Thematic Concentration and Mutual Fund Performance

John (Jianqiu) Bai, Yuehua Tang, Chi Wan, H. Zafer Yüksel*

August 2025

ABSTRACT

This study examines whether mutual fund managers generate alpha through thematic investment strategies that select stocks poised to benefit from specific themes. Using textual analysis of 10-K filings, we identify stocks' thematic exposures and construct each fund's thematic concentration index (TCI) from its holdings. High TCI funds significantly outperform, with a top-minus-bottom decile spread of 4.26% in annualized four-factor alpha. Managers' thematic expertise is related to their undergraduate field of study. Outperformance arises from superior stock selection rather than theme-related timing, with an informational advantage on firm earnings, particularly in stocks exposed to themes related to their academic background.

Keywords: Mutual Funds, Thematic Investing, Fund Performance, Undergraduate Specialization, Textual Analysis

JEL classification: G11, G23, J24

---

* John (Jianqiu) Bai is from the D'Amore-McKim School of Business at Northeastern University, Boston, MA 02115 and can be reached at j.bai@northeastern.edu; Yuehua Tang, the corresponding author, is from the Warrington College of Business at the University of Florida, Gainesville, FL 32611 and can be reached at yuehua.tang@warrington.ufl.edu; Chi Wan is from the Fowler College of Business, San Diego State University, San Diego, CA 92182 and can be reached at cwan@sdsu.edu. H. Zafer Yuksel is from the College of Business at the University of Rhode Island and can be reached at hzyuksel@uri.edu. We have benefited from comments and suggestions from Clemens Sialm, Vikas Agarwal, Kent Daniel, Shingo Gato, Jerry Hoberg, Zoran Ivković, Wei Jiang, Nitish Kumar, Yukun Liu, Alejandro Lopez-Lira, Linlin Ma, Pedro Matos, Jun Pan, Jinfei Sheng, Lu Zheng, Russ Wermers, and seminar participants at FIU, SAIF, Texas Tech, and University of Rhode Island. We also thank John Dorney, Alimursal Ibrahimov, Boyuan Li, Miles Nathan, Paige Santiago, Yuan Wang, and Fei Yu for excellent research assistance.

Thematic Concentration and Mutual Fund Performance

August 2025

ABSTRACT

This study examines whether mutual fund managers generate alpha through thematic investment strategies that select stocks poised to benefit from specific themes. Using textual analysis of 10-K filings, we identify stocks' thematic exposures and construct each fund's thematic concentration index (TCI) from its holdings. High TCI funds significantly outperform, with a top-minus-bottom decile spread of 4.26% in annualized four-factor alpha. Managers' thematic expertise is related to their undergraduate field of study. Outperformance arises from superior stock selection rather than theme-related timing, with an informational advantage on firm earnings, particularly in stocks exposed to themes related to their academic background.

***Keywords*:** Mutual Funds, Thematic Investing, Fund Performance, Undergraduate Specialization, Textual Analysis

***JEL classification*:** G11, G23, J24

1. Introduction

Understanding the sources of investment skills of professional money managers is an enduring topic in financial economics. One increasingly popular yet still under-studied strategy is thematic investing—actively tilting portfolios toward firms whose future cash flows are unusually sensitive to broad structural trends such as generative-AI diffusion, population aging, or the transition to a low-carbon economy.[1] Thematic investing starts with a top-down hypothesis (i.e., a "theme") about a technological, regulatory, demographic, or social shift poised to reshape corporate risks and opportunities, and then deliberately selects stocks whose cash flows are positioned to benefit from or withstand that shift. Because themes cut across conventional region, sector, and style boxes, a single "AI infrastructure" portfolio can span semiconductor companies, cloud-software firms, and power-grid operators. This transcending scope sets thematic investing apart from traditional factor, sector, or geographic approaches and allows managers to capitalize on proprietary insights that span market capitalizations, industries, and regions.

Although researchers have extensively examined style-, region-, and sector-based investment strategies, we know little about whether active fund managers can generate alpha by concentrating their portfolios on themes in which they hold specialized expertise. This paper closes that gap in three ways. First, we develop a novel textual analysis method that measures each stock's exposure to themes and aggregates those exposures to the portfolio level, sidestepping reliance on fund marketing materials or ex post industry labels. Second, we show that funds with tightly focused thematic portfolios generate economically and statistically significant abnormal returns relative to their less concentrated peers. Third, we find that thematic investment skill is linked to managers' human capital, most notably their undergraduate field of study. Together, our findings provide novel evidence that thematic expertise is a distinct and important dimension of active-management skills.

[1] Recently, thematic investing has attracted attention from practitioners and academics as investment products related to thematic investing have gained popularity among investors. See, e.g., https://insight.factset.com/thematic-investing-catches-fire; https://www.cnbc.com/2021/06/29/thematic -investing-has-taken-off-how-to-capitalize-on-trends-.html.

Identifying themes ex ante requires a disclosure source that is comprehensive, comparable, and forward-looking. We analyze the "Risk Factors" section (Item 1A) of Form 10-K filings and apply Latent Dirichlet Allocation (LDA), an unsupervised machine learning algorithm, to the complete corpus of public firm reports.[2] Three attributes make Item 1A ideal for extracting themes across firms. First, it is forward-looking. Regulation S-K requires firms to disclose "the most significant factors" that could impact future performance, prompting managers to discuss factors such as supply-chain reshoring, climate regulation, and cybersecurity before they materially affect results. Second, the section's highly standardized format yields a large, homogeneous text corpus on which topic models such as LDA perform reliably. Third, the litigation risk for omissions encourages thorough disclosure, ensuring that even niche, yet economically meaningful, themes appear in the text.

While Item 1A is titled "Risk Factors", two attributes make this section a suitable setting to capture investment themes. First, the common factors or forces identified by many firms often hurt some firms but benefit others. For instance, a new climate-policy mandate is a risk to carbon-intensive producers but an opportunity for suppliers of low-carbon technology. Likewise, the diffusion of generative-AI tools threatens some labor-intensive service firms while rewarding semiconductor designers and cloud providers. Second, firms frequently disclose certain factors as "risks" while simultaneously undertaking proactive measures to mitigate them. In particular, firms that disclose a given "risk" often invest in technologies or business model adaptations to mitigate it, which means that common Item 1A factors can have heterogeneous effects across firms depending on their strategic responses. Consequently, when LDA groups these disclosures, it produces cross-firm topics that we label "themes," because they can encompass both downside risk and upside potential from an investment perspective. Sophisticated investors who understand how a particular factor or

[2] Beginning in 2005, Item 105 of Regulation S-K requires firms to disclose in words in their 10-K Item 1A the most significant factors that make investing in the company speculative or risky. We analyze this textual data using the LDA model proposed by Blei, Ng, and Jordan (2003) to extract underlying semantic themes across the stock universe, with each topic identified by the LDA model representing an investment theme. In our baseline analysis, we set the number of topics to be 100, though our results are not sensitive to this choice. LDA has been widely used to analyze various types of textual data in the literature (e.g., Campbell et al., 2014; Dyer, Lang, and Stice-Lawrence, 2017; Hansen, McMahon, and Prat, 2018; Hanley and Hoberg, 2019; Bybee et al., 2021).

theme will differentially affect firms can overweight future winners and underweight losers, translating superior assessment of idiosyncratic risks into abnormal returns.

We aggregate the stock-level theme scores to the fund level and examine a comprehensive sample of more than 2,400 actively managed U.S. equity mutual funds from 2006 to 2023. In the spirit of Kacperczyk, Sialm, and Zheng (2005), we construct a thematic concentration index (TCI) at the fund level by summing the squared differences between a fund's portfolio weight in each theme and that of the market portfolio. Because TCI is built from actual holdings using a bottom-up approach (rather than fund marketing labels), it captures thematic investing, broadly defined, by selecting stocks based on their thematic exposures. As a fund's holdings become more concentrated in stocks influenced by certain investment themes, the TCI value increases. We hypothesize that skilled managers with an information edge in particular themes will have a portfolio concentrated in stocks positively influenced by these themes and deliver superior fund performance.

To examine how thematic concentration relates to fund performance, we sort mutual funds into deciles based on their thematic concentration index. We find a strong, positive relation between TCI and fund performance. For instance, funds in the top TCI decile outperform those in the bottom decile by an average of 4.26% per year in net Carhart (1997) four-factor alpha.[3] Fama–MacBeth (1973) regressions controlling for fund size, age, expense ratio, turnover, and other characteristics confirm this positive relation. The TCI coefficient remains positive and highly significant whether abnormal performance is measured by Carhart four-factor alpha or by the Daniel, Grinblatt, Titman, and Wermers (1997, DGTW) characteristic-selectivity (CS) measure. Results are also robust when we use (i) Sharpe ratio, (ii) information ratio, (iii) six-factor alpha (i.e., Fama and French (2015) five-factor plus the UMD factor), (iv) the conditional alpha measure of Cederburg et al. (2018), or (v) the firm-identified risk model of Lopez-Lira (2023) to evaluate fund performance, suggesting that the outperformance of high-TCI funds is not simply compensation for omitted risk factors.

[3] We obtain similar results if we examine gross fund returns (i.e., net return plus expense ratio).

Importantly, the TCI measure captures information distinct from existing proxies for managerial skill and fund activeness, including the industry concentration index (ICI, Kacperczyk, Sialm, and Zheng, 2005), active share (Cremers and Petajisto, 2009), R^2 (Amihud and Goyenko, 2013), and the offshore concentration index (OCI, Bai, Tang, Wan, and Yüksel, 2022). After controlling each of these metrics, the positive relation between TCI and fund performance remains robust. Conceptually, TCI also differs from ICI of Kacperczyk, Sialm, and Zheng (2005): stocks within the same industry can load on multiple themes, while stocks from different industries may share a common theme. Consistent with this idea, the sample correlation between TCI and ICI is below 0.31.[4]

Despite the positive relation we document between TCI and fund performance, one could argue that the pattern could arise because stocks with high thematic concentration earn higher future returns, and high TCI funds simply hold more of those stocks, exploiting a return anomaly rather than displaying thematic investment skill. To test this possibility, we conduct both portfolio sorts and Fama-MacBeth regressions to examine whether stocks with greater thematic concentration deliver higher returns. We find no evidence of such an anomaly. Hence, the outperformance of high TCI funds cannot be attributed to harvesting a stock-level anomaly associated with thematic concentration.

High TCI funds' alpha can be generated from superior market timing (e.g., overweighting themes that later deliver higher returns) or better stock selection. To test this idea, we add a theme-timing variable (i.e., the return implied by a fund's theme weights without within-theme stock selection) to our Fama-MacBeth regressions. Its coefficient is small and insignificant, while the coefficient on TCI remains unchanged. Thus, high TCI outperformance of high TCI funds appears to be driven by their stock selection rather than theme-level market-timing ability. Furthermore, our analysis of fund trades (i.e., changes in holdings) suggests that high TCI managers possess an informational advantage regarding

[4] Moreover, the bivariate distributions of TCI and ICI in our sample show that within each category (e.g., quintile) of TCI or ICI, there is considerable variation in the other measure. For instance, funds in the top ICI quintile can be classified into any quintile based on TCI (with the majority not being ranked in the top TCI quintile), and vice versa. Thus, it is important to distinguish between the two dimensions of active management.

firm fundamentals such as corporate earnings. These managers earn a significantly larger spread between the abnormal returns of stocks purchased and those sold than do low-TCI managers, and their trading activity positively predicts future earnings surprises.

If high thematic concentration reflects a manager's information advantage in certain investment themes, a natural question is where that advantage originates. One plausible source is the manager's educational background, particularly the field of undergraduate study. Prior work shows that educational training plays a critical role in generating performance in the asset-management industry.[5] Managers with bachelor's or higher degrees in economics, business, science, technology, or related fields are more likely to possess the knowledge and skills needed to gain an edge in themes tied to those disciplines.[6] Moreover, theoretical work by van Nieuwerburgh and Veldkamp (2009) shows that an initial advantage in a given field promotes further specialization through continuous learning and information acquisition, thereby translating into profitable, theme-concentrated investment strategies. In their model, specialization leads to concentrated portfolios tilted toward stocks where investors already hold an information edge, delivering higher risk-adjusted returns.

In our final set of analyses, we investigate whether a fund manager's educational background is related to expertise in specific investment themes. We first compile a comprehensive dataset on the undergraduate majors of managers. Using textual analysis, we then assess whether a theme aligns with a manager's major by comparing the theme's keywords with course descriptions from a database of more than 81,000 undergraduate courses across 235 disciplines. Three findings support our conjecture. First, a difference-in-differences test around portfolio manager turnovers shows that a fund's TCI in themes related to the departing manager's major experiences a significant decline, while its TCI in themes related to the new manager's major significantly increases after degree-changing turnovers. Second, a fund's TCI in themes related to its manager's major has particularly

[5] See, e.g., Chevalier and Ellison (1999), Cohen, Frazzini, and Malloy (2008), Li, Zhang, and Zhao, (2011), Chaudhuri, Ivković, Pollet, and Trzcinka (2020).

[6] For instance, a portfolio manager with a computer science bachelor's degree is likely to develop expertise in that area and thus possess proprietary views about investment themes related to disruptive changes in information technology such as big data and artificial intelligence.

strong predictive power for future performance, suggesting that managers' educational background is likely a key contributor to their specialization in certain themes and thematic investment skills. Finally, an analysis of fund trades shows that the buy-minus-sell difference in stock performance and future earnings surprises for high-TCI funds is significantly larger when those trades involve stocks exposed to themes connected to the manager's major. These results suggest that field-specific training facilitates further skill development and specialization in related themes among high-TCI managers, enabling them to construct more theme-concentrated portfolios and deliver superior performance.

Taken together, our findings suggest that fund managers possess theme-specific expertise and tailor their portfolios accordingly. Interpreting complex economic, technological, political, and social shifts and assessing their impact on individual firms' valuations represent investment opportunities for managers who can leverage specialized knowledge to derive an edge in stocks exposed to particular long-term trends. Importantly, we also show that thematic expertise correlates with a manager's educational background, especially the field of undergraduate study, echoing van Nieuwerburgh and Veldkamp's (2009) prediction that an initial (disciplinary) advantage induces continued learning and deeper specialization.

Our paper adds to several strands of literature. First, we contribute to the literature on the sources of managerial ability to generate alpha. Prior studies show that managers with an information edge concentrate their portfolios in specific industries, local firms, or firms with overseas exposure (Coval and Moskowitz, 2001; Kacperczyk, Sialm, and Zheng, 2005; Huang and Kale, 2013; Choi et al., 2017; Cici et al., 2018; Bai et al., 2022).[7] We uncover a new source of fund manager skill: an edge derived from expertise in particular investment themes. The magnitude of the performance spread across thematic concentration levels highlights the importance of thematic investment, a fast-growing yet understudied strategy, in asset management. More broadly, our evidence supports theoretical predictions that

[7] Other studies on mutual fund managers' ability to generate alpha include, among others, Jiang, Yao, and Yu (2007), Kacperczyk and Seru (2007), Cohen, Frazzini, and Malloy (2008), Kacperczyk, Sialm, and Zheng (2008), Pool, Stoffman, and Yonker (2015), Kempf, Manconi, and Spalt (2017), Hwang, Titman, and Wang (2018), Jiang and Zheng (2018), Hoberg, Kumar, and Prabhala (2018), and Busse et al. (2021).

asymmetric information leads to disparate returns among market participants (e.g., Grossman and Stiglitz, 1976, 1980; Levy and Livingston, 1995) and investor specialization persists over time (e.g., van Nieuwerburgh and Veldkamp, 2009). Our study also adds to the literature on active portfolio management by introducing a new measure that exploits novel textual information from corporate filings and is distinct from the existing skill proxies.[8]

Our paper is also related to the recent ETF study by Ben-David, Franzoni, Kim, and Moussawi (2023). They show that 95 thematic ETFs underperform in their first five years because sponsors exploit investor attention by launching products when the underlying stocks are overvalued. Our analysis differs in methodology, sample, and research question. First, we study whether thematic investing skill exists in active mutual funds and, if so, what its sources are, whereas they focus on investor attention in passive ETFs, where managerial skill plays no role. Second, we build bottom-up measures of funds' thematic exposures by applying a textual-analysis approach, while they rely on the ETFs' own marketing labels. Because of these fundamental differences, we find that skilled active managers use thematic investment to gain an edge, and high-TCI funds deliver superior performance.

Lastly, our study contributes to the literature on how managerial characteristics shape investment performance. Prior research shows that educational background plays a pivotal role in managerial skills in the asset management industry (e.g., Chevalier and Ellison, 1999; Cohen, Frazzini, and Malloy, 2008; Li, Zhang, and Zhao, 2011; Chaudhuri et al., 2020). We add new evidence that a manager's undergraduate field of study cultivates expertise in thematically related investments. In doing so, we highlight an important channel through which educational attainment translates into portfolio management skills. Our evidence suggests that undergraduate training does not directly confer an investment edge; rather, it provides a foothold that managers can build on through ongoing learning and specialization in related areas.

[8] See, e.g., Cremers and Petajisto (2009), Amihud and Goyenko (2013), Doshi, Elkamhi, and Simutin (2015), and Cremers et al. (2016).

2. Hypothesis Development

Active mutual fund managers constantly seek an edge to generate alpha. Prior work has shown that active fund managers tend to hold concentrated portfolios that take advantage of their information edge in specific industries (Kacperczyk, Sialm, and Zheng, 2005), sectors in which they have worked previously (Cici et al., 2018), countries (Choi et al., 2017), firms located nearby (Coval and Moskowitz, 2001), firms with vertical supply chain relationships (Huang and Kale, 2013), and firms with offshore operating activities (Bai et al., 2022). A common theme in these studies is that for managers with an information advantage in certain areas, concentrating their portfolios on stocks where they have an edge is optimal as it delivers superior performance, consistent with portfolio theory with asymmetric information (e.g., Levy and Livingston, 1995; Van Nieuwerburgh and Veldkamp, 2009, 2010).

This paper advances this line of economic inquiry by exploring whether certain fund managers can attain a competitive edge through an important, yet distinct strategy from traditional approaches: thematic investing. Skilled managers with domain expertise can better understand how a given force differentially affects firms and therefore overweight likely winners while underweighting potential losers. We hypothesize that managers with an information edge in specific investment themes will hold portfolios concentrated in stocks that benefit from those themes, thereby delivering superior risk-adjusted performance. If this conjecture is correct, fund performance should increase with a portfolio's thematic concentration. Thus, our first hypothesis is:

H1: Funds with a higher thematic concentration index (TCI) deliver a higher risk-adjusted return.

Since thematic concentration reflects fund managers' expertise in specific investment themes, it is crucial to understand the underlying sources of such skills. There could be various potential explanations for why fund managers possess theme-related specialization. Themes and trends can arise from a wide array of factors and fields, including economic, technological, social, or political forces. Considering the diversity of themes spanning many different disciplines, we focus on the influence of educational background, examining whether a manager's field of undergraduate study (i.e., the point at which most professionals first specialize) helps explain theme-related skill.

There exists a rich literature that examines the impact of education on various economic outcomes. Chevalier and Ellison (1999) find that fund managers who attend undergraduate institutions with higher average student SAT scores tend to achieve higher returns. In the education literature, Allmendinger (1989) contends that the education system and educational attainment significantly influence long-term labor market outcomes, while others argue that undergraduate field of study plays a crucial role in determining students' subsequent career choices and line of work (e.g., Kim and Kim, 2003; Van de Werfhorst, 2004).

We hypothesize that fund managers' expertise in specific themes is related to their educational background, particularly their undergraduate field of study, which provides an initial advantage in themes related to that discipline. While the exposure of an undergraduate study is unlikely to directly translate to investment skills, it offers the fundamental knowledge in and familiarity with that field that facilitate further expertise development later in their career, if desired. For instance, a manager with a science or engineering degree is more likely to have the knowledge and skills to gain an edge when analyzing firms exposed to disruptive technological changes. Importantly, theoretical work by van Nieuwerburgh and Veldkamp (2009) predicts that an initial informational edge prompts continued specialization through additional learning and information acquisition, leading to profitable, theme-concentrated investment strategies. To test this idea, we split the TCI measure into a component related to a manager's undergraduate major and a residual component, using textual analysis to measure the similarity between each theme's keywords and the course descriptions of 235 academic majors. If a manager's informational advantage truly stems from formal training, the major-related TCI component should display stronger predictive power for future performance than the unrelated component.

H2: The portion of thematic concentration index (TCI) that is related to a manager's undergraduate degree has a larger effect on funds' risk-adjusted return.

3. Data, the Construction of TCI, and Summary Statistics

In this section, we first discuss the data sources and the construction of the thematic concentration index. We then present summary statistics of and correlation structures between the key variables.

3.1. Data Sources

We construct our data set from several sources. First, we obtain fund returns and characteristics from the CRSP survivorship-bias-free mutual fund database. We then use WRDS MFLINKS to merge the CRSP fund data with the mutual fund holdings data from Thomson-Reuters. Following the literature (e.g., Kacperczyk, Sialm, and Zheng, 2008), we filter out balanced, bond, money market, international, sector, and index funds, and focus primarily on actively-managed domestic equity mutual funds.[9] Moreover, we apply the following selection criteria: We remove funds with less than 10 stocks in order to compute a meaningful fund-level measure of thematic concentration; remove the first two years of return data to eliminate incubation bias (Evans, 2010), and exclude funds with total net assets (TNA) less than $15 million. For funds with multiple share classes, we follow Wermers (2000) and compute value-weighted fund characteristics, except for fund age, which is based on the oldest share class. Finally, we obtain information on fund managers' names and their educational background from Morningstar. In addition, we also collect a comprehensive data set on the majors of managers' undergraduate studies from their LinkedIn pages. We use this manager-level data in our analysis that links managers' educational background to their expertise in certain investment themes. We obtain firms' 10-K filings from the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database.

---

[9] Specifically, we use funds with the Lipper objectives codes of: EIEI, G, LCCE, LCGE, LCVE, MCCE, MCGE, MCVE, MLCE, MLGE, MLVE, SCCE, SCGE, SCVE, CA, EI, G, GI, MC, MR, SG. If Lipper objective codes are not available, we select funds with Strategic Insights codes of AGG, GMC, GRI, GRO, ING, SCG. If a fund has neither the Lipper nor the SI objective, then we use the Wiesenberger Fund Type Code to select funds with the following objectives: G, G-I, AGG, GCI, GRI, GRO, LTG, MCG, SCG. If none of these objectives are available and the fund has a "CS" policy code, then the fund is included. These are common filters used in the prior literature (e.g., Kacperczyk, Sialm, and Zheng, 2008; Huang, Sialm, and Zhang, 2011; Doshi, Elkamhi, and Simutin, 2015; and Hoberg, Kumar, and Prabhala, 2018). Index funds are identified based on their names. We also manually remove from our sample the index funds that are misclassified as active funds.

3.2 Firm's Narrative Risk Disclosures (NRD), LDA, and Themes

One of our main objectives is to develop a new, text-based method for quantifying a fund's exposure to investment themes and then to link those exposures to fund performance. Neither fund managers nor regulators require mutual funds to disclose the themes underlying their trades. We therefore infer themes from corporate disclosures that affect future firm value and then aggregate these stock-level exposures up to the portfolio level.

A key challenge is to locate a source of firm-specific, forward-looking information that is comprehensive, comparable, and sufficiently granular to support large-scale text analysis. We therefore turn to Item 1A ("Risk Factors") in annual 10-K filings and apply Latent Dirichlet Allocation (LDA) to analyze the complete corpus of public firms. Three features make Item 1A uniquely suitable for this purpose. First, it is forward-looking. Regulation S-K requires firms to discuss the "most significant factors" that could impact future performance, compelling managers to disclose emerging opportunities and threats before they register in earnings.[10] Second, the risk-factor section follows a highly standardized format, giving topic models a rich yet homogeneous linguistic substrate on which to operate. Third, the possibility of shareholder litigation for omitted risks encourages firms to be exhaustive, ensuring that even niche but economically consequential themes are captured in the text. Because of these unique features, the theme-extraction approach based on Item 1A disclosures is more effective than relying on other textual sources such as the Management's Discussion and Analysis (MD&A, item 7) in 10-K filings, earnings conference calls, or news articles.[11]

Also, prior research has confirmed that Item 1A disclosures carry meaningful information. Campbell et al. (2014) show that firms in riskier environments list more firm-specific risk factors, while Brown et al. (2018) find that companies promptly revise these

[10] Specifically, Item 105 of Regulation S-K requires firms to disclose in words in Item 1A of their 10-K fillings the material factors that make investing in the company speculative or risky. Form 10-K instructions can be found at https://www.sec.gov/about/forms/form10-k.pdf.

[11] Previous literature also highlights the informativeness of Management's Discussion and Analysis (MD&A, Item 7) of 10-K filings (e.g., Hoberg and Lewis, 2017; Ball, Hoberg, and Maksimovic, 2015). In robustness analysis, we apply LDA to analyze MD&A and repeat our analysis. Our results remain similar.

disclosures after the Securities and Exchange Commission (SEC) comment-letter scrutiny. Consistent with their informativeness, the factors disclosed in Item 1A matter for stock returns (Campbell et al. 2019; Lopez-Lira 2023).

The choice of method for analyzing firms' Item 1A content also requires careful consideration. Firms operate in distinct industries, have unique products and services, face different clientele, and likely operate in different geographic areas. As a result, a systematic analysis of the content included in different firms' narrative risk disclosures seems to be a daunting task. Moreover, the average length of firm-level Item 1A is about 5,000 words (Campbell et al., 2014) and continues to increase, making the understanding and interpretation of related information at a large scale a highly demanding task. To circumvent this issue, we employ LDA, a topic modeling technique in the field of natural language processing. An unsupervised Bayesian linguistic tool, LDA is conceptually similar to factor analysis but is applicable to text (Blei, Ng, and Jordan, 2003). It can be viewed as a dimension reduction tool that extracts topics present in a collection of documents, and in each document, infers the proportion of discussion dedicated to each topic.[12] Because LDA discovers topics without researcher intervention, it is well suited to identifying each stock's exposure to semantic themes that firms themselves deem important for their future performance.

One key result of the LDA procedure is a set of semantically coherent topics that can be mapped onto investable themes and evolve over time. While Item 1A is titled "Risk Factors," the disclosed "risks" are firm-specific or idiosyncratic in nature. Two features make Item 1A a suitable setting to capture investment themes. First, the common forces disclosed (e.g., carbon policy, generative-AI adoption, supply-chain reshoring) could move a firm's cash

[12] LDA has been widely used to analyze textual information in the literature. For example, several studies have used LDA to analyze the Item 1A "Risk Factor" section (e.g., Hanley and Hoberg, 2019; Lopez-Lira, 2023; Campbell et al., 2014; Bao and Datta, 2014). In addition, Huang, Lehavy, Zang, and Zheng (2018) analyze conference call transcripts and analyst reports using LDA to investigate the financial intermediary role of analysts in capital markets; Hoberg and Lewis (2017) and Ball, Hoberg and Maksimovic (2015) apply this technique to Management's Disclosure and Analysis (MD&A); Dyer, Lang, and Stice-Lawrence (2017) apply LDA to the entire 10-K to analyze trends in annual reports. In addition, Hansen, McMahon, and Prat (2018) use LDA to analyze FOMC transcripts to identify policy-related topics, Bybee et al. (2021) use this technique to analyze the structure of business news captured by Wall Street Journal articles, and Liu, Sheng, and Wang (2021) use LDA to analyze the Initial Coin Offering (ICO) whitepapers to capture technological sophistication.

flows away from the status quo positively or negatively. For example, a new climate regulation could hurt carbon-intensive producers but benefit suppliers of low-carbon technology. Second, firms often disclose certain factors as "risks" while at the same time taking proactive actions to mitigate them (e.g., hedging, investing in new technologies, or adapting business models). As a result, common Item 1A factors can have different effects across firms depending on their strategic responses. For these two reasons, when LDA groups these disclosures, it produces cross-firm topics that we label investment themes because they can entail either downside risk or upside exposure from an investment perspective. As Jensen (1968) pointed out, to earn positive alpha, a manager must deliberately select stocks whose idiosyncratic return component is expected to be positive, i.e., take non-systematic bets that pay off. In our setting, a skilled manager who understands how a particular force or theme will differentially affect firms can overweight future winners and underweight losers, translating their superior interpretation of idiosyncratic risks into alpha.

3.3 Implementation of LDA

Essentially, LDA reduces the dimensionality of each document from thousands of words to a distribution of topics; each topic is then mapped to a cluster of keywords. We calibrate LDA to identify a large spectrum of semantic topics that capture firms' Item 1A profiles. We accomplish this by (i) focusing on bigrams (i.e., two adjacent words), the meaning of which is often less ambiguous than that of individual words (Aiken and Lee, 2022); and (ii) annually updating topics to gauge the dynamic evolution of a firm's profile.

3.3.1. Use of Bigrams to Identify Semantic Themes

Conventional unigram LDA is a "bag of words model" that ignores the order of words in a given corpus. To better capture the semantic themes discussed in the text, we extract topics based on bigrams (i.e., pairs of adjacent words) rather than unigrams. To illustrate the usefulness of bigrams in extracting topics, consider "capital expenditures". Taken separately, the context of "capital" and "expenditures" can generate spurious topics, especially in the financial context. While some words may suffice individually (such as "bankruptcy"), combining two words would generate clearer and more coherent topics.

3.3.2. Selecting the Number of Semantic Themes

The key manual input in the implementation of LDA is the total number of topics in the corpus, which depends on the researchers' objective: for example, Lopez-Lira (2023) chooses 25 in his analysis to manually identify several key topics that systematically affect many firms. Dyer, Lang, and Stice-Lawrence (2017) suggest that the corpus of 10-Ks, including Item 1A and all other sections, can be classified into 150 topics.

We determine the number of topics by fitting LDA models for a range of candidate values and computing the perplexity for each choice. We then employ the perplexity score to guide us in determining the total number of topics following the prior literature (e.g., Blei, Ng, and Jordan, 2003; Dyer, Lang, and Stice-Lawrence, 2017). Specifically, given a pre-specified number of topics, perplexity diagnoses the performance of LDA estimated using the a randomly selected 90% subset of documents to predict the topic mixtures of the remaining held-out documents (the "validation set").[13] As an example, Figure 1 presents the perplexity score of bigram LDA with the number of topics varying from 25 to 150 estimated using all firms' NRD in 2009 and 2013 (Panels A and B, respectively). The perplexity score decreases as the number of topics increases, indicating better generalization of topics obtained from the estimation dataset to the validation data. However, as the number of topics increases further, the improvement in model fit diminishes, often at the expense of a loss in topic interpretability (Chang et al., 2009; Dyer, Lang, and Stice-Lawrence, 2017).

[Insert Figure 1 Here]

Given this tradeoff, we choose 100 topics (often referred to as the "elbow" point, where the rate of perplexity change begins to level off) to implement our baseline analysis. Beyond

[13] Perplexity is a widely used evaluation metric in probabilistic language modeling. It is defined as the exponential of the negative normalized log-likelihood of the test set under the fitted model. And intuitively, perplexity is related to a monotonic transformation of the (average) log-likelihood per word in the held-out test sample. Following Dyer, Lang, and Stice-Lawrence (2017), we estimate the model on 90% of the data and use a random hold-out sample of 10% as the testing data to calculate perplexity.

the elbow point, additional topics yielded diminishing improvements in perplexity. This choice balances statistical performance with topic interpretability and is also in line with Dyer, Lang, and Stice-Lawrence (2017). It assumes that there are 100 pertinent topics firms discuss annually in aggregate and is consistent with our goal to accommodate the spectrum of topics in the NRD collection, which may span various technological, regulatory, demographic, or social factors. In robustness checks, we vary the number of topics from 25 to 150 and find that our results remain similar.

3.3.3. Annual Update of Firm-level Semantic Themes

The SEC requires firms to revise their Item 1A annually to provide up-to-date risk-related information to capital market participants. Hence, to ensure that our extracted topics capture a firm's dynamically evolving profile, we separately estimate bigram LDA on all firms' Item 1A in each year, effectively updating the topics annually to track the dynamic risk environment that each firm faces. Therefore, the topics extracted by LDA from Item 1A reflect the risks identified by the firms in that year.

This procedure produces two outputs for each firm annually: the first data set (i.e., topics) produces a set of bigrams and their associated frequencies that make up each topic. The second data set (i.e., topic loadings) describes the distribution of topics discussed in a firm's NRD. In essence, topic loadings indicate the relative importance of each topic in a firm's NRD. When the loading on a topic is closer to one, it reflects the high relevance of that particular factor or theme for a firm. Conversely, a smaller topic loading implies a low prevalence of the corresponding topic discussed by the firm in its Item 1A.

In June of each year t from 2006 to 2023, LDA identifies 100 topics that best represent the distribution of topics that the universe of public firms has discussed in their Item 1A in the previous fiscal year ending in year t-1. To identify the relevant exposure of a particular firm in a given year, we extract the probabilistic distribution of those topics present in the firm's Item 1A. Specifically, the distribution of topics in firm i's NRD is presented as $T_i = (T_{i,1}, T_{i,2}, \dots, T_{i,k}, \dots, T_{i,100})'$, where its k^{th} element, $T_{i,k}$, gauges the relative importance of topic k in firm i's Item 1A disclosure in a particular year. Elements of T_i sum up to one. Both the value of $T_{i,k}$ and the corresponding topic are updated annually, capturing the dynamic evolution of firms' environments. The year subscript is omitted for brevity.

3.3.4. Illustrative Examples of LDA Topics

To illustrate the extracted topics using bigram LDA, in Figure 2, we graph word clouds to visualize six topics identified from the corpus of Item 1A disclosures in 10-K filings across different years: three from 2010, one from 2015, and two from 2023.

[Insert Figure 2 Here]

Word Cloud 1 is on climate change and greenhouse gas. This important topic would be difficult to quantify using firm fundamentals but is well captured by our methodology. Word Cloud 2 concerns technological innovation and competition (i.e., patent applications and protections). Word Cloud 3 is on an increasingly important topic—artificial intelligence and its potential disruptive impact on firms and businesses. Word Cloud 4 is related to discussions about exchange rates and foreign currencies. Given that nearly half of the total sales of S&P 500 firms are generated in foreign markets, it is not surprising that this risk topic is discussed by many firms. Word Cloud 5 is about the rules and regulatory requirements. It focuses on compliance and legal obligations that affect a wide range of companies and sectors. Word Cloud 6 is about renewable energy, illustrating a thematic focus on renewable energy investments and the related costs and valuation issues.

These examples demonstrate that the topics extracted through LDA could represent themes crucial for stock returns across style, sector, and region, on which sophisticated investors can capitalize and profit from their proprietary insights.

3.4 Construction of Fund-level Thematic Concentration Index (TCI)

Based on the firm-level distribution of exposure to various semantic themes, we aggregate across all firms in a fund's portfolio to construct a fund-month level metric that gauges the concentration of a fund's portfolio across themes. In June of each year t over the sample period from 2006 to 2023, we combine the LDA-identified topics and firms' loadings, derived from Item 1A disclosures in public firms' 10-K filings for the previous fiscal year ending in year $t-1$, with the latest fund holdings data to calculate the monthly thematic concentration measure from July of year t to June of $t+1$. This approach ensures at least a six-month gap between a firm's fiscal year-end and the use of the topic distribution identified from its Item 1A disclosure in our fund-level calculation, thereby guaranteeing that the information in 10-K filings is publicly available to capital market participants. The monthly TCI measure is based on each fund's portfolio holdings observed at the end of each calendar quarter. For each set of holdings at a given quarter-end, the measure is computed for the subsequent three months. For example, the March quarter-end holdings of a fund are used to calculate the TCI measure for April, May, and June (i.e., monthly changes in stock prices changing portfolio weights and thus the fund-level TCI measure).

As a direct application of the Herfindahl-Hirschman index (HHI), we consider a similar construction to capture the extent to which a fund's portfolio is concentrated in the space spanned by the full list of semantic themes. Specifically, at the fund level, the value-weighted theme distribution for fund p across the 100 themes is as follows:

$$V_p = \left(v_{p,1}, \dots, v_{p,k}, \dots v_{p,100}\right)' = \left(\sum\nolimits_{i \in P} \omega_i \times T_{i,1}, \dots, \sum\nolimits_{i \in P} \omega_i \times T_{i,k}, \dots\right)', \tag{1}$$

where the combined exposure of fund p to theme k, $v_{p,k}$, is simply the weighted sum of individual stocks' exposure to that theme (i.e., $v_{p,k} = \sum_{i \in P} \omega_i \times T_{i,k}$). Similarly, we define the value-weighted theme distribution of the market portfolio (i.e., all stocks in CRSP/Compustat universe) as $V_M = \left(v_{M,1}, \dots, v_{M,k}, \dots v_{M,100}\right)'$. Then, similar to the ICI measure of Kacperczyk, Sialm, and Zheng (2005), the market-adjusted HHI-type thematic concentration measure at the fund level is calculated as follows:

$$TCI_P \equiv \left(V_p - V_M\right)'\left(V_p - V_M\right) = \sum\nolimits_{k=1}^{100} \left(v_{p,k} - v_{M,k}\right)^2. \tag{2}$$

This thematic concentration index, TCI_P, increases as fund p becomes more concentrated in firms with significant exposure to specific semantic themes, thereby deviating from the composition of the market portfolio. Throughout our empirical analysis, we multiply TCI by 100 to ease presentation.

3.5 Summary Statistics, and TCI's Correlations with Fund Characteristics

Panel A of Table 1 reports summary statistics at the fund-month level for our final sample during the period from 2006Q3 to 2023Q4. Our sample consists of 2,449 distinct funds, for which the average TNA and fund age are approximately $1,527 million and 16.8 years, respectively. The average annual expense ratio is 1.01%, while the average portfolio turnover is about 67.6%. The monthly Carhart (1997) four-factor alpha (i.e., the intercept of the four-factor model estimated using 36 monthly net fund returns) for the sample is -0.082%. All continuous variables are winsorized at the 1% and 99% levels to reduce the impact of outliers.

[Insert Table 1 Here]

Turning to *TCI*, its average is 0.131 with a standard deviation of 0.053. Moreover, we find that the managers of growth funds tend to have higher thematic concentration than those of blend or value funds (see Panel A of Figure 3). However, as shown in Panel B of Figure 3, there are substantial variations (with similar magnitude) in TCI across all size × value style categories.[14] Finally, we plot the TCI measure's percentiles over time and find that TCI is generally stable (see Figure A1 of the Internet Appendix).

In subsequent analyses, we also control for other existing measures of fund activeness, including *ICI*, *Active Share*, R^2, and *OCI*.[15] Panel B reports the correlation matrix. We find that *TCI* is positively correlated with four-factor alpha, *ICI*, *Active Share*, and *OCI*, but is

[14] We follow Nanda, Wang, and Zhang (2004) to classify funds into 3 × 3 "size × value" styles based on funds' rolling 36-month four-factor loadings. Specifically, we run the Carhart (1997) four-factor model using fund monthly returns from the previous 36 months (requiring a minimum of 20 monthly returns) and obtain the factor loadings. Each month, we assign all funds into three groups based on the SMB and HML loadings. Mutual funds ranked in the top tercile of SMB (HML) loading are labeled as small-cap (value-style), those in the bottom tercile are labeled as large-cap (growth-style), and those in the middle tercile of SMB (HML) are labeled as mid-cap (blend-style).

[15] Specifically, *ICI* measures the extent of portfolio concentration across 10 broadly defined industries by summing the squared differences between the industry weights of a fund and the industry weights of the total market portfolio (Kacperczyk, Sialm, and Zheng, 2005). R^2 is defined as the proportion of the variance of the fund return explained by the Carhart (1997) four-factor model (Amihud and Goyenko, 2013). Active share measures the percentage of fund holdings that is different from the benchmark holdings (Cremers and Petajisto, 2009), and its data is obtained from the University of Notre Dame academic research database (https://activeshare.nd.edu/). *OCI* measures the extent of portfolio concentration across offshore markets by summing the squared differences between a fund's weighting of each foreign country relative to that of the market portfolio (Bai et al., 2022).

negatively associated with R^2. Importantly, none of the correlations between TCI and other measures of activeness exceeds 0.37, which suggests that *TCI* likely captures a dimension of investment skill that is distinct from existing fund activeness proxies.

We emphasize that TCI is conceptually different from the industry concentration index (Kacperczyk, Sialm, and Zheng, 2005) because stocks in the same industry can be exposed to a variety of different themes, and stocks of different industries can be exposed to the same theme (e.g., artificial intelligence). Further analysis of the bivariate distributions of *TCI* and *ICI* in our sample shows that distinguishing between the two dimensions of active management is also empirically important. While there is a positive correlation between the two measures, within each category (e.g., quintile) of *TCI* or *ICI*, there is still considerable variation in the other measure (see Table A1 of the Internet Appendix). For example, funds in the top *ICI* quintile can be classified into any quintile based on *TCI* (with the majority not being ranked in the top *TCI* quintile), and vice versa. Thus, there are important differences between the two measures.

In Table 2, Panel A, we examine how fund characteristics such as fund size, fund age, turnover, and expense ratios are related to *TCI*, while controlling for fund style fixed effects. Since TCI is constructed from funds' quarterly holdings, we conduct this analysis at quarterly frequency. Quarterly TCI is measured as the average monthly TCI in a quarter. Column (1) shows that *TCI* is negatively related to fund age and turnover and positively related to past fund alpha. In columns (2) through (4), we augment the regression model with several known measures of managerial skills. Specifically, we include a fund's industry concentration index (ICI), R^2, *Active Share*, and offshore concentration index (OCI). Overall, the results consistently show that *TCI* is positively related to proxies of active management.

[Insert Table 2 Here]

We further examine the time persistence of *TCI*. To this end, we regress each fund's quarterly average *TCI* on its lagged values up to four quarters. The results of this exercise are presented in Table 2, Panel B. In columns (1) through (4), we regress a fund's TCI on its first, second, third, and fourth lag. Our results show strong persistence in TCI measure. In particular, the past values of *TCI* have significant predictive power for its future value, although this predictability decreases gradually over time (coefficients of 0.881 in column (1) vs. 0.768 in column (4)). Our evidence on persistence over time suggests that investing in stocks exposed to a small set of themes reflects deliberate fund strategies rather than random chance.

4. Empirical Results on TCI and Fund Performance

In this section, we first study the performance implications of fund thematic concentration using both portfolio sorting and Fama-MacBeth regressions. After that, we examine the potential reasons that could explain the superior performance of high TCI funds.

4.1. TCI and Fund Performance: Portfolio Evidence

In the portfolio analysis, we sort mutual funds by TCI and evaluate fund performance over the subsequent period. In particular, each month, we sort mutual funds into decile portfolios based on their lagged TCI. For each decile portfolio, we compute both the equal- and TNA-weighted average returns for each month. To measure abnormal performance, we calculate four-factor alpha as the intercept of the Carhart (1997) four-factor model based on the time series of the monthly average returns for each decile portfolio:

$$r_{p,t} = \alpha_p^{4F} + \beta_{1,p} MKT_t + \beta_{2,p} SMB_t + \beta_{3,p} HML_t + \beta_{4,p} UMD_t + \varepsilon_p, \tag{3}$$

where $r_{p,t}$ is the monthly portfolio return in excess of the one-month T-bill rate; MKT is the excess return on a value-weighted market portfolio; and SMB, HML, and UMD are the returns on the zero-investment factor mimicking portfolios for size, book-to-market, and momentum, respectively. In our analysis, we examine both net- and gross-of-fee returns in our analysis, as return can be viewed as a more appropriate measure than net return to evaluate the investment skills of fund managers because those with better skills may charge higher expenses (e.g., Berk and Green (2004) and Berk and van Binsbergen (2015)). Gross returns are created by adding back 1/12 of the annual expense ratio to each monthly net return.

The results are presented in Table 3. The first column of Panel A reports the equal-weighted average net returns for funds in each *TCI* decile in the next month. Columns (2) and (3) report Fama-French (1993) three-factor and Carhart (1997) four-factor monthly alphas based on net returns, respectively. Results are similar if we use gross four-factor alpha or the DGTW CS measure to measure performance in columns (4) and (5). We find consistent evidence that funds with higher TCI meaningfully outperform those with lower TCI. Fund performance decreases (almost monotonically) from the top to the bottom deciles of TCI, and the difference between two extreme deciles is positive and statistically significant (labeled as "D10-D1"): the monthly return difference is 0.355% for net four-factor alpha, 0.368% for gross four-factor alpha, and 0.193% for the DGTW CS measure. All return differences are statistically significant at the 1% level.

[Insert Table 3 Here]

In Panel B of Table 3, we repeat our analysis using TNA-weighted portfolios. The performance difference between funds in the top and bottom *TCI* deciles remains positive and significant across all performance measures, which suggests that our baseline results are not entirely driven by the small TNA funds in the sample. In fact, the magnitude of the top-minus-bottom difference becomes larger for TNA-weighted portfolios, suggesting that the effect of TCI on fund performance is stronger among larger funds. Lastly, we also calculate the Spearman rank correlations for each of our sorting analyses and report them in the bottom row of the table. The Spearman correlations are high (with a value of 1 and a *p*-value below 0.1% in all columns), underscoring the monotonic relation between TCI and fund performance.

We also evaluate fund performance using the six-factor model that extends Carhart (1997) by adding the Fama and French (2015) profitability and investment factors. The outperformance of high-TCI funds remains significant and of similar magnitude (see Table A2 in the Internet Appendix). The results are also robust to using the conditional alpha measure of Cederburg et al. (2018) (see Table A3 in the Internet Appendix). Third, Lopez-Lira (2023) constructs mimicking portfolios for Item 1A risks and shows they are priced beyond common multi-factor models. Our findings are unchanged when we use these risk factors to evaluate performance (see Table A4 in the Internet Appendix). Finally, we split the sample into two subperiods: 2006-2014 and 2015-2023 and our results continue to hold in both subsamples (see Tables A5 in the Internet Appendix). Therefore, our findings are not driven by a specific market environment.

One might argue that the TCI simply reflects "risk" concentration since it is constructed from the Risk Factors section of 10-Ks and that less diversified risk exposure mechanically produces higher returns. To address this concern, we examine information ratios (abnormal return divided by its standard deviation) as well as Sharpe ratios for TCI-sorted portfolios. If high-TCI outperformance were due to omitted risk factors, high- and low-TCI funds would exhibit similar information and Sharpe ratios. We find that the top-TCI decile exhibits a substantially higher information ratio than the bottom decile under both the four- and six-factor models (see Table A6 in the Internet Appendix). In addition, we also reconstruct the TCI measure by applying LDA to the MD&A section of 10-K filings and continue to find a positive TCI-performance relation, with the magnitude being slightly lower compared to our results in Table 3 (see Table A7 of the Internet Appendix).

In short, our results consistently show that funds with concentrated thematic exposures in their portfolios significantly outperform those with less concentrated thematic exposures. Importantly, the outperformance of high TCI funds is not due to concentrated exposure to certain "risk" factors that are not captured in the performance evaluation models.

4.2. TCI and Fund Performance: Regression Analysis

In this section, we examine the relation between fund *TCI* and future performance using Fama-MacBeth (1973) regressions, which control for various fund characteristics that predict future fund performance. Specifically, we estimate the following specifications:

$$\alpha_{i,t}^{4F}\left(CS_{i,t}\right) = \lambda + \beta_1 TCI_{i,t-1} + \beta_2 \log(TNA)_{i,t-1} + \beta_3 Log(Age)_{i,t-1} + \beta_4 Expenses_{i,t-1} + \beta_5 Turnover_{i,t-1} + \beta_6 \sigma_{t-12:t-1} + \beta_7 Flow_{i,t-12:t-1} + \varepsilon_{i,t}, \tag{4}$$

where i indexes fund, and t is the month subscript. In each month, we regress fund abnormal performance on the most recent *TCI* and other fund characteristics, all lagged by one month. We then calculate the time-series averages of the monthly coefficient estimates, and standard errors are adjusted for serial autocorrelations with three lags (Newey and West, 1987).

The dependent variables are future fund performance as measured by Carhart (1997) four-factor alpha net ($\alpha_{i,t}^{4F,Net}$) and gross ($\alpha_{i,t}^{4F,Gross}$) of fees, and DGTW CS measure. To compute four-factor alpha ($\alpha_{i,t}^{4F}$) for fund i in month t, we first estimate factor loadings by running the Carhart (1997) four-factor model on the prior 36 months of monthly fund returns (requiring at least 20 observations). We then calculate $\alpha_{i,t}^{4F}$ as the fund's actual return minus the expected return implied by these loadings and the contemporaneous factor realizations. We construct the CS measure following Daniel et al. (1997), which analyzes fund holdings to evaluate stock-selection ability relative to a benchmark portfolio that matches each holding on size, book-to-market, and momentum. In essence, the CS measure assesses whether managers outperform a portfolio of stocks with similar characteristics on a gross of fee basis. Following previous literature (e.g., Chen et al., 2004; Pollet and Wilson, 2008; Jordan and Riley, 2015; Busse, Jiang, and Tang, 2021), we control for a comprehensive set of fund performance determinants: the logarithm of fund TNA ($Log(TNA)$), the logarithm of one plus fund age ($Log(Age)$), the fund expense ratio ($Expenses$), the portfolio turnover ratio ($Turnover$), fund return volatility ($\sigma_{t-12:t-1}$), and past fund flows ($Flow_{t-12:t-1}$).

We report the estimation results in Table 4. Across all specifications, we find a positive and statistically significant relation between *TCI* and future fund performance. For example, in column (3) with lagged fund characteristics as controls and the fund style fixed effects, the coefficient on *TCI* is 1.143, and the t–statistic is 4.16. It suggests that a one-standard-deviation increase in *TCI* leads to an increase in net four-factor alpha of 6.1 basis points a month or 72.7 (=12 × 1.143 × 0.053) basis points a year. This effect is economically significant since Table 1 shows that the average equity mutual fund underperforms the Carhart four-factor model by

98.4 basis points per year (=−0.082%×12).[16] Moreover, columns (7) through (9) show that *TCI* is positively and significantly related to the *CS* measure, which indicates that funds with high *TCI* also outperform a benchmark portfolio that consists of stocks with similar characteristics.

[Insert Table 4 Here]

In additional analyses, we first re-estimate our TCI measure after varying the LDA topic count to 25, 50, 75, and 150. Our results remain unchanged (see Table A8 of the Internet Appendix), suggesting that our findings are not sensitive to the number of topics in the LDA model. Second, we examine the TCI-performance relation across style categories. We find it is not driven by any single style, and the effect is stronger in Small-Cap and Growth funds (see Table A9 of the Internet Appendix). Third, because several mega-cap stocks performed exceptionally well during the 2010s, we test whether they drive our results.[17] We find that the TCI-performance relation persists for funds with both high and low mega-cap exposure, suggesting that mega-cap holdings do not explain our findings (see Table A10 of the Internet Appendix). Fourth, we repeat our analysis using the value-added measure (the product of fund AUM and gross alpha) of Berk and van Binsbergen (2015) and continue to find a positive relation (see Table A11 of the Internet Appendix). Fifth, one may be concerned that high TCI funds are likely to shift their risk levels over time. We add risk-shifting measures of Huang, Sialm, and Zhang (2011) as additional controls and our results on TCI continue to hold (see Table A12 of the Internet Appendix). Finally, we split the sample into two subperiods: 2006-2014 and 2015-2023 and find results in both subsamples (see Table A13 of the Internet Appendix). Overall, these findings provide consistent support for our first hypothesis: higher TCI funds earn higher risk-adjusted returns.

[16] Average alpha being negative is consistent with the prior literature (e.g., Jensen (1968), Gruber (1996), Carhart (1997), Wermers (2000), Fama and French (2010)). For instance, Gruber (1996) also finds that the average equity mutual fund underperforms a four-factor model by about 65 basis points per year during the period of 1985-1994.
[17] Specifically, each quarter we rank funds by their aggregate weight in the "MegaCap-8" stocks (Alphabet (GOOG/GOOGL), Amazon (AMZN), Apple (AAPL), Meta (META), Microsoft (MSFT), Netflix (NFLX), Nvidia (NVDA), and Tesla (TSLA)) and classify the top quintile as mega-cap-focused funds. On average, these funds allocate 16.9% of assets to the eight stocks, whereas other funds allocate only 4.68%.

Although TCI predicts fund performance, it is also correlated with other activeness metrics, including industry concentration, active share, low R^2, and offshore concentration, each known to influence returns. To test whether TCI adds information beyond these measures, we regress future returns on TCI and all four activeness controls in a multivariate framework. We present the results in Table 5. We use net four-factor alpha, gross four-factor alpha, and DGTW *CS* measure as dependent variables in Panels A, B, and C, respectively. In column (1) of Panel A, we add ICI as an additional regressor in the regression. The coefficient of *ICI* is 0.594 (*t*-statistic = 2.26). Importantly, the coefficient of *TCI* is 1.239, which is positive and significant at the 1% level with a *t*-statistic of 3.80. This finding highlights that *TCI* is fundamentally different from ICI of Kacperczyk, Sialm, and Zheng (2005). We then repeat the exercise by separately including Active Share (column (2)), R^2 (column (3)), *OCI* (column (4)), and all four fund activeness proxies (column (5)). Across all specifications, *TCI* remains a strong predictor of future fund returns.[18]

[Insert Table 5 Here]

4.3. Alternative Explanation based on Stock-level Thematic Concentration

Our results show a strong correlation between fund-level *TCI* and fund performance. However, one could argue that the relation could be driven by an anomaly in stock returns: stocks with higher concentration in their exposure to certain semantic themes are associated with higher future returns. If so, funds with high *TCI* could simply pick these high *TCI* stocks to exploit such an anomaly, rather than having an information advantage about firms with certain semantic themes. We thus investigate whether our findings so far are an artifact of high TCI funds exploiting such a stock-level anomaly.

To this end, we conduct both portfolio sorting and Fama-MacBeth regression analyses to examine whether *stocks* with higher *TCI* are associated with higher future returns. This

[18] Our findings are also robust if we use the orthogonalized TCI measures computed as the residuals of cross-sectional regressions of TCI on ICI, Active Share, R^2, OCI, and style fixed effects (see Table A14 of the Internet Appendix).

analysis is similar to our main analysis in Tables 3 and 4 but uses the *stock-level TCI* ($TCI^{Stock} = \sum_{k=1}^{100} {T_{i,k}}^2$) and stock returns instead. The results are presented in Table 6. Specifically, Panel A of this table reports various metrics of returns to decile portfolios of stocks sorted on stock-level *TCI*. Panel B reports the results from Fama and MacBeth (1973) regressions of monthly future stock returns on lagged stock-level *TCI*. As shown in Panel A, the difference between the top decile and bottom decile portfolio returns ranked by stock-level *TCI* (TCI^{Stock}) is small in magnitude and statistically insignificant. Similarly, in the Fama-MacBeth regressions in Panel B, the coefficients on TCI_{t-1}^{Stock} are small in magnitude (all close to zero) and statistically insignificant. Taken together, this evidence suggests that our findings are not driven by harvesting a stock-level anomaly related to thematic concentration but instead reflect fund managers' informational advantage.

[Insert Table 6 Here]

4.4. How Do High TCI Funds Achieve Superior Performance

In this section, we first examine whether high TCI funds' superior performance stems from their theme-timing ability or stock selection abilities, and then investigate how TCI-related trades affect both fund performance and the underlying stocks' fundamentals.

4.4.1. The TCI Advantage: Thematic Timing vs. Stock Selection

High TCI managers can generate alpha through superior theme-related market timing, superior stock selection, or both. Superior timing (i.e., the ability to overweight themes that later deliver higher returns) suggests that managers can identify emerging, unpriced themes, whereas superior stock selection reflects an edge in assessing how a given theme affects individual firms differently. In this section, we investigate whether high TCI managers derive their edge from theme-timing skill.

If a manager has superior timing ability, she will strategically allocate a greater weight of their portfolio to themes that later deliver higher returns and thus earn positive alpha. To test this idea, we first construct, for each topic k, Mkt_t^k to capture market return in month t attributable to exposure to topic k in corporate Item 1A disclosures:

$$Mkt_t^k = \sum_i \omega_{i,t}^{Mkt}\left(T_{i,k} \times Ret_{i,t}\right), \tag{5}$$

where Ret is firm i's stock return during the same period and $T_{i,k}$ represents the relative importance of topic k in the firm's risk factor disclosure as defined in Section 3.3.3. The term $\left(T_{i,k} \times Ret_{i,t}\right)$ reflects the individual stock return driven by the firm's exposure to theme k. We aggregate this return component using firms' market weight, $\omega_{i,t}^{Mkt}$, to obtain the market return component that captures firms' collective exposure to each theme. Next, we calculate, for each fund, $vRet$ to capture the portion of fund return predictability that can be attributed to the portfolio firms' exposure to various themes:

$$vRet_{p,t} = \sum_k v_{p,k} \times Mkt_t^k, \tag{6}$$

where $v_{p,k}$ is the combined exposure of fund p's portfolio to theme k (i.e., $v_{p,k} = \sum_{i \in p} \omega_i \times T_{i,k}$). Essentially, $vRet_{p,t}$ measures fund p's hypothetical returns in a month based on its portfolio weight distribution across all themes without selecting stocks within each theme. If high TCI managers generate alpha mainly through theme-related timing, then this measure will be positively and significantly related to their performance.

We next re-estimate our baseline regressions in Table 4, except that we add the $vRet_{p,t}$ measure as an additional regressor. Intuitively, if high TCI managers' superior performance is partially attributable to their theme-related market timing ability, we should observe a positive and significant coefficient on $vRet_{p,t}$. We report the results in Table 7. Similar to Table 4, we examine three performance variables: Carhart (1997) four-factor alpha net ($\alpha_t^{4F,Net}$) and gross ($\alpha_t^{4F,Gross}$) of fees, and the DGTW CS measure. For each variable, we augment our original regressions with two versions of $vRet_{p,t}$ capturing portfolio firms' theme-related returns over one month ($vRet_{p,t}$) and three months from $t-2$ to t ($vRet_{p,t-2:t}$). Across all specifications, the coefficients on theme-timing variables ($vRet_{p,t}$ and $vRet_{p,t-2:t}$) remain small in magnitude and statistically insignificant at conventional levels. Importantly, the coefficients of TCI remain positive and significant, with a similar magnitude to those in Table 4. Overall, our results suggest that the superior performance of high TCI managers is more likely to be driven by their stock selection rather than theme-level market timing ability.

[Insert Table 7]

4.4.2. Holding Changes, Performance, and Earnings

To better analyze stock selection skills, we next examine the performance implications of funds' trades in individual stocks that change their thematic concentration index. This test allows us to observe *TCI* and performance at a much more refined level, i.e., the stock level rather than the portfolio level.

Each quarter, we sort funds into decile portfolios based on the TCI measure at the beginning of quarter q. Within each fund, we break down fund trades into buy and sell trades in quarter q as follows. We first calculate the changes in portfolio weights resulting from active portfolio rebalancing in a quarter and categorize these changes into buy and sell trades. Specifically, we begin by computing the fund's hypothetical portfolio weights in a given quarter absent trading, denoted as $\widetilde{\omega}$:

$$\widetilde{\omega}_{i,q} = \frac{\omega_{i,q-1}(1 + r_{i,q})}{\sum_{i \in P} \omega_{i,q-1}(1 + r_{i,q})}, \tag{7}$$

where $\omega_{i,q-1}$ is fund P's weight in stock i at the end of quarter $q-1$ and $r_{i,q}$ is stock i's return in quarter q. Our calculation closely follows prior literature (e.g., Jiang, Yao, and Yu, 2007; Wermers, 2012; Wei, Wermers, and Yao, 2015, McLemore et al., 2021), and reflects the notion that, when no trades occur in quarter q, changes in portfolio weights are entirely driven by stock returns during the same period. Then, we calculate the difference between $\omega_{i,q}$ and $\widetilde{\omega}_{i,q}$, which by design captures only the effects due to active portfolio rebalancing in quarter q. Specifically, $\omega_{i,q} > \widetilde{\omega}_{i,q}$ means a "buy" trade in stock i; and $\omega_{i,q} < \widetilde{\omega}_{i,q}$ indicates a "sell" trade in stock i. In the final step of the analysis, we calculate and compare the time-series averages of subsequent performance measures for buy and sell trades across TCI deciles.

In Panel A of Table 8, we compare the next-quarter performance of stocks bought and sold by funds in each TCI decile. For each fund, we calculate the average quarterly DGTW-adjusted returns in quarter q+1 for the stocks bought or sold in quarter q. We find that stocks purchased tend to outperform those sold in high *TCI* decile portfolios in the subsequent quarter. In the top *TCI* decile (i.e., D10), the difference in quarterly DGTW-adjusted returns is 1.091%. Further, the difference between the buy and sell portfolios generally increases with TCI, primarily driven by the top two deciles. Moreover, the difference in the buy–sell spreads between the top and bottom deciles (i.e., D10−D1) amounts to 1.339%, both economically and statistically significant. Finally, the superior performance of high *TCI* funds is mainly driven by their buy trades, consistent with earlier work (e.g., Chan and Lakonishok, 1993; Chen, Jegadeesh, and Wermers, 2000; Baker et al., 2010; Puckett and Yan, 2011).

[Insert Table 8 Here]

Next, we investigate what type of information high *TCI* funds have and whether it is related to firms' fundamentals, such as corporate earnings. In particular, we examine whether the trading of top *TCI*-decile funds is positively related to *future* earnings surprises. If high *TCI* fund managers have an information advantage about their portfolio firms' earnings and make profitable trades accordingly, we expect stocks bought (sold) by high TCI funds in quarter q to be followed by positive (negative) earnings surprises in quarter q+1.

We present the results on earnings surprises in Panel B of Table 8. Each number represents the cumulative market-adjusted return (i.e., CAR) over the three-day window, [-1, +1], around the *next* earnings announcement. A positive CAR indicates a positive earnings surprise, and vice versa. Columns (4)-(6) of Panel B show that stocks bought by top TCI funds (i.e., D10) experience a significantly higher CAR than stocks sold by these funds over the three-day window around the next earnings announcement (0.861% vs. 0.200%), with the difference (0.662%) being significant at the 1% level. We do not find the same patterns for funds in the bottom TCI decile (i.e., D1). The buy-sell difference for the funds in D1 is small and insignificant, suggesting that their trades cannot correctly forecast the information revealed in subsequent earnings announcements. The difference in the buy-sell CAR spreads between the top and bottom deciles (i.e., D10−D1) is 0.636%, statistically significant at the

5% level. In short, our results suggest that high TCI fund managers tend to have an information advantage regarding future corporate earnings.[19]

Finally, we conduct an additional analysis to examine the performance of stocks held or purchased in the "themes" that are overweighted by mutual funds, following the portfolio approach used by Pool et al. (2015) and Kumar et al. (2020) (see Section A.2 of the Internet Appendix for more details). We form four distinct portfolios at the beginning of each quarter based on whether the fund bought or sold a stock during the previous quarter and whether the trade was of a stock with over-weighted themes. We find that the long-short portfolio based on funds' trades in stocks in overweighted themes significantly outperforms the long-short portfolio of other stocks, with the difference in four-factor alpha and DGTW being 0.419% and 0.323% per month, respectively (see Table A16 of the Internet Appendix). It suggests that fund managers tend to overweight topics in which they have expertise, resulting in superior performance of the trades in stocks in their over-weighted themes.

5. Thematic Skills and Managers' Educational Background

While we have shown that a fund's TCI is positively related to performance, two questions remain: (i) what underlies the superior stock-selection skill of managers who run thematically concentrated portfolios, and (ii) why don't all managers pursue the same strategy? Our final set of analyses tackles these questions by probing the sources of thematic investment skill.

5.1. TCI Changes around Manager Turnovers: A Difference-in-Differences Approach

Previous literature finds that various managerial characteristics contribute to the professional performance of portfolio managers. For instance, managers who attend undergraduate institutions with higher average student SAT scores tend to earn higher

[19] In an additional analysis, we also decompose the stock purchases and sales into those that increase the TCI and those that decrease it. As reported in the Internet Appendix (Section A.1 and Table A15, Panel A), we find that TCI-increasing stock purchases by top TCI-decile funds are likely the underlying reason for their superior abnormal performance. Consistent with this result, in Panel B of Table A15, we find that the relationship between TCI and firm fundamentals (e.g., corporate earnings) is driven by TCI-increasing purchases.

returns (Chevalier and Ellison, 1999). More broadly, the notion that educational system and educational attainment matter for long-term labor market outcomes has long been the focus of academic interest in sociology and education literature (Allmendinger, 1989). Specifically, undergraduate specialization appears to play a significant role in explaining students' career choices and line of work when they join the labor force (Kim and Kim, 2003; Van de Werfhorst, 2004). Building on this line of work, we posit that a manager's undergraduate major provides the foundational domain knowledge that later enables her to further learn and specialize as a portfolio manager, constructing portfolios focused on related themes. We note that undergraduate study is unlikely to confer investment skill directly, but it equips managers with core knowledge and familiarity that can be deepened later in their careers. For example, a science or engineering graduate is better positioned to gain an edge when analyzing firms exposed to disruptive technological change.

While this conjecture is intuitively appealing, it is empirically challenging to operationalize. To this end, we take a multistep approach to map fund managers' educational backgrounds to the potential sources of expertise in specific investment themes. First, we collect a comprehensive data set on the majors of managers' undergraduate studies from their LinkedIn pages. Second, we extract all descriptions of 81,885 courses from 235 departments/academic units publicized in Stanford University's Course Bulletin for the academic years 2006 – 2023 (18 years).[20] We then combine the descriptions of each individual course offered by the same department over the years (2006–2023). The combined description of a course provides a concise overview of the knowledge base, skill sets, and learning objectives that the course intends to impart.

Finally, we employ textual analysis to cross-analyze the semantic similarity between the text of investment themes and that of course descriptions. Specifically, we first create a

[20] Ideally, we should obtain the course descriptions from the schools that each manager went to, but such data is not available. Our assumption is that the course descriptions from Stanford is a reasonable text corpus for the purpose of identifying the different contents across academic disciplines. The most recent Stanford Course Description Bulletin is available at https://explorecourses.stanford.edu/browse?academicYear=20242025. Check all four terms on the right and click on each department to view detailed courses offering information in a year. Beginning with all descriptions of 128,632 courses, we remove graduate-level courses, thesis courses, honors courses, and courses with very short descriptions (less than 150 chars).

list of 100 themes (i.e., topics) covered by all firms' Item 1A disclosures in a year. Since each topic is associated with several most frequently used keywords, we measure the incidence that one or more of the top 10 keywords (bigrams) are mentioned in the combined description of courses offered by a degree. If this is the case, we then classify a given topic as being covered by a degree. Topics and degrees have a many-to-many relationship: a single topic may be covered by multiple degrees, and a single degree may encompass multiple topics. We update both topics and degree coverage annually. When we look at different types of courses offered by different degree programs, we find that economics courses cover the most topics (41/100) in 2006, followed by the degree of "management science & engineering." At the same time, "structural biology" covers none of the themes. This systematic approach allows us to construct a reasonable, albeit imperfect and possibly noisy, proxy that measures the overlap between a manager's field of undergraduate study and her choice of investment theme.[21]

With this constructed measure for the subset of managers for whom we have complete education data, we examine whether a given fund changes its TCI when the previous manager departs and a new manager with a different undergraduate degree steps in. Our empirical strategy is similar to a difference-in-differences (DID) framework often employed in corporate finance and can be illustrated with the following example. Imagine Fund A and Fund B both experience portfolio manager turnover. The difference between these two funds is that while Fund A's new manager holds a different undergraduate degree than its departing manager, Fund B's new manager holds the same undergraduate degree as its departing manager. In our empirical tests, the groups of funds similar to Fund A and Fund B are our treated group and control group, respectively. We expect Fund A's investment themes to change significantly due to their new manager's different undergraduate degree, but smaller or insignificant changes in investment themes in the case of Fund B.

[21] In Figure A2 of the Internet Appendix, we provide four examples of topics extracted through LDA, representing themes that are closely associated with (or covered by) specific undergraduate disciplines such as Computer Science, Biochemistry, Economics, or Finance. These examples demonstrate that the semantic similarity between themes and undergraduate majors that we capture based on textual information aligns intuitively with expectations.

[Insert Table 9 Here]

The results of this exercise are presented in Table 9. First, we have 2,156 manager turnovers with a change in the bachelor's degree between the departing manager and the incoming manager. In addition, we identify 4,821 turnovers without such a change. We track the *TCI* for the degree-change turnovers (i.e., treated group) and degree-no-change (i.e., control group) samples separately in one, two, three, and four quarters after the turnovers. We analyze fund TCI in themes related to the previous manager's undergraduate major in Panel A and fund TCI in themes related to the new manager's undergraduate major in Panel B. The results show a significant decrease (increase) in the future TCI associated with the departing (new) manager's degree following the degree-change turnovers. In contrast, we find almost no change in *TCI* in the non-degree changing turnovers (i.e., control group). Table 9 also reports the difference-in-differences results between degree-changing and degree-no-change turnovers in the bottom two rows. Consistent with our conjecture that educational background (i.e., the field of the bachelor's degree) is related to theme-related skills, we find a significant difference in the TCI changes between the treatment and control groups over time.[22] That is, compared to the control group of turnovers with no degree change, the change in thematic concentration is significantly larger in themes related to the manager's bachelor's degree for the treated group. Our evidence suggests that fund managers appear to be making a conscious decision to form portfolios that concentrate on investment themes related to their field of undergraduate study.

5.2. A Multivariate Regression Approach

Thus far, we have shown that managers with field-specific training, who tend to possess expertise in obtaining value-relevant information in related themes, are more likely

[22] As an additional analysis, in Table A17 of the Internet Appendix, we use a matched fund sample without manager turnover as the control group. Specifically, we require funds in the matched sample to have a similar size ($\mathrm{TNA}^{\mathrm{treated}} \times 0.9 \leq \mathrm{TNA}^{\mathrm{control}} \leq \mathrm{TNA}^{\mathrm{treated}} \times 1.1$) and the same investment objective as the control funds (i.e., those with degree changing turnover) during the quarter when the manager departs or joins. As shown in Table A17, there is no change in the TCI of matched funds over time.

to construct a thematic-concentrated portfolio. We posit, in our second hypothesis, that if managers' initial information advantage is due to their undergraduate field of study, then the portion of *TCI* that is semantically related to their bachelor's degree should be more important in helping to generate superior performance. To further shed light on whether managers' undergraduate degree is a likely contributing factor for the *TCI* and the subsequent superior performance, we repeat our multivariate Fama-Macbeth regression in Table 4 but decompose our *TCI* measures into two components: (i) the part that is related to the manager's degree and (ii) the part that is unrelated to the manager's degree. Specifically, we decompose Equation (2) into the portion related to the manager's education (i.e., $\mathrm{TCI}_{t-1}^{\text{Degree Related}} = \sum_{k\in edu}\sum_{i\in P}\left(\omega_i \times T_{i,k}\right)^2$), and the portion that is unrelated to a manager's education (i.e., $\mathrm{TCI}_{t-1}^{\text{Degree Unrelated}} = \sum_{k\notin edu}\sum_{i\in P}\left(\omega_i \times T_{i,k}\right)^2$). If expertise or domain knowledge acquired during a manager's undergraduate education is a significant contributing factor to the fund's *TCI*, we would expect the portion of *TCI* related to a fund manager's degree to be particularly strong in predicting the fund's outperformance.

We present the results of this exercise in Table 10. Due to the availability of managers' educational background, these tests are conducted on a smaller sample compared to our baseline regression in Table 4. However, in columns (1), (3), and (5), the results show that in this subsample, the positive and significant effect of *TCI* on funds' abnormal return continues to hold. In columns (2), (4), and (6), we re-estimate these regressions with the decomposed *TCI* measures. We find that, irrespective of the performance measure we use, the coefficients on $\mathrm{TCI}_{t-1}^{\text{Degree Related}}$ is significantly greater than those on $\mathrm{TCI}_{t-1}^{\text{Degree Unrelated}}$. The difference between these coefficients is large in magnitude and statistically significant at the 5% level or lower. Thus, consistent with *Hypothesis 2*, these results provide further support for the idea that managers' educational background is an important contributing factor to managers' *TCI* choice and superior fund performance.

[Insert Table 10 Here]

5.3. A Manager's Bachelor's Degree and Trade-based Performance of TCI Sorted Funds

In the final set of tests, we analyze the stock-level semantic similarity between stocks traded by the fund and the fund manager's degree. Based on the median level of semantic similarity between the stocks traded and the manager's degree, we classify each trade into two categories: degree-related and degree-unrelated, separately for buy and sell trades. We then investigate the performance implications of these trades on fund performance. If high TCI managers' initial information advantage is due to their undergraduate field of study, we expect that trades in degree-related stocks will be more profitable than trades in degree-unrelated stocks.

Our findings show a significant difference in stock performance between buy and sell trades among high *TCI* funds, particularly when funds trade stocks exposed to themes related to their undergraduate major. Specifically, in Panel A of Table 11, we compare the future performance of buy vs. sell trades in each *TCI* decile, computing the average quarterly DGTW benchmark-adjusted returns. We find that the superior trade performance of high *TCI* funds is mainly driven by trades fund managers made in degree-related stocks. For example, in column (3), the difference between the buy and sell trades is 1.998% (*t*-stat = 3.40) when stocks traded are degree-related. In contrast, this difference is 0.587% (*t*-stat = 1.30) for degree-unrelated stocks. In addition, we do not observe buys outperforming sells in degree-related stocks in low TCI funds. It suggests that undergraduate training itself does not directly confer investment skill; instead, it provides a foothold that some managers build on through continued learning and information acquisition to develop an investment edge. The D10-D1 differences in the bottom two rows of Panel A show that trades in degree-related stocks of high *TCI* managers are significantly more profitable compared to those of low TCI managers. Thus, only the high TCI managers, those with thematically concentrated portfolios, are deriving an edge from their undergraduate field of study.

[Insert Table 11]

Our earlier results show that the trading of top *TCI*-decile funds is positively related to future earnings surprises. If fund managers have an information advantage or expertise that is related to their undergraduate major, we expect that the positive relation between the trading of top *TCI*-decile funds and future earnings surprises should be more pronounced among degree-related stocks. Indeed, we find that the difference in CAR between the buy and sell trades in degree-related stocks of the top decile funds (D10) is 1.223% and significant at the 1% level, as shown in Panel B of Table 11. On the other hand, the difference in CAR between buy and sell trades of the top decile funds in degree-unrelated stocks in column (6) is 0.581% and insignificant at conventional levels. Moreover, Panel B shows similar patterns to those in Panel A when we calculate the D10-D1 differences. Finally, we also split the sample into two subperiods: 2006-2014 and 2015-2023 and find results in both subsamples (see Table A18 of the Internet Appendix). Altogether, these results suggest that managers' educational background is related to their expertise and investment skills in specific themes, providing further support for *Hypothesis 2*.

6. Conclusion

The world of active investment management is fast evolving, and mutual fund managers' source of investment skills, if any, has long been the focus of attention for academic research and financial practitioners. In this paper, we document and propose a new measure of investment skill that is based on investment themes. Using textual analysis of 10-K filings to identify stocks exposed to different themes, we construct a new measure to gauge a mutual fund's thematic exposure concentration. Our results provide strong and consistent evidence that mutual funds with higher thematic concentration significantly outperform their peers. The outperformance of high TCI managers appears to stem from superior stock selection rather than theme-related timing.

We further investigate the source of the differential TCI between different funds. We posit that a manager's major during her undergraduate studies provides early exposure to certain domain knowledge, which allows her to subsequently specialize in certain investment themes and capture related investment opportunities. By looking into managers' educational background and mapping each degree program's typical courses to funds' investment themes, we provide novel evidence that fund managers' expertise in investment themes appears to be closely linked to their undergraduate major. Specifically, the portion of TCI that is related to a manager's undergraduate major contributes significantly more to the fund's superior performance than the portion that is unrelated. Our results also suggest that undergraduate training does not directly translate into investment skill; rather, it offers a foundation that managers can build upon through ongoing learning and specialization.

Our study provides support to theoretical models that asymmetric information can lead to disparate returns among market participants (e.g., Grossman and Stiglitz, 1980) and investor specialization that generates superior returns and persists over time (e.g., van Nieuwerburgh and Veldkamp, 2009). Moreover, by documenting new evidence on the link between the undergraduate field of study and areas of investment expertise, our paper also adds to the paper by Chevalier and Ellison (1999), in which the authors provide some of the first evidence on the importance of education on shaping the investment performance of fund managers. We view further investigation of how educational background and attainment, as well as early life experiences in general, shape the professional behavior of mutual fund managers as a continually underexplored and fruitful area for future research.

References

Aiken, A.L., Lee, C., 2022, Attention to Detail: Learning About Mergers, Working Paper.

Allmendinger, J., 1989, Educational Systems and Labor Market Outcomes. European Sociological Review 5(3), 231-250.

Amihud, Y., Goyenko, R., 2013. Mutual Fund's R2 as Predictor of Performance. Review of Financial Studies 26(3), 667-694.

Bai, J.J., Tang, Y., Wan, C., Yüksel, H.Z., 2022. Fund Manager Skill in the era of Globalization: Offshore Concentration and Fund Performance. Journal of Financial Economics 145(2), 18-40.

Ball, C., Hoberg, G., Maksimovic, V., 2015. Disclosure and Firm Separation: A Text-Based Examination. Working Paper.

Baker, M., Litov, L., Wachter, J.A., Wurgler, J., 2010. Can Mutual Fund Managers Pick Stocks? Evidence from Their Trades Prior to Earnings Announcements. Journal of Financial and Quantitative Analysis 45(5), 1111-1131.

Bao, Y., Datta, A., 2014. Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures. Management Science 60(6), 1351-1616.

Ben-David, I., Franzoni, F., Kim, B., Moussawi, R., 2023. Competition for Attention in the ETF Space. Review of Financial Studies 36(3), 987-1042.

Berk, J.B., Green, R.C., 2004. Mutual Fund Flows and Performance in Rational Markets. Journal of Political Economy 112(6), 1269-1295.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003, Latent Dirichlet Allocation. Journal of Machine Learning Research 3(1), 993-1022.

Berk, J., van Binsbergen, J., 2015, Measuring skill in the mutual fund industry. Journal of Financial Economics 118(1), 1-20.

Brown, V.B., Tian, X.S., Tucker, J.W., 2018. The Spillover Effect of SEC Comment Letter on Qualitative Corporate Disclosure: Evidence from the Risk Factor Disclosure, Contemporary Accounting Research 35(2), 622-656.

Busse, J., Chordia T., Jiang L., Tang Y., 2021. Transaction Costs, Portfolio Characteristics, and Mutual Fund Performance. Management Science 67(2), 1227-1248.

Busse, J. A., Jiang, L., Tang, Y., 2021. Double-adjusted mutual fund performance. Review of Asset Pricing Studies 11(1), 169-208.

Bybee, L., Kelly, B. T., Manela, A., Xiu, D., 2021. Business news and business cycles. NBER Working Paper No. w29344.

Campbell J.L., Cecchini M., Cianci, A.M., Ehinger, A.C., Werner, E.M., 2019. Tax-related Mandatory Risk Factor Disclosures, Future Profitability, and Stock Returns. Review of Accounting Studies 24, 264-308.

Campbell, J.L., Chen, H., Dhaliwal, D., Lu, H., Steele, L., 2014. The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings. Review of Accounting Studies, 19(1), 396-445.

Carhart, M.M., 1997. On Persistence in Mutual Fund Performance. Journal of Finance 52(1), 57-82.

Cederburg, S., O'Doherty, M.S., Savin, N. E., Tiwari, A., 2018. Conditional Benchmarks and Predictors of Mutual Fund Performance, Critical Finance Review 7(2), 331-372.

Chan, L.K., Lakonishok, J., 1993. Institutional Trades and Intraday Stock Price Behavior. Journal of Financial Economics 33(2), 173-199.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., Blei, D., 2009. Reading tea leaves: How humans interpret topic models. Advances in Neural Processing Systems, (pp. 288-296). New York.

Chaudhuri, R., Ivkovic, Z., Pollet J., Trzcinka C., 2020. A Tangled Tale of Training and Talent: PhDs in Institutional Asset Management, Management Science 66(12), 5485-6064.

Chen, J., Hong H., Huang M., Kubik, J., 2004. Does Fund Size Erode Performance? Organizational Diseconomies and Active Money Management. American Economic Review 94(5), 1276–1302.

Chen, H.L., Jegadeesh, N., Wermers, R., 2000. The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of Fund Managers. Journal of Financial and Quantitative Analysis 35(3), 343-368.

Chevalier, J., Ellison, G., 1999, Career Concerns of Mutual Fund Managers. Quarterly Journal of Economics 114(2), 389-432.

Choi, N., Fedenia, M., Skiba, H., Sokolyk, T., 2017. Portfolio concentration and performance of institutional investors worldwide. Journal of Financial Economics 123(1), 189-208.

Cici, G., Gehde-Trapp, M., Göricke, M. A., Kempf, A., 2018. The investment value of fund managers' experience outside the financial sector. Review of Financial Studies 31(10), 3821-3853.

Cohen L., Frazzini, A., Malloy, C., 2008, The Small World of Investing: Board Connections and Mutual Fund Returns, Journal of Political Economy 116(5), 951-979.

Coval, J., Moskowitz T., 2001. The Geography of Investment: Informed Trading and Asset Prices. Journal of Political Economy 109(4), 811-841.

Cremers K., Petajisto A., 2009. How Active is Your Fund Manager? A New Measure that Predicts Performance. Review of Financial Studies 22(9), 3329-3365.

Cremers, M., Ferreira, M. A., Matos, P., Starks, L., 2016. Indexing and active fund management: International evidence. Journal of Financial Economics 120(3), 539-560.

Daniel, K., Grinblatt M., Titman S., Wermers, R., 1997. Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. Journal of Finance 52(3), 1035–1058.

Doshi, H., Elkamhi, R., Simutin, M., 2015. Managerial activeness and mutual fund performance. Review of Asset Pricing Studies 5(2), 156-184.

Dyer, T., Lang, M., Stice-Lawrence, L. 2017. The Evolution of 10-K Disclosure: Evidence from Latent Dirichlet Allocation. Journal of Accounting and Economics 64(2-3), 221-245.

Evans, R. B., 2010. Mutual fund incubation. Journal of Finance 65(4), 1581-1611.

Fama, E., French, K., 1993. Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33(1), 3-56.

Fama, E.F., French, K.R., 2010. Luck Versus Skill in the Cross-section of Mutual Fund Returns. Journal of Finance 65(5), 1915-1947.

Fama, E.F., French, K.R., 2015. A Five-factor Asset Pricing Model. Journal of Financial Economics 116(1), 1-22.

Fama, E., MacBeth, J., 1973. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy 81(3), 607-636.

Grossman, S. J., Stiglitz, J. E., 1976. Information and competitive price systems. American Economic Review 66(2), 246-253.

Grossman, S. J., Stiglitz, J. E., 1980. On the impossibility of informationally efficient markets. American Economic Review 70(3), 393-408.

Gruber, M., 1996. Another Puzzle: The Growth in Actively Managed Mutual Funds. Journal of Finance 5(3), 783–810.

Hanley, K., Hoberg, G., 2019. Dynamic Interpretation of Emerging Risks in the Financial Sector. Review of Financial Studies 32(12), 4543-4603.

Hansen, S., McMahon, M., Prat, A., 2018. Transparency and deliberation within the FOMC: A computational linguistics approach. Quarterly Journal of Economics 133(2), 801-870.

Hoberg, G., Kumar, N., Prabhala, N., 2018. Mutual fund competition, managerial skill, and alpha persistence. Review of Financial Studies 31(5), 1896-1929.

Hoberg, G., Lewis, C., 2017. Do Fraudulent Firms Produce Abnormal Disclosure? Journal of Corporate Finance 43, 58-85.

Huang, A., Lehavy, R., Zhang, A., Zheng, R., 2018. Analyst information discovery and interpretation roles: A topic modeling approach. Management Science, 64(6), 2833-2855.

Huang, L., Kale, J.R., 2013. Product Market Linkages, Manager Quality, and Mutual Fund Performance. Review of Finance 17(6), 1895-1946.

Huang, J., Sialm, C., Zhang, H., 2011. Risk Shifting and Mutual Fund Performance. Review of Financial Studies 24(8), 2575-2616.

Hwang, C.Y., Titman, S., Wang, Y., 2018. Is it Who You Know or What You Know? Evidence from IPO Allocations and Mutual Fund Performance. Journal of Financial and Quantitative Analysis 53(6), 2491-2523.

Jensen, M.C., 1968. The Performance of Mutual Funds in the Period 1945–1964. Journal of Finance 23(2), 389-416.

Jiang, G.J., Yao, T., Yu, T., 2007. Do Mutual Funds Time the Market? Evidence from Portfolio Holdings. Journal of Financial Economics 86(3), 724-758.

Jiang, H., Zheng, L., 2018. Active Fundamental Performance. Review of Financial Studies 31(12), 4688-4719.

Jordan, B., Riley, T., 2015. Volatility and Mutual Fund Manager Skill. Journal of Financial Economics 118(2), 289-298.

Kacperczyk, M., Sialm C., Zheng, L., 2005. On Industry Concentration of Actively Managed Equity Mutual Funds. Journal of Finance 60(4), 1983–2011.

Kacperczyk, M., Sialm, C., Zheng, L., 2008. Unobserved actions of mutual funds. Review of Financial Studies, 21(6), 2379-2416.

Kacperczyk, M., Seru, A., 2007. Fund Manager Use of Public Information: New Evidence on Managerial Skills. Journal of Finance 62(2), 485-528.

Kempf, E., Manconi, A. and Spalt, O.G., 2017. Learning by doing: The value of experience and the origins of skill for mutual fund managers. Working Paper.

Kim, A. and Kim, K.W., 2003. Returns to tertiary education in Germany and the UK: Effects of fields of study and gender, Working Paper.

Kumar, N., Mullally, K., Ray, S., Tang, Y., 2020. Prime (information) brokerage. Journal of Financial Economics 137(2), 371-391.

Levy, A., Livingston, M., 1995. The gains from diversification reconsidered: Transaction costs and superior information. Financial Markets, Institutions, and Instruments 4(3), 1–60.

Li, H., Zhang, X., Zhao, R., 2011. Investing in Talents: Manager Characteristics and Hedge Fund Performances. Journal of Financial and Quantitative Analysis 46(1), 59-82.

Liu, Y., Sheng, J., Wang, W., 2021. Technology and cryptocurrency valuation: evidence from machine learning. Working Paper.

Lopez-Lira, A. (2023). Risk Factors that Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns. Working Paper.

McLemore, P., Sias, R., Wan, C., Yüksel, H.Z., 2021. Active Technological Similarity and Mutual Fund Performance. Journal of Financial and Quantitative Analysis 57(5), 1862-1884.

Nanda, V., Wang Z., Zheng L., 2004. Family Values and the Star Phenomenon: Strategies of Mutual Fund Families. Review of Financial Studies 17(3), 667-698.

Newey, W., West, K., 1987. Hypothesis Testing with Efficient Method of Moments Estimation. International Economic Review 28(3), 777-787.

Pollet, J.M., Wilson, M., 2008. How does Size Affect Mutual Fund Behavior? Journal of Finance 63(6), 2941-2969.

Pool, V. K., Stoffman, N., Yonker, S. E., 2015. The people in your neighborhood: Social interactions and mutual fund portfolios. Journal of Finance 70(6), 2679-2732.

Puckett, A., Yan, X., 2011. The Interim Trading Skills of Institutional Investors. Journal of Finance 66(2), 601-633.

Van de Werfhorst, H. G., 2004, Systems of Educational Specialization and Labor Market Outcomes in Norway, Australia, and the Netherlands. International Journal of Comparative Sociology 45(5), 315-335.

Van Nieuwerburgh, S., Veldkamp, L., 2009. Information Immobility and the Home Bias Puzzle. Journal of Finance 64(3), 1187-1215.

Van Nieuwerburgh, S., Veldkamp, L., 2010. Information Acquisition and Under-Diversification, Review of Economic Studies 77(2), 779-805.

Wei K., Wermers, R., Yao, T., 2015. Uncommon Value: The Characteristics and Investment Performance of Contrarian Funds. Management Science 61(10) 2281-2547.

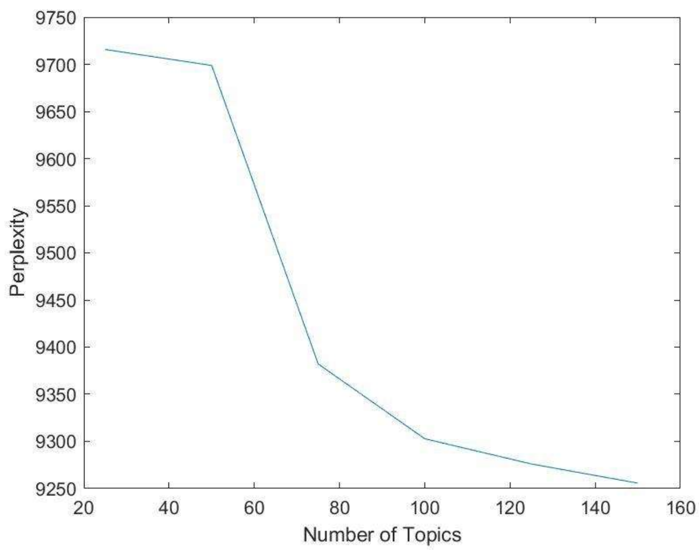
Wermers, R., 2000. Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses. Journal of Finance 55(4), 1655-1695.

Wermers, R., 2012. Matter of Style: The Causes and Consequences of Style Drift in Institutional Portfolios. Working Paper.

Figure 1. Perplexity by number of topics

Figure 1 presents the perplexity score of bigram LDA with the number of topics varying from 25 to 150 estimated using all firms' NRD in 2009 and 2013 (Panels A and B, respectively). The perplexity score decreases as the number of topics increases, indicating better generalization of topics obtained from the training data to the testing data. However, as the number of topics increases further, the improvement in model fit diminishes, which is often at the expense of loss of topic interpretability (Chang et al., 2009; Dyer, Lang, and Stice-Lawrence, 2017). We therefore choose 100 topics (often referred as the "elbow" point as the rate of perplexity change begins leveling off) to implement our baseline analysis. This choice assumes that there are 100 pertinent risk-related topics firms discuss annually in the aggregate and is also consistent with our goal to accommodate the spectrum of topics in the collection of NRD that may span various firm-specific, geographic, or technological risks.

Panel A. Perplexity of bigram LDA estimated using all NRD in 2009



Panel B. Perplexity of bigram LDA estimated using all NRD in 2013

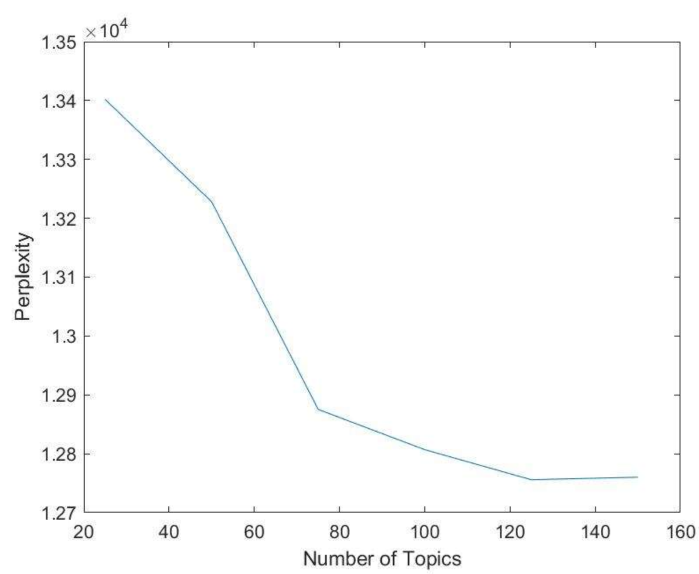
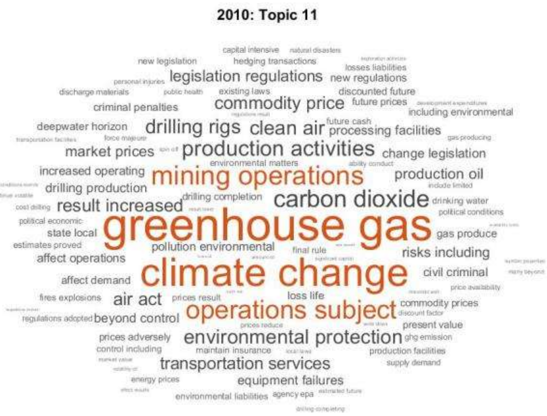


Figure 2. Word Clouds of some NRD topics

This figure presents six topics extracted from the bigram LDA analysis of the corpus of Item 1A disclosures in 10-K filings. The font size of bigrams represents the relative frequency of bigrams to a particular topic.

Word Cloud 1



Word Cloud 2

Word Cloud 3

Word Cloud 4

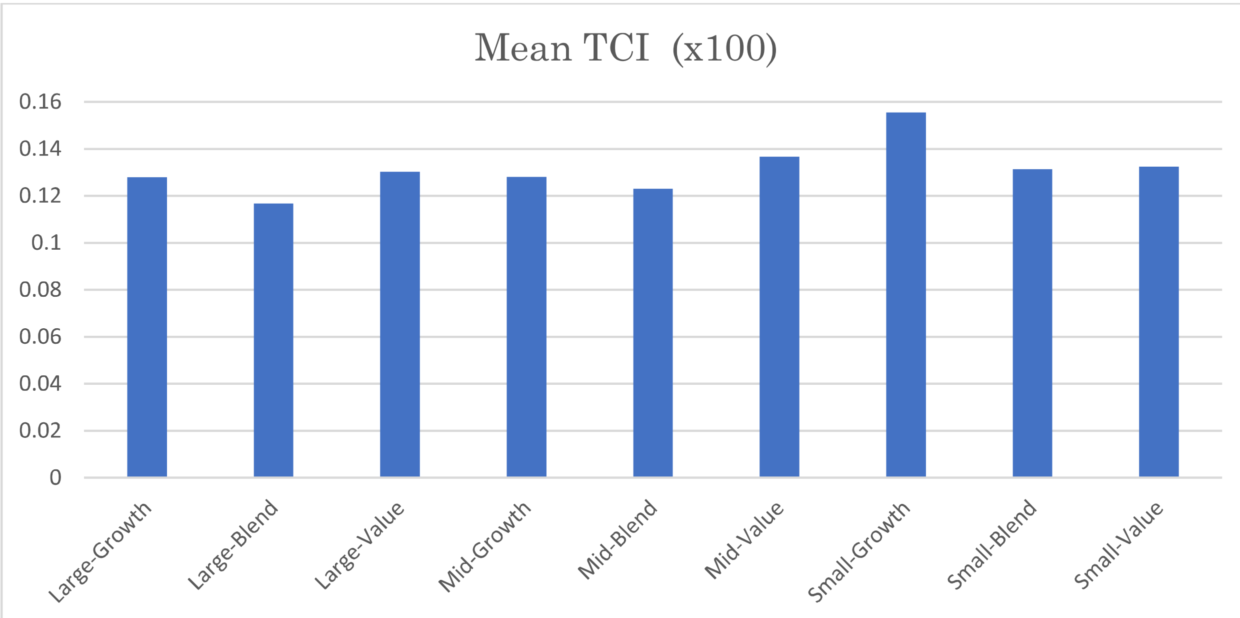
Word Cloud 5

Word Cloud 6

Figure 3. Thematic Concentration Index and Mutual Fund Style

This figure presents the time-series averages of cross-sectional mean and standard deviation of Thematic Concentration Index (TCI) within each style/value category. The construction of TCI is described in Section 3.4. Mutual funds are classified into size/value style categories based on a fund's four-factor loadings as described in Section 3.5. The sample period is from 2006Q3 to 2023Q4.

Panel A. Mean TCI across fund style categories



Panel B. Standard deviation of TCI across fund style categories

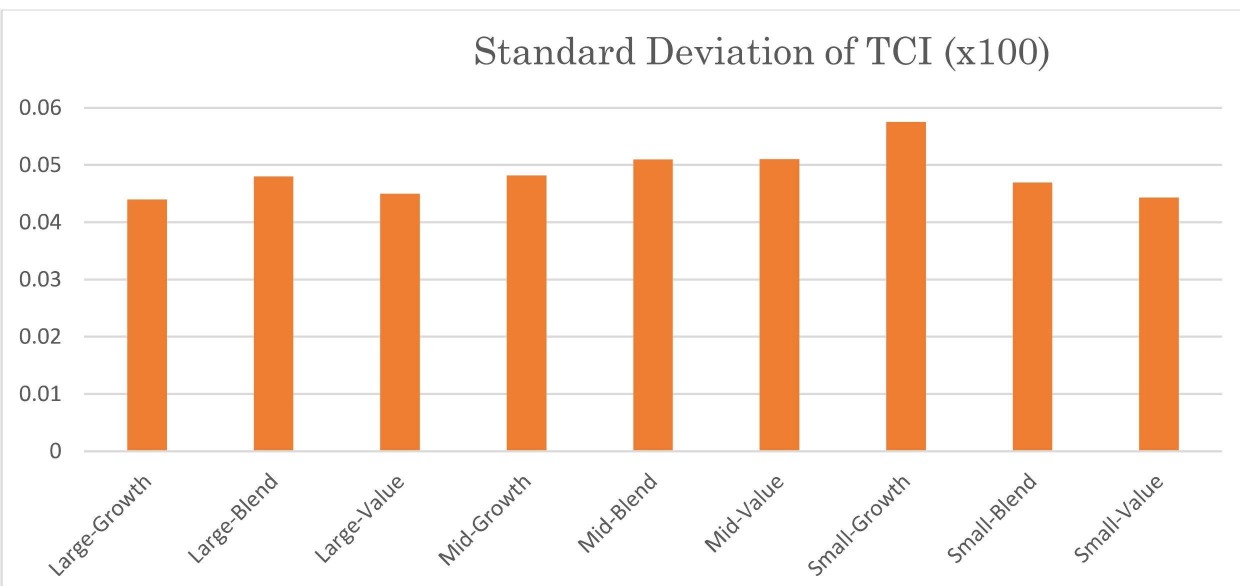


Table 1. Summary Statistics

This table reports the summary statistics for fund characteristics at the fund-month level. TCI is the monthly Thematic Concentration Index as described in Section 3.4; TNA ($million) is the total net assets under fund management (TNA) at the beginning of the month; Fund Age is in years since inception; Expenses (%) is the percentage of total investment that shareholders pay for a fund's expenses; Turnover (%) is defined as the minimum of aggregate purchases or sales of securities during the year, divided by the average TNA; σ^{Ret} is the return volatility of a fund measured as the standard deviation of monthly fund return over the prior twelve months; Flow (%) is the prior twelve-month normalized net flow into a fund and defined as $(TNA_{i,t} - TNA_{i,t-12}(1 + R_{t-1,t-12}))/TNA_{i,t-12}$; ICI is the industry concentration index (Kacperczyk, Sialm, and Zheng, 2005); ActvShr is the active share measure (Cremers and Petajisto, 2009). R^2 (α^{4F}) is the fund R^2 (the intercept) from a time-series regression of fund returns on market, size, value, and momentum factors over the previous 24 months (Amihud and Goyenko, 2013). OCI is the offshore concentration index (Bai, Tang, Wan, and Yuksel, 2022). Panel A presents the time-series averages of cross-sectional mean, median, standard deviation, 25th percentile and 75th percentile of fund characteristics. ICI, ActvShr., and OCI at the quarterly frequency, and other fund characteristics are at a monthly frequency. Panel B presents the time-series averages of cross-sectional correlations between TCI and three measures of fund activeness (ICI, R^2, ActvShr, OCI), and past fund alpha measured as the fund's Carhart (1997) four-factor alpha at the end of the prior month estimated over the previous 36 months (minimum of 20 months observation). Spearman (Pearson) correlations are reported above (below) the diagonal. The sample period is from 2006Q3 to 2023Q4 except for OCI (2006Q3 to 2017Q2).

Panel A. Summary statistics

	Mean	Median	St. Dev.	25th Pctl.	75th Pctl.	N
TCI (×100)	0.131	0.119	0.053	0.095	0.153	256,893
TNA ($ millions)	1,526.5	376.7	3189.4	106.9	1,349.3	256,893
Age (years)	16.83	15.32	10.39	9.84	21.03	256,893
Turnover (%)	67.56	53.12	55.41	30.16	87.71	256,893
Expenses (%)	1.012	1.040	0.383	0.834	1.239	256,893
σ^{Ret} (%)	4.743	4.644	0.911	4.133	5.298	256,893
α^{4F} (%)	-0.082	-0.084	0.463	-0.266	0.100	256,893
Flow (%)	4.080	-7.385	54.066	-17.390	7.026	256,893
ICI	0.069	0.063	0.032	0.046	0.084	256,893
R^2	0.859	0.872	0.064	0.841	0.895	256,893
ActvShr	0.834	0.883	0.150	0.721	0.970	187,736
OCI (×100)	0.277	0.212	0.268	0.113	0.342	155,400

Panel B. Correlation matrix of fund activeness

	TCI	α^{4F}	ICI	R^2	ActvShr	OCI
TCI	1	0.039	0.288	-0.185	0.216	0.367
α^{4F}	0.023		0.038	-0.015	0.000	0.076
ICI	0.306	0.026	1	-0.267	0.287	0.370
R^2	-0.183	-0.007	-0.232	1	-0.255	-0.351
ActvShr	0.197	-0.022	0.282	-0.200	1	0.675
OCI	0.313	0.035	0.277	-0.217	0.478	1

Table 2. Determinants and Persistence of TCI

This table reports results from Fama-MacBeth regressions of quarterly TCI (measured as the average monthly TCI in a quarter) on lagged fund characteristics in Panel A and lagged TCI measures in Panel B. The construction of the TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. Panel A includes style fixed effects. Mutual funds are classified into size/value style categories based on their four-factor loadings as described in Section 3.5. Newey-West (1987) *t*-statistics with a lag of 3 are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. We multiply the coefficients of Log(TNA) and Log(Age) by 100 to ease presentation. The sample period is from 2006Q3 to 2023Q4, except for OCI (2006Q3 to 2017Q2).

Panel A. TCI and Fund Characteristics

	(1)	(2)	(3)	(4)	(5)
Log(TNA)$_{q-1}$	-0.038	-0.045*	-0.024	-0.035	-0.014
	(-1.56)	(-1.85)	(-1.00)	(-1.04)	(-0.38)
Log(Age)$_{q-1}$	-0.287***	-0.286***	-0.176***	-0.082	-0.299***
	(-3.86)	(-4.25)	(-3.43)	(-0.68)	(-3.94)
Turnover $_{q-1}$	-0.358***	-0.082	-0.445***	-0.233***	-0.250***
	(-4.60)	(-1.04)	(-5.30)	(-2.73)	(-2.73)
Expenses $_{q-1}$	0.018***	0.012***	0.013***	0.013***	0.015***
	(8.63)	(6.76)	(6.58)	(6.12)	(6.20)
$\sigma^{Ret}{}_{q-1}$	0.012***	0.009***	0.013***	0.011***	0.011***
	(4.82)	(4.88)	(5.63)	(4.48)	(3.89)
Flow $_{q-1}$	0.001	0.001	0.001	0.001	0.001
	(1.41)	(0.37)	(0.59)	(0.87)	(0.82)
$\alpha^{4F}{}_{q-1}$	0.008***	0.004	0.009**	0.009***	0.010***
	(2.83)	(1.59)	(2.65)	(2.75)	(2.92)
ICI $_{q-1}$		0.584***			
		(7.08)			
R^2_{q-1}			-0.154***		
			(-5.68)		
ActvShr $_{q-1}$				0.085***	
				(5.59)	
OCI $_{q-1}$ (x100)					0.057***
					(8.97)
Style FEs	Yes	Yes	Yes	Yes	Yes
N. of Obs.	79,529	79,523	79,529	59,763	47,962
N. of Quarters	69	69	69	69	44
Adj. R^2	0.188	0.324	0.250	0.244	0.251

Panel B. Persistence of TCI

	(1)	(2)	(3)	(4)
TCI_{q-1}	0.881***			
	(7.59)			
TCI_{q-2}		0.827***		
		(9.23)		
TCI_{q-3}			0.804***	
			(8.76)	
TCI_{q-4}				0.768***
				(7.58)
N. of Obs.	79,444	77,413	75,003	73,000
N. of Quarters	69	68	67	66
Adj. R^2	0.848	0.686	0.603	0.540

Table 3. TCI and Mutual Fund Performance: Portfolio Analysis

This table reports the equal-weighted (TNA-weighted) future returns of mutual funds sorted into decile portfolios according to TCI in Panel A (Panel B). Each month, all funds in the sample are ranked into decile portfolios based on the previous month's TCI. For each decile portfolio, we then calculate both the equal- and TNA-weighted average returns for each month. R^{Net} is the net return. $\alpha^{4F,Net}$ ($\alpha^{4F,Gross}$) is the net (gross) alpha measured as the intercept from the Carhart (1997) four-factor model regression based on the corresponding time series of the monthly average returns for each decile portfolio. CS is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). The table also reports differences in fund returns and four-factor alphas between the top and bottom decile portfolios. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. All returns are at monthly frequency and expressed in percentage points. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. Spearman rank correlations are also included. The sample period is from 2006Q3 to 2023Q4.

	Panel A. Equal-Weighted					Panel B. TNA-weighted				
	R^{Net}	$\alpha^{3F,Net}$	$\alpha^{4F,Net}$	$\alpha^{4F,Gross}$	CS	R^{Net}	$\alpha^{3F,Net}$	$\alpha^{4F,Net}$	$\alpha^{4F,Gross}$	CS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
D10	0.953***	0.075	0.077	0.171**	0.108***	1.017***	0.101	0.106	0.182**	0.120***
	(2.79)	(1.01)	(1.01)	(2.24)	(3.07)	(2.94)	(1.39)	(1.42)	(2.46)	(3.13)
D9	0.870**	-0.024	-0.024	0.064	0.059*	0.955***	0.027	0.032	0.107*	0.093***
	(2.52)	(-0.39)	(-0.37)	(0.99)	(1.65)	(2.78)	(0.45)	(0.53)	(1.76)	(3.01)
D8	0.846**	-0.041	-0.041	0.043	0.051	0.908***	-0.023	-0.021	0.052	0.080***
	(2.50)	(-0.80)	(-0.77)	(0.82)	(1.56)	(2.68)	(-0.48)	(-0.43)	(1.07)	(2.75)
D7	0.826**	-0.064	-0.064	0.019	0.023	0.884***	-0.046	-0.045	0.027	0.059*
	(2.45)	(-1.27)	(-1.25)	(0.37)	(0.73)	(2.61)	(-1.00)	(-0.97)	(0.58)	(1.97)
D6	0.800**	-0.091*	-0.093*	-0.011	0.022	0.841**	-0.093*	-0.096*	-0.025	0.031
	(2.38)	(-1.89)	(-1.91)	(-0.23)	(0.71)	(2.47)	(-1.86)	(-1.90)	(-0.49)	(0.87)
D5	0.783**	-0.107**	-0.108**	-0.026	-0.004	0.827**	-0.100**	-0.101**	-0.030	0.008
	(2.33)	(-2.38)	(-2.38)	(-0.58)	(-0.13)	(2.47)	(-2.08)	(-2.05)	(-0.62)	(0.23)
D4	0.745**	-0.155***	-0.156***	-0.075	-0.030	0.784**	-0.139**	-0.139**	-0.068	-0.020
	(2.15)	(-3.06)	(-3.03)	(-1.45)	(-0.94)	(2.27)	(-2.34)	(-2.30)	(-1.13)	(-0.59)
D3	0.717**	-0.183***	-0.185***	-0.104*	-0.044	0.722**	-0.185***	-0.184***	-0.113*	-0.050
	(2.08)	(-3.35)	(-3.32)	(-1.86)	(-1.36)	(2.13)	(-3.13)	(-3.04)	(-1.86)	(-1.45)
D2	0.686*	-0.222***	-0.224***	-0.142**	-0.069*	0.706**	-0.217***	-0.219***	-0.148***	-0.086**
	(1.96)	(-3.65)	(-3.64)	(-2.31)	(-1.94)	(2.03)	(-3.84)	(-3.84)	(-2.60)	(-2.05)
D1	0.632*	-0.277***	-0.278***	-0.197***	-0.085**	0.644*	-0.302***	-0.302***	-0.231***	-0.106**
	(1.94)	(-5.42)	(-3.98)	(-2.69)	(-2.17)	(2.04)	(-4.82)	(-3.77)	(-2.63)	(-2.06)
	Difference: D10 – D1									
	0.320***	0.352***	0.355***	0.368***	0.193***	0.373***	0.402***	0.408***	0.414***	0.226***
	(3.36)	(3.41)	(3.39)	(3.52)	(4.67)	(3.59)	(3.45)	(3.45)	(3.51)	(4.20)
Spearman rank corr.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4. TCI and Fund Performance: Fama-MacBeth Regressions

This table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent month's TCI and other fund characteristics. The construction of TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. In Panel A (Panel B), the dependent variable is $\alpha_t^{4F,Net}$ ($\alpha_t^{4F,Gross}$), the fund's one-month Carhart (1997) four-factor alpha net (gross) of fees. To compute four-factor alpha ($\alpha_{i,t}^{4F}$) for fund i in month t, we first calculate factor loadings by estimating the Carhart (1997) four-factor model using the prior 36 monthly fund returns (requiring at least 20 observations). We then calculate $\alpha_{i,t}^{4F}$ as the fund's actual return minus the expected return implied by these loadings and the contemporaneous factor realizations. In Panel C, $CS_{i,t}$ is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Mutual funds are classified into size/value style categories based on fund's four-factor loadings described in Section 3.5. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. The sample period is from 2006Q3 to 2023Q4.

	Panel A. $\alpha_{i,t}^{4F,Net}$			Panel B. $\alpha_{i,t}^{4F,Gross}$			Panel C. $CS_{i,t}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
TCI_{t-1}	1.712***	1.424***	1.143***	1.764***	1.429***	1.148***	2.118***	1.801***	1.110***
	(4.03)	(4.31)	(4.16)	(4.21)	(4.33)	(4.18)	(4.42)	(4.17)	(3.64)
$Log(TNA)_{t-1}$		0.006*	0.004		0.006	0.004		0.001	0.000
		(1.65)	(1.58)		(1.55)	(1.44)		(0.32)	(0.02)
$Log(Age)_{t-1}$		0.003	0.005		0.001	0.003		0.014*	0.004
		(0.42)	(0.84)		(0.17)	(0.52)		(1.85)	(0.81)
$Expenses_{t-1}$		-0.029**	-0.032***		0.049***	0.046***		0.012	0.009
		(-2.11)	(-2.68)		(3.51)	(3.86)		(0.89)	(1.02)
$Turnover_{t-1}$		-0.066***	-0.055***		-0.066***	-0.055***		-0.010	-0.012
		(-4.67)	(-4.29)		(-4.64)	(-4.26)		(-0.75)	(-1.11)
$\sigma^{Ret}_{t-12:t-1}$		-0.041**	-0.049**		-0.041**	-0.050**		-0.016	-0.003
		(-2.24)	(-2.33)		(-2.24)	(-2.35)		(-0.67)	(-0.13)
$Flow_{t-12:t-1}$		0.038***	0.039***		0.039***	0.040***		0.004	0.005
		(3.91)	(4.55)		(3.95)	(4.59)		(0.37)	(0.81)
Style FE	No	No	Yes	No	No	Yes	No	No	Yes
N. of Obs.	256,893	256,893	256,893	256,893	256,893	256,893	252,701	252,701	252,701
N. of Months	210	210	210	210	210	210	210	210	210
Adj. R^2	0.021	0.075	0.182	0.021	0.076	0.183	0.019	0.069	0.187

Table 5. TCI, Fund Activeness, and Future Performance

This table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent quarter's TCI, ICI, ActvShr, R^2, OCI and other fund characteristics. The construction of TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. In Panel A (Panel B), $\alpha_t^{4F,Net}(\alpha_t^{4F,Gross})$ is the fund's one-month Carhart (1997) four-factor net (gross) alpha. In Panel C, CS_t is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Newey-West (1987) $t-$ statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. The sample period is from 2006Q3 to 2023Q4 except for OCI (2006Q3 to 2017:Q2).

	Panel A. $\alpha_{i,t}^{4F,Net}$				
	(1)	(2)	(3)	(4)	(5)
TCI_{t-1}	1.239***	1.396***	1.141***	1.643***	1.843***
	(4.22)	(3.70)	(4.04)	(3.90)	(3.19)
ICI_{t-1}	0.594**				0.948**
	(2.26)				(2.41)
$ActvShr_{t-1}$		0.226			0.029
		(1.62)			(0.21)
R_{t-1}^2			-0.274*		-0.451**
			(-1.91)		(-2.58)
OCI_{t-1}				0.066*	0.059**
				(1.94)	(2.24)
Adj. R^2	0.187	0.205	0.189	0.174	0.199
	Panel B. $\alpha_{i,t}^{4F,Gross}$				
TCI_{t-1}	1.244***	1.400***	1.147***	1.653***	1.851***
	(4.24)	(3.71)	(4.06)	(3.93)	(3.21)
ICI_{t-1}	0.594**				0.959**
	(2.26)				(2.45)
$ActvShr_{t-1}$		0.226			0.029
		(1.63)			(0.21)
R_{t-1}^2			-0.273*		-0.454**
			(-1.91)		(-2.60)
OCI_{t-1}				0.065*	0.059**
				(1.92)	(2.22)
Adj. R^2	0.187	0.205	0.189	0.174	0.199
	Panel C. $CS_{i,t}$				
TCI_{t-1}	1.162***	1.489***	1.123***	1.515***	1.827**
	(3.31)	(3.24)	(3.34)	(3.04)	(2.40)
ICI_{t-1}	0.529**				0.929***
	(2.18)				(2.70)
$ActvShr_{t-1}$		0.035			-0.126
		(0.43)			(-1.28)
R_{t-1}^2			-0.224*		-0.559***
			(-1.92)		(-3.38)
OCI_{t-1}				0.035	0.036
				(1.36)	(1.24)
Adj. R^2	0.191	0.199	0.191	0.159	0.172
No of Obs.	256,869	187,736	256,893	155,400	122,447
Control Variables	Yes	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	Yes	Yes	Yes
No of Months	210	210	209	132	132

Table 6. Stock-level Thematic Exposure Concentration and Stock Returns

Panel A of this table reports monthly equal-weighted and value-weighted average excess return (R_e), CAPM (α^{CAPM}), the Fama and French three-factor alpha (α^{3F}), and the Carhart (1997) four-factor alpha (α^{4F}) to decile portfolios of stocks sorted on stock-level thematic concentration measure, TCI^{Stock}, the stock-level analog of fund TCI. The construction of $TCI^{Stock}$is described in Section 4.4. Panel B reports the results from Fama-MacBeth (1973) regressions of future stock returns over the next month (Ret_t) on lagged TCI^{Stock}. In Panel B, we control for size (Log(size)), book-to-market (Log(B/M)), operating profitability (Operating Prof.), past performance prior one month (Ret_{t-1}), past return from twelve to two months ($Ret_{t-12:t-2}$). Industry fixed effects are based on industries defined by the Fama and French 48 industry classifications. Newey-West (1987) $t-$ statistics with a lag of 3 are reported in parentheses. All returns are expressed in %. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

Panel A. Portfolio Sorting Analysis

	Equal-Weighted				Size-Weighted			
	R_e	α^{CAPM}	α^{3F}	α^{4F}	R_e	α^{CAPM}	α^{3F}	α^{4F}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D10	0.813	-0.330	-0.201	-0.162	0.755	-0.284	-0.164	-0.154
	(1.32)	(-0.99)	(-0.90)	(-0.77)	(1.55)	(-0.87)	(-0.64)	(-0.62)
D9	0.813	-0.250	-0.091	-0.058	0.939*	0.031	0.076	0.093
	(1.19)	(-0.62)	(-0.30)	(-0.20)	(1.94)	(0.12)	(0.31)	(0.40)
D8	0.741	-0.262	-0.134	-0.088	0.815**	-0.032	-0.031	-0.018
	(1.34)	(-1.01)	(-0.69)	(-0.49)	(2.04)	(-0.22)	(-0.24)	(-0.14)
D7	0.865*	-0.141	-0.019	0.019	1.051***	0.215	0.213	0.213
	(1.71)	(-0.68)	(-0.14)	(0.17)	(2.97)	(1.30)	(1.39)	(1.38)
D6	0.746	-0.271	-0.148	-0.109	0.844***	0.055	0.056	0.071
	(1.39)	(-1.15)	(-0.89)	(-0.70)	(2.70)	(0.59)	(0.61)	(0.81)
D5	0.979**	-0.000	0.141	0.172	0.918***	0.118	0.133	0.158
	(1.97)	(-0.00)	(0.89)	(1.17)	(2.73)	(0.85)	(0.93)	(1.16)
D4	0.761	-0.247	-0.110	-0.054	0.714*	-0.152	-0.155	-0.126
	(1.34)	(-0.88)	(-0.49)	(-0.26)	(1.92)	(-1.14)	(-1.16)	(-1.04)
D3	0.859	-0.169	-0.042	-0.000	0.875**	0.067	0.074	0.092
	(1.41)	(-0.53)	(-0.17)	(-0.00)	(2.38)	(0.43)	(0.49)	(0.65)
D2	0.730	-0.339	-0.190	-0.155	0.744*	-0.161	-0.140	-0.132
	(1.31)	(-1.17)	(-0.87)	(-0.73)	(1.80)	(-0.94)	(-0.91)	(-0.82)
D1	0.944*	-0.103	0.051	0.091	0.680**	-0.158	-0.135	-0.121
	(1.68)	(-0.32)	(0.21)	(0.43)	(2.03)	(-0.73)	(-0.61)	(-0.55)
	Difference: D10 – D1							
D10-D1	-0.131	-0.227	-0.252	-0.253	0.075	-0.126	-0.028	-0.033
	(-0.47)	(-0.83)	(-0.95)	(-0.94)	(0.22)	(-0.39)	(-0.10)	(-0.12)

Panel B. Fama-MacBeth (1973) regression analysis

	(1)	(2)	(3)
TCI^{Stock}_{t-1}	-0.022	0.010	0.038
	(-0.34)	(0.15)	(0.68)
Log(Size) $_{t-1}$	0.030	-0.040	0.015
	(0.36)	(-0.60)	(0.20)
Log(B/M) $_{t-1}$	0.214*	0.158	0.283***
	(1.76)	(1.38)	(2.85)
Ret_{t-1}	-0.023***	-0.020***	-0.005
	(-3.66)	(-3.45)	(-0.27)
$Ret_{t-12:t-2}$	-0.001	-0.002	-0.008
	(-0.21)	(-0.37)	(-1.06)
Profitability $_{t-1}$		1.878***	1.454
		(3.49)	(1.49)
Investment $_{t-1}$		-0.247*	-0.467**
		(-1.90)	(-2.04)
Industry FE	No	No	Yes
Avg. N	2,031	2,031	2,031
Adj. R^2	0.029	0.037	0.056

Table 7. TCI, Thematic Timing Ability, and Fund Performance: Fama-MacBeth Regressions

This table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent month's TCI, fund-level thematic timing ability measure ($vRet_t$ or $vRet_{t-2:t}$), and other fund characteristics. The construction of TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. $vRet_t$ ($vRet_{t-2:t}$) is one- (thee-) month theme-related market timing measure described in Section 4.4. In Panel A (Panel B) $\alpha_t^{4F,Net}$ ($\alpha_t^{4F,Gross}$) is the fund's one-month Carhart (1997) four-factor alpha and is obtained from the fund's excess net (gross) return less the sum of the products of each of the four-factor realizations estimated using the preceding 36 monthly fund returns. In Panel C CS_t is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Mutual funds are classified into size/value style categories based on fund's four-factor loadings described in Section 3.5. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. The sample period is from 2006Q3 to 2023Q4.

	Panel A. $\alpha_{i,t}^{4F,Net}$		Panel B. $\alpha_{i,t}^{4F,Gross}$		Panel C. $CS_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
TCI_{t-1}	0.438***	0.438***	0.445***	0.445***	0.389***	0.389***
	(3.24)	(3.24)	(3.29)	(3.29)	(2.73)	(2.73)
$vRet_t$	0.004		0.017		-0.011	
	(0.02)		(0.07)		(-0.08)	
$vRet_{t-2:t}$		-0.104		-0.103		-0.026
		(-1.57)		(-1.56)		(-0.54)
$Log(TNA)_{t-1}$	0.004	0.004	0.004	0.004	-0.000	-0.000
	(1.56)	(1.56)	(1.42)	(1.42)	(-0.01)	(-0.01)
$Log(Age)_{t-1}$	0.007	0.007	0.005	0.005	0.007	0.007
	(1.14)	(1.14)	(0.80)	(0.80)	(1.29)	(1.29)
$Expenses_{t-1}$	-0.032***	-0.032***	0.046***	0.046***	0.009	0.009
	(-2.68)	(-2.68)	(3.86)	(3.86)	(1.07)	(1.07)
$Turnover_{t-1}$	-0.055***	-0.055***	-0.055***	-0.055***	-0.012	-0.012
	(-4.29)	(-4.29)	(-4.27)	(-4.27)	(-1.11)	(-1.11)
$\sigma^{Ret}{}_{t-12:t-1}$	-0.049**	-0.049**	-0.049**	-0.049**	-0.003	-0.003
	(-2.31)	(-2.31)	(-2.33)	(-2.33)	(-0.12)	(-0.12)
$Flow_{t-12:t-1}$	0.039***	0.039***	0.040***	0.040***	0.005	0.005
	(4.50)	(4.50)	(4.55)	(4.55)	(0.74)	(0.74)
Style FE	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	256,893	256,893	256,893	256,893	252,701	252,701
N. of Months	210	210	210	210	210	210
Adj. R^2	0.183	0.183	0.183	0.183	0.188	0.188

Table 8. Trade-based Performance of TCI Sorted Funds

This table reports the subsequent DGTW characteristic-adjusted performance and earnings announcement abnormal returns of stocks purchased and sold by funds sorted on TCI as described in Section 3.4. At the end of quarter q-1, we sort funds into decile portfolios based on the TCI measure. Within each fund, we break down fund trades into buy and sell trades as described in Section 4.5. Panel A reports the time-series mean quarterly DGTW benchmark-adjusted returns of trades by mutual funds in quarter q. Panel B reports the time-series mean earnings announcement abnormal return of trades by mutual funds. The earnings announcement abnormal returns are defined as the cumulative market-adjusted return over a three-day window [-1, +1] around the next earnings announcement date. The table also reports differences (shown in columns labeled as "Difference") in DGTW benchmark-adjusted returns and earnings announcement abnormal returns between Buy and Sell trade portfolios. Newey-West (1987) t – statistics with the lag of 3 are reported in parentheses. All returns are expressed in %. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

	Panel A. DGTW Benchmark-adjusted returns			Panel B. Cumulative Market-adjusted Return over [-1, +1] around Earnings Announcement		
	Buy (1)	Sell (2)	Difference (3)	Buy (4)	Sell (5)	Difference (6)
All Funds	0.277	-0.011	0.289	0.385***	0.117	0.268
	(0.87)	(-0.04)	(0.65)	(3.02)	(0.73)	(1.30)
D10	0.855*	-0.236	1.091**	0.861***	0.200	0.662***
	(1.73)	(-0.64)	(2.42)	(3.38)	(0.88)	(2.68)
D9	0.668	-0.243	0.911***	0.859***	0.279	0.580**
	(1.54)	(-0.64)	(3.16)	(4.06)	(1.15)	(2.42)
D8	0.349	0.139	0.210	0.555***	0.372**	0.183
	(0.92)	(0.44)	(0.82)	(3.37)	(2.16)	(0.92)
D7	0.353	-0.006	0.359	0.374**	0.474**	-0.121
	(0.87)	(-0.02)	(1.43)	(2.01)	(2.15)	(-0.64)
D6	0.040	0.343	-0.302	0.173	0.360*	-0.140
	(0.10)	(1.00)	(-1.24)	(0.87)	(1.83)	(-0.78)
D5	0.086	0.129	-0.043	0.128	0.313	-0.185
	(0.22)	(0.36)	(-0.15)	(0.64)	(1.58)	(-0.93)
D4	0.026	0.026	0.000	0.550**	0.211	0.260
	(0.07)	(0.08)	(0.00)	(2.07)	(1.10)	(0.96)
D3	0.006	-0.265	0.271	-0.085	0.112	-0.197
	(0.02)	(-0.71)	(0.90)	(-0.50)	(0.62)	(-1.46)
D2	0.271	0.113	0.159	0.139	0.118	0.021
	(0.77)	(0.31)	(0.71)	(0.69)	(0.67)	(0.12)
D1	0.112	0.359	-0.248	0.088	0.002	0.086
	(0.29)	(0.91)	(-0.90)	(0.62)	(0.01)	(0.56)
			Difference: D10 – D1			
	0.743*	-0.596	1.339**	0.758***	0.122	0.636**
	(1.75)	(-1.33)	(2.47)	(2.71)	(0.49)	(2.12)

Table 9. Portfolio Manager Turnover, Change in Bachelor's Degree, and Fund-level TCI

This table reports the difference in the future TCI of funds associated with manager turnover with degree change and manager turnover without degree change. In Panel A, columns (1), (2), (3), and (4) report the differences in the future TCI of funds associated with the departing manager's degree over next one-, two-, three-, and four-quarter, respectively. In Panel B, columns (5), (6), (7), and (8) report the differences in the future TCI of funds associated with the new manager's degree over next one-, two-, three-, and four-quarter, respectively. This table also reports the difference-in-difference results between manager turnover with degree change and without degree change. *p*–values are reported in parentheses. The sample period is 2006Q3 to 2023Q4.

	Panel A. Themes Associated with the Departing Manager's Degree				Panel B. Themes Associated with the New Manager's Degree			
	Quarters After Degree Change				Quarters After Degree Change			
	(t+1) – t	*(t+2) – t*	*(t+3) – t*	*(t+4) – t*	*(t+1) – t*	*(t+2) – t*	*(t+3) – t*	*(t+4) – t*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Change in Degree	0.0267	-0.0560**	-0.0681***	-0.0757***	-0.0001	0.0183***	0.0202***	0.0245***
(N=2,156)	(0.50)	(-1.97)	(-3.00)	(-3.86)	(-0.07)	(10.13)	(10.51)	(12.18)
No Change in Degree	0.0025	0.0002	0.0020	0.0014	0.0025	0.0002	0.0020	0.0014
(N=4,821)	(0.55)	(0.06)	(0.82)	(0.70)	(0.55)	(0.06)	(0.82)	(0.70)
Differences								
	0.0242	-0.0562***	-0.0701***	-0.0771***	-0.0026	0.0182***	0.0182***	0.0231***
	(0.67)	(-2.87)	(-4.49)	(-5.73)	(-0.37)	(3.77)	(4.70)	(7.16)

Table 10. A Manager's Bachelor's Degree, TCI, and Fund Performance

This table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent month's TCI, $TCI^{Degree\ Related}$, and $TCI^{Degree\ Unrelated}$ and other fund characteristics. The construction of TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. TCI is decomposed into two components based on manager's related degree ($TCI^{Degree\ Unrelated}$) and manager's outside degree expertise ($TCI^{Degree\ Related}$) as described in Section 5.2. In Panel A (Panel B) $\alpha_t^{4F,Net}$ ($\alpha_t^{4F,Gross}$) is the fund's one-month Carhart (1997) four-factor alpha and is obtained from the fund's excess net (gross) return less the sum of the products of each of the four-factor realizations estimated using the preceding 36 monthly fund returns. In Panel C CS_t is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Mutual funds are classified into size/value style categories based on fund's four-factor loadings described in the paper. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. Columns (2), (4), and (6) of this table also report *p-value* of the hypothesis that the difference in $TCI^{(Degree\ Related)}$ and $TCI^{(Degree\ Unrelated)}$ is equal to zero. The sample period is from 2006Q3 to 2023Q4.

	Panel A. $\alpha_{i,t}^{4F,Net}$		Panel B. $\alpha_{i,t}^{4F,Gross}$		Panel C. $CS_{i,t}$	
	(1)	(2)	(3)	(4)	(5)	(6)
TCI_{t-1}	1.383***		1.371***		1.327***	
	(3.99)		(3.89)		(2.88)	
$TCI_{t-1}^{Degree\ Related}$		1.672***		1.663***		1.571***
		(4.35)		(4.24)		(3.13)
$TCI_{t-1}^{Degree\ Unrelated}$		1.381***		1.369***		1.317***
		(3.93)		(3.83)		(2.82)
$Log(TNA)_{t-1}$	0.004	0.004	0.004	0.004	0.003	0.003
	(1.37)	(1.44)	(1.26)	(1.32)	(0.79)	(0.83)
$Log(Age)_{t-1}$	0.011	0.011	0.009	0.010	0.009	0.010
	(1.52)	(1.57)	(1.28)	(1.35)	(1.33)	(1.49)
$Expenses_{t-1}$	-0.036***	-0.035***	0.043***	0.044***	0.014	0.016
	(-3.23)	(-3.09)	(3.85)	(3.91)	(1.00)	(1.05)
$Turnover_{t-1}$	-0.045***	-0.045***	-0.048***	-0.048***	-0.019	-0.019
	(-3.20)	(-3.24)	(-3.23)	(-3.28)	(-1.20)	(-1.27)
$\sigma^{Ret}{}_{t-12:t-1}$	-0.054**	-0.053**	-0.052**	-0.051**	0.006	0.008
	(-2.33)	(-2.29)	(-2.27)	(-2.22)	(0.25)	(0.31)
$Flow_{t-12:t-1}$	0.050***	0.050***	0.051***	0.051***	0.009	0.009
	(3.99)	(3.96)	(4.02)	(3.99)	(0.77)	(0.83)
Style FE	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	165,500	165,500	165,500	165,500	163,003	163,003
N. of Months	210	210	210	210	210	210
Adj. R^2	0.195	0.195	0.196	0.195	0.202	0.204
Difference: $TCI_{t-1}^{Degree\ Related} - TCI_{t-1}^{Degree\ Unrelated}$						
		0.291***		0.294***		0.254**
(*p-value*)		*(0.002)*		*(0.002)*		*(0.022)*

Table 11. A Manager's Bachelor's Degree, TCI, and Trade-based Performance

This table reports the subsequent DGTW characteristic-adjusted performance and earnings announcement abnormal returns of degree-related and degree-unrelated stocks purchased and sold by funds sorted on TCI as described in Section 3.4. At the end of quarter q-1, we sort funds into decile portfolios based on the TCI measure. Within each fund, we break down fund trades into buy and sell trades as described in Section 4.5, and classify fund's trades during quarter q as Degree-Related or Degree-Unrelated as described in Section 5.3. Panel A reports the time-series mean quarterly DGTW benchmark-adjusted returns of trades by mutual funds in quarter q. Panel B reports the time-series mean earnings announcement abnormal return of trades by mutual funds. The earnings announcement abnormal returns are defined as the cumulative market-adjusted return over a three-day window [-1, +1] around the next earnings announcement date. The table also reports differences (shown in columns labeled as "Difference") in DGTW benchmark-adjusted returns and earnings announcement abnormal returns between Buy and Sell trade portfolios. Newey-West (1987) t – statistics with the lag of 3 are reported in parentheses. All returns are expressed in %. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

Panel A. DGTW benchmark-adjusted returns

	Degree-Related Stocks			Degree-Unrelated Stocks		
	Buys	Sells	Difference	Buys	Sells	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
D10	1.322**	-0.676	1.998***	0.575	-0.012	0.587
	(2.15)	(-1.20)	(3.40)	(1.07)	(-0.03)	(1.30)
D9	0.886	-0.758	1.645***	0.493	0.057	0.436
	(1.48)	(-1.49)	(3.49)	(1.03)	(0.13)	(0.97)
D8	0.671	-0.172	0.844*	0.377	0.264	0.113
	(1.34)	(-0.41)	(1.97)	(0.94)	(0.61)	(0.26)
D7	-0.004	0.183	-0.187	0.364	-0.347	0.712**
	(-0.01)	(0.52)	(-0.46)	(0.79)	(-0.96)	(2.12)
D6	-0.029	0.003	-0.032	-0.045	0.272	-0.317
	(-0.07)	(0.01)	(-0.10)	(-0.11)	(0.65)	(-0.94)
D5	-0.283	-0.113	-0.171	0.194	-0.087	0.281
	(-0.58)	(-0.26)	(-0.43)	(0.41)	(-0.23)	(0.82)
D4	0.056	-0.051	0.107	-0.447	0.034	-0.481
	(0.13)	(-0.13)	(0.38)	(-0.83)	(0.08)	(-1.29)
D3	0.153	-0.326	0.478	0.292	-0.290	0.582
	(0.37)	(-0.70)	(1.19)	(0.59)	(-0.69)	(1.37)
D2	0.293	-0.149	0.442	-0.028	0.319	-0.346
	(0.73)	(-0.34)	(1.08)	(-0.07)	(0.78)	(-1.28)
D1	0.054	0.511	-0.456	0.210	0.352	-0.142
	(0.11)	(1.23)	(-0.94)	(0.39)	(0.76)	(-0.37)
	Difference: D10 – D1					
	1.268*	-1.187*	2.455***	0.365	-0.364	0.729
	(1.97)	(-1.94)	(2.88)	(0.56)	(-0.74)	(1.33)

Panel B. Returns around earnings announcements

	Degree-Related Stocks			Degree-Unrelated Stocks		
	Buys	Sells	Difference	Buys	Sells	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
D10	1.112***	-0.111	1.223***	0.932***	0.351	0.581
	(3.30)	(-0.35)	(3.33)	(2.89)	(1.32)	(1.64)
D9	1.065***	0.228	0.837*	0.747***	0.285	0.461
	(3.59)	(0.51)	(1.80)	(2.89)	(1.21)	(1.60)
D8	0.622**	0.488**	0.134	0.552**	0.341	0.210
	(2.47)	(2.27)	(0.59)	(2.29)	(1.41)	(0.65)
D7	0.374	0.435*	-0.060	0.109	0.431*	-0.321
	(1.55)	(1.75)	(-0.27)	(0.43)	(1.81)	(-1.04)
D6	0.250	0.285	-0.035	0.232	0.465*	-0.233
	(1.14)	(1.31)	(-0.15)	(0.97)	(1.97)	(-0.97)
D5	0.128	0.070	0.058	0.239	0.669**	-0.430
	(0.48)	(0.33)	(0.23)	(1.05)	(2.61)	(-1.53)
D4	0.533*	0.171	0.362	0.552**	0.393*	0.159
	(1.78)	(0.73)	(0.96)	(2.32)	(1.79)	(0.59)
D3	-0.097	-0.038	-0.059	-0.063	0.182	-0.245
	(-0.47)	(-0.16)	(-0.26)	(-0.28)	(1.32)	(-1.41)
D2	-0.189	0.123	-0.312	0.021	0.086	-0.065
	(-0.67)	(0.57)	(-1.22)	(0.09)	(0.38)	(-0.25)
D1	-0.074	0.001	-0.076	0.116	0.089	0.027
	(-0.36)	(0.01)	(-0.36)	(0.64)	(0.40)	(0.12)
	Difference: D10 – D1					
	1.186***	-0.112	1.299***	0.837**	0.262	0.575
	(3.09)	(-0.29)	(2.93)	(2.34)	(0.78)	(1.25)

Internet Appendix

Thematic Concentration and Mutual Fund Performance

August 2025

Section A. Additional Tests

A.1. TCI Increasing and Decreasing Trading

One key analysis in the paper is examining the portfolio holding changes (i.e., trading activities) of theme-concentrated funds to gauge the source of their outperformance. The buy vs. sell differences in Panel A of Table 7 suggest that managers of high *TCI* funds are more successful in selecting stocks than those of less concentrated funds. To better understand the role of *TCI* in shaping fund performance, we further partition stock trading activities into *TCI*-increasing and *TCI*-decreasing trades.

As portfolio weight ω enters into both the numerator and denominator of $\Theta_{P,k}$, *TCI* is highly nonlinear with respect to ω. To determine the impact of a stock purchase (i.e., $\omega_i > \widetilde{\omega}_i$) on fund-level *TCI*, we proximate the numerical differentiation of *TCI* with respect to ω_i as follows.

$$\Delta TCI_i \equiv \frac{\partial TCI(\omega)}{\partial \omega_i} \approx \frac{TCI(\omega_i, \widetilde{\omega}_{-i}) - TCI(\widetilde{\omega}_i, \widetilde{\omega}_{-i})}{\omega_i - \widetilde{\omega}_i} = \frac{TCI(\omega_i, \widetilde{\omega}_{-i}) - TCI(\widetilde{\omega})}{\omega_i - \widetilde{\omega}_i}, \qquad \text{(A1)}$$

where ω_i is the portfolio weight associated with stock i, and ω denotes the vector of portfolio weights assigned to each stock in the fund portfolio. Correspondingly, $\widetilde{\omega}$ is the portfolio weight under the assumption of no trading. The numerator, $TCI(\omega_i, \widetilde{\omega}_{-i}) - TCI(\widetilde{\omega}_i, \widetilde{\omega}_{-i})$, isolates the impact of stock i's increased weight (from $\widetilde{\omega}_i$ to ω_i) on fund *TCI* while assuming no trading activities for the rest stocks in the portfolio (denoted as $-i$).[23]

We divide the sample of buy trades ($\omega_i - \widetilde{\omega}_i > 0$) into *TCI*-increasing ($\Delta TCI_i > 0$) and *TCI*-decreasing ($\Delta TCI_i < 0$) trades, and report their average DGTW-adjusted performance in columns (4) to (6) of Panel A, Table A15. As reported in column (6) the difference in TCI-increasing and TCI-decreasing stock buys create significantly more value for the funds in the top TCI decile than those in the bottom TCI decile (0.828 with t-statistics of 1.88) consistent with the notion that fund managers in the top TCI decile presumably possess investment expertise in certain investment themes.

In columns (7) to (9) of Panel A, we focus on stocks sold ($\omega_i - \widetilde{\omega}_i < 0$) and contrast the DGTW-adjusted stock performance of *TCI*-increasing ($\Delta TCI_i > 0$) vs. *TCI*-decreasing sells

[23] The sum of ω_i and $\widetilde{\omega}_{-i}$ may not be one. However, rescaling both (by the sum of ω_i and $\widetilde{\omega}_{-i}$) is not necessary as ω appears linearly in both the numerator and denominator of $\Theta_{P,k}$. We note that this analysis is computationally intensive as we need to calculate TCI twice [$TCI(\omega_i, \widetilde{\omega}_{-i})$ and $TCI(\widetilde{\omega}_i, \widetilde{\omega}_{-i})$] for each stock in a portfolio.

($\Delta TCI_i < 0$). No clear pattern is detected as the return difference remains statistically insignificant in both the top and bottom *TCI* deciles. This finding is consistent with Chen, Jegadeesh, and Wermers (2000). Since mutual funds do not short-sell stocks, their sell activities are less informative for uncovering their managers' potential investment skills.

In addition, we examine whether the trading of top *TCI*-decile funds is positively related to firms' fundamentals such as future earnings surprises. Like the tests above, in columns (4)-(9) of Panel B of Table A15, we analyze stock purchases and sales separately by breaking them down into *TCI*-increasing and *TCI*-decreasing trades. In column (4), we show that among stock purchased and sold by funds in the top *TCI* decile, those that increase *TCI* experience a statistically higher CAR than those in the bottom TCI decile. In summary, the results in Table A15 suggest that the outperformance from the trading activities of high *TCI* managers mainly come from *TCI*-increasing rather than *TCI*-decreasing trades.

A.2. Portfolio Analysis on Performance of Stocks in Overweighted Themes

We also conduct tests using a portfolio approach to examine the performance of securities held or purchased within "themes" that are over-weighted, i.e., stocks in over-weighted topics. If fund managers overweigh certain themes in which they have expertise, we expect that stocks in these over-weighted themes will perform particularly well. We identify overweighted topics as follows:

At the fund level, our initial task is to identify semantic themes that are held in higher proportions than expected. For example, using the theme distribution outlined in Equation (1), a specific theme is considered overweight when the fund's exposure ($\sum_{i \in P} \omega_i \times T_{i,k}$) surpasses the naïve exposure level of 1/100, assuming there are a total of 100 underlying themes. We then compute, at the individual stock level, each stock's cumulative exposure to the themes that were identified as over-weighted at the fund level. For instance, if a stock's exposure to 10 overweighted themes exceeds 0.1 (calculated as 10 times 1/100), the stock is categorized as having a thematic overweighting. This categorization is based on its contribution to the thematic concentration at the fund level within those overweighted themes.

Finally, we follow the portfolio approach of Pool et al. (2015) and Kumar et al. (2020). Specifically, in the portfolio tests, we form four distinct portfolios at the beginning of each quarter based on whether the fund bought or sold a stock during the previous quarter and whether the trade was of a stock with over-weighted topics (i.e., themes). Stocks that are bought are aggregated into the buy portfolio, while those that are sold are placed in the sell portfolio. We create two subgroups within the buy and sell portfolios: stocks with over-weighted themes and other stocks. We calculate the average monthly returns of these portfolios for each fund in each quarter, weighing each stock's return in the portfolios by the dollar trade value during the previous quarter. We rebalance at the end of the quarter.

Finally, we average the returns of each sub-portfolio across the funds in our sample using the dollar assets (TNA) of each fund in the previous quarter as weights, producing value-weighted average monthly returns for each of the four portfolios. We report the results in Table A16 of the Internet Appendix. Regardless of the performance measures used, the long-short portfolio of stocks in overweighted themes significantly outperforms the long-short portfolio of other stocks. Specifically, the difference-in-differences portfolio has a positive four-factor alpha and DGTW of 0.419% and 0.323% per month, respectively, both significant at the 5% level. This finding supports the idea that fund managers tend to overweight themes in which they have expertise, resulting in superior performance of stocks in these overweighted themes.

Section B. Results for Robustness Tests

This appendix provides additional results for the robustness tests discussed in the paper.

Figure A1. Thematic Concentration Index Over Time

This figure displays time-series averages of cross-sectional 10th percentile, 25th percentile, mean, 75th percentile, and 90th percentile values of the Thematic Concentration Index (TCI). The TCI is measured as the average monthly value for each year throughout our sample period from 2006Q3 to 2023Q4. The construction methodology for TCI is detailed in Section 3.4.

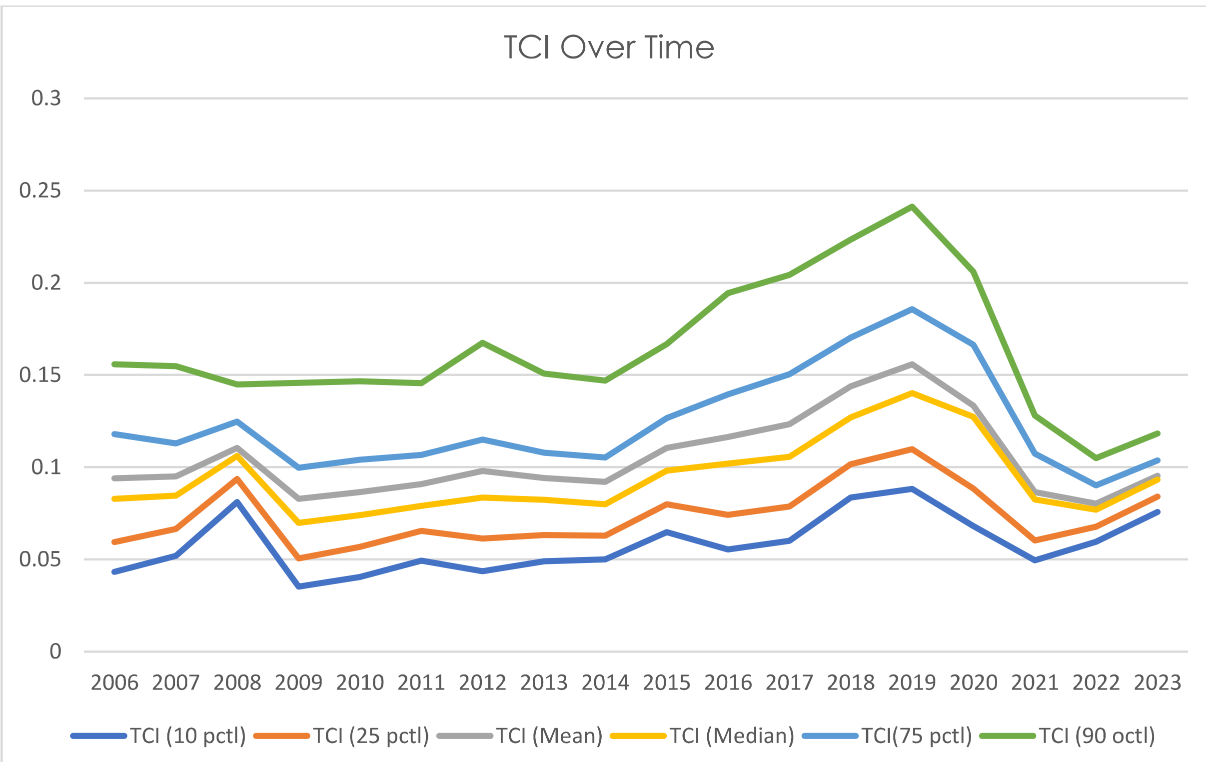


Figure A2. Themes Closely Related to Specific Disciplines

This figure illustrates four examples of topics extracted through LDA, representing themes that are closely associated with (or "covered" by) specific undergraduate disciplines such as Computer Science, Biochemistry, Economics, or Finance. See Section 5.1 for a detailed definition of a topic being covered by a particular field. The font size of bigrams represents the relative frequency of bigrams to a particular topic.

Panel A. Word Cloud covered by Computer Science Dept.

Panel B. Word Cloud covered by Biochemistry Dept

Panel C. Word Cloud covered by Economics Dept.

Panel D. Word Cloud covered by Finance but not Economics Dept.

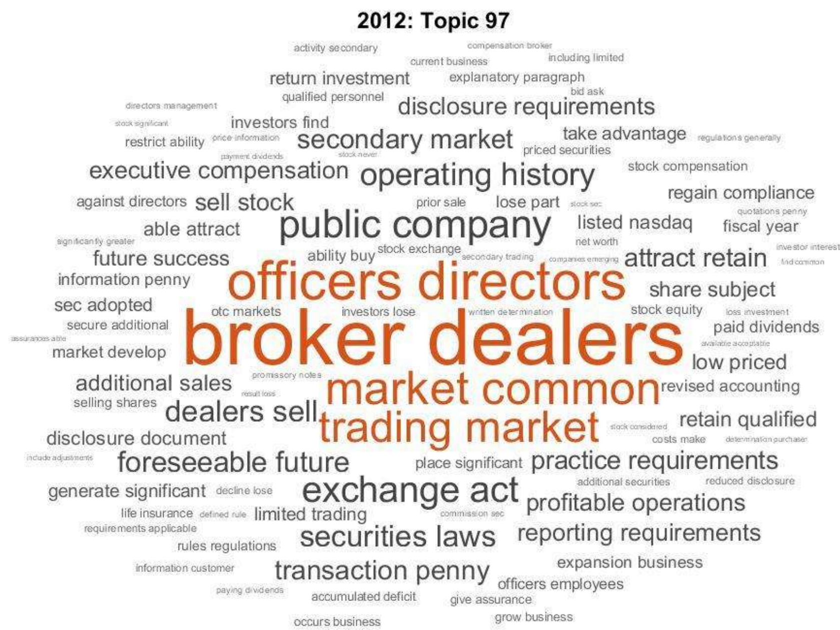


Table A1. The Number of Funds sorted by TCI and ICI dimensions

This table reports the time series averages of the number of funds as bivariate distributions along the Thematic Concentration Index (TCI) we propose and the Industry Concentration Index (ICI) of Kacperczyk, Sialm, and Zheng (2005). Mutual funds are independently sorted according to the most recent month's TCI and ICI into quintiles. The sample period is from 2006Q3 to 2023Q4.

		ICI Quintiles					
		Q1	Q2	Q3	Q4	Q5	Total
TCI Quintiles	Q5	88	54	40	32	31	245
	Q4	54	58	52	45	36	245
	Q3	42	54	54	54	41	245
	Q2	34	49	56	57	49	245
	Q1	27	30	44	57	88	245
	Total	246	245	245	245	245	1,225

Table A2. TCI and Mutual Fund Performance: Portfolio Analysis

This table reports the future returns (equal-weighted and TNA-weighted) of mutual funds sorted into decile portfolios according to the most recent month's TCI as described in Section 3.4. At the end of month, all funds in the sample are ranked into ten deciles based on TCI. $\alpha^{6F,Net}$ ($\alpha^{6F,Gross}$) the one-month net (gross) alpha measured as the intercept of the six-factor model [i.e., Carhart (1997) four factors plus the Fama and French (2015) profitability and investment factors]. The table also reports differences in performance between the top and bottom decile portfolios. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. All returns are expressed in percentage points. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

	$\alpha^{6F,Net}$		$\alpha^{6F,Gross}$	
	Equal-Weighted	TNA-Weighted	Equal-Weighted	TNA-Weighted
	(1)	(2)	(3)	(4)
D10	0.066	0.094	0.159**	0.171**
	(0.90)	(1.30)	(2.17)	(2.38)
D9	-0.046	0.039	0.042	0.113*
	(-0.74)	(0.67)	(0.67)	(1.96)
D8	-0.056	-0.009	0.028	0.064
	(-1.11)	(-0.19)	(0.56)	(1.36)
D7	-0.083*	-0.038	-0.000	0.034
	(-1.66)	(-0.85)	(-0.00)	(0.76)
D6	-0.103**	-0.079	-0.020	-0.008
	(-2.18)	(-1.61)	(-0.43)	(-0.16)
D5	-0.113**	-0.088*	-0.032	-0.018
	(-2.59)	(-1.97)	(-0.72)	(-0.40)
D4	-0.145***	-0.111**	-0.064	-0.040
	(-2.88)	(-1.99)	(-1.27)	(-0.71)
D3	-0.163***	-0.147**	-0.082	-0.076
	(-2.88)	(-2.40)	(-1.45)	(-1.23)
D2	-0.190***	-0.173***	-0.108*	-0.103*
	(-3.04)	(-2.99)	(-1.73)	(-1.77)
D1	-0.227***	-0.247***	-0.147**	-0.176**
	(-3.32)	(-3.26)	(-2.15)	(-2.33)
	Difference: D10 – D1			
	0.293***	0.341***	0.306***	0.347***
	(2.92)	(3.14)	(3.06)	(3.20)

Table A3. TCI and Mutual Fund Performance: Portfolio Analysis using Cederburg, O'Doherty, Savin, and Tiwari (2018) Conditional Alpha Measure

This table reports the future returns (equal-weighted and TNA-weighted) of mutual funds sorted into decile portfolios according to the most recent month's TCI as described in Section 3.4. At the end of month, all funds in the sample are ranked into ten deciles based on TCI. $\alpha^{4F,Net}$ ($\alpha^{4F,Gross}$) the one-month net (gross) conditional alpha measured as the intercept of the Carhart (1997) four-factor model from the Cederburg, O'Doherty, Savin, and Tiwari (2018) conditional model. The table also reports differences in performance between the top and bottom decile portfolios. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. All returns are expressed in percentage points. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

	$\alpha^{4F,Net}$		$\alpha^{4F,Gross}$	
	Equal-Weighted	TNA-Weighted	Equal-Weighted	TNA-Weighted
	(1)	(2)	(3)	(4)
D10	0.038	0.058	0.135**	0.132**
	(0.61)	(0.89)	(2.07)	(2.13)
D9	-0.044	0.001	0.075	0.044
	(-0.87)	(0.01)	(1.45)	(0.85)
D8	-0.057	-0.038	0.036	0.027
	(-1.40)	(-0.87)	(0.82)	(0.66)
D7	-0.080**	-0.055	0.018	0.004
	(-2.01)	(-1.34)	(0.43)	(0.10)
D6	-0.090**	-0.088*	-0.016	-0.008
	(-2.49)	(-1.91)	(-0.35)	(-0.21)
D5	-0.106***	-0.087*	-0.017	-0.024
	(-2.88)	(-1.95)	(-0.38)	(-0.65)
D4	-0.154***	-0.138**	-0.066	-0.072*
	(-3.61)	(-2.52)	(-1.21)	(-1.69)
D3	-0.175***	-0.167***	-0.096*	-0.094*
	(-3.64)	(-3.10)	(-1.77)	(-1.94)
D2	-0.215***	-0.196***	-0.125**	-0.133**
	(-4.00)	(-3.60)	(-2.30)	(-2.47)
D1	-0.262***	-0.280***	-0.209***	-0.181***
	(-4.35)	(-3.72)	(-2.79)	(-3.02)
	Difference: D10 - D1			
	0.300***	0.338***	0.344***	0.313***
	(3.41)	(3.39)	(3.47)	(3.57)

Table A4. TCI and Mutual Fund Performance: Portfolio Analysis using Firm-identified Risk Factors

This table reports the future returns (equal-weighted and TNA-weighted) of mutual funds sorted into decile portfolios according to the most recent month's TCI as described in Section 3.4. At the end of month, all funds in the sample are ranked into ten deciles based on TCI. In Panel A $\alpha^{4FIRF,Net}$ ($\alpha^{4FIRF,Gross}$) the one-month net (gross) alpha measured as the intercept of the four systematic Firm-Identified Risk Factors (FIRFs) based on Lopez-Lira (2023). In Panel B $\alpha^{4FIRF+1,Net}$ ($\alpha^{4FIRF+1,Gross}$) the one-month net (gross) alpha measured as the intercept of the four systematic Firm-Identified Risk Factors (FIRFs) plus the orthogonal component based on Lopez-Lira (2023). The FIRFs quantify firms' self-identified risks from their annual reports (10-K) in the section called Item 1A Risk Factors, along with an orthogonal component that explains the average returns unrelated to systematic risk (Lopez-Lira (2023)). The table also reports differences in performance between the top and bottom decile portfolios. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. All returns are expressed in percentage points. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

Panel A. Portfolio Analysis based on the Lopez-Lira's (2023) four systematic Firm-Identified Risk Factors (FIRFs).

	$\alpha^{4FIRF,Net}$		$\alpha^{4FIRF,Gross}$	
	Equal-Weighted	TNA-Weighted	Equal-Weighted	TNA-Weighted
	(1)	(2)	(3)	(4)
D10	0.145	0.196	0.241**	0.277**
	(1.18)	(1.64)	(1.98)	(2.33)
D9	0.071	0.148	0.162	0.225*
	(0.59)	(1.20)	(1.35)	(1.83)
D8	0.070	0.137	0.157	0.213*
	(0.60)	(1.12)	(1.35)	(1.74)
D7	0.073	0.149	0.160	0.224*
	(0.63)	(1.19)	(1.38)	(1.80)
D6	0.057	0.118	0.142	0.192
	(0.50)	(0.94)	(1.24)	(1.54)
D5	0.051	0.099	0.135	0.172
	(0.48)	(0.88)	(1.27)	(1.53)
D4	0.016	0.058	0.100	0.132
	(0.15)	(0.54)	(0.94)	(1.23)
D3	-0.002	0.022	0.082	0.096
	(-0.02)	(0.20)	(0.77)	(0.86)
D2	-0.026	0.019	0.059	0.092
	(-0.24)	(0.16)	(0.54)	(0.80)
D1	-0.071	-0.095	0.012	-0.021
	(-0.59)	(-0.69)	(0.10)	(-0.16)
Difference: D10 – D1				
	0.216**	0.291**	0.229**	0.298**
	(2.11)	(2.40)	(2.25)	(2.48)

Panel B. Portfolio Analysis based on the Lopez-Lira's (2023) four systematic Firm-Identified Risk Factors (FIRFs) plus an orthogonal component.

	$\alpha^{4FIRF+1,Net}$		$\alpha^{4FIRF+1,Gross}$	
	Equal-Weighted	TNA-Weighted	Equal-Weighted	TNA-Weighted
	(1)	(2)	(3)	(4)
D10	0.138	0.191	0.235**	0.272**
	(1.18)	(1.64)	(2.00)	(2.34)
D9	0.064	0.142	0.155	0.220*
	(0.56)	(1.19)	(1.37)	(1.84)
D8	0.062	0.131	0.150	0.207*
	(0.58)	(1.12)	(1.38)	(1.76)
D7	0.066	0.143	0.152	0.218*
	(0.60)	(1.19)	(1.39)	(1.81)
D6	0.050	0.114	0.135	0.188
	(0.47)	(0.93)	(1.26)	(1.54)
D5	0.045	0.094	0.130	0.167
	(0.44)	(0.86)	(1.28)	(1.53)
D4	0.011	0.054	0.095	0.128
	(0.11)	(0.52)	(0.93)	(1.23)
D3	-0.006	0.018	0.078	0.092
	(-0.06)	(0.17)	(0.76)	(0.85)
D2	-0.029	0.015	0.056	0.089
	(-0.27)	(0.14)	(0.52)	(0.79)
D1	-0.073	-0.097	0.009	-0.024
	(-0.61)	(-0.71)	(0.08)	(-0.18)
Difference: D10 – D1				
	0.212**	0.288**	0.225**	0.295**
	(2.09)	(2.37)	(2.24)	(2.44)

Table A5. TCI and Mutual Fund Performance: Portfolio Analysis – Subsample Analysis

This table reports the equal-weighted (TNA-weighted) future returns of mutual funds sorted into decile portfolios according to the most recent month's TCI during the sample period from 2006Q3 to 2014Q4 in Panel A and 2015Q1 to 2023Q4 in Panel B. At the end of month, all funds in the sample are ranked into ten deciles based on TCI. R^{Net} is the one-month net return. $\alpha^{4F,Net}$ ($\alpha^{4F,Gross}$) is the one-month net (gross) alpha measured as the intercept of the Carhart (1997) four-factor model. CS is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). The table also reports differences in fund returns and four-factor alphas between the top and bottom decile portfolios. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. All returns are expressed in percentage points. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively.

Panel A. Sample Period from 2006Q3 to 2014Q4.

	Equal-Weighted					TNA-weighted				
	R^{Net}	$\alpha^{3F,Net}$	$\alpha^{4F,Net}$	$\alpha^{4F,Gross}$	CS	R^{Net}	$\alpha^{3F,Net}$	$\alpha^{4F,Net}$	$\alpha^{4F,Gross}$	CS
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
D10	0.886	0.022	0.019	0.122	0.075	0.893	0.011	0.007	0.095	0.090
	(1.62)	(0.27)	(0.25)	(1.60)	(1.28)	(1.61)	(0.12)	(0.08)	(1.19)	(1.61)
D9	0.830	-0.044	-0.045	0.053	0.045	0.856	-0.019	-0.022	0.064	0.071
	(1.49)	(-0.73)	(-0.78)	(0.90)	(0.72)	(1.56)	(-0.28)	(-0.37)	(1.05)	(1.48)
D8	0.780	-0.074	-0.076	0.018	0.024	0.775	-0.088	-0.090	-0.007	0.048
	(1.42)	(-1.34)	(-1.40)	(0.34)	(0.39)	(1.42)	(-1.34)	(-1.43)	(-0.10)	(0.97)
D7	0.771	-0.079	-0.080	0.012	-0.005	0.761	-0.094	-0.095	-0.014	0.024
	(1.42)	(-1.59)	(-1.63)	(0.24)	(-0.09)	(1.39)	(-1.54)	(-1.57)	(-0.23)	(0.46)
D6	0.746	-0.106*	-0.106*	-0.015	0.002	0.758	-0.104	-0.104	-0.023	0.012
	(1.38)	(-1.90)	(-1.90)	(-0.27)	(0.04)	(1.39)	(-1.42)	(-1.41)	(-0.32)	(0.21)
D5	0.745	-0.099*	-0.099*	-0.009	-0.016	0.771	-0.086	-0.087	-0.008	0.008
	(1.39)	(-1.95)	(-1.95)	(-0.17)	(-0.25)	(1.44)	(-1.39)	(-1.39)	(-0.13)	(0.13)
D4	0.715	-0.140**	-0.141**	-0.052	-0.028	0.758	-0.119	-0.120	-0.041	0.005
	(1.29)	(-2.50)	(-2.50)	(-0.92)	(-0.49)	(1.35)	(-1.51)	(-1.52)	(-0.51)	(0.09)
D3	0.692	-0.163**	-0.163**	-0.074	-0.037	0.718	-0.135*	-0.137*	-0.058	-0.014
	(1.26)	(-2.62)	(-2.60)	(-1.18)	(-0.68)	(1.31)	(-1.86)	(-1.94)	(-0.82)	(-0.28)
D2	0.660	-0.194***	-0.194***	-0.104	-0.045	0.712	-0.137**	-0.136**	-0.058	-0.052
	(1.19)	(-2.93)	(-2.90)	(-1.56)	(-0.77)	(1.29)	(-2.21)	(-2.20)	(-0.94)	(-0.81)
D1	0.570	-0.272***	-0.272***	-0.186***	-0.078	0.554	-0.322***	-0.322***	-0.245**	-0.109*
	(1.04)	(-4.69)	(-3.93)	(-2.72)	(-1.60)	(0.95)	(-3.02)	(-2.66)	(-2.03)	(-1.77)
	Difference: D10 - D1									
	0.316**	0.294**	0.291**	0.308***	0.154***	0.339**	0.333**	0.329**	0.340**	0.199***
	(2.59)	(2.46)	(2.51)	(2.68)	(3.49)	(2.41)	(2.10)	(2.11)	(2.20)	(3.04)

Panel B. Sample Period from 2015Q1 to 2023Q4.

	Equal-Weighted					TNA-weighted				
	R^{Net}	$\alpha^{3F,Net}$	$\alpha^{4F,Net}$	$\alpha^{4F,Gross}$	CS	R^{Net}	$\alpha^{3F,Net}$	$\alpha^{4F,Net}$	$\alpha^{4F,Gross}$	CS
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
D10	1.019**	0.052	0.051	0.135	0.141***	1.142***	0.128	0.129	0.194**	0.150***
	(2.49)	(0.62)	(0.61)	(1.61)	(3.65)	(2.74)	(1.41)	(1.38)	(2.09)	(2.89)
D9	0.910**	-0.075	-0.072	0.005	0.073**	1.053**	0.031	0.036	0.099	0.114***
	(2.20)	(-1.00)	(-0.94)	(0.07)	(1.98)	(2.54)	(0.42)	(0.47)	(1.28)	(2.93)
D8	0.913**	-0.064	-0.061	0.014	0.079***	1.041**	0.010	0.012	0.074	0.113***
	(2.28)	(-1.12)	(-1.04)	(0.24)	(3.01)	(2.59)	(0.19)	(0.21)	(1.34)	(3.55)
D7	0.882**	-0.104	-0.102	-0.028	0.051*	1.006**	-0.028	-0.025	0.038	0.094***
	(2.19)	(-1.66)	(-1.62)	(-0.44)	(1.89)	(2.52)	(-0.47)	(-0.42)	(0.65)	(3.46)
D6	0.855**	-0.128**	-0.132**	-0.058	0.043*	0.923**	-0.106*	-0.117*	-0.056	0.050
	(2.13)	(-2.43)	(-2.52)	(-1.12)	(1.78)	(2.26)	(-1.67)	(-1.80)	(-0.86)	(1.24)
D5	0.821**	-0.160***	-0.160***	-0.086	0.007	0.882**	-0.142**	-0.145**	-0.084	0.008
	(2.02)	(-3.02)	(-2.94)	(-1.59)	(0.29)	(2.21)	(-2.23)	(-2.18)	(-1.26)	(0.25)
D4	0.774*	-0.208***	-0.210***	-0.136**	-0.032	0.810**	-0.195***	-0.196***	-0.132*	-0.044
	(1.86)	(-3.43)	(-3.35)	(-2.17)	(-1.11)	(2.01)	(-2.96)	(-2.84)	(-1.93)	(-1.31)
D3	0.741*	-0.234***	-0.236***	-0.163**	-0.050	0.725*	-0.257***	-0.264***	-0.200**	-0.086*
	(1.76)	(-3.35)	(-3.21)	(-2.21)	(-1.44)	(1.80)	(-3.32)	(-3.21)	(-2.45)	(-1.82)
D2	0.713*	-0.268***	-0.272***	-0.198**	-0.093**	0.700	-0.302***	-0.301***	-0.239**	-0.120**
	(1.66)	(-3.03)	(-2.92)	(-2.13)	(-2.25)	(1.66)	(-3.45)	(-3.23)	(-2.57)	(-2.26)
D1	0.695	-0.292***	-0.295***	-0.220**	-0.091*	0.735*	-0.286***	-0.293***	-0.228**	-0.103*
	(1.63)	(-3.74)	(-2.73)	(-2.03)	(-1.98)	(1.75)	(-3.58)	(-2.63)	(-2.06)	(-1.77)
Difference: D10 - D1										
	0.324**	0.344**	0.346**	0.354**	0.231***	0.407***	0.414**	0.422**	0.422**	0.253***
	(2.21)	(2.35)	(2.26)	(2.32)	(3.35)	(2.65)	(2.60)	(2.50)	(2.50)	(2.97)

Table A6. Information Ratios of TCI-sorted Portfolios of Mutual Funds

This table reports the information ratio and Sharpe Ratio of the equal-weighted or TNA-weighted portfolios of funds sorted according to the most recent month's TCI using net fund return and gross fund return in Panel A and Panel B, respectively. In columns (1) and (2) ((5) and (6)) of Panel A, IR is computed as the net alpha divided by the tracking error where the net (gross) alpha is measured as the intercept of the Carhart (1997) four-factor model. In columns (3) and (4) ((7) and (8)) of Panel B, a six-factor model is used by augmenting the Fama and French (2015) five-factor model with the momentum factor. Tracking error is the standard deviation of the error term from the factor model. The sample period is from 2006Q3 to 2023Q4.

Panel A. Information Ratio of Funds sorted on TCI.

	Information using Net Fund Return				Information Ratio using Gross Fund Return			
	Carhart four-factor model		Fama and French five-factor + UMD		Carhart four-factor model		Fama and French five-factor + UMD	
	Equal-Weight	TNA-Weight	Equal-Weight	TNA-Weight	Equal-Weight	TNA-Weight	Equal-Weight	TNA-Weight
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D10	0.083	0.106	0.073	0.096	0.183	0.184	0.176	0.174
D9	-0.031	0.041	-0.060	0.051	0.081	0.136	0.056	0.146
D8	-0.062	-0.033	-0.087	-0.014	0.066	0.082	0.046	0.104
D7	-0.101	-0.072	-0.137	-0.062	0.030	0.043	0.001	0.056
D6	-0.153	-0.154	-0.175	-0.129	-0.018	-0.040	-0.034	-0.013
D5	-0.187	-0.162	-0.208	-0.149	-0.045	-0.049	-0.058	-0.030
D4	-0.248	-0.202	-0.243	-0.166	-0.118	-0.098	-0.107	-0.060
D3	-0.262	-0.234	-0.245	-0.195	-0.147	-0.143	-0.123	-0.101
D2	-0.293	-0.277	-0.261	-0.229	-0.186	-0.187	-0.148	-0.136
D1	-0.324	-0.326	-0.280	-0.278	-0.231	-0.250	-0.181	-0.199

Panel B. Sharpe Ratio of Funds sorted on TCI

	Sharpe Ratio using Net Fund Return		Sharpe Ratio using Gross Fund Return	
	Equal-Weighted	TNA-Weighted	Equal-Weighted	TNA-Weighted
	(1)	(2)	(3)	(4)
D10	0.174	0.187	0.193	0.202
D9	0.156	0.175	0.173	0.190
D8	0.153	0.168	0.171	0.183
D7	0.149	0.163	0.166	0.178
D6	0.144	0.153	0.160	0.167
D5	0.140	0.150	0.157	0.165
D4	0.131	0.141	0.147	0.156
D3	0.125	0.129	0.142	0.144
D2	0.119	0.124	0.135	0.139
D1	0.109	0.110	0.126	0.124

Table A7. TCI constructed based on Management's Discussion & Analysis (Item 7) and Mutual Fund Performance

Panel A of this table reports the equal-weighted and TNA-weighted future returns of mutual funds sorted into decile portfolios according to the most recent month's $TCI^{Item\ 7}$ based on the Management's Discussion & Analysis (Item 7). $TCI^{Item\ 7}$ is computed similar to TCI measure described in Section 3.4. At the end of month, all funds in the sample are ranked into ten deciles based on TCI. R^{Net} is the one-month net return. $\alpha^{4F,Net}$ ($\alpha^{4F,Gross}$) is the one-month net (gross) alpha measured as the intercept of the Carhart (1997) four-factor model. CS is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Panel A also reports differences in fund returns and four-factor alphas between the top and bottom decile portfolios. Panel B of this table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent month's $TCI^{Item\ 7}$ based on Management's Discussion & Analysis (Item 7) and other fund characteristics. $\alpha_t^{4F,Net}$ ($\alpha_t^{4F,Gross}$) is the fund's one-month Carhart (1997) four-factor alpha and is obtained from the fund's excess net (gross) return less the sum of the products of each of the four-factor realizations estimated using the preceding 36 monthly fund returns. CS_t is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Mutual funds are classified into size/value style categories based on fund's four-factor loadings described in Section 3.5. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. The sample period is from 2006Q3 to 2023Q4.

Panel A. TCI based on MD&A (Item 7) and Mutual Fund Performance: Portfolio Results

	Equal-Weighted					TNA-weighted				
	R^{Net}	$\alpha^{3F,Net}$	$\alpha^{4F,Net}$	$\alpha^{4F,Gross}$	CS	R^{Net}	$\alpha^{3F,Net}$	$\alpha^{4F,Net}$	$\alpha^{4F,Gross}$	CS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
D10	0.888***	0.009	0.006	0.098	0.078**	0.966***	0.033	0.037	0.114	0.094**
	(2.60)	(0.13)	(0.09)	(1.43)	(2.18)	(2.65)	(0.34)	(0.38)	(1.18)	(2.14)
D9	0.849**	-0.033	-0.040	0.049	0.031	0.934***	0.014	0.017	0.092	0.080**
	(2.47)	(-0.57)	(-0.68)	(0.84)	(0.90)	(2.72)	(0.22)	(0.27)	(1.46)	(2.55)
D8	0.811**	-0.075	-0.080	0.007	0.015	0.887***	-0.038	-0.038	0.035	0.074**
	(2.35)	(-1.47)	(-1.55)	(0.14)	(0.49)	(2.60)	(-0.73)	(-0.70)	(0.65)	(2.48)
D7	0.800**	-0.091*	-0.094*	-0.009	0.006	0.874**	-0.057	-0.056	0.016	0.053*
	(2.32)	(-1.76)	(-1.80)	(-0.17)	(0.21)	(2.56)	(-1.13)	(-1.10)	(0.32)	(1.76)
D6	0.788**	-0.106**	-0.107**	-0.023	0.014	0.856**	-0.075*	-0.078*	-0.007	0.039
	(2.32)	(-2.42)	(-2.42)	(-0.51)	(0.45)	(2.54)	(-1.77)	(-1.81)	(-0.15)	(1.14)
D5	0.777**	-0.119***	-0.116***	-0.033	-0.009	0.812**	-0.118**	-0.119**	-0.049	0.000
	(2.27)	(-2.74)	(-2.60)	(-0.73)	(-0.33)	(2.40)	(-2.18)	(-2.16)	(-0.88)	(0.01)
D4	0.761**	-0.139***	-0.139***	-0.057	-0.014	0.794**	-0.128**	-0.128**	-0.057	-0.014
	(2.24)	(-3.06)	(-2.98)	(-1.21)	(-0.44)	(2.31)	(-2.31)	(-2.27)	(-1.01)	(-0.41)
D3	0.759**	-0.145***	-0.142***	-0.061	-0.014	0.743**	-0.170***	-0.168***	-0.096*	-0.043
	(2.23)	(-2.90)	(-2.76)	(-1.19)	(-0.41)	(2.20)	(-3.14)	(-3.05)	(-1.75)	(-1.32)
D2	0.726**	-0.179***	-0.178***	-0.099*	-0.024	0.726**	-0.203***	-0.204***	-0.134**	-0.074*
	(2.16)	(-3.46)	(-3.34)	(-1.86)	(-0.66)	(2.09)	(-3.79)	(-3.76)	(-2.46)	(-1.82)
D1	0.699**	-0.211***	-0.207***	-0.131**	-0.053	0.695**	-0.234***	-0.233***	-0.163***	-0.080**
	(2.04)	(-4.86)	(-3.23)	(-2.05)	(-1.65)	(2.04)	(-4.82)	(-3.77)	(-2.63)	(-2.06)
	Difference: D10 – D1									
	0.188**	0.220***	0.213***	0.228***	0.130***	0.271***	0.266**	0.270**	0.276**	0.173***
	(2.45)	(2.77)	(2.64)	(2.85)	(4.00)	(2.67)	(2.30)	(2.28)	(2.34)	(3.07)

Panel B. TCI based on MD&A (Item 7) and Mutual Fund Performance: Fama-MacBeth Regressions

	$\alpha_t^{4F,Net}$			$\alpha_t^{4F,Gross}$			CS_t		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$TCI_{t-1}^{Item\ 7}$	0.400**	0.547***	0.448***	0.486***	0.548***	0.449***	0.543***	0.546***	0.347***
	(2.49)	(3.71)	(3.79)	(3.02)	(3.70)	(3.78)	(3.24)	(3.34)	(3.11)
$Log(TNA)_{t-1}$		0.006*	0.005*		0.006	0.004		0.002	0.000
		(1.68)	(1.69)		(1.57)	(1.55)		(0.36)	(0.14)
$Log(Age)_{t-1}$		0.004	0.006		0.002	0.004		0.014*	0.004
		(0.56)	(0.90)		(0.30)	(0.58)		(1.93)	(0.83)
$Expenses_{t-1}$		-0.030**	-0.033***		0.047***	0.044***		0.013	0.009
		(-2.10)	(-2.72)		(3.28)	(3.63)		(0.94)	(1.02)
$Turnover_{t-1}$		-0.064***	-0.053***		-0.064***	-0.053***		-0.010	-0.011
		(-4.46)	(-4.14)		(-4.43)	(-4.11)		(-0.71)	(-1.07)
$\sigma^{Ret}{}_{t-12:t-1}$		-0.042**	-0.050**		-0.042**	-0.050**		-0.017	-0.002
		(-2.24)	(-2.31)		(-2.23)	(-2.32)		(-0.68)	(-0.11)
$Flow_{t-12:t-1}$		0.037***	0.039***		0.037***	0.039***		0.003	0.005
		(3.76)	(4.49)		(3.81)	(4.53)		(0.29)	(0.83)
Style FE	No	No	Yes	No	No	Yes	No	No	Yes
N. of Obs.	256,893	256,893	256,893	256,893	256,893	256,893	252,701	252,701	252,701
N. of Months	210	210	210	210	210	210	210	210	210
Adj. R^2	0.011	0.069	0.178	0.012	0.069	0.178	0.010	0.062	0.184

Table A8. TCI and Fund Performance: LDA with Different Number of Topics

This table repeats the Fama-MacBeth (1973) analysis in Table 4, except that we reconstruct the TCI measure by varying the number of topics in the LDA model from 25 to 150. Panel A presents results based on LDA with 25 topics, Panel B with 50 topics, Panel C with 75 topics, and Panel D with 150 topics. Note that in our baseline analysis in Table 4, we use 100 topics as input in the LDA model. The construction of TCI measure is described in Section 2.4. Newey-West (1987) t – statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. The sample period is from 2006Q3 to 2023Q4.

Panel A. LDA with 25 Topics

	$\alpha_t^{4F,Net}$	$\alpha_t^{4F,Gross}$	CS_t
	(1)	(2)	(3)
TCI	0.247***	0.248***	0.263***
	(3.89)	(3.91)	(3.57)
Controls + Style FE	Yes	Yes	Yes
N of Obs.	256,893	256,893	252,701
N of Months	210	210	210
Adj. R^2	0.181	0.181	0.186

Panel B. LDA with 50 Topics

	$\alpha_t^{4F,Net}$	$\alpha_t^{4F,Gross}$	CS_t
	(1)	(2)	(3)
TCI	0.476***	0.478***	0.507***
	(3.98)	(4.00)	(3.66)
Controls + Style FE	Yes	Yes	Yes
N of Obs.	256,893	256,893	252,701
N of Months	210	210	210
Adj. R^2	0.181	0.181	0.186

Panel C. LDA with 75 Topics

	$\alpha_t^{4F,Net}$	$\alpha_t^{4F,Gross}$	CS_t
	(1)	(2)	(3)
TCI	0.682***	0.685***	0.731***
	(3.75)	(3.77)	(3.51)
Controls + Style FE	Yes	Yes	Yes
N of Obs.	256,893	256,893	252,701
N of Months	210	210	210
Adj. R^2	0.180	0.180	0.186

Panel D. LDA with 150 Topics

	$\alpha_t^{4F,Net}$	$\alpha_t^{4F,Gross}$	CS_t
	(1)	(2)	(3)
TCI	1.336***	1.341***	1.449***
	(3.81)	(3.83)	(3.52)
Controls + Style FE	Yes	Yes	Yes
N of Obs.	256,893	256,893	252,701
N of Months	210	210	210
Adj. R^2	0.180	0.180	0.185

Table A9. TCI and Fund Performance Across Size and Value Dimensions: Fama-MacBeth Regressions

This table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent month's TCI and other fund characteristics across Size and Value/Growth Dimensions. The construction of TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. $\alpha_t^{4F,Net}$ is the fund's one-month Carhart (1997) four-factor alpha and is obtained from the fund's excess net return less the sum of the products of each of the four-factor realizations estimated using the preceding 36 monthly fund returns. Mutual funds are classified into size/value style categories based on fund's four-factor loadings described in Section 3.5. Newey-West (1987) *t* – statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. The sample period is from 2006Q3 to 2023Q4.

	Size Dimension			Growth/Value Dimension		
	Large	Mid	Small	Growth	Blend	Value
	(1)	(2)	(3)	(4)	(5)	(6)
TCI	0.493*	1.373***	2.569***	2.053***	0.668***	0.923***
	(1.91)	(3.49)	(3.31)	(3.77)	(3.15)	(3.44)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	83,257	85,680	87,956	84,970	86,730	85,193
Adj. R^2	0.109	0.080	0.083	0.111	0.072	0.059

Table A10. TCI and Fund Performance: MegaCap Funds versus Other Funds

This table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent month's TCI and other fund characteristics for funds with and without a focus on mega cap stocks. The construction of TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. MegaCap funds are defined as follows: each quarter, mutual funds are ranked based on their aggregate portfolio weight in the so-called "MegaCap-8" stocks: Alphabet (GOOG and GOOGL), Amazon (AMZN), Apple (AAPL), Meta (META), Microsoft (MSFT), Netflix (NFLX), Nvidia (NVDA), and Tesla (TSLA), and funds in the top quintile are categorized as MegaCap Funds. $\alpha_t^{4F,Net}$ ($\alpha_t^{4F,Gross}$) is the fund's one-month Carhart (1997) four-factor alpha and is obtained from the fund's excess net (gross) return less the sum of the products of each of the four-factor realizations estimated using the preceding 36 monthly fund returns. CS_t is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Mutual funds are classified into size/value style categories based on fund's four-factor loadings. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. The sample period is from 2006Q3 to 2023Q4.

	MegaCap Funds			Other Funds		
	$\alpha_t^{4F,Net}$	$\alpha_t^{4F,Gross}$	CS_t	$\alpha_t^{4F,Net}$	$\alpha_t^{4F,Gross}$	CS_t
	(1)	(2)	(3)	(4)	(5)	(6)
TCI	2.172***	2.183***	2.954***	1.005***	1.008***	0.821***
	(4.19)	(4.19)	(3.71)	(3.76)	(3.78)	(2.93)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	51,338	51,338	50,730	205,555	205,555	201,971
N. of Month	210	210	210	210	210	210
Adj. R^2	0.223	0.222	0.221	0.165	0.165	0.173

Table A11. TCI and Fund Performance: Fama-MacBeth Regressions – Value-added Measure

This table reports results from Fama-MacBeth (1973) regressions of dollar value-added (Berk and van Binsbergen, 2015) of funds on the most recent month's TCI and other fund characteristics. The construction of the TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. Fund dollar value-added is measured as the estimated gross alpha multiplied by the previous month's TNA. Mutual funds are classified into size/value style categories based on the fund's four-factor loadings described in Section 3.5. Newey-West (1987) *t*-statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. Sample period is from 2006Q3 to 2023Q4.

	(1)	(2)	(3)
TCI_{t-1}	23.305***	17.586***	14.556**
	(3.52)	(2.82)	(2.48)
$Log(TNA)_{t-1}$		-0.278	-0.331
		(-0.76)	(-0.96)
$Log(Age)_{t-1}$		0.080	0.065
		(0.44)	(0.45)
$Expenses_{t-1}$		-0.224	-0.232
		(-0.84)	(-0.93)
$Turnover_{t-1}$		-0.550***	-0.430**
		(-3.47)	(-2.47)
$\sigma^{Ret}_{t-12:t-1}$		-0.635*	-0.678*
		(-1.70)	(-1.70)
$Flow_{t-12:t-1}$		0.315**	0.332**
		(2.04)	(2.28)
Style FE	No	No	Yes
N. of Obs.	256,893	256,893	256,893
N. of Months	210	210	210
Adj. R^2	0.003	0.054	0.088

Table A12. The relation between TCI and Risk Shifting

Panel A of this table reports various risk-shifting characteristics of mutual funds sorted according to the most recent month's TCI. At the end of each month, all funds in the sample are ranked into ten deciles based on TCI. Following Huang, Sialm, and Zhang (2011), risk shifting is measured as the changes in the difference between the most recent characteristics and the average characteristics over the prior 12 months (RS^{12M}) and 24 months (RS^{24M}). Panel B of this table reports the results from Fama-MacBeth regressions of quarterly risk-shifting measures at the end of quarter q on quarterly TCI, computed as the average monthly TCI at the end of q-1, and the lagged fund characteristics. Panel C of this table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent month's TCI, various shifting measures and other fund characteristics. $\alpha_t^{4F,Net}$($\alpha_t^{4F,Gross}$) is the fund's one-month Carhart (1997) four-factor net (gross) alpha and is obtained from the fund's excess return less the sum of the products of each of the four-factor realizations estimated using the preceding 36 monthly fund net (gross) returns. CS_t is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). The construction of TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. Newey-West (1987) t-statistics with a lag of 3 are reported in parentheses. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

Panel A. TCI and Risk Shifting Measure – Portfolio Approach

	RS^{12M}	RS^{24M}
D10	0.079	0.031
D9	0.069	0.069
D8	0.048	0.064
D7	0.043	0.064
D6	0.028	0.045
D5	0.054	0.066
D4	0.060	0.075
D3	0.037	0.032
D2	0.038	0.084
D1	0.058	0.070
	D10 – D1	
	0.020	-0.039
	(0.88)	(-1.58)

Panel B. TCI and Risk Shifting Measure – Fama-MacBeth Regressions

	RS^{12M}	RS^{24M}
	(1)	(2)
TCI_{q-1}	-0.001	-0.003
	(-0.86)	(-1.46)
$Log(TNA)_{q-1}$	0.000	0.000
	(1.34)	(0.83)
$Log(Age)_{q-1}$	-0.000	0.000
	(-0.24)	(1.04)
$Turnover_{q-1}$	0.000	0.000
	(1.05)	(1.37)
$Expenses_{q-1}$	0.000	-0.000
	(0.26)	(-0.51)
$\sigma^{Ret}{}_{q-1}$	-0.000	0.000*
	(-1.04)	(1.69)
$\alpha^{4F}{}_{q-1}$	0.001*	0.001***
	(1.91)	(3.72)
$Flow_{q-1}$	-0.000	-0.000**
	(-1.33)	(-2.01)
Style FEs	Yes	Yes
N. of Obs.	79,646	79,646
N. of Quarters	69	69
Adj. R^2	0.071	0.093

Panel C. Cross-sectional regressions of fund performance on TCI controlling risk shifting measures

	$\alpha_t^{4F,Net}$		$\alpha_t^{4F,Grross}$		CS_t	
	(1)	(2)	(3)	(4)	(5)	(6)
TCI	1.135***	1.146***	1.142***	1.153***	1.079***	1.095***
	(4.25)	(4.36)	(4.26)	(4.38)	(3.56)	(3.67)
RS^{12M}	-2.413		-2.354		0.577	
	(-1.25)		(-1.22)		(0.35)	
RS^{24M}		-1.391		-1.349		-0.793
		(-0.68)		(-0.66)		(-0.42)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	252,924	253,723	252,924	253,723	248,867	249,640
Adj R^2	0.192	0.194	0.192	0.194	0.193	0.193

Table A13. TCI and Fund Performance: Fama-MacBeth Regressions – Subsample Analysis

This table reports results from Fama-MacBeth (1973) regressions of future fund performance on the most recent month's TCI and other fund characteristics from 2006Q3 to 2014Q4 in Panel A and 2015Q1 to 2023Q4 in Panel B. The construction of TCI measure is described in Section 3.4, and other fund characteristics are defined in Table 1. $\alpha_t^{4F,Net}$ ($\alpha_t^{4F,Gross}$) is the fund's one-month Carhart (1997) four-factor alpha and is obtained from the fund's excess net (gross) return less the sum of the products of each of the four-factor realizations estimated using the preceding 36 monthly fund returns. CS_t is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Mutual funds are classified into size/value style categories based on fund's four-factor loadings described in Section 3.5. Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

	Panel A. Sample period: 2006Q3 - 2014Q4			Panel B. Sample period: 2015Q1 - 2023Q4		
	$\alpha_t^{4F,Net}$	$\alpha_t^{4F,Gross}$	CS_t	$\alpha_t^{4F,Net}$	$\alpha_t^{4F,Gross}$	CS_t
	(1)	(2)	(3)	(4)	(5)	(6)
TCI_{t-1}	1.566***	1.580***	1.394***	0.720**	0.716**	0.827**
	(3.71)	(3.75)	(2.90)	(2.15)	(2.14)	(2.24)
$Log(TNA)_{t-1}$	0.002	0.002	-0.000	0.006	0.006	0.001
	(0.59)	(0.46)	(-0.12)	(1.64)	(1.58)	(0.21)
$Log(Age)_{t-1}$	0.022**	0.020**	0.017**	-0.011	-0.013	-0.008
	(2.58)	(2.23)	(2.26)	(-1.35)	(-1.50)	(-1.38)
$Expenses_{t-1}$	-0.036**	0.040**	0.012	-0.028	0.051***	0.006
	(-2.23)	(2.50)	(1.02)	(-1.58)	(2.94)	(0.44)
$Turnover_{t-1}$	-0.055***	-0.055***	-0.018	-0.055***	-0.055***	-0.005
	(-3.54)	(-3.52)	(-1.14)	(-2.69)	(-2.67)	(-0.37)
$\sigma^{Ret}_{t-12:t-1}$	-0.082***	-0.082***	-0.015	-0.017	-0.017	0.009
	(-2.82)	(-2.83)	(-0.43)	(-0.56)	(-0.57)	(0.31)
$Flow_{t-12:t-1}$	0.029***	0.030***	-0.000	0.049***	0.050***	0.011
	(3.08)	(3.08)	(-0.05)	(3.47)	(3.52)	(1.04)
Style FE	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	128,879	128,879	126,087	128,014	128,014	126,614
N. of Months	105	105	105	105	105	105
Adj. R^2	0.163	0.164	0.140	0.202	0.201	0.234

Table A14. Orthogonalized TCI and Future Fund Performance

This table reports the results from Fama-MacBeth (1973) regressions of future fund performance on the most recent quarter's orthogonalized TCI ($TCI_1^{\perp}$, $TCI_2^{\perp}$, $TCI_3^{\perp}$). $TCI_1^{\perp}$ is the residual of cross-sectional regressions of TCI on ICI, R^2, *Past* α^{4F} and style fixed effects. $TCI_2^{\perp}$ is the residual of cross-sectional regression of TCI on ICI, R^2, *Past* α^{4F}, *ActvShr,* and style fixed effects. $TCI_3^{\perp}$ is the residual of cross-sectional regressions of TCI on ICI, R^2, *Past* α^{4F}, *ActvShr, OCI* and style fixed effects. In Panel A, α_t^{4F} is the fund's one-month Carhart (1997) four-factor alpha and is obtained from the fund's excess return less the sum of the products of each of the four-factor realizations estimated using the preceding 36 monthly fund returns. In Panel B, CS_t is the monthly Characteristic Selectivity measure computed following Daniel, Grinblatt, Titman, and Wermers (1997). Newey-West (1987) $t-$statistics with a lag of 3 are reported in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. The sample period is from 2006Q3 to 2023Q4 except OCI (2006Q3 to 2017:Q2).

	Panel A. $\alpha_t^{4F,Net}$			Panel B. CS_t		
	(1)	(2)	(3)	(4)	(5)	(6)
$TCI_1^{\perp}$	1.234***			1.137***		
	(4.15)			(3.04)		
$TCI_2^{\perp}$		1.356***			1.418***	
		(3.34)			(2.70)	
$TCI_3^{\perp}$			1.864***			1.802**
			(3.22)			(2.40)
Controls + Style FE	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	256,869	187,736	122,447	252,701	187,502	122,342
N. of Months	210	209	132	210	209	132
Adj. R^2	0.182	0.189	0.172	0.187	0.192	0.157

Table A15. TCI Increasing and Decreasing Trade Portfolio Performance of TCI Sorted Funds

This table reports the DGTW characteristic-adjusted performance and earnings announcement abnormal returns of TCI-increasing and TCI-decreasing stocks purchased and sold by funds sorted on TCI. At the end of quarter q-1, we sort funds into decile portfolios based on the TCI measure. Within each fund, we break down fund trades into buy and sell trades as described in Section 4.5, and classify each fund's trades during quarter q as either TCI-increasing or TCI-decreasing as described in Appendix Section A.1. Panel A reports the time-series mean quarterly DGTW benchmark-adjusted returns of trades by mutual funds in quarter q. Panel B reports the time-series mean earnings announcement abnormal return of trades by mutual funds. The earnings announcement abnormal returns are defined as the cumulative market-adjusted return over a three-day window [-1, +1] around the earnings announcement date in quarter q. The table also reports differences (shown in columns labeled as "Difference") in DGTW benchmark-adjusted returns and earnings announcement abnormal returns between Buy and Sell trade portfolios. Newey-West (1987) t – statistics with the lag of 3 are reported in parentheses. All returns are expressed in %. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

Panel A. DGTW benchmark-adjusted returns

	All Trades			Stock Buys			Stock Sells		
				TCI	TCI		TCI	TCI	
	Buy	Sell	Difference	Increasing	Decreasing	Difference	Increasing	Decreasing	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	DGTW Benchmark-adjusted Return								
All Funds	0.277	-0.011	0.289	0.286	0.091	0.195	-0.002	-0.073	0.070
	(0.87)	(-0.04)	(0.65)	(0.87)	(0.27)	(0.42)	(-0.01)	(-0.24)	(0.15)
D10	0.855*	-0.236	1.091**	1.267*	0.761	0.506	-0.020	-0.198	0.178
	(1.73)	(-0.64)	(2.42)	(1.79)	(1.55)	(0.80)	(-0.04)	(-0.53)	(0.52)
D9	0.668	-0.243	0.911***	0.704	0.976	-0.272	-0.054	-0.266	0.212
	(1.54)	(-0.64)	(3.16)	(1.59)	(1.58)	(-0.49)	(-0.13)	(-0.64)	(0.64)
D8	0.349	0.139	0.210	0.239	0.477	-0.238	0.253	0.154	0.099
	(0.92)	(0.44)	(0.82)	(0.60)	(1.01)	(-0.59)	(0.61)	(0.49)	(0.31)
D7	0.353	-0.006	0.359	0.071	0.554	-0.483	0.016	-0.066	0.082
	(0.87)	(-0.02)	(1.43)	(0.17)	(1.16)	(-1.26)	(0.04)	(-0.21)	(0.24)
D6	0.040	0.343	-0.302	0.093	-0.060	0.152	0.199	0.256	-0.057
	(0.10)	(1.00)	(-1.24)	(0.26)	(-0.12)	(0.38)	(0.52)	(0.74)	(-0.18)
D5	0.086	0.129	-0.043	0.140	-0.174	0.314	-0.107	0.115	-0.222
	(0.22)	(0.36)	(-0.15)	(0.38)	(-0.39)	(1.00)	(-0.24)	(0.34)	(-0.71)
D4	0.026	0.026	0.000	0.154	-0.204	0.358	-0.405	0.115	-0.519**
	(0.07)	(0.08)	(0.00)	(0.31)	(-0.47)	(0.69)	(-1.09)	(0.32)	(-2.27)
D3	0.006	-0.265	0.271	-0.016	-0.023	0.008	-0.264	-0.311	0.048
	(0.02)	(-0.71)	(0.90)	(-0.04)	(-0.05)	(0.02)	(-0.61)	(-0.83)	(0.15)
D2	0.271	0.113	0.159	-0.003	0.070	-0.073	0.010	0.110	-0.100
	(0.77)	(0.31)	(0.71)	(-0.01)	(0.15)	(-0.18)	(0.03)	(0.27)	(-0.31)
D1	0.112	0.359	-0.248	-0.117	0.204	-0.321	0.399	0.370	0.029
	(0.29)	(0.91)	(-0.90)	(-0.29)	(0.41)	(-0.76)	(0.85)	(0.93)	(0.09)
	Difference: D10 - D1								
	0.743*	-0.596	1.339**	1.384**	0.557	0.828*	-0.419	-0.567	0.148
	(1.75)	(-1.33)	(2.47)	(2.15)	(1.00)	(1.88)	(-0.74)	(-1.26)	(0.34)

Panel B. Returns around earnings announcements

	All Trades			Stock Buys			Stock Sells		
	Buy	Sell	Difference	TCI Increasing	TCI Decreasing	Difference	TCI Increasing	TCI Decreasing	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Cumulative Market-adjusted Return over [-1, +1] around Earnings Announcement								
All Funds	0.385***	0.117	0.268	0.471***	0.249**	0.576	0.247	0.095	0.152
	(3.02)	(0.73)	(1.30)	(3.53)	(2.08)	(1.24)	(1.49)	(0.62)	(0.67)
D10	0.861***	0.200	0.662***	0.912***	0.891**	0.021	0.351	0.220	0.131
	(3.38)	(0.88)	(2.68)	(3.85)	(2.00)	(0.05)	(1.16)	(0.97)	(0.54)
D9	0.859***	0.279	0.580**	0.960***	0.761**	0.175	0.461	0.273	0.188
	(4.06)	(1.15)	(2.42)	(4.64)	(2.09)	(0.45)	(1.07)	(1.25)	(0.47)
D8	0.555***	0.372**	0.183	0.557***	0.707***	-0.150	0.446**	0.371**	0.083
	(3.37)	(2.16)	(0.92)	(3.23)	(2.75)	(-0.58)	(2.41)	(2.03)	(0.50)
D7	0.374**	0.474**	-0.121	0.485**	0.537*	-0.053	0.531**	0.507**	-0.080
	(2.01)	(2.15)	(-0.64)	(2.30)	(1.79)	(-0.15)	(2.22)	(2.42)	(-0.39)
D6	0.173	0.360*	-0.140	0.229	0.422	-0.145	0.444*	0.330*	0.114
	(0.87)	(1.83)	(-0.78)	(1.17)	(1.38)	(-0.49)	(1.93)	(1.71)	(0.55)
D5	0.128	0.313	-0.185	0.185	0.095	0.074	0.138	0.358*	-0.142
	(0.64)	(1.58)	(-0.93)	(0.81)	(0.48)	(0.31)	(0.65)	(1.90)	(-0.65)
D4	0.550**	0.211	0.260	0.614**	0.480	0.134	0.374*	0.221	0.074
	(2.07)	(1.10)	(0.96)	(2.42)	(1.66)	(0.52)	(1.67)	(1.16)	(0.38)
D3	-0.085	0.112	-0.197	-0.070	-0.038	-0.031	0.311*	0.083	0.151
	(-0.50)	(0.62)	(-1.46)	(-0.34)	(-0.20)	(-0.15)	(1.69)	(0.44)	(0.89)
D2	0.139	0.118	0.021	0.292	-0.247	0.538**	0.343	0.050	0.217
	(0.69)	(0.67)	(0.12)	(1.29)	(-1.09)	(2.27)	(1.53)	(0.27)	(1.15)
D1	0.088	0.002	0.086	0.314	-0.284	0.588**	0.278	-0.014	0.217
	(0.62)	(0.01)	(0.56)	(1.59)	(-1.43)	(2.05)	(1.12)	(-0.07)	(0.97)
	Difference: D10 - D1								
	0.758***	0.122	0.636**	0.599**	1.165**	-0.567	0.073	0.158	-0.086
	(2.71)	(0.49)	(2.12)	(2.00)	(2.28)	(-0.97)	(0.21)	(0.66)	(-0.30)

Table A16. Performance of Stocks in Overweighted Themes

This table reports the performance of portfolios based on mutual funds' trades in stocks in over-weighted themes versus portfolios based on other trades in the control group. For each fund, we form four distinct portfolios at the beginning of each quarter based on (1) whether the fund bought or sold a stock during the previous quarter and (2) whether the trade was of a stock with over-weighted topics (i.e., themes). The definition of overweighted-topic stock is described in Section A.2 of the Internet Appendix. Stocks that are bought are aggregated into the buy portfolio, while those that are sold are placed in the sell portfolio. We create two subgroups within the buy and sell portfolios: stocks with over-weighed themes and other stocks. We calculate the average monthly returns of these portfolios for each fund in each quarter, weighing each stock's return in the portfolios by the dollar trade value during the previous quarter. We rebalance at the end of the quarter. We then average the returns of each sub-portfolio across the funds in our sample using the dollar assets (TNA) of each fund in the previous quarter as weights, producing value-weighted average monthly returns for each of the four portfolios. Columns (1) and (2) report the risk-adjusted average monthly returns of the overweighted buy and sell portfolios, respectively. Risk adjustment is based on both DGTW benchmark returns and the Carhart (1997) four-factor alpha. Columns (4) and (5) report the corresponding results for the buy and sell portfolios for other stocks. Finally, columns (3) and (6) describe the difference in the returns of the buy and sell portfolios for the overweighted stocks and other stocks, respectively, and column (7) provides the difference-in-difference estimate. *t* - statistics are reported in parentheses. Significance levels for tests of difference in means are denoted by ∗∗∗, ∗∗, and ∗, which correspond to 1%, 5%, and 10% levels, respectively.

	Stocks in Overweight Themes			Other Stocks			
	Buys	Sells	Diff.	Buys	Sells	Diff.	Diff.-Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Raw return excess risk free	0.834*	0.556	0.278*	0.471	0.548	-0.077	0.355**
	(1.84)	(1.41)	(1.87)	(1.00)	(1.29)	(-0.75)	(2.30)
DGTW-adj Return	0.044	-0.226***	0.270**	-0.296***	-0.243***	-0.053	0.323**
	(0.44)	(-2.64)	(2.29)	(-3.23)	(-3.89)	(-0.75)	(2.45)
Four-factor alpha	-0.042	-0.338***	0.297**	-0.456***	-0.333***	-0.123	0.419***
	(-0.27)	(-3.11)	(2.08)	(-3.00)	(-3.82)	(-1.09)	(3.12)

Table A17. Portfolio Manager Turnover, Change in Bachelor's Degree, and Fund-level TCI

This table reports the difference in the future TCI of funds associated with manager turnover with degree change and the matched funds that do not experience manager turnover. Matched funds are required to be similar size and have the same investment objective as the funds that experience manager turnover with a degree change in a given quarter. In Panel A, columns (1), (2), (3), and (4) report the differences in the future TCI of funds associated with the departing manager's degree over next one-, two-, three-, and four-quarter, respectively. In Panel B, columns (5), (6), (7), and (8) report the differences in the future TCI of funds associated with the new manager's degree over next one-, two-, three-, and four-quarter, respectively. This table also reports the difference-in-difference results between manager turnover with degree change and the matched fund sample. $p-values$ are reported in parentheses. The sample period is 2006Q3 to 2023Q4.

	Panel A. Themes Associated with the Departing Manager's Degree				Panel B. Themes Associated with the New Manager's Degree			
	Quarters After Degree Change				Quarters After Degree Change			
	$(t+1)-t$	$(t+2)-t$	$(t+3)-t$	$(t+4)-t$	$(t+1)-t$	$(t+2)-t$	$(t+3)-t$	$(t+4)-t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Change in Degree	0.0267	-0.0560**	-0.0681***	-0.0757***	-0.0001	0.0183***	0.0202***	0.0245
(N=2,156)	(0.50)	(-1.97)	(-3.00)	(-3.86)	(-0.07)	(10.13)	(10.51)	(12.18)
Matched Funds w/o Mgr. Turnover	-0.0001	0.0004	-0.0001	0.0003	-0.0001	0.0004	-0.0001	0.0003
(N=6,580)	(-0.40)	(1.37)	(-0.11)	(0.98)	(-0.40)	(1.37)	(-0.11)	(0.98)
	Differences							
	0.0269	-0.564**	-0.0681***	-0.0760***	0.0001	0.0179***	0.0202***	0.0243***
	(0.52)	(-2.03)	(-3.07)	(-3.96)	(0.37)	(10.00)	(10.64)	(12.21)

Table A18. A Manager's Bachelor's Degree, TCI, and Trade-based Performance - Subsample Analysis

This table reports the DGTW characteristic-adjusted performance and earnings announcement abnormal returns of degree-related and degree-unrelated stocks purchased and sold by funds sorted on TCI. from 2006Q3 to 2014Q4 and 2015Q1 to 2023Q4. At the end of quarter q-1, we sort funds into decile portfolios based on the TCI measure. Within each fund, we break down fund trades into buy and sell trades as described in Section 4.5, and classify fund's trades during quarter q as Degree-Related or Degree-Unrelated as described in Section 5.3. Panel A reports the time-series mean quarterly DGTW benchmark-adjusted returns of trades by mutual funds in quarter q. Panel B reports the time-series mean earnings announcement abnormal return of trades by mutual funds. The earnings announcement abnormal returns are defined as the cumulative market-adjusted return over a three-day window [-1, +1] around the next earnings announcement date. The table also reports differences (shown in columns labeled as "Difference") in DGTW benchmark-adjusted returns and earnings announcement abnormal returns between Buy and Sell trade portfolios. Newey-West (1987) t – statistics with the lag of 3 are reported in parentheses. All returns are expressed in %. ***, **, * denote statistical significance at the 1%, 5%, or 10% level, respectively. The sample period is from 2006Q3 to 2023Q4.

Panel A. DGTW benchmark-adjusted returns

	Period: 2006Q3 to 2014Q4						Period: 2015Q1 to 2023Q4					
	Degree Related			Degree Unrelated			Degree Related			Degree Unrelated		
	Buys	Sells	Difference	Buys	Sells	Difference	Buys	Sells	Difference	Buys	Sells	Difference
D10	1.477**	-0.089	1.566**	0.713	0.575	0.138	1.158	-1.299	2.457**	0.428	-0.636	1.064
	(2.04)	(-0.11)	(2.66)	(0.99)	(1.18)	(0.25)	(1.13)	(-1.61)	(2.36)	(0.53)	(-1.17)	(1.46)
D9	1.166	-0.724	1.890***	0.899	0.229	0.671	0.590	-0.795	1.385*	0.062	-0.124	0.186
	(1.50)	(-0.95)	(3.16)	(1.34)	(0.39)	(1.60)	(0.63)	(-1.17)	(1.87)	(0.09)	(-0.18)	(0.23)
D8	1.943***	0.202	1.740***	0.560	0.771*	-0.211	-0.677	-0.570	-0.107	0.182	-0.274	0.456
	(2.98)	(0.36)	(3.27)	(0.88)	(1.78)	(-0.55)	(-0.97)	(-0.92)	(-0.17)	(0.38)	(-0.36)	(0.57)
D7	0.276	0.686*	-0.410	1.155**	0.133	1.022***	-0.301	-0.350	0.048	-0.474	-0.856*	0.383
	(0.39)	(1.71)	(-0.74)	(2.13)	(0.26)	(2.78)	(-0.42)	(-0.60)	(0.08)	(-0.64)	(-1.73)	(0.67)
D6	0.542	0.307	0.235	0.569	0.622	-0.052	-0.635	-0.319	-0.316	-0.696	-0.099	-0.597
	(0.92)	(0.50)	(0.60)	(0.89)	(1.09)	(-0.15)	(-0.96)	(-0.54)	(-0.64)	(-1.35)	(-0.16)	(-1.02)
D5	-0.076	-0.037	-0.039	0.639	0.724	-0.085	-0.503	-0.193	-0.310	-0.278	-0.947*	0.670
	(-0.11)	(-0.06)	(-0.11)	(0.93)	(1.43)	(-0.22)	(-0.72)	(-0.29)	(-0.42)	(-0.42)	(-1.80)	(1.16)
D4	0.498	0.077	0.420	0.315	0.501	-0.186	-0.413	-0.188	-0.225	-1.255	-0.462	-0.793
	(0.75)	(0.14)	(1.13)	(0.50)	(1.01)	(-0.47)	(-0.80)	(-0.35)	(-0.52)	(-1.42)	(-0.67)	(-1.23)
D3	0.402	0.075	0.327	0.952	0.486	0.465	-0.538	-0.830	0.292	-0.408	-1.113*	0.705
	(1.41)	(0.24)	(0.96)	(1.39)	(0.89)	(1.00)	(-0.93)	(-1.22)	(0.39)	(-0.58)	(-1.82)	(0.96)
D2	0.994*	0.180	0.814	0.093	0.466	-0.374	-0.449	-0.497	0.048	-0.254	-0.333	0.079
	(1.73)	(0.29)	(1.56)	(0.32)	(0.74)	(-0.99)	(-0.84)	(-0.83)	(0.07)	(-0.43)	(-0.67)	(0.21)
D1	0.605	0.715	-0.110	0.435	0.845	-0.410	-0.530	0.294	-0.823	-0.029	-0.171	0.142
	(0.99)	(1.23)	(-0.32)	(0.67)	(1.45)	(-1.51)	(-0.64)	(0.49)	(-0.89)	(-0.03)	(-0.23)	(0.19)
	0.872	-0.804	1.676**	0.279	-0.269	0.548	1.688	-1.592	3.280**	0.457	-0.465	0.922
	(1.18)	(-1.14)	(2.27)	(0.43)	(-0.50)	(0.99)	(1.57)	(-1.56)	(2.08)	(0.39)	(-0.55)	(0.95)

Panel B. Returns around earnings announcements

	Period: 2006Q3 to 2014Q4						Period: 2015Q1 to 2023Q4					
	Degree Related			Degree Unrelated			Degree Related			Degree Unrelated		
	Buys	Sells	Difference	Buys	Sells	Difference	Buys	Sells	Difference	Buys	Sells	Difference
D10	1.525***	0.277	1.248**	0.790*	-0.075	0.865**	1.476**	-0.151	1.627**	0.424	0.415	0.009
	(2.78)	(0.69)	(2.10)	(1.88)	(-0.17)	(2.04)	(2.74)	(-0.33)	(2.64)	(1.19)	(1.15)	(0.02)
D9	0.868*	0.443	0.425	0.974**	-0.069	1.043**	1.168***	0.563	0.605	0.640**	0.145	0.494
	(1.84)	(1.23)	(0.84)	(2.15)	(-0.17)	(2.47)	(3.09)	(0.67)	(0.69)	(2.48)	(0.47)	(1.57)
D8	0.287	0.818**	-0.531	0.402	0.265	0.137	0.871**	0.740**	0.131	0.786***	-0.082	0.868***
	(0.67)	(2.10)	(-0.90)	(1.29)	(1.00)	(0.54)	(2.14)	(2.14)	(0.33)	(3.14)	(-0.29)	(3.10)
D7	-0.092	0.664	-0.755	0.071	0.118	-0.047	0.717**	0.792**	-0.076	0.282	0.231	0.051
	(-0.19)	(1.42)	(-1.22)	(0.21)	(0.36)	(-0.19)	(2.11)	(2.10)	(-0.19)	(1.19)	(1.21)	(0.25)
D6	0.684*	0.936**	-0.252	-0.165	-0.063	-0.101	0.718**	0.678*	0.040	-0.168	0.047	-0.216
	(1.70)	(2.28)	(-0.63)	(-0.58)	(-0.27)	(-0.40)	(2.21)	(1.82)	(0.09)	(-0.65)	(0.19)	(-0.75)
D5	0.394	1.195**	-0.801	-0.119	0.179	-0.298	0.397	-0.050	0.447	0.102	0.204	-0.102
	(0.93)	(2.51)	(-1.49)	(-0.38)	(0.59)	(-0.94)	(0.90)	(-0.17)	(1.13)	(0.48)	(0.95)	(-0.44)
D4	0.749	0.520	0.229	0.081	0.258	-0.178	1.044*	0.072	0.972	0.383	0.284	0.099
	(1.70)	(1.28)	(0.45)	(0.30)	(0.77)	(-0.47)	(1.90)	(0.22)	(1.45)	(1.66)	(1.32)	(0.39)
D3	-0.027	0.718*	-0.745**	0.249	-0.042	0.291	-0.488*	-0.033	-0.455	-0.095	-0.104	0.010
	(-0.06)	(1.86)	(-2.07)	(0.87)	(-0.13)	(0.98)	(-1.70)	(-0.10)	(-1.29)	(-0.49)	(-0.52)	(0.05)
D2	-0.062	0.136	-0.198	0.108	0.442*	-0.334	-0.514	-0.226	-0.287	0.097	0.041	0.056
	(-0.16)	(0.33)	(-0.43)	(0.40)	(1.90)	(-1.40)	(-1.00)	(-0.62)	(-0.61)	(0.38)	(0.19)	(0.21)
D1	0.187	0.398	-0.211	0.106	0.006	0.100	-0.278	-0.003	-0.274	0.053	-0.185	0.238
	(0.70)	(1.26)	(-0.62)	(0.48)	(0.02)	(0.42)	(-0.77)	(-0.01)	(-0.74)	(0.21)	(-0.59)	(0.74)
	1.380**	-0.132	1.512**	0.684	-0.081	0.765	1.753**	-0.148	1.901**	0.371	0.600	-0.229
	(2.26)	(-0.25)	(2.11)	(1.58)	(-0.17)	(1.58)	(2.70)	(-0.23)	(2.48)	(0.92)	(1.39)	(-0.40)