citations, textual descriptions, and related characteristics. Mariani, Medo and Lafond (2019) uses the patent citation network to identify important technologies; Kelly, Papanikolaou, Seru and Taddy (2021) uses a text-based method to create new indicators on important patents; Bowen III, Frésard and Hoberg (2023) also use patent text to position startups in business cycles); and Kogan, Papanikolaou, Seru and Stoffman (2017) combines patent data with stock return data to calculate the financial valuation of each patent. In addition, I argue that the financial valuation may extend beyond these mechanisms, as patents are shaped by successive knowledge that increase technological complexity, which may contribute to industry-level technological complexity.

In Arthur (2009), technology is defined as a system of interconnected components used for practical purposes. To operationalize this concept empirically, I rely on the Cooperative Patent Classification (CPC) system, which organizes patented inventions into detailed technological categories. As discussed in Broekel (2019), knowledge components (10-digit CPC subclass) form the fundamental building blocks for technologies, which are usually embedded in patents as "claims" and classified under a 4-digit CPC classification system. Each patent therefore embeds a network of knowledge components. By aggregating the networks of all patents within the same 4-digit CPC class, I construct a larger technological network whose linkages among knowledge components

[2]The metaphor "standing on the shoulders of giants" expresses that new discoveries extend prior ones. Contemporary discussions use this to emphasize that innovation almost never comes from nowhere; each advance opens new "adjacent possible" avenues for further advances. See Farnam Street, "Standing on the Shoulders of Giants," and Wikipedia: Standing on the shoulders of giants.

are used to compute the Network Diversity Score (NDS), a proxy for industry-level technological complexity. This measure is based on four structural network indicators, each capturing a distinct property of the technological network: (1) module density - the number of densely connected groups relative to the overall size; (2) variation in module size - how different in the size of the densely connected groups are from each other; (3) overall network connectivity - how well the nodes are connected with each other; and (4) the growth of small subgroups within the network - which compares the number of patterns with three nodes to the number of patterns with four nodes.

This approach of measuring technological complexity outperforms other existing complexity measures in two important ways. First, when comparing with other existing complexity measures that do not use network theory, this method performs better in terms of its linkage to R&D intensity (Broekel (2019)). Since, long-term corporate valuation is critically dependent on R&D investments and R&D intensity (see, Coad and Rao (2008)), we expect technological complexity to be an important determinant of corporate valuation. Second, the proposed measure, rather than relying on a single network feature, combines four structural indicators, enabling classification of networks by appropriately capturing the degree of heterogeneity in interconnections between nodes. This is demonstrated by comparing it with three standard network measures: the Bertz index, Wiener index, and mean distance deviation (MDD). In environmental technologies, Bertz, Wiener, and NDS all trend upward steadily, whereas MDD is noisy with little informative pattern. In AI, NDS clearly reflects the post-2021 surge in complexity consistent with the recent boom, while among the other metrics, only the Wiener index shows a similar but weaker uptick. Overall, the evidence indicates a sustained performance for NDS relative to the other methods.

One of the contributions of this paper is developing a two-step conversion process and associated concordance file to merge our CPC-based complexity scores with NAICS-classified economic and financial data. First, CPC codes are mapped to the initial three NAICS versions using the ALP probabilistic links from Nathan, Lybbert and Jason (2019). Since this study spans seven NAICS revisions, I then chain together adjacent NAICS versions using probability weights from the Census Bureau's concordance files. Because complexity scores must stay on the same scale during conversions and the patent/citation count should be preserved at an aggregated level, separate weighting schemes and formulas are applied for each measure. This conversion framework is a key contribu-

tion of this analysis, as it can be applied to translate any CPC-based variable to NAICS industries across multiple NAICS versions.

After mapping patent data and their corresponding CPC codes to NAICS industries, I first examine the distributions of complexity scores, which show that overall complexity increases. With more technologies scoring high and fewer scoring zero, this fits the cumulative nature of innovation. Breaking the trend down by sector shows a steady increase across all sectors until 2020, followed by a post-2020 pullback most pronounced in fixed construction. Besides, I find that sectors with fewer usable observations have limited opportunities or needs for innovation and thus generate fewer patents. By focusing on 8 selected industries, we find that the post-pandemic decline is concentrated in labor-intensive fields, while high-tech industries show no significant fall. Also, the industry-level graph shows that complexity scores are not comparable across industries, making industry-level comparisons and analysis challenging. To address this, I standardize the score by de-meaning it within each industry, generating an "excess" complexity score that is comparable across all industries.

The primary contribution of this paper lies in establishing the linkages between technological complexity, product market share, and subsequent economic impacts. Because technological knowledge components, as captured by CPC codes, form the foundation of our complexity measure and are likely to be shared within sectoral and industry categories, we conduct our analysis at the NAICS industry level.

I then examine how industry-level technological complexity relates to firms' operating characteristics and patent activity. The main hypothesis is that firms' incentives to pursue technological complexity depend on their capital investment intensity and the competitive environment in which they operate. The benefits of richer capital expenditure and the harms of defensive patenting to the complexity score are revealed. In relatively competitive industries, greater capital investment intensity is more strongly associated with subsequent technological complexity, indicating that greater market concentration may reduce the extent to which additional investment is translated into more complex technological structures. This highlights competition as an important source of heterogeneity in the determinants of technological complexity.

The economic consequences of technological complexity extend beyond the structure of innovation itself, and this paper makes a main contribution by establishing who benefits from increasingly complex technologies. I first examine labor-market outcomes and find that greater technological complexity is associated with higher employment and total payroll but lower average weekly wages. This pattern is consistent with an expansion in labor demand that is disproportionately concentrated among relatively lower-paid workers, or wage compression. The labor-market effects are particularly pronounced in more concentrated industries, where complexity is associated with stronger employment and payroll growth as well as larger declines in average wages.

I then examine whether firms themselves benefit from this expansion in technologically complex activity. The profitability results indicate that greater technological complexity is generally associated with lower subsequent ROA and profit margins, consistent with the possibility that complex technologies require substantial implementation and commercialization costs. The combination of workforce expansion and weaker profitability suggests that firms may need to devote broader organizational resources to develop and operate complex technological systems without receiving improvements in operating performance. An important exception emerges in highly concentrated industries, where firms with moderate levels of excess technological complexity experience improved profitability over part of the complexity distribution before profitability declines again at higher levels.

Finally, I examine whether these operating consequences are reflected in market valuation. Technological complexity is positively and robustly associated with the economic value of patented innovation, indicating that complex technologies are associated with more valuable innovative output. However, this higher innovation value does not translate into superior firm-level abnormal stock returns. Moreover, industry concentration does not systematically alter the relationship between complexity and abnormal returns.

Taken together, the results suggest that the gains from technological complexity are devoted more to employment and to the value of innovative output than to firm shareholders or operating performance. Firms generally do not appear to capture these gains through higher profitability or abnormal market returns, except where market power and moderate levels of complexity may al-

low some firms to better absorb the costs of complex technological production.

The rest of this paper is organized as follows. Section 2 explains how to construct the technological complexity score, compares it with other measures, and describes the conversions to and between NAICS versions. Section 3 presents the trends and key properties of the complexity scores over time. Section 4 details our data and regression models and explores the drivers of complexity across five sample splits. Section 5 shows how does technological complexity relate to economic outcomes. Finally, Section 6 offers our conclusions.

2 Construction of Complexity Score

2.1 Idea

In network theory, networks are broadly classified into three types: ordered networks, which exhibit predictable connections between nodes and present low heterogeneity; random networks, in which the connections are determined purely by chances, so the heterogeneity degree is the largest; and complex networks, which lie between those extremes, is a mixture of ordered & random networks. Building on this classification, the industrial complexity score in this paper is calculated using the Network Diversity Score (NDS) — initially proposed by Emmert-Streib and Dehmer (2012) in a biological context. Unlike other network-based metrics, NDS simultaneously uses four structural indicators, enabling it to distinguish between ordered, complex, and random networks.

Emmert-Streib and Dehmer (2012) originally proposed NDS as a purely theoretical metric to quantify the complexity of a network. Broekel (2019) later adapted this measure to the patent world. Technologies have been viewed as combinations of interconnected elements (Arthur (2009), McNerney, Farmer, Redner and Trancik (2011)). These elements — ranging from physical parts and functional modules to individual steps in an industrial process — are considered linked whenever a change in one alters the behavior or performance of another. In Broekel (2019), a technology is described as a recipe, in which the information includes the knowledge components and the combination of how these components are combined. For example, the same basic inputs (flour,

eggs, milk, oil) can produce something as simple as scrambled eggs or something as elaborate as a multi-layer cake. Metaphoring knowledge components as ingredients and technologies as recipes illustrate that complexity depends less on the ingredients themselves and more on how knowledge components are linked with each other, specifically on the heterogeneity of their connections. By applying NDS to technological networks, we can rigorously quantify how diversely knowledge components interact.

2.2 Construction

By modeling each 4-digit CPC code as a network, the NDS represents the diversity score of the networks, and each technology (4-digit CPC code) is regarded as a combination of knowledge components (Broekel (2019)). Therefore, the complexity of each depends on how these knowledge components are linked to each other. The more diverse and complex the direct or indirect connection, the more complex this technology is. The author considered the lowest level of CPC classes (10-digit CPC subclasses) as knowledge components and their co-occurrence on patent as linkages (Fleming and Sorenson (2001), Sorenson and Fleming (2004)).

The data files used to construct the networks are from USPTO's PatentsView as of July 2024. For each 4-digit CPC code, a network is built annually using all patents classified into this 4-digit CPC class with a moving window of 3 years. The nodes are the 10-digit CPC subclasses belonging to this 4-digit class and have existed in the patent files of the time range. The edges between two subclasses are defined as existing if these two subclasses ever co-occurred in one of the patent files. For each 4-digit CPC class, we first build its network and extract the giant component — the largest connected cluster of nodes and edges. Following Broekel (2019), we then randomly pick 50 nodes from that component. From each chosen node, we run a 150-step random walktrap to generate a subnetwork. Computing the iNDS on each of these 50 subnetworks by using four structural indicators and averaging the results gives the annual complexity score for the CPC class. The equations for iNDS and NDS are the followings with $S = 50$:

$$iNDS(G_T) = \frac{\alpha_{module} \cdot r_{graphlet}}{v_{module} \cdot v_\lambda} \tag{1}$$

$$NDS(\{G_T^S | G_M\}) = \frac{1}{S} \sum_{G_T \in G_M}^{S} iNDS(G_T) \tag{2}$$

$$Score(G_M) = -\log NDS(\{G_T^S | G_M\}) \tag{3}$$

The four network structural indicators in equation 1 represent different properties of a network. $\alpha_{module} = \frac{M}{n}$ with M being the number of modules (densely connected subgroups) and n being the number of nodes. This measurement represents the module density of a network, namely, how many densely connected groups there are compared to the overall size. It gives a higher value for a less heterogeneous network and helps distinguish between ordered, complex, and random networks. $v_{module} = \frac{var(m)}{mean(m)}$ with m being a vector containing sizes of all modules, showing how different in size the modules are from each other. This variable mainly distinguishes the networks within the three types of networks and gives higher values for higher heterogeneity. $V_\lambda = \frac{var(\Lambda(L))}{mean(\Lambda(L))}$ with $\Lambda(L)$ representing the eigenvalues of Laplacian matrix for the network tells how well a graph is connected. So V_λ is also higher for more heterogeneous networks within the three types of networks. The last indicator $r_{graphlet} = \frac{N_{graphlet(3)}}{N_{graphlet(4)}}$ and $N_{graphlet(a)}$ is number of graphlets of size a. This captures how do small subgroup in the network growth, or compares the number to patterns with three nodes to the number of patterns with 4 nodes. It helps to classify networks into the three types since $r_{graphlet}$ is smallest for random networks, largest for ordered networks and moderate for complex networks.

Figure 1 shows the changes in complexity score at the level of 1-digit CPC industry by averaging the 4-digit level score. The complexity score shows a tendency of increase across years for all industries before 2020. In contrast, decreases occur in all industries after the pandemic, and the sharpest drop happens in Fixed Constructions. This is mirrored by an increase in work-from-home cases and a decrease in the need for transportation and office buildings. Considering the fact that patent application is a revenue-driven activity from the perspective of firms, uncertainty about future policies and market demand might cause fewer patents to be applied and the technological complexity score to decline correspondingly.

The change in the big-picture of complexity score is clearer in figure 2. It shows the distribution of the industrial complexity scores of almost 672 technologies every year from 1982 to 2024. The shading darkens as time goes on. For each year, the distribution is bimodal with a peak at zero and bell-shaped at moderate values. The peaks at zero are caused by the fact that many technologies have too few patents to construct a complicated network and to calculate a score. As years pass, growing patent numbers lead the density at zero to become lower. Furthermore, the maximum density moving to the right side is consistent with technologies being built upon existing ones. This figure shows a similar trend with EU technologies' changes as described in Broekel (2019) using EU patent data.

2.3 Comparison

The advantages of NDS become clear both when compared with traditional complexity measures that ignore network topology and in head-to-head comparisons with other network-based metrics, where it consistently outperforms.

The first part was already discussed and studied by Dr. Broekel in Broekel (2019). The author carried out a direct comparison of NDS against Fleming's modular complexity (FS modular; Fleming and Sorenson (2001)) and Balland's technological complexity index (KCI; Balland and Rigby (2017)) to see how each relates to R&D intensity. NDS correlates positively with patent counts and with collaborative R&D (measured by inventor numbers), assigns higher scores to high-tech fields in every period, and gives older technologies lower complexity scores. In contrast, FS modular increases complexity for older technologies, which contrasts the stylized fact. KCI not only relates negatively to patenting and collaborative R&D but also ranks older technologies as more complex. This comparison highlights that of these three, the structural diversity method (NDS) reflects certain features commonly associated with technological complexity.

The comparison between structural diversity (NDS) and other network-based metrics is theoretically mentioned in Emmert-Streib and Dehmer (2012). Other 16 network complexity measures are also studied in Emmert-Streib and Dehmer (2012) and compared with NDS. These measures differ

from NDS in two main ways. First, NDS combines four distinct structural indicators, whereas the others each rely on a single principle. Second, it constructs the estimation by sampling a finite set of networks from the population and applying the central limit theorem, which helps to eliminate ambiguity from one simple network. The authors contended that NDS is the only system capable of distinguishing ordered, complex, and random networks unambiguously by capturing the heterogeneity in how nodes are connected. In this paper, three traditional network diversity measures are applied to patent data and evaluated alongside the NDS. The three measures are the Bertz index, Wiener index, and mean distance deviation.

Bertz index expresses the total structural information content of a graph and takes the formula (Bertz (1983)):

$$B(G) = 2n\log(n) - \sum_{i=1}^{k} |N_i| \log(|N_i|),\ n = \text{number of nodes} \tag{4}$$

Wiener index is one of the most widely used topological indices in graph theory. It captures the overall compactness by summing up the pairwise length of the shortest path between nodes. The formula for this measurement is (Wiener (1947)):

$$W(G) = \sum_{\{u,v\}\subset V} d(u,v),\ d(u,v) = \text{distance between nodes } u \text{ and } v \tag{5}$$

Mean distance deviation quantifies how spread out the pairwise shortest-path distances are around their average. It complements the Wiener index (which sums distances) by capturing dispersion rather than the total sum using the distances between nodes. In consequence, it offers insight into heterogeneity, information that neither the Bertz index nor the Wiener index can provide. It is defined as (Skorobogatov and Dobrynin (1988), Todeschini and Consonni (2008)):

$$MDD(G) = \frac{1}{n}\sum_{i=1}^{n} |\mu(v_i) - \frac{2W(G)}{n}| \tag{6}$$

where

$$\mu(v_i) = \sum_{j=1}^{n} d(v_i, v_j) \tag{7}$$

When complexity over time for two areas – environmental area (Figure 3) and AI tech (Figure 4) – are reported, apparent differences can be shown. In the environmental sector, the structural diversity score, Bertz index, and Wiener index all follow a similar, steadily increasing path, while the mean distance deviation (MDD) jumps around with no clear pattern. In contrast, for AI technologies, the structural diversity score rises gently up to 2020 and then spikes up sharply from 2021 onward, matching the recent AI boom. Of the other three metrics, only the Wiener index shows a noticeable uptick after 2021 and a gradual rise before that.

The contrasts become even clearer when we aggregate the data up to the 1-digit CPC level (Figure 1 for structural diversity score). The Wiener index (Figure 5) and MDD (Figure 6) stay flat over the years, and their scales differ broadly across sectors to be comparable. The Bertz index (Figure 7) does increase across every sector, but is not yet comparable, because the scores in the electrical sector significantly exceed the rest.

2.4 Conversion to NAICS

The classification system used by the USPTO's dataset is the Cooperative Patent Classification (CPC). When calculating the industrial level complexity score using the patent file from USPTO, the classification is documented under CPC, so one score is generated for each four-digit CPC code in each year. Conversion to other forms of classification is needed since the CPC is used to manage patent documents. The North American Industry Classification System (NAICS) is the standard code used by federal agencies to classify business establishments for the purpose of analyzing statistical data related to the U.S. business economy. Therefore, in order to link the industrial complexity score with Compustat and CRSP data, the score needs to be converted from the CPC to the NAICS form. The detailed conversion processes for the complexity score can be found in Appendices A.1 and B.1.

The North American Industry Classification System (NAICS) was first adopted in 1997 to replace the Standard Industrial Classification (SIC) system. Since then, NAICS has been scheduled to be reviewed every five years for potential revisions, and will be for the foreseeable future, so that it can keep pace with the emerging industries and make clarifications for select industry definitions. Since the time range in my dataset is 1985 - 2023, the relevant NAICS versions include NAICS 1997, NAICS 2002, NAICS 2007, NAICS 2012, NAICS 2017, and NAICS 2022. Historical NAICS codes (naicsh) from Compustat, which represent each firm's NAICS classification for each fiscal year, were used, applying the relevant NAICS version for that year. The crosswalks from CPC to NAICS are only feasible for the first three NAICS versions, which are the NAICS 1997, NAICS 2002, and NAICS 2007, so the conversion between different versions of NAICS is also necessary. In this paper, two types of conversion will be applied. If the analysis needs to link the industrial complexity score with Compustat, conversion to the corresponding version will be applied. If only comparison within complexity score data across years is needed, conversion to the NAICS 2007 version is exercised for easier interpretation.

3 Properties

From this section on, the complexity score will be shown at 4-digit NAICS level by averaging the affiliated 6-digit NAICS codes' scores. This conversion is necessary because further analysis will be based on industry-level by aggregating firm-level data from Compustat or CRSP, and the firm number categorized in each 6-digit NAICS industry is limited in each fiscal year. In order to increase the research power at the industry level, 4-digit NAICS are the better choice because they include more firms each year, facilitating a more comprehensive industry-level view.

Figure 8 shows the distribution of the complexity scores at the level of 4-digit NAICS, in which the outlier represents the year 2024 with incomplete granted patent data. It shows a similar trend with figure 2; majority density moves to the right side, and the left tail becomes thinner, indicating that the overall technologies are becoming more complex over the years. The disappearance of peaks at zero is caused by the conversion from CPC to NAICS, and there are few cases where there are

not enough patents to calculate for all 4-digit CPC classes one NAICS code mapped to.

Compared to figure 8, which presents the change in the technological complexity score on a larger scale, figure 9 selects eight industries and shows the annual complexity scores and employment for each industry. They are, from left to right and top to bottom: Agriculture Group; Air and Rail Transportation; Computer and Electronic Product Manufacturing; Data Processing; Motor Vehicle Manufacturing; Oil and Gas Extraction; Software; and Textile Mills. The chosen industries, ranging from labor-intensive to high-tech areas, all exhibit climbing scores until 2020. After the Covid-19 pandemic struck, labor-intensive industries suffered reductions in complexity scores, while high-tech ones remained stable. Given that labor is one of the main input capital resources, in order to find out whether the drops in labor-intensive industries are paralleled with a decrease in employment, employment data from the Bureau of Labor Statistics's OEWS dataset is used and represented by red lines in the graphs. It turns out that the employment statistics remained steady for these industries and can not account for the drop in complexity scores.

Another observation from Figure 9 is that the complexity score varies across industries, making direct comparisons challenging. For example, the Agriculture Group industry has scores ranging from 12.79 (in 1984) to 33.82 (in 2018), while the Data Processing industry spans from 25.60 (in 1984) to 46.97 (in 2020). High-tech industries' technological complexity levels from forty years ago are estimated to be similar to those of labor-intensive industries today (Agriculture Group scored 25.90 in 2023). This difference arises because each industry is inherently distinct, and the initial scoring system captures this because of the variances in the patent data. Consequently, it is necessary to establish an industry-specific benchmark to enable comparisons of technological complexity scores across industries. Each benchmark is determined by averaging the industry's scores over the entire time period, and the excess score is calculated by subtracting this benchmark from the initial complexity score. Figure 10 presents the same eight industries' excess scores from 1982 to 2023, and after adjusting by subtracting the benchmark, the scores become more comparable across industries.

Table 1 summarizes excess score statistics at the 2-digit NAICS level (the system's broadest category). Column 1 is the 2-digit NAICS sector. Column 2 reports, for each 2-digit NAICS sector,

the total number of panel observations in the data set from 1982 to 2023. Column 3 shows the total count of 4-digit NAICS codes according to the Census Bureau, and Column 4 reports how many of those have valid excess scores over the same period. Finally, Columns 5 and 6 present the median and standard deviation of the excess scores for each sector. There are substantial gaps between Columns 3 and 4 in several sectors, most notably 42 (Wholesale Trade), 48 (Transportation), 53 (Real Estate and Rental and Leasing), and 81 (Other Services). In these industries, fewer 4-digit codes have measurable complexity scores because they generate relatively few patents – reflecting inherently limited opportunities for innovation, either because technological improvement isn't central to their processes or because they've largely reached an innovation plateau. Figure **??** presents a scatter plot of each 2-digit NAICS sector over time, serving as a graphical counterpart to Table 1. The consistent upward trends in every sector further confirm that technological complexity has increased over the years.

4 Drivers of Complexity

4.1 Data

In this section, regressions are used to discuss the determinants of complexity at the industry level (4-digit NAICS). Patent activity is incorporated by counting, for each CPC code in year t quarter q, the growth of the number of patents filed ("*Patent Growth*") and the citations they received within three years of filing ("*3-Year Citation Count*"). These data come from the USPTO's PatentsView database. The annual growth rates of those measures are calculated. In the regressions, lagged versions of these variables are employed. For example, "*L4.3-Year Citation Count*" in year t measures the 3-year citations accrued by patents filed in year $t-4$ (citations are received up to the end of year $t-1$). So that all patent measures reflect prior activity when explaining current complexity.

For this section, all analyses will be shown and discussed in conversion to the corresponding NAICS version since the patent-related data (complexity score, patent count, and citation count)

need to be linked with Compustat. For the data within 1985-2011, the ALP probabilistic linkages from Nathan et al. (2019) are used to map data to the relevant NAICS version. For 2012–2023, when later NAICS versions lack direct CPC links, data are first converted to NAICS 2007 using the CPC-NAICS07 linkages, then conversions between adjacent NAICS versions are applied. The full details on converting patent and citation counts from CPC to NAICS and between adjacent NAICS versions are in Appendices A.2 and B.2, respectively. Once patent data are mapped to 6-digit NAICS codes, we roll them up to the 4-digit level by summing the values for all 6-digit codes that fall under each 4-digit industry.

Compustat provides annual firm-level accounting data along with each firm's historical 6-digit NAICS code. The accounting variables considered include capital expenditure intensity, total assets, return on assets, leverage ratio, market-to-book ratio, and the Herfindahl-Hirschman Index (HHI). The annual accounting figures are then obtained at the industry level by taking the mean of all affiliated firms. A panel of accounting data for each 4-digit NAICS ranging from 1985 (the first year with historical 6-digit NAICS records) to 2023 was constructed. Then a one-to-one merge with excess complexity scores and patent activity indicators was performed to prepare for the regressions.

Table **??** and Table **??** provide definitions and summary statistics for all potential explanatory variables. The statistics show each variable in its original units and in its natural-log form for use in the regressions.

4.2 Models

The baseline analysis begins by examining whether the intensity of capital expenditure is associated with subsequent excess technological complexity, as shown in 8. The dependent variable, $C_{i,t}$, is the natural logarithm of the excess technological complexity score for 4-digit NAICS industry i in year t. The intensity of capital expenditure, measured as capital expenditures relative to total assets, is lagged by three years to allow the time for investment to translate into patent applications as innovation output. To assess whether this relationship is robust to other financial characteris-

tics, additional controls are introduced sequentially. These controls include lagged total assets, leverage, return on assets, market-to-book ratio, and industry concentration. Patent activity controls are included throughout the analysis. $g(patent)_{i,t-1}$ captures the lagged growth in patenting activity, while $Citation^{3}_{i,t-4}$ measures the subsequent citation impact of prior patenting activity. This allows us to examine whether the estimated association between capital expenditure intensity and technological complexity remains stable after accounting for differences in firm size, financial condition, valuation, industry concentration, and innovation activity.

$$\begin{aligned} C_{i,t} = \beta^{0} + \beta^{Capx} \cdot CpaxInt_{i,t-3} + \beta^{A} \cdot Asset_{i,t-1} + \beta^{ROA} \cdot ROA_{i,t-1} + \beta^{L} \cdot Leverage_{i,t-1} \\ + \beta^{MB} \cdot MB_{i,t-1} + \beta^{HHI} \cdot HHI_{i,t-1} \\ + \beta^{g} \cdot g(patent)^{1}_{i,t-1} + \beta^{C} \cdot Citation^{3}_{i,t-4} + \gamma_{i} + \alpha_{t} + u_{i,t} \end{aligned} \quad (8)$$

We further examine whether the relationship of capital expenditure intensity with technological complexity depend on the competitive environment in which firms operate, as shown in Equation 9. In particular, industry concentration may affect firms' incentives and ability to undertake investments that lead to more complex technologies. Greater concentration can provide firms with additional market power and potential rents from innovation, while at the same time reducing competitive pressure to innovate. To examine this heterogeneity, we divide industries according to the median HHI and define $HighConcentration_{i,t}$ as an indicator equal to one when industry i's HHI is at or above the 67th or 50th percentile of that year. We then interact this indicator with the intensity of capital expenditure and the other financial variables. The interaction terms test whether the association between these firm characteristics and subsequent technological complexity differs between relatively competitive and relatively concentrated industries.

$$C_{i,t} = \beta_0 + \beta^{H_3} HighConcentration_{i,t-3} + \beta^{Capx} CapxInt_{i,t-3} + \beta^{Capx,H} \left(HighConcentration_{i,t-3} \times CapxInt_{i,t-3}\right) + \beta^{g} g(patent)_{i,t-1} + \beta^{C} Citation_{i,t-4} + \gamma_i + \alpha_t + u_{i,t}. \quad (9)$$

4.3 Results

Table 6 summarizes seven regressions. The regression in Column (1) include the $CAPX/Assets$ with lags from 1-year to 3-year in the same regressions and only 3-year lagged one shows significant coefficients and will be used in later regressions. Column (2)-(7) sequentially introduce additional financial controls to examine the determinants of industry excess technological complexity. Across all columns, the intensity of capital expenditure remains positively and statistically significant. Importantly, the estimated coefficient on $CAPX/Assets$ is very stable in all columns, remaining close to 1.2 even when controls are added for firm size, leverage, profitability, market-to-book ratio and industry concentration. This stability suggests that the positive association between capital investment intensity and technological complexity is not materially attenuated by including additional financial characteristics. In contrast, most of the other accounting variables are statistically insignificant in the baseline regressions. For patent activity indicators, patent growth is generally insignificant, while the lagged three-year citation count is consistently negative and highly significant. This suggests that industries receiving more citations tend to exhibit lower subsequent excess technological complexity, conditional on patenting activity and the included financial controls.

The positive association between $CAPX/Assets$ and excess complexity score suggests that firms undergoing greater capital renewal or expansion tend to exhibit higher subsequent technological complexity. It shows that capital investment enables the adoption of specialized equipment, production facilities, and other complementary technologies that are necessary to support more sophisticated technological systems. In addition, new capital equipment and production processes may require firms to integrate previously separate technological capabilities. This process of technolog-

ical recombination can generate denser and more heterogeneous relationships among technologies, consistent with the structural features captured by excess technological complexity.

The negative association between citation counts and subsequent excess technological complexity suggests that technological impact and technological complexity capture distinct dimensions of innovation. Highly cited patents may represent influential technological advances, whereas greater complexity may arise from the integration of a larger number of technological components, for which a large number of citations is not needed. This matches defensive patenting, which is often associated with a mismatch between patenting effort and citation impact. Firms engaging in defensive patenting may file multiple closely related patents around a core invention, introducing only incremental modifications for intellectual property protection. Because these patents tend to rely on relatively homogeneous technological components, defensive patenting may reduce the structural diversity of the underlying technology network, while patenting activity increases without a proportional increase in citations.

Table **??** shows that the drivers of excess technological complexity differ substantially between relatively competitive and more concentrated industries. In industries below the 67th or 50th percentile of HHI, capital expenditure intensity is positively and significantly associated with subsequent excess complexity. However, the interaction terms indicate that these relationships are substantially attenuated in highly concentrated industries. In particular, the estimated combined effects are statistically insignificant. These results show that in relatively competitive industries, greater investment intensity is associated with the development of more complex technologies, whereas this relationship largely disappears when firms operate in more concentrated markets.

5 Related to Economic Impact

This section examines the broader economic outcomes associated with technological complexity. Three dimensions through which greater complexity may matter for industries and firms are considered. First, I study labor-market outcomes to assess how technological complexity is associated with the scale and composition of labor demand. Second, firm operating performance is exam-

ined, focusing on profitability, to assess whether greater technological complexity is associated with improved economic performance. Finally, for the market reaction part, KPSS and buy-and-hold abnormal returns (BHAR) are used to investigate how technological complexity is reflected in subsequent stock market performance. Together, these analyses provide a broader view of the economic relevance of technological complexity beyond its determinants.

5.1 Labor Market

For the labor-market analysis, I examine whether technological complexity is associated with changes in industry employment and compensation outcomes. Using quarterly QCEW data as dependent variables, the model is shown as Equation 10. I regress the logarithm of employment (in thousands), total quarterly wages (in millions $), and average weekly wages (in thousands $) on lagged technological complexity and its squared term, while controlling for patent activity and including industry, year, and quarter fixed effects. The baseline specification uses technological complexity lagged by one quarter, and results are shown in Table 8. Additional specifications using four- and eight-quarter lags are reported as robustness tests in Table 13.

The results show a consistent pattern that higher excess technological complexity is positively associated with employment and total quarterly wages, but negatively associated with average weekly wages. The positive and significant squared term for employment further suggests that the relationship with employment becomes stronger at higher levels of complexity, whereas the negative squared term for average weekly wages indicates a more pronounced decline in average wages as complexity increases. Overall, these findings are consistent with technological complexity being associated with an expansion in labor demand accompanied by a shift in workforce composition toward relatively lower-paid workers.

$$L_{i,t.q} = \beta^0 + \beta^C \cdot C_{k,t,q-1} + \beta^{C^2} \cdot C^2_{k,t,q-1} + \beta^p \cdot g(patent)^1_{i,t,q-1} + \beta^C \cdot Citation^3_{i,t-3,q-1} + \gamma_k + \alpha_t + \sigma_q + u_{i,t,q} \quad (10)$$

To examine whether the labor-market consequences of technological complexity vary with industry concentration, I augment Equation 10 by including the interactions of *High Concentration* with both the linear and quadratic terms of technological complexity. The resulting model is Equation 11. The results are shown in Table 9 and indicate substantial heterogeneity across market structures. For employment and total quarterly wages, the interaction between high industry concentration and technological complexity is positive and statistically significant. Further checks show that the combined effects of excess complexity in concentrated industries are still positive and significant. This suggests that the positive association between technological complexity and labor demand is concentrated primarily in concentrated industries. In contrast, the interaction terms for average weekly wages are negative and significant, indicating that the decline in average wages associated with complexity is present across different industry structures but stronger in highly concentrated industries.

$$
\begin{aligned}
L_{i,t,q} = \beta^0 &+ \beta^H \cdot HighConcentration_{i,t,q-1} \\
&+ \beta^C \cdot C_{i,t,q-1} + \beta^{C,HHI} \cdot (HighConcentration_{i,t,q-1} \times C_{i,t,q-1}) \\
&+ \beta^{C^2} \cdot C^2_{i,t,q-1} + \beta^{C^2,HHI} \cdot \left(HighConcentration_{i,t,q-1} \times C^2_{i,t,q-1}\right) \\
&+ \beta^p \cdot g(patent)^1_{i,t,q-1} + \beta^{Cit} \cdot Citation^3_{i,t-3,q-1} + \gamma_i + \alpha_t + \sigma_q + u_{i,t,q}. \quad (11)
\end{aligned}
$$

5.2 Profitability

In the previous subsection, a shift toward a larger share of relatively lower-paid workers in the workforce is found. This raises a natural question about the firm-level consequences of such labor adjustments. In particular, if firms expand employment while relying more heavily on lower-paid labor, the resulting change in labor costs and organizational structure may affect the profitability generated from complex technologies. Therefore, this part examines whether technological complexity is associated with firms' subsequent operating profitability.

Profitability is measured using return on assets and profit margin, as specified in the models in

Equation 12. Because the economic consequences of patented technologies may take time to materialize, the baseline specification uses technological complexity lagged by three years, while shorter- and longer-horizon lags are reported as robustness tests. The results are shown in 10 Column (1)-(2). Across both profitability measures, greater technological complexity is generally associated with lower subsequent profitability, although the strength and nonlinear shape of the relationship vary across time horizons. The coefficients are either negative or statistically insignificant across different complexity lags, providing strong evidence that greater technological complexity worsens operating performance. This pattern is consistent with the possibility that more complex technologies impose higher implementation, coordination, and integration costs that deteriorate future profitability.

$$P_{i,t,q} = \beta^0 + \beta^C \cdot C_{i,t-2,q} + \beta^{C^2} \cdot C^2_{i,t-2,q} + \beta^p \cdot g(patent)_{i,t-2,q} + \beta^{Cit} \cdot Citation^3_{i,t-5,q-1} + \gamma_i + \alpha_t + \sigma_q + u_{i,t,q}. \quad (12)$$

To examine whether the profitability relationship varies with industry concentration, the interactions with *High Concentration* are included. The model is shown in Equation 13 and the results are shown in Table 10 Columns (3)-(4). The predicted value of ROA and profit margins at different level of industrial excess complexity for concentrated industries and competitive industries are shown in Graph **??**. The predicted-profitability plots reveal substantial heterogeneity by industry concentration. Greater technological complexity is generally associated with lower subsequent profitability, although the relationship varies with industry concentration and the level of complexity. In highly concentrated industries, the relationship is nonlinear, with profitability improving over part of the complexity distribution before declining at higher levels.

$$
\begin{aligned}
P_{i,t,q} = \beta^0 &+ \beta^H \cdot HighConcentration_{i,t-2,q} \\
&+ \beta^C \cdot C_{i,t-2,q} + \beta^{C,HHI} \cdot (HighConcentration_{i,t-2,q} \times C_{i,t-2,q}) \\
&+ \beta^{C^2} \cdot C^2_{i,t-2,q} + \beta^{C^2,HHI} \cdot \left(HighConcentration_{i,t-2,q} \times C^2_{i,t-2,q}\right) \\
&+ \beta^p \cdot g(patent)^1_{i,t-2,q} + \beta^{Cit} \cdot Citation^3_{i,t-5,q-1} + \gamma_i + \alpha_t + \sigma_q + u_{i,t,q}. \quad (13)
\end{aligned}
$$

5.3 Market Reaction

The previous section clearly shows that greater technological complexity worsens operating profitability performance with heterogeneity existed between competitive and concentrated industries. In this section, I examine whether industry-level technological complexity is associated with the market valuation of innovation and subsequent firm-level stock-market performance. Two conceptually different types of valuation outcome are considered. The first is quarterly buy-and-hold abnormal returns (BHARs) calculated using the CAPM, Fama-French three-factor, Fama-French/Carhart four-factor benchmarks, and DGTW characteristic-matched benchmarks. The second part focuses on innovation value, using the economic value of patents developed by Kogan et al. (2017), which estimates the dollar value of patents assigned to publicly traded firms based on stock-market reactions around patent grants. We aggregate the KPSS patent values of patents granted to each firm within a quarter to construct a firm-quarter measure of the market value of its innovative output. KPSS measure captures the value associated specifically with patented innovations, and the BHAR measures capture abnormal changes in the market value of the firm as a whole.

The empirical regression is shown in Equation 14. The dependent variable $V_{i,t,q}$ represents the valuation outcome for firm i in year t and quarter q. The main explanatory variable is the industry excess technological complexity score, constructed from the CPC technology network and mapped to each firm through its primary 4-digit NAICS industry. Because the complexity score for quarter q is constructed using patents applied for between $q-11$ and q, while patent applications generally become publicly available only after a delay (about 18 months after the earliest effective filing ex-

cept in special circumstances), the baseline regression uses technological complexity lagged by six quarters. This timing allows even the most recently filed applications to become publicly observable and for the complexity measure to enter the regression with sufficient time to become publicly observable. Three-year lags, matched with the regressions on profitability, are additionally examined as robustness tests. The regressions include both the linear and squared terms of technological complexity to allow for potential nonlinearities, as well as controls for firm patenting activity and citation impact. Industry, year, and quarter fixed effects are included throughout. The linkage between patents and publicly traded firms follows the updated data from Kogan et al. (2017).

$$V_{i,t,q} = \beta^0 + \beta^C \cdot C_{k,t-1,q-2} + \beta^{C^2} \cdot C^2_{k,t-1,q-2} + \beta^p \cdot g(patent)_{k,t-1,q-2} + \beta^{Cit} \cdot Citation^3_{i,t-4,q-3} + \gamma_k + \alpha_t + \sigma_q + u_{i,t,q}. \quad (14)$$

Table 11 reports the baseline relationship between industry-level technological complexity and firm-level valuation outcomes using a six-quarter lag. The results reveal a clear distinction between the valuation of patented innovation and the valuation of the firm as a whole.In Columns (1)-(4), the BHAR regressions provide little evidence of a systematic relationship between technological complexity and subsequent firm-level abnormal returns. Under the six-quarter baseline specification, the linear complexity coefficient is insignificant at the 5% significance level. Moreover, the 3-year robustness regressions, presented in Table 14, show a stable, negative pattern across the three BHAR measures, consistent with the results in the profitability section. In contrast, In Column (5), technological complexity is positively and significantly associated with the KPSS measurement. The coefficient on the linear complexity term is 0.083 and statistically significant at the 1% level. Thus, the evidence indicates a positive relationship between industry technological complexity and the economic value of patented innovation, which remains robust with the 3-year lagged regressions. These results suggest that the higher value associated with technologically complex innovations does not transform into superior subsequent excess stock-market performance at the firm level, at both the risk-adjusted level and compared to similar firms.

$$
\begin{aligned}
V_{i,t,q} = \beta^0 &+ \beta^H \cdot HighConcentration_{k,t-1,q-2} \\
&+ \beta^C \cdot C_{k,t-1,q-2} + \beta^{CH} \cdot (HighConcentration_{k,t-1,q-2} \times C_{k,t-1,q-2}) \\
&+ \beta^{C^2} \cdot C^2_{k,t-1,q-2} + \beta^{C^2H} \cdot \left(HighConcentration_{k,t-1,q-2} \times C^2_{k,t-1,q-2}\right) \\
&+ \beta^p \cdot g(patent)_{k,t-1,q-2} + \beta^{Cit} \cdot Citation^3_{k,t-4,q-3} + \gamma_k + \alpha_t + \sigma_q + u_{i,q}. \quad (15)
\end{aligned}
$$

To examine whether the valuation implications of technological complexity vary with market structure, an indicator for concentrated industries is added both by itself and as interaction terms. The model is shown in Equation 15 and the results are shown in Table 12. The results indicate no heterogeneity across different concentration regimes.

5.4 Summary of Economic Impacts

Taken together, the economic-impact results suggest that the benefits of technological complexity accrue more clearly to labor than to firms. Greater technological complexity is associated with expanding employment and higher total payroll, while average weekly wages decline, a pattern consistent with workforce growth being disproportionately concentrated among relatively lower-paid workers. From the firm perspective, however, the benefits are much less apparent. Greater complexity is generally associated with lower subsequent profitability, while the market-reaction analysis provides little evidence of improved firm-level abnormal returns despite a positive association with the value of patented innovation. An important exception appears in highly concentrated industries, where firms with moderate levels of excess technological complexity can experience improved profitability before performance declines again at higher complexity levels.

Overall, the evidence suggests that complex technologies are associated with broader employment gains and increased demand for relatively lower-paid labor, while the corresponding gains to firms are limited and depend on market structure and the degree of technological complexity.

6 Conclusion

Using patent data's CPC classification from 1985 to 2023, networks representing the knowledge component's connection can be built annually for each technology, and the structural diversity score measuring the heterogeneity degree of the network is proved to be better than other existing network metrics or complexity measurements in quantifying industrial complexity level.

This paper lays out a step-by-step method for converting complexity scores, patent counts, and citation counts from CPC to NAICS and between NAICS versions. The same procedure can be used to translate any other variables from CPC to NAICS or across different NAICS versions.

By examining the industrial complexity score, this paper reveals that complexity rose steadily across all sectors until 2020. When the Covid-19 pandemic hit, complexity fell in labor-intensive industries – compared with the stable scores in tech-related areas, though employment didn't drop by the same amount. To make scores comparable across industries, we set each industry's own benchmark and calculated excess complexity. It has also been found that, for some sectors, more affiliated industries lack measurable complexity scores because they lack patents.

The drivers analysis shows that higher capital expenditure intensity is consistently associated with greater subsequent technological complexity, even after progressively adding financial and industry controls. The heterogeneity results further indicate that this relationship varies with market concentration, with the positive association between investment intensity and complexity substantially weaker in more concentrated industries.

The benefits of greater technological complexity fall more in labor than in firms. Wage compression, with greater distribution toward lower-paid workers, and expansion in the total labor force are found in industries with high excess technological complexity scores. However, positive impacts toward firms are not found, except that firms with moderate levels of excess technological complexity in concentrated industries can experience higher profitability with higher technological complexity.

References

Arthur, W. Brian, "The Nature of Technology : What It Is and How It Evolves," *Free Press*, 2009.

Balland, Pierre-Alexandre and David Rigby, "The Geography of Complex Knowledge," *Economic Geography*, 2017, *93* (1), 1–23.

Bertz, S.H., "On the complexity of graphs and molecules," *Bltn Mathcal Biology*, 1983, *45*, 849–855.

Broekel, Tom, "Using structural diversity to measure the complexity of technologies," *PLOS ONE*, 05 2019, *14* (5), 1–23.

Coad, Alex and Rekha Rao, "Innovation and Firm Growth in High-Tech Sectors: A Quantile Regression Approach," *Research Policy*, 2008, *37* (4), 633–648.

Emmert-Streib, Frank and Matthias Dehmer, "Exploring Statistical and Population Aspects of Network Complexity," *PLOS ONE*, 05 2012, *7* (5), 1–17.

Fleming, Lee and Olav Sorenson, "Technology as a complex adaptive system: evidence from patent data," *Research Policy*, 2001, *30* (7), 1019–1039.

III, Donald E Bowen, Laurent Frésard, and Gerard Hoberg, "Rapidly evolving technologies and startup exits," *Management Science*, 2023, *69* (2), 940–967.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring technological innovation over the long run," *American Economic Review: Insights*, 2021, *3* (3), 303–320.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological Innovation, Resource Allocation, and Growth*," *The Quarterly Journal of Economics*, 03 2017, *132* (2), 665–712.

Mariani, Manuel Sebastian, Matúš Medo, and François Lafond, "Early identification of important patents: Design and validation of citation network metrics," *Technological forecasting and social change*, 2019, *146*, 644–654.

McNerney, James, J. Doyne Farmer, Sidney Redner, and Jessika E. Trancik, "Role of design complexity in technology improvement," *Proceedings of the National Academy of Sciences*, 2011, *108* (22), 9008–9013.

Mewes, Lars and Tom Broekel, "Technological Complexity and Economic Growth of Regions," *Research Policy*, 2022, *51* (8), 104156.

Nathan, Travis J. Lybbert, and Zolas Nikolas Jason, "An 'Algorithmic Links with Probabilities' Crosswalk for USPC and CPC Patent Classifications with an Application Towards Industrial Technology Composition," *Economics of Innovation and New Technology*, 2019.

Nepelski, Daniel and Giuditta De Prato, "Technological complexity and economic development," *Review of Development Economics*, 2020, *24* (2), 448–470.

Skorobogatov, VA and AA Dobrynin, "Metrical analysis of graphs," *Commun. Math. Comp. Chem*, 1988, *23*, 105–155.

Sorenson, Olav and Lee Fleming, "Science and the diffusion of knowledge," *Research Policy*, 2004, *33* (10), 1615–1634.

Todeschini, Roberto and Viviana Consonni, *Handbook of molecular descriptors*, John Wiley & Sons, 2008.

Wiener, Harry, "Structural Determination of Paraffin Boiling Points," *Journal of the American Chemical Society*, 1947, *69* (1), 17–20.

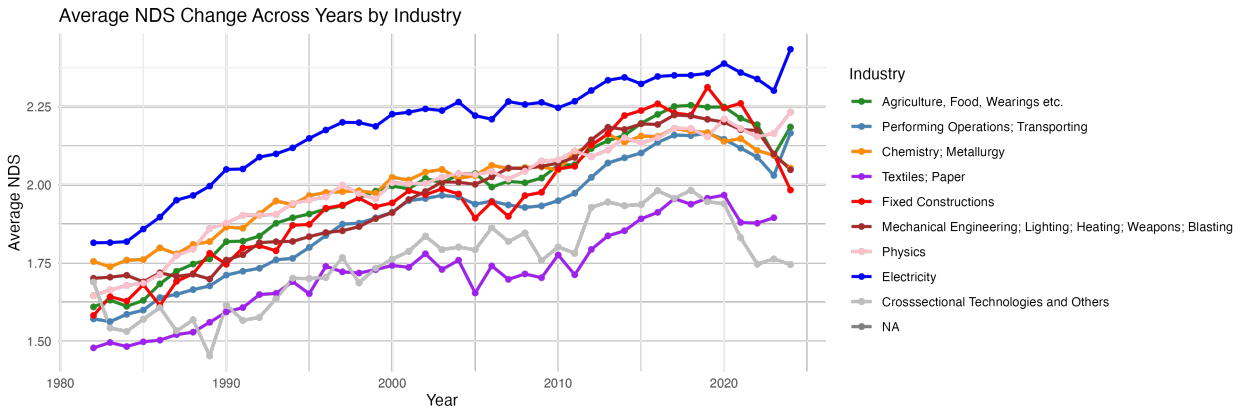


Figure 1: Complexity Score Across Years by 1-digit CPC Industry

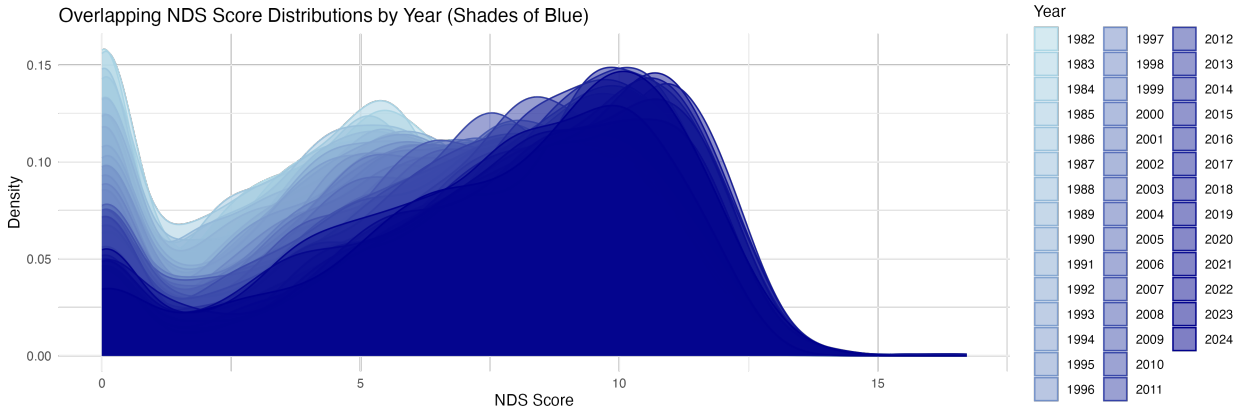


Figure 2: Overlapping NDS Score Distributions by Year

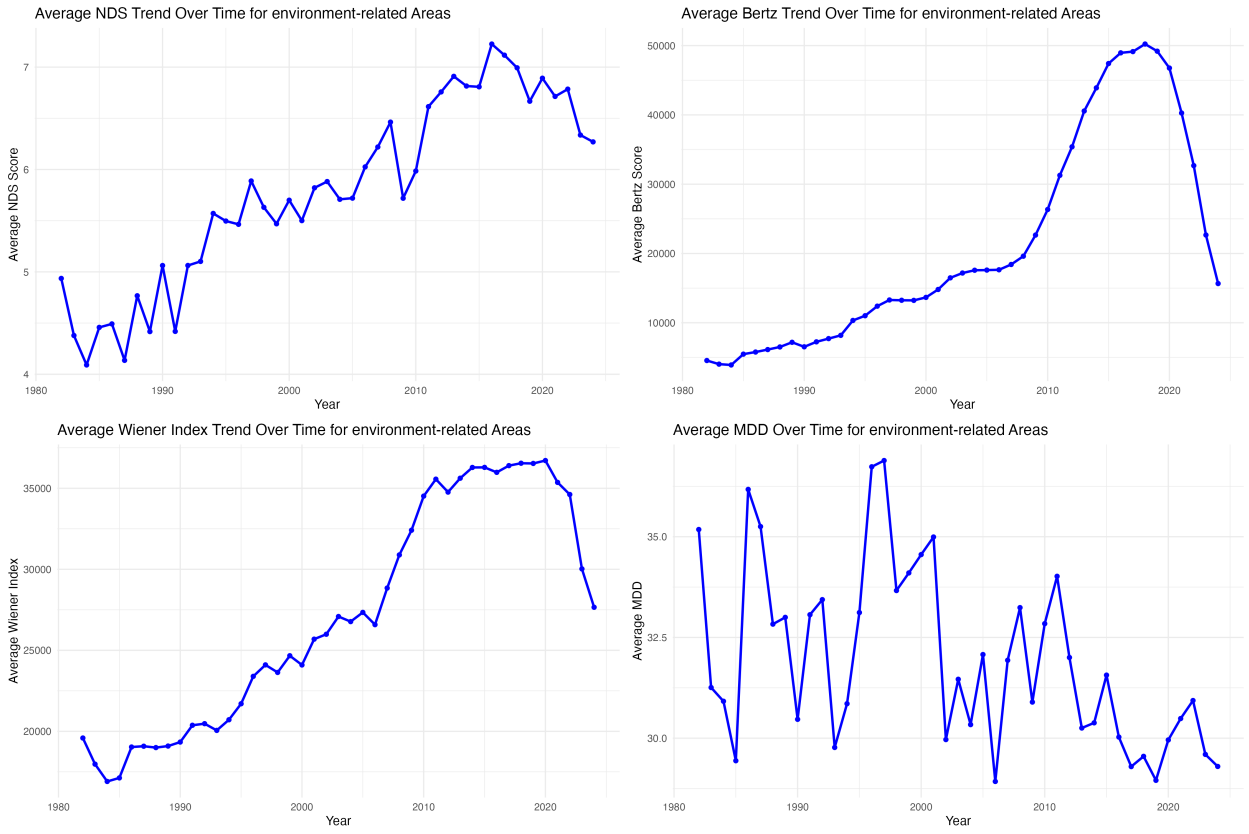


Figure 3: Annual Complexity Score of Environmental-related Areas

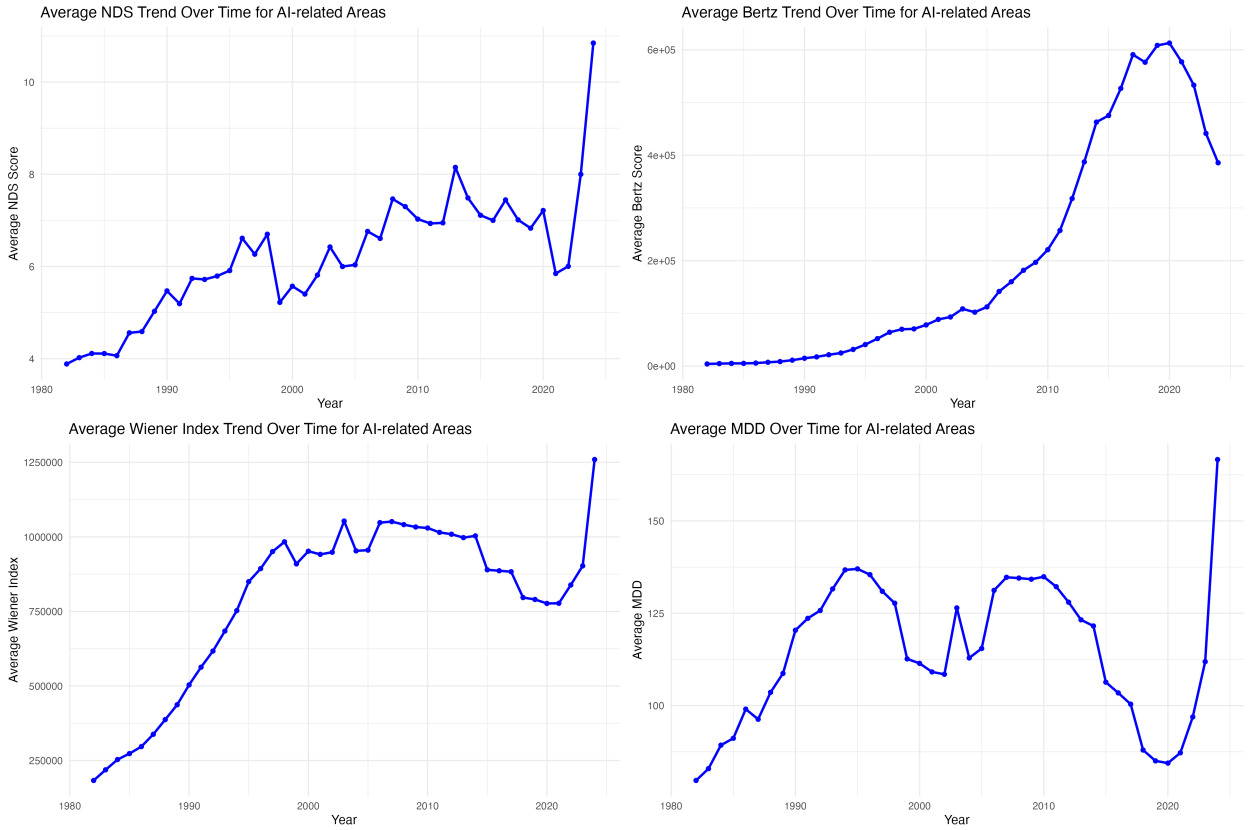


Figure 4: Annual Complexity Score of AI-related Areas

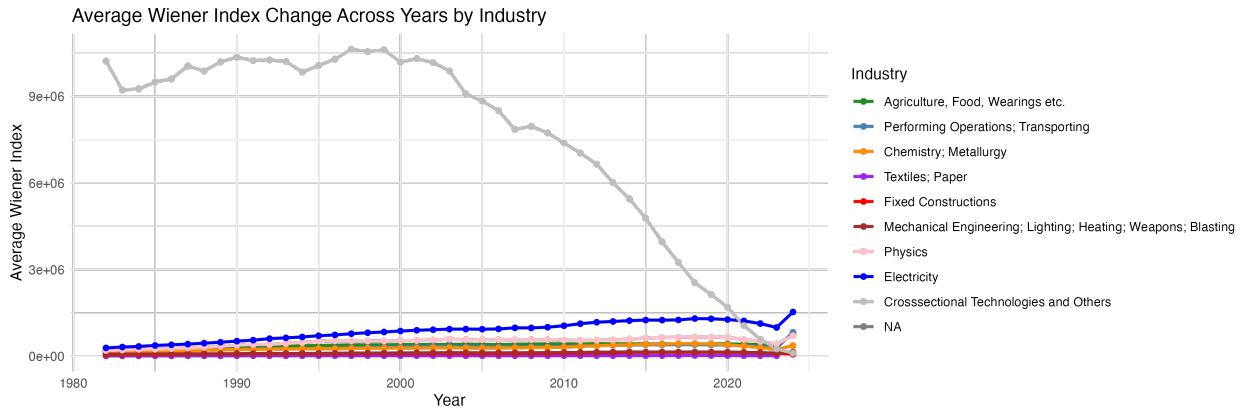


Figure 5: Complexity Wiener Index Across Years by 1-digit CPC Industry

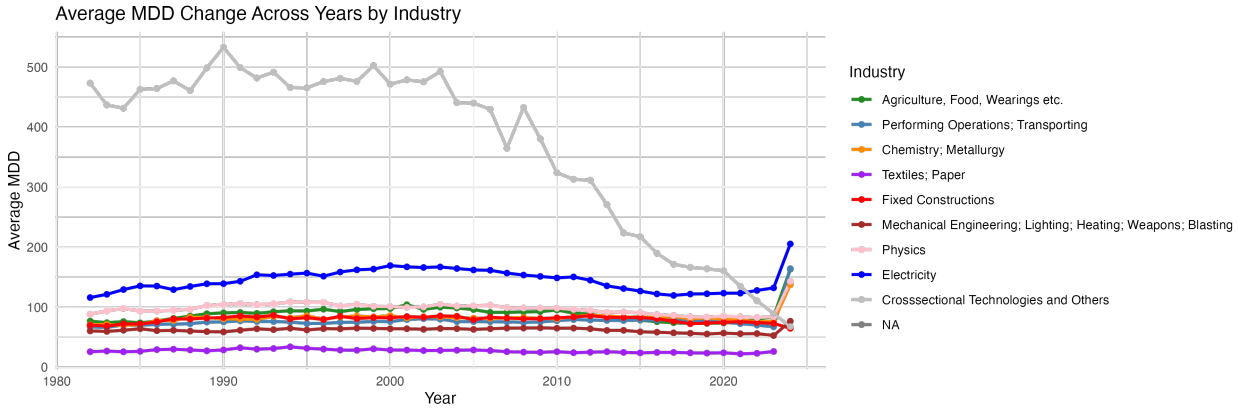


Figure 6: Complexity MDD Across Years by 1-digit CPC Industry

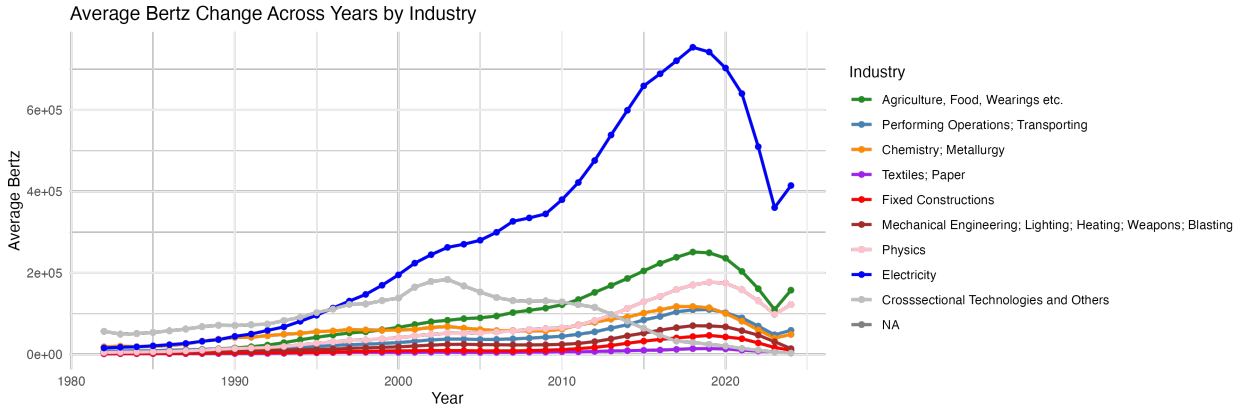


Figure 7: Complexity Bertz Index Across Years by 1-digit CPC Industry

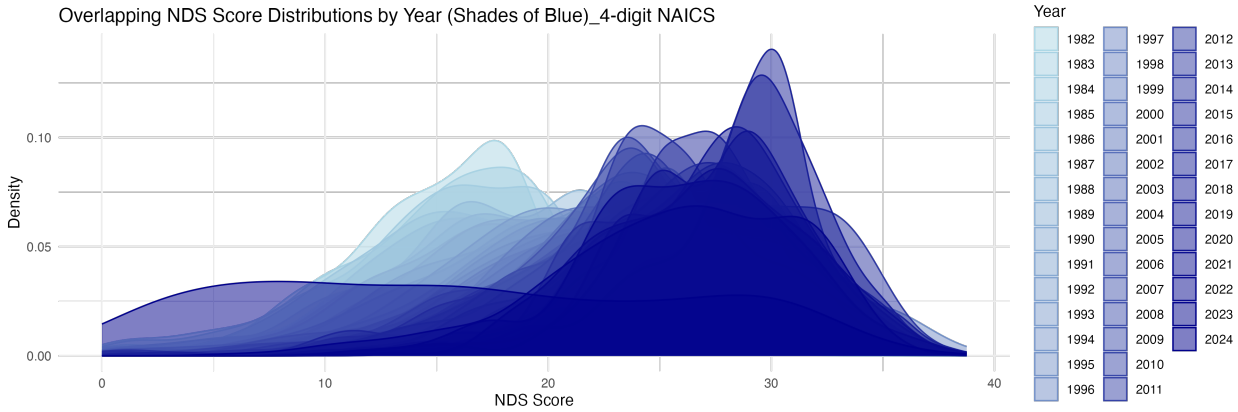


Figure 8: Overlapping NDS Score Distributions by Year at the level of 4-digit NAICS

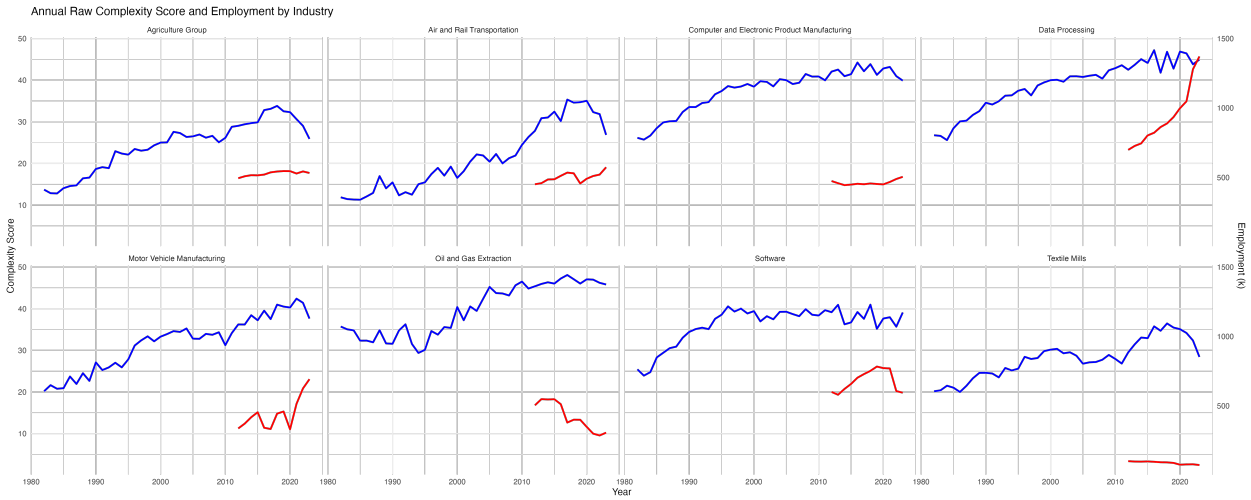


Figure 9: Annual Complexity Score and Employment by Industry

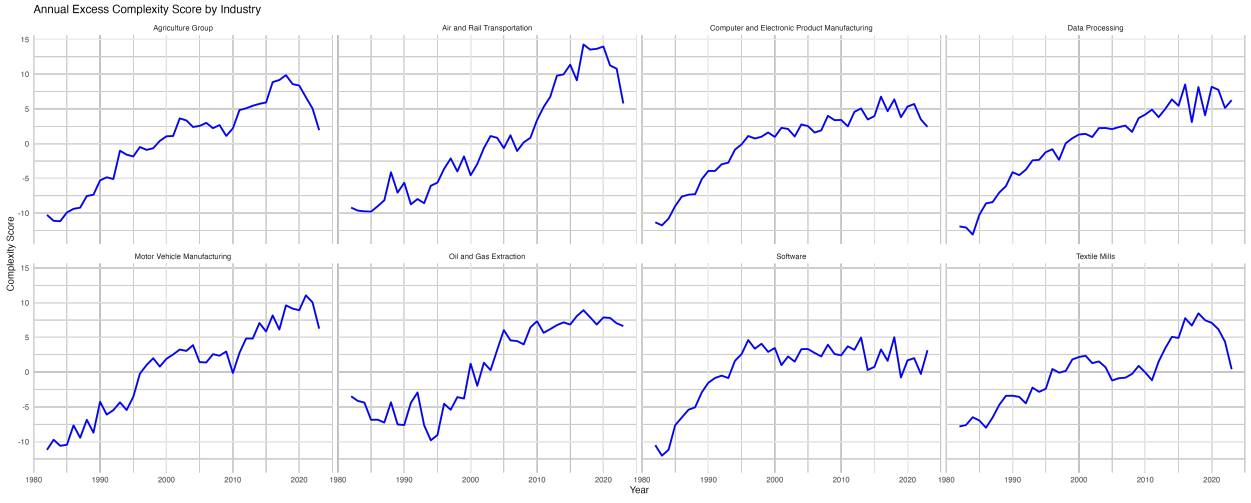
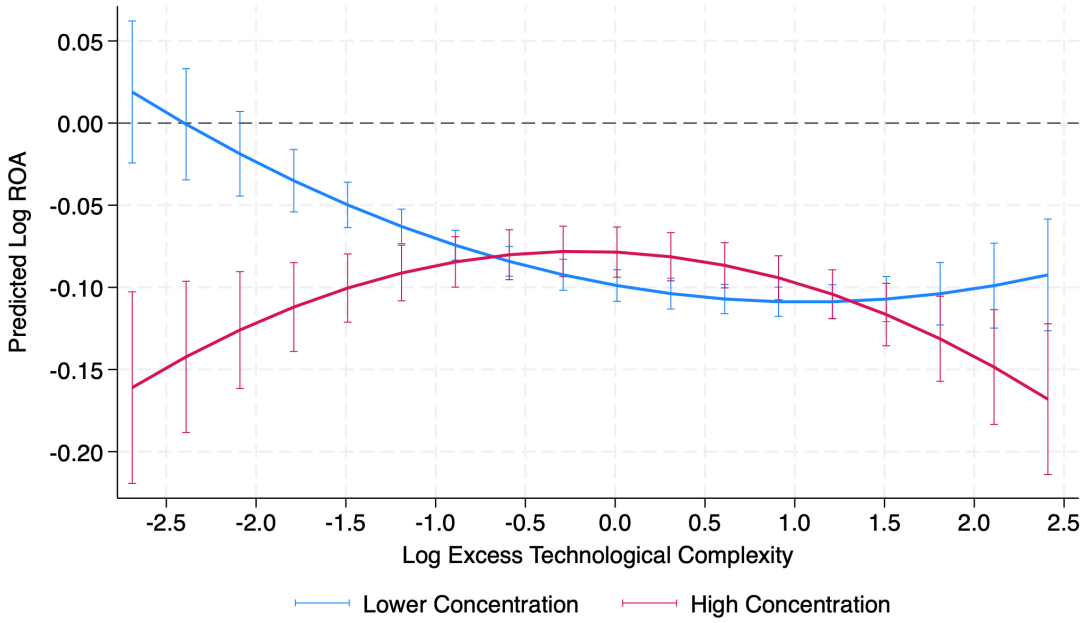
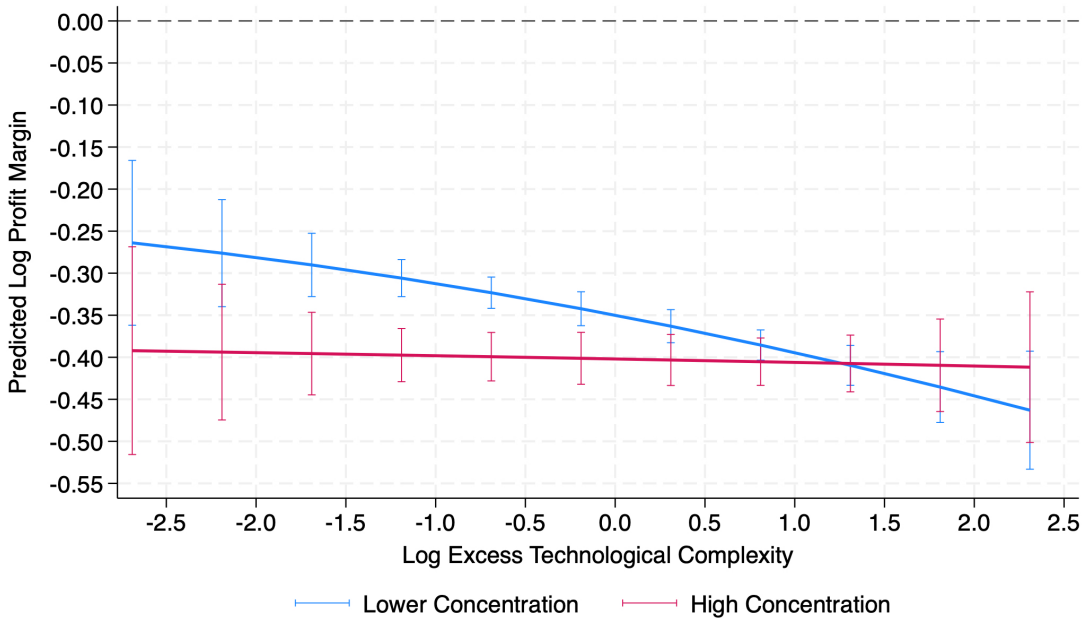


Figure 10: Annual Excess Complexity Score by Industry



(a) Return on Assets



(b) Profit Margin

Notes. This figure plots predicted firm profitability against lagged excess technological complexity. Panel (a) reports predicted log return on assets (ROA), and Panel (b) reports predicted log profit margin. *High Concentration* denotes industries with HHI at or above the 67th percentile of the sample distribution, while *Lower Concentration* denotes industries below this threshold. Predicted values are based on the profitability regressions using technological complexity lagged by 12 quarters. Error bars represent 95% confidence intervals.

Table 1: Statistics for Excess Score at 4-digit NAICS level

2-digit NAICS	Obs	Count of 4-digit NAICS	Count w/ score data	Median	Std. dev.
11 (Agriculture, Forestry, Fishing & Hunting)	773	19	18	0.83	6.57
21 (Mining, Quarrying, Oil & Gas Extraction)	215	5	5	0.51	5.79
22 (Utilities)	86	3	2	1.16	5.69
23 (Construction)	430	10	10	-0.47	6.57
31 (Food, Beverage, Tabacco & Cloth Manufacturing)	938	22	22	0.59	5.95
32 (Wood, Paper, Printing, Coal Products, Chemical Manufacturing)	901	21	21	0.54	4.80
33 (Metal Prod, Machinery, Electronic Prod & Transportation Euip)	1,803	43	42	0.58	6.11
42 (Wholesale Trade)	429	19	10	0.04	6.50
44 (Retail 1)	686	16	16	0.69	7.07
45 (Retail 2)	387	11	9	0.97	6.30
48 (Transportation)	902	25	21	0.07	7.66
49 (Postal Service & Warehousing)	171	4	4	1.57	6.61
51 (Information)	473	12	11	1.75	6.76
52 (Finance & Insurance)	473	11	11	-0.26	6.02
53 (Real Estate & Rental & Leasing)	129	8	3	-1.56	5.99
54 (Professional, Scientific & Technical Services)	387	9	9	0.87	5.49
55 (Management of Companies & Enterprises)	43	1	1	1.85	6.23
56 (Administrative & Support, Waste Mgmt & Remediation Service)	387	11	9	0.40	5.24
61 (Educational Services)	215	7	5	0.76	4.95
62 (Health Care & Social Assistance)	687	18	16	0.83	5.78
71 (Arts, Entertainment, & Recreation)	344	9	8	-0.94	6.51
72 (Accommodation & Food Services)	258	7	6	1.14	7.27
81 (Other Services)	386	14	9	0.58	5.89
92 (Public Administration)	344	8	8	0.90	6.05
24 Sectors	11,847	313	276	0.64	6.23

Notes. This table summarizes excess score statistics at the 2-digit NAICS level (the system's broadest category). Column 1 is the 2-digit NAICS sector. Column 2 reports, for each 2-digit NAICS sector, the total number of panel observations in the data set from 1985 to 2023. Column 3 shows the total count of 4-digit NAICS codes according to the Census Bureau, and Column 4 reports how many of those have valid excess scores over the same period. Finally, Columns 5 and 6 present the median and standard deviation of the excess scores for each sector.

Table 2: Variable Definitions

Variable	Definition	Source
Panel A: Technological Complexity and Patent Activity		
Excess Technological Complexity	Industry-level technological complexity measure constructed from the CPC technology-class network.	USPTO
Log Excess Technological Complexity	Natural logarithm of the excess technological complexity score.	USPTO
Patent Growth	Growth in the number of patent applications relative to the preceding period. The logarithmic transformation is used in the regression analyses.	USPTO
3-Year Citation Count	Number of forward citations received by patents filed during the specified period within three years after filing.	USPTO
Panel B: Financial and Industry Characteristics		
CAPX/Assets	Capital expenditures divided by total assets.	Compustat
Total Assets	Total book value of firm assets.	Compustat
Leverage Ratio	Firm leverage measured relative to the firm's asset base.	Compustat
Return on Assets (ROA)	Net income scaled by total assets.	Compustat
Profit Margin	Net income relative to sales.	Compustat
Market-to-Book Ratio	Ratio of the market value of the firm to its book value.	Compustat
HHI	Herfindahl–Hirschman Index measuring industry concentration.	Compustat
Concentrate	Indicator equal to one when an industry's HHI is at or above the 67th (or 50th) percentile of the sample HHI distribution, and zero otherwise.	Compustat
Panel C: Labor Market Outcomes		
Employment	Number of employees in the industry during the quarter.	QCEW
Total Quarterly Wages	Total wages paid to workers in the industry during the quarter.	QCEW
Average Weekly Wage	Average weekly wage paid to workers in the industry during the quarter.	QCEW
Panel D: Firm Performance and Market Valuation		
KPSS Patent Value	Private economic value of patented innovation based on the market-value measure developed by Kogan et al. (2017).	KPSS
CAPM BHAR	Quarterly buy-and-hold abnormal return calculated relative to expected returns from the Capital Asset Pricing Model.	CRSP
FF3 BHAR	Quarterly buy-and-hold abnormal return calculated relative to expected returns from the Fama-French three-factor model.	CRSP
FF4 BHAR	Quarterly buy-and-hold abnormal return calculated relative to expected returns from the Fama-French/Carhart four-factor model.	CRSP

Notes. Unless otherwise stated, continuous variables entering the regression analyses are transformed using the natural logarithm. Patent variables are dated using patent application dates. Industry-level variables are defined at the 4-digit NAICS level. Variable lags differ across empirical specifications and are reported in the corresponding regression tables.

Table 3: Summary Statistics

Variable	Obs.	Mean	P25	Median	P75	Std. Dev.
Panel A: Determinants of Technological Complexity — Industry-Year Level						
Log Excess Technological Complexity	8,406	0.050	-1.499	0.276	1.515	1.583
Log CAPX/Assets	8,385	0.052	0.028	0.044	0.066	0.039
Log Total Assets	8,401	7.163	6.199	7.264	8.214	1.591
Log Leverage Ratio	8,292	0.673	0.339	0.528	0.850	0.503
Log Return on Assets	8,400	-0.014	-0.043	0.013	0.046	0.116
Log Market-to-Book Ratio	8,388	1.088	0.843	1.147	1.478	0.994
Log HHI	8,406	8.027	7.510	8.058	8.647	0.856
Log Patent Growth	8,162	-0.115	-0.093	0.025	0.108	0.897
Log 3-Year Citation Count	8,406	5.766	3.182	5.781	8.374	3.298
Panel B: Labor Market — Industry-Quarter						
Log Excess Technological Complexity	35,153	0.006	-1.017	0.089	1.004	1.131
Log Employment (000s)	36,445	5.176	4.383	5.226	6.197	1.471
Log Quarterly Wages ($ mil.)	36,445	7.492	6.602	7.522	8.573	1.572
Log Average Weekly Wage ($000s)	36,445	0.625	0.429	0.588	0.775	0.268
Panel C: Profitability — Industry-Quarter						
Log Return on Assets	34,768	-0.081	-0.048	-0.004	0.010	0.392
Log Profit Margin	34,732	-0.343	-0.350	-0.029	0.034	0.873
Panel D: Market Valuation — Firm-Quarter						
Log KPSS Patent Value	124,205	3.310	1.421	2.886	4.820	2.296
Log CAPM BHAR	949,051	-0.002	-0.107	-0.008	0.079	0.208
Log FF3 BHAR	949,051	-0.005	-0.106	-0.009	0.076	0.206
Log FF4 BHAR	949,051	-0.005	-0.105	-0.008	0.077	0.207
Log Patent Growth	87,136	0.112	-0.268	0.000	0.405	0.502
Log 3-Year Citation Count	949,286	0.383	0.000	0.000	0.000	1.265

Notes. This table reports summary statistics for variables used in the empirical analyses. Panel A reports annual 4-digit NAICS industry-level variables used in the determinants analysis. Panel B reports quarterly industry-level labor-market outcomes. Panel C reports quarterly profitability measures, and Panel D reports firm-quarter market-valuation outcomes. The annual and quarterly technological complexity measures are calculated separately at their corresponding frequencies. Continuous variables are reported using the transformed measures employed in the regressions. For variables that may take negative values, the transformation is defined as $\text{sgn}(x)\ln(1+|x|)$.

Table 4: Robustness Checks for the Determinants of Industrial Technological Complexity

	(1)	(2)	(3)
Panel A: Firm and industry characteristics			
	Dependent variable: Industrial Technological Complexity Excess Score		
Log Total Assets (lag 1)	-0.008**	-0.007	-0.008
	(-2.03)	(-0.61)	(-0.66)
Log Total Assets (lag 2)		-0.004	-0.004
		(-0.38)	(-0.31)
Log Total Assets (lag 3)			-0.003
			(-0.23)
Log Return on Assets (lag 1)	-0.088***	-0.088***	-0.062*
	(-3.34)	(-2.76)	(-1.84)
Log Return on Assets (lag 2)		-0.020	-0.001
		(-0.61)	(-0.03)
Log Return on Assets (lag 3)			-0.038
			(-1.14)
Log Leverage Ratio (lag 1)	-0.014	-0.018	-0.014
	(-1.09)	(-0.88)	(-0.63)
Log Leverage Ratio (lag 2)		-0.006	0.009
		(-0.28)	(0.36)
Log Leverage Ratio (lag 3)			-0.031
			(-1.41)
Log Market-to-Book Ratio (lag 1)	-0.001	0.025*	0.024*
	(-0.17)	(1.93)	(1.88)
Log Market-to-Book Ratio (lag 2)		-0.029**	-0.028
		(-2.27)	(-1.64)
Log Market-to-Book Ratio (lag 3)			0.002
			(0.12)
Log Herfindahl–Hirschman Index (lag 1)	-0.014***	-0.009	-0.009
	(-2.84)	(-1.26)	(-1.19)
Log Herfindahl–Hirschman Index (lag 2)		-0.007	-0.009
		(-1.01)	(-1.12)
Log Herfindahl–Hirschman Index (lag 3)			0.001
			(0.12)
Industry fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes
Panel B: Patent activity indicators			
	Dependent variable: Industrial Technological Complexity Excess Score		
Log Patent Growth (lag 1)	0.033*	0.059***	0.073***
	(1.90)	(3.00)	(3.61)
Log Patent Growth (lag 2)		0.054***	0.082***
		(2.74)	(3.59)
Log Patent Growth (lag 3)			0.041*
			(1.89)
Log 1-year Citation (lag 5)	-0.056***	-0.028***	-0.019***
	(-9.97)	(-4.14)	(-2.68)
Log 1-year Citation (lag 6)		-0.054***	-0.045***
		(-8.07)	(-6.36)
Log 1-year Citation (lag 7)			-0.026***
			(-3.58)
Industry fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes

Notes. This table reports separate robustness checks for the determinants of industrial technological complexity at the four-digit NAICS level. Panel A examines firm and industry characteristics, and Panel B examines patent activity indicators. Each set of three columns represents three separate regressions with the displayed lag structure. Column (1) includes the first displayed lag, column (2) adds the second displayed lag, and column (3) adds the third displayed lag. The variables are expressed in logarithmic terms. Parentheses report t-statistics. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 5: Correlations Between Accounting Variables and Patent-Activity Indicators

	Total Assets	ROA	Lvg. Ratio	MB Ratio	HHI	Patent Growth	1-yr Citation
Total Assets	1.0000						
ROA	0.1141	1.0000					
Lvg. Ratio	0.0324	-0.0637	1.0000				
MB Ratio	0.0292	0.0383	-0.0662	1.0000			
HHI	-0.6597	-0.0800	0.0571	0.0442	1.0000		
Patent Growth	-0.0268	0.0031	-0.0175	-0.0007	-0.0195	1.0000	
1-yr Citation	0.2712	0.0574	-0.2012	-0.0111	-0.3017	0.0944	1.0000

Notes. This table reports the pairwise correlation matrix for the log-transformed accounting variables, industry characteristics, and patent-activity measures used in the analysis. The matrix is reported in lower-triangular form. Patent Growth is the log-transformed growth rate of the patent count from quarter $q-2$ to quarter $q-1$. The 1-yr Citation variable is the log-transformed number of citations received within one year by patents filed in quarter $q-5$; its citation window ends in quarter $q-1$.

Table 6: Regression Results for Determinants of Industry Technological Complexity

	Log Excess Technological Complexity Score						
Logged Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
L. CAPX/Assets	-0.255						
	(-0.47)						
L2. CAPX/Assets	-0.326						
	(-0.58)						
L3. CAPX/Assets	1.312***	1.168**	1.139**	1.225**	1.228**	1.236**	1.212**
	(2.90)	(2.38)	(2.31)	(2.48)	(2.49)	(2.51)	(2.47)
L. Total Assets			-0.024	-0.025	-0.020	-0.020	-0.021
			(-1.22)	(-1.25)	(-0.97)	(-0.96)	(-1.00)
L. Leverage Ratio				-0.035	-0.045	-0.046	-0.054
				(-0.62)	(-0.79)	(-0.81)	(-0.94)
L. Return on Assets					-0.195	-0.220	-0.222
					(-1.29)	(-1.45)	(-1.46)
L. Market-to-Book Ratio						0.021	0.022
						(1.42)	(1.48)
L. HHI							-0.056
							(-1.37)
Patent Activity Controls							
L. Patent Growth		0.069	0.070	0.059	0.059	0.059	0.055
		(1.21)	(1.23)	(1.05)	(1.05)	(1.04)	(0.98)
L4. 3-Year Citation Count		-0.154***	-0.151***	-0.140***	-0.140***	-0.140***	-0.145***
		(-5.28)	(-5.15)	(-4.73)	(-4.72)	(-4.71)	(-4.85)
Constant	-1.943***	-0.917***	-0.778***	-0.830***	-0.859***	-0.882***	-0.397
	(-28.95)	(-4.82)	(-3.49)	(-3.70)	(-3.80)	(-3.90)	(-0.94)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,346	7,086	7,084	7,004	7,003	7,000	7,000
Adjusted R-squared	0.466	0.447	0.447	0.450	0.450	0.450	0.450

Notes. This table reports regression estimates of the determinants of industry technological complexity using annual data. The dependent variable is the log-transformed excess technological complexity score at the 4-digit NAICS industry level. All specifications include 4-digit NAICS industry fixed effects and year fixed effects. Column (1) jointly includes capital expenditure intensity, measured as CAPX divided by total assets, lagged by one, two, and three years to examine the timing of the relationship between capital investment and subsequent technological complexity. Columns (2)-(7) use the three-year lag of CAPX/Assets and sequentially add controls for firm size, leverage, profitability, market-to-book ratio, and industry concentration. Patent Growth is lagged by one year, and the 3-Year Citation Count is lagged by four years. Parentheses report heteroskedasticity-robust t-statistics. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 7: Heterogeneity in Determinants of Industry Technological Complexity by Industry Concentration

Logged Variables	Log Excess Technological Complexity Score (1) HHI Median	(2) HHI 67th Percentile
L3.High Concentration	0.192***	0.156**
	(2.59)	(2.34)
L3. CAPX/Assets	4.124***	2.402***
	(4.38)	(2.97)
L3.High Concentration ×	-3.696***	-1.860*
L3. CAPX/Assets	(-3.68)	(-1.95)
Patent Activity Controls		
L. Patent Growth	0.068	0.070
	(1.20)	(1.23)
L4. 3-Year Citation Count	-0.151***	-0.150***
	(-5.13)	(-5.12)
Constant	-1.119***	-1.038***
	(-5.48)	(-5.28)
Industry FE	Yes	Yes
Year FE	Yes	Yes
Observations	7,086	7,086
Adjusted R-squared	0.448	0.447

Notes. This table examines whether the relationship between capital expenditure intensity and subsequent technological complexity varies with industry concentration. The dependent variable is the log-transformed excess technological complexity score at the 4-digit NAICS industry level. In Column (1), *High Concentration* equals one for industries with HHI at or above the sample median. In Column (2), *High Concentration* equals one for industries with HHI at or above the 67th percentile. Capital expenditure intensity and the concentration indicator are lagged by three years. Patent Growth is lagged by one year, and the 3-Year Citation Count is lagged by four years. All specifications include 4-digit NAICS industry fixed effects and year fixed effects. Parentheses report heteroskedasticity-robust t-statistics. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

Table 8: Technological Complexity and Labor Market Outcomes

Logged Variables	Labor Market Outcomes		
	Employment	Total Quarterly Wages	Average Weekly Wage
	(1)	(2)	(3)
L. Exc Complexity	0.013***	0.009***	-0.007***
	(6.47)	(3.92)	(-10.25)
L. Exc Complexity^2	0.008***	0.003	-0.006***
	(3.97)	(1.29)	(-9.59)
Patent Activity Controls			
L.Patent Growth	-0.002	-0.003	-0.001
	(-0.30)	(-0.44)	(-0.57)
L13. 3-Year Citation Count	0.056***	0.076***	0.015***
	(17.88)	(21.56)	(15.62)
Constant	4.873***	6.643***	0.371***
	(234.19)	(290.38)	(59.37)
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Observations	31,399	31,399	31,399
Adjusted R-squared	0.976	0.973	0.929

Notes. This table reports regressions examining the association between technological complexity and labor market outcomes at the 4-digit NAICS industry level. The dependent variables are the logarithms of employment, total quarterly wages, and average weekly wages, respectively. Technological complexity and its squared term are lagged by one quarter. Patent activity controls are included in all specifications. All regressions include industry, year, and quarter fixed effects. Parentheses report heteroskedasticity-robust t-statistics. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 9: Technological Complexity and Labor Market Outcomes: Heterogeneity by Industry Concentration

Logged Variables	Labor Market Outcomes		
	Employment	Total Quarterly Wages	Average Weekly Wage
	(1)	(2)	(3)
L.High Concentration	-0.063***	-0.076***	-0.004**
	(-10.30)	(-11.68)	(-2.30)
L. Exc Complexity	-0.001	-0.005*	-0.005***
	(-0.28)	(-1.77)	(-5.08)
L.High Concentration ×	0.018***	0.016***	-0.008***
L. Exc Complexity	(4.50)	(3.73)	(-7.49)
L. Exc Complexity^2	0.001	-0.004*	-0.005***
	(0.55)	(-1.66)	(-5.22)
L.High Concentration ×	0.001	0.005	-0.003***
L. Exc Complexity^2	(0.16)	(1.13)	(-2.85)
Patent Activity Controls			
L. Patent Growth	-0.003	0.000	0.002
	(-0.53)	(0.04)	(0.78)
L13. 3-Year Citation Count	0.053***	0.078***	0.018***
	(14.98)	(19.28)	(16.84)
Constant	5.045***	6.818***	0.361***
	(212.83)	(258.61)	(50.51)
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Observations	26,268	26,268	26,268
Adjusted R-squared	0.974	0.969	0.936

Notes. This table reports regressions examining whether the relationship between technological complexity and labor market outcomes differs across industries with different levels of concentration. The dependent variables are the logarithms of employment, total quarterly wages, and average weekly wages, respectively. *High Concentration* equals one for industries with HHI above the 67th percentile of the sample distribution. Technological complexity, the HHI indicator, and their interaction terms are lagged by one quarter. Patent activity controls are included in all regressions. All regressions include 4-digit NAICS industry fixed effects, year fixed effects, and quarter fixed effects. Parentheses report heteroskedasticity-robust t-statistics. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 10: Technological Complexity and Firm Profitability

Logged Variables	Baseline		HHI Heterogeneity	
	ROA	Profit Margin	ROA	Profit Margin
	(1)	(2)	(3)	(4)
L12. High Concentration			0.020**	-0.052***
			(2.10)	(-2.65)
L12. Exc Complexity	-0.015***	-0.030***	-0.019***	-0.041***
	(-4.43)	(-4.65)	(-5.33)	(-5.59)
L12. High Concentration ×			0.014***	0.037***
L12. Exc Complexity			(3.44)	(3.79)
L12. Exc Complexity^2	0.002	-0.003	0.009***	-0.003
	(0.61)	(-0.44)	(2.82)	(-0.46)
L12. High Concentration ×			-0.022***	0.003
L12. Exc Complexity^2			(-4.78)	(0.30)
Patent Activity Controls				
L12. Patent Growth	-0.006	0.032	-0.005	0.031
	(-0.67)	(1.50)	(-0.65)	(1.49)
L25. 3-Year Citation Count	-0.004	-0.021**	-0.004	-0.022**
	(-1.07)	(-2.20)	(-1.13)	(-2.25)
Constant	-0.020	-0.178***	-0.026	-0.161***
	(-0.93)	(-3.00)	(-1.17)	(-2.70)
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Observations	27,533	27,475	27,533	27,475
Adjusted R-squared	0.205	0.291	0.206	0.291

Notes. This table examines the relationship between technological complexity and subsequent firm profitability. Columns (1) and (2) report the baseline specifications using return on assets (ROA) and profit margin as the dependent variables, respectively. Columns (3) and (4) examine whether these relationships vary with industry concentration. *High Concentration* equals one for industries with HHI at or above the 67th percentile of the sample distribution, and zero otherwise. Technological complexity, its squared term, the concentration indicator, their interaction terms, and Patent Growth are lagged by 12 quarters. The 3-Year Citation Count is lagged by 25 quarters. All regressions include 4-digit NAICS industry, year, and quarter fixed effects. Parentheses report heteroskedasticity-robust t-statistics. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

Table 11: Technological Complexity and Market Valuation

	Market Valuation Outcomes				
	BHARs				Innovation
Logged Variables	CAPM	FF3	FF4	DGTW	KPSS
	(1)	(2)	(3)	(4)	(5)
L6. Exc Complexity	-0.007	0.028*	0.013	-0.006	0.081***
	(-0.42)	(1.74)	(0.77)	(-0.35)	(6.15)
L6. Exc Complexity^2	-0.015	-0.004	0.007	-0.017	-0.004
	(-0.98)	(-0.23)	(0.44)	(-1.13)	(-0.34)
Patent Activity Controls					
L6. Patent Growth	0.059***	0.053***	0.055***	0.066***	0.284***
	(3.51)	(3.17)	(3.24)	(3.87)	(22.18)
L19. 3-Year Citation Count	0.030***	0.028***	0.029***	0.025***	0.513***
	(8.02)	(7.51)	(7.59)	(6.56)	(175.48)
Constant	-0.472***	0.089	0.100	-0.223*	1.134***
	(-3.87)	(0.73)	(0.82)	(-1.83)	(13.12)
Industry FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	85,942	85,942	85,942	81,823	62,761
Adjusted R-squared	0.016	0.008	0.007	0.005	0.491

Notes. This table reports regressions examining the relationship between industry technological complexity and firm-level valuation outcomes. Columns (1)-(4) use quarterly buy-and-hold abnormal returns based on the CAPM, Fama-French three-factor, Fama-French/Carhart four-factor, and DGTW characteristic-matched benchmarks, respectively. Column (5) uses the KPSS patent-value measure. Excess Complexity and *Patent Growth* are lagged by six quarters. *3-Year Citation Count* is lagged by 19 quarters to allow the citation window to be fully observed. All regressions include 4-digit NAICS industry, year, and quarter fixed effects. Parentheses report heteroskedasticity-robust t-statistics. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 12: Technological Complexity and Market Valuation: Heterogeneity by Industry Concentration

	Market Valuation Outcomes				
	BHARs				Innovation
	CAPM	FF3	FF4	DGTW	KPSS
Logged Variables	(1)	(2)	(3)	(4)	(5)
L6. High Concentration	-0.011	0.060	0.047	0.090	0.102
	(-0.13)	(0.69)	(0.53)	(1.01)	(1.64)
L6. Exc Complexity	-0.001	0.026	0.012	-0.006	0.074***
	(-0.07)	(1.51)	(0.71)	(-0.33)	(5.40)
L6. High Concentration ×	0.040	0.060	0.055	0.014	0.025
L6. Exc Complexity	(0.78)	(1.19)	(1.08)	(0.27)	(0.71)
L6. Exc Complexity^2	-0.017	-0.003	0.008	-0.019	-0.007
	(-1.03)	(-0.17)	(0.49)	(-1.15)	(-0.60)
L6. High Concentration ×	-0.008	-0.022	-0.017	-0.031	0.094**
L6. Exc Complexity^2	(-0.15)	(-0.39)	(-0.30)	(-0.53)	(2.39)
Patent Activity Controls					
L6. Patent Growth	0.058***	0.053***	0.053***	0.064***	0.280***
	(3.37)	(3.04)	(3.07)	(3.68)	(21.65)
L19. 3-Year Citation Count	0.031***	0.029***	0.030***	0.026***	0.517***
	(7.85)	(7.54)	(7.61)	(6.59)	(173.83)
Constant	-0.460***	0.087	0.094	-0.221*	1.065***
	(-3.74)	(0.71)	(0.77)	(-1.80)	(12.30)
Industry FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	82,074	82,074	82,074	78,095	59,922
Adjusted R-squared	0.017	0.009	0.007	0.005	0.495

Notes. This table examines whether the relationship between technological complexity and market valuation varies with industry concentration. *High Concentration* equals one for industries with HHI at or above the 67th percentile of the sample distribution, and zero otherwise. Excess Complexity, its squared term, the concentration indicator, their interaction terms, and *Patent Growth* are lagged by six quarters. *3-Year Citation Count* is lagged by 19 quarters. Columns (1)-(4) use quarterly buy-and-hold abnormal returns based on the CAPM, Fama-French three-factor, Fama-French/Carhart four-factor, and DGTW characteristic-matched benchmarks, respectively. Column (5) uses the KPSS patent-value measure. All regressions include 4-digit NAICS industry, year, and quarter fixed effects. Parentheses report heteroskedasticity-robust t-statistics. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

Data Appendix

A Conversion from CPC to NAICS

A.1 Conversion of Complexity Score

As mentioned in Broekel (2019), when calculating the NDS for the network of each 4-digit CPC code, 50 nodes are selected, and a subnetwork is drawn by a random walktrap of 150 steps starting from each of the node. These 50 subnetworks are used to calculate the iNDS and the final NDS is the average of the 50 iNDS. This process emphasizes the importance of topologies' diversity rather than the size of the network, so when the case is that multiple 4-digit CPC industries are mapped into one 6-digit NAICS industry, the scale complexity score should not be aggregated.

The probability weight needed for the conversion comes from Nathan et al. (2019) which utilized the 'Algorithmic Links with Probabilities' method and provided direct probabilistic linkages between classification systems and supports joint analysis of patent and economic data. The files that map 6-digit NAICS to 4-digit CPC were used, so for each 6-digit NAICS, the sum of probability weight across all the CPC codes it is mapped to is 1. The original score was calculated under each 6-digit NAICS code i might be mapped to n 4-digit CPC code js with the respective probability weight and original score for each j as $Pr(j|i)$ and C^j_{CPC-4}. Then the 6-digit NAICS code's complexity score should be the sum of all the 4-digit CPC code's scores times the respective weight with the weight summing up to 1 as the equation 16.

$$C^i_{NAICS-6} = \sum_j C^j_{CPC-4} \cdot Pr(j|i) \quad \text{with} \sum_j Pr(j|i) = 1 \tag{16}$$

A.2 Conversion of Patent Count and Citation Data

After calculating the number of patents for each 4-digit CPC code in each year and the count of citations these patents receive in the subsequent one-, two-, and three- year periods, conversion to NAICS is needed for future analysis. The difference from the previous conversion is that the total

patent numbers and citation counts should be preserved after the crosswalk, so equation 16 cannot be applied to the patent and citation number.

As the complexity score, the original patent, or citation count is calculated under the 4-digit CPC level, a crosswalk to the 6-digit NAICS level is necessary for economic analysis. For this conversion, the files used mapped one or multiple NAICS industries to one CPC code, so the sum of weight assigned to these NAICS industries would be 1, and in consequence, the number of patents and citations does not change during the conversion.

For each 6-digit NAICS industry i, it may show up in the mapping into m 4-digit CPC code js with the respective probability weight for each i as $Pr(j|i)$ and the patent or citation number as N^j_{CPC-4}. Then, the 6-digit NAICS code's patent or citation number should be the sum of all the 4-digit CPC code's patent or citation number, times the respective weight, as in equation 17.

$$N^i_{NAICS-6} = \sum_j N^j_{CPC-4} \cdot Pr(j|i) \tag{17}$$

B Conversion between difference versions of NAICS

The crosswalks from CPC to NAICS are only feasible for the first three NAICS versions, which are the NAICS 1997, NAICS 2002, and NAICS 2007, so the conversion between different versions of NAICS is necessary between the recent 4 NAICS versions, which are NAICS 2007, NAICS 2012, NAICS 2017, and NAICS 2022. Once the conversion between different versions of NAICS are available, the conversion from CPC to the later versions of NAICS can be achieved.

B.1 Conversion of Complexity Score

To convert complexity scores between NAICS versions, we use the Census Bureau's concordance files to assign probability weights. For instance, when mapping from 2007 to 2012 codes, all corresponding 6-digit NAICS 2007 entry is equally weighted across the 6-digit NAICS 2012 codes they map to. As a result, the weights for each 2012 code — summed over all its linked 2007 codes

— add up to 1. This ensures that when multiple 2007 codes merge into a single 2012 code, the aggregated complexity score remains on the same scale.

The complexity scores are first calculated or converted to one version. To move from one version to the next, we use equal probability weights from the Census concordance. Specifically, for each 6-digit code i in the new version that maps to m industries j in the old version, we give each link a weight of $\frac{1}{m}$ with the respective complexity score for each j as $C^j_{NAICS_n}$. Then the new version's 6-digit NAICS complexity score is the average weighted of the old version's 6-digit NAICS complexity scores as the equation 18 .

$$C^i_{NAICS_{n+1}} = \frac{1}{m} \sum_j C^j_{NAICS_n} \tag{18}$$

B.2 Conversion of Patent Count and Citation Data

As mentioned in Appendix A.2, we need to keep the total patent and citation counts unchanged during any version conversion, so we can't use Equation 18 for these variables.

The probability weights are also assigned by using the the Census Bureau's concordance files between two adjacent versions. To convert from old version to the new one, equal probability weights are assigned. Specifically, for each 6-digit code j in the old version, it may show up in the concordance with m industries i of the new version, we give each link a weight of $Pr(j) = \frac{1}{m}$ with the respective patent count or citation count for each j as $N^j_{NAICS_n}$. The count for each new code i is then the sum of all linked old counts multiplied by their respective weights. This approach, shown in Equation 19, guarantees that the aggregate number of patents or citations stays the same through every step of the conversion process.

$$N^i_{NAICS_{n+1}} = \sum_j C^j_{NAICS_n} \cdot Pr(j) \tag{19}$$

Table 13: Technological Complexity and Labor Market Outcomes: Robustness to Alternative Lag Structures

	Employment		Total Quarterly Wages		Average Weekly Wage	
	L4	L8	L4	L8	L4	L8
	(1)	(2)	(3)	(4)	(5)	(6)
Exc Complexity	0.016***	0.016***	0.013***	0.014***	-0.006***	-0.005***
	(7.46)	(7.95)	(5.48)	(6.03)	(-9.09)	(-8.14)
Exc Complexity^2	0.008***	0.008***	0.001	-0.000	-0.008***	-0.009***
	(4.08)	(3.95)	(0.34)	(-0.04)	(-11.49)	(-12.36)
Patent Activity Controls						
Patent Growth	-0.003	-0.002	-0.004	-0.005	-0.002	-0.002
	(-0.48)	(-0.45)	(-0.57)	(-0.79)	(-0.65)	(-0.70)
3-Year Citation Count	0.047***	0.038***	0.068***	0.056***	0.015***	0.014***
	(14.65)	(11.37)	(18.61)	(15.02)	(15.74)	(14.69)
Constant	4.942***	5.005***	6.736***	6.837***	0.378***	0.392***
	(234.96)	(236.56)	(289.61)	(292.36)	(59.35)	(62.26)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30,359	29,286	30,359	29,286	30,359	29,286
Adjusted R-squared	0.977	0.978	0.974	0.975	0.930	0.933

Notes. This table reports robustness tests examining the association between technological complexity and labor market outcomes using alternative lag structures. Columns (1), (3), and (5) use technological complexity lagged one year, while Columns (2), (4), and (6) use excess technological complexity lagged two years. The dependent variables are the logarithms of employment, total quarterly wages, and average weekly wages, respectively. Patent activity controls are included in all specifications with lag structures corresponding to each robustness check. All regressions include industry, year, and quarter fixed effects. Parentheses report heteroskedasticity-robust t-statistics. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

Table 14: Technological Complexity and Market Valuation: Robustness to a Twelve-Quarter Lag

	Market Valuation Outcomes				
	BHARs				Innovation
Logged Variables	CAPM	FF3	FF4	DGTW	KPSS
	(1)	(2)	(3)	(4)	(5)
L12. Exc Complexity	-0.040**	-0.030*	-0.036**	-0.039**	0.062***
	(-2.28)	(-1.74)	(-2.10)	(-2.21)	(4.36)
L12. Exc Complexity^2	-0.024	-0.026	-0.017	-0.030*	-0.001
	(-1.50)	(-1.60)	(-1.05)	(-1.81)	(-0.06)
Patent Activity Controls					
L12. Patent Growth	0.027	0.027	0.027	0.025	0.266***
	(1.51)	(1.52)	(1.51)	(1.39)	(18.82)
L25. 3-Year Citation Count	0.021***	0.019***	0.020***	0.019***	0.504***
	(5.29)	(4.93)	(5.12)	(4.64)	(160.47)
Constant	0.000	-0.025	-0.063	-0.098	1.826***
	(0.00)	(-0.28)	(-0.72)	(-1.11)	(28.96)
Industry FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	75,336	75,336	75,336	71,622	54,483
Adjusted R-squared	0.015	0.007	0.006	0.004	0.479

Notes. This table reports robustness regressions examining the relationship between industry technological complexity and firm-level valuation outcomes using a twelve-quarter lag. Columns (1)-(4) use quarterly buy-and-hold abnormal returns based on the CAPM, Fama-French three-factor, Fama-French/Carhart four-factor, and DGTW characteristic-matched benchmarks, respectively. Column (5) uses the KPSS patent-value measure. Excess Complexity and *Patent Growth* are lagged by twelve quarters. *3-Year Citation Count* is lagged by 25 quarters to allow the citation window to be fully observed. All regressions include 4-digit NAICS industry, year, and quarter fixed effects. Parentheses report heteroskedasticity-robust t-statistics. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

Table 15: Technological Complexity and Market Valuation: Heterogeneity by Industry Concentration — Twelve-Quarter Lag

	Market Valuation Outcomes				
	BHARs				Innovation
	CAPM	FF3	FF4	DGTW	KPSS
Logged Variables	(1)	(2)	(3)	(4)	(5)
L12. High Concentration	0.013	0.073	0.043	0.159*	0.253***
	(0.14)	(0.82)	(0.48)	(1.75)	(3.81)
L12. Exc Complexity	-0.033*	-0.024	-0.030	-0.031*	0.049***
	(-1.80)	(-1.34)	(-1.63)	(-1.65)	(3.33)
L12. High Concentration ×	0.049	0.075	0.061	0.017	0.021
L12. Exc Complexity	(0.94)	(1.43)	(1.16)	(0.31)	(0.58)
L12. Exc Complexity^2	-0.029*	-0.034**	-0.027	-0.033*	-0.001
	(-1.73)	(-1.99)	(-1.60)	(-1.91)	(-0.10)
L12. High Concentration ×	-0.018	-0.031	-0.018	-0.060	0.058
L12. Exc Complexity^2	(-0.30)	(-0.53)	(-0.30)	(-1.00)	(1.44)
Patent Activity Controls					
L12. Patent Growth	0.032*	0.030*	0.030	0.028	0.261***
	(1.73)	(1.66)	(1.62)	(1.52)	(18.34)
L25. 3-Year Citation Count	0.020***	0.020***	0.021***	0.019***	0.509***
	(4.99)	(4.87)	(5.06)	(4.62)	(159.17)
Constant	0.024	0.003	-0.035	-0.081	1.745***
	(0.27)	(0.03)	(-0.39)	(-0.90)	(27.55)
Industry FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	72,006	72,006	72,006	68,424	51,990
Adjusted R-squared	0.015	0.007	0.006	0.004	0.485

Notes. This table examines whether the relationship between technological complexity and market valuation varies with industry concentration using a twelve-quarter lag. *High Concentration* equals one for industries with HHI at or above the 67th percentile of the sample distribution, and zero otherwise. Excess Complexity, its squared term, the concentration indicator, their interaction terms, and *Patent Growth* are lagged by twelve quarters. *3-Year Citation Count* is lagged by 25 quarters. Columns (1)-(4) use quarterly buy-and-hold abnormal returns based on the CAPM, Fama-French three-factor, Fama-French/Carhart four-factor, and DGTW characteristic-matched benchmarks, respectively. Column (5) uses the KPSS patent-value measure. All regressions include 4-digit NAICS industry, year, and quarter fixed effects. Parentheses report heteroskedasticity-robust t-statistics. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.