Corporate Filings in the Age of Natural Language Processing

Analyzing the impact of language complexity on investor outcomes

Morningstar Quantitative Research
15 July 2021
Version 1.0

Contents

Lee Davidson, CFA
Head of Manager and Quantitative Research
+1 312 244-7541
lee.davidson@morningstar.com

Sabeeh Ashhar
Director of Quantitative Research
sabeeh.ashhar@morningstar.com

Harshal Dalvi
Principal Data Scientist
harshal.dalvi@morningstar.com

Executive Summary

The financial community has predominantly focused on proven quantitative factors to evaluate short- and long-term performance of the company. With proliferation of alternative data and rapid adoption of machine-learning techniques and cloud computing, it is now possible to uncover a new generation of factors that could improve investor outcomes. We attempt to exploit corporate disclosures data here, which plays an important role in communicating financial health, promoting culture and brand, and engaging an entire range of stakeholders. While the information present in financial disclosures is of considerable importance, it is also unstructured in nature. In this paper, we will use modern natural language processing, or NLP, techniques to build factors from corporate disclosures. We then venture to uncover relationship between fund performance, fund attributes, and these new factors.

In this paper, we focus on creating one such language complexity factor based on textual information present in business, risk factors, and management discussion and analysis, or MD&A, sections of corporate disclosures, with the hypothesis being that an increase in language complexity (high complexity) over time is indicative of suboptimal investor outcomes. To create the language complexity factor, we build a framework from the ground up for securities by equi-weighting the scores across mentioned three sections and aggregating insights at the fund level. Thus, we propose one of the first cross-sectional examinations of managed products' exposure to factors derived from language analysis of corporate filings. To maximize fund universe coverage of our U.S. equity asset class, we build the factors for companies present in Russell 3000 Index, which represents 98% of investable U.S. equity market.

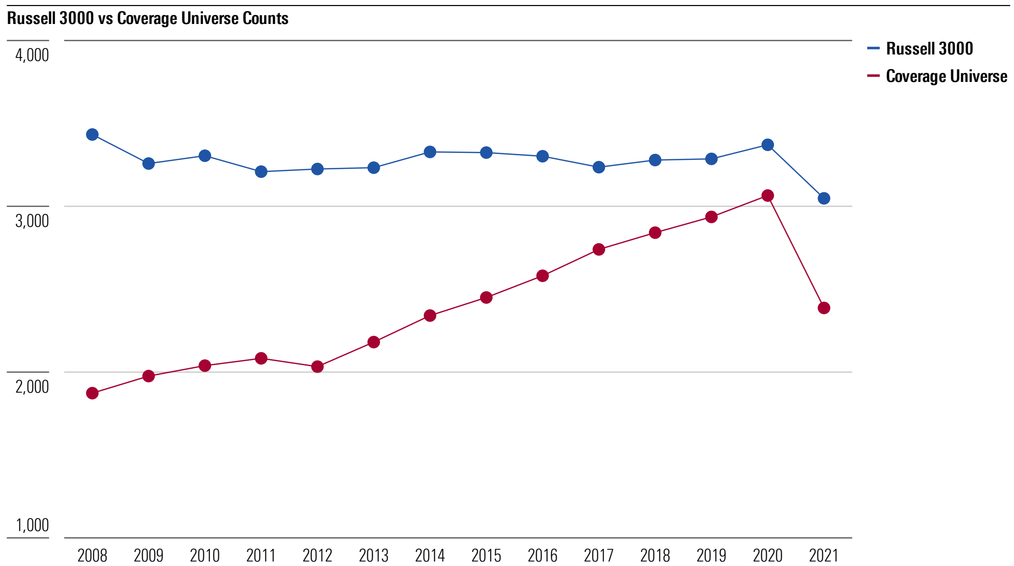
We find that:

- The language complexity factor is unique and differentiated compared with other popular style factors, such as value, size, and momentum over analysis period of 2009-20.
- The complexity factor is investable, having a positive forward relationship with future returns alongside a moderate turnover.
- The complexity factor attributes for fund returns whereby the inclusion in Fama-MacBeth regression improved explainability by 7%.
- Funds holding companies with low complexity factor are generally cheaper and attract more inflows.
- In our performance attribution case study, we found that positive exposure to complexity factor can help improve investor outcomes.

Data

In this study, we select Russell 3000 Index as a representative universe, considering it has 98% of investable U.S. equities and would help to maximize the coverage of U.S equity funds. We download the 10-K filings from the SEC EDGAR website for all the available companies from Russell 3000 Index from 2008 to 2021. In Exhibit 1, we show the number of companies from Russell 3000 Index and our coverage universe from 2008 to 2021. Coverage of companies increases from 2012 onward. Based on the data availability, we cover, on average, 2,400 companies and 2,900 U.S. equity funds in our analysis.

Exhibit 1 Russell 3000 vs Coverage Universe Counts



We subsequently perform our analysis based on the companies present in the coverage universe. Further, to account for these differences, we consider equi-weighted portfolio of all the companies, in a given time, as our benchmark portfolio.

Introduction

Corporate disclosures are key to providing investors with accurate and required information. They serve as a cornerstone of a fair and transparent capital market. In order to facilitate this flow of information in an organized manner, the U.S. Securities and Exchange Commission requires every publicly traded company to file a 10-K report, which provides a comprehensive description of the company's financial performance. The 10-K report plays a key role in keeping investors informed of a company's financial situation and providing them with enough information before making investment decisions. Considering the depth and nature of the information in 10-Ks, it can be tedious for investors to go through them. Consequently, analysts and investors traditionally depended on accounting ratios and market-based models to perform security analysis that could aid them in investment decision-making.

With the proliferation of artificial intelligence/machine earning, and cloud computing in today’s world, a substantial amount of stock buying and selling is triggered by the assessment and recommendations made by robots and algorithms. The SEC estimates that "as much as 85% of the documents visited are by internet bot" (Bauguess, 2018). The major users of this data are predominantly quantitative hedge funds (“How to Talk When a Machine is Listening,” Sean Cao, Wei Jiang et al.). Hence, it is essential to understand the importance of factors derived from corporate disclosures for improving investment decision-making. Our research examines the relationship of complexity factor on stock and fund performance. Because the 10-K filing is specific to the U.S. region, we limit our scope to U.S.-domiciled companies. We specifically focus on the three key sections of annual 10-K filings—business, risk factors, and MD&A. The business section requires a description of the company’s business, including its main products and services, what subsidiaries it owns, and what markets it operates in. The risk factors section includes information about the most significant risks that apply to the company or its securities. The management's discussion and analysis of financial condition and results of operations section, or MD&A, gives the company’s perspective on the business results of the past financial year. The detailed description of these sections is provided in the Appendix.

This paper is divided into the following sections: Illustration, Factor Construction, Factor Differentiation, Factor Efficacy Evaluation, Factor Incisiveness Evaluation, Fund Level Interest Evaluation, and Case Study. In the Illustration section, we provide the example of how language change in filings adversely affected the stock performance of the company. Next, in the Factor Construction section, we provide the methodology for constructing the complexity factor. In the Factor Differentiation section, we evaluate whether the factor is differentiated compared with other popular factors. In the Factor Efficacy Evaluation section, we analyze predictive power of factor alongside turnover. The Factor Incisiveness Evaluation section covers whether the factor helps to distinguish returns by creating long-short portfolios. Subsequently, in the Fund Level Interest Evaluation section, we perform Fama-MacBeth cross-sectional regression to understand the "explainability" of the factor. Finally, we conclude with a few remarks on the complexity factor.

Illustration

What Is Language Complexity Factor?

In an ideal and unchanging state of affairs, companies are not likely to make significant changes to information in three sections—business, risk factors, and MD&A—when evaluated against previous filings. However, if there are significant changes, it indicates that managers may have identified new risks that may impact their future activities. There may also be cases where these changes may be beneficial to the business as well as to investors, but we believe that such changes have a negative impact on their future stock performance, on average.

In order to better understand this factor, we examine one such example of "Bed Bath & Beyond," a home merchandise retailing company that is included among the Fortune 500 companies. First, we compared the 10-K filing of 2015 with that of 2014. In order to understand how complexity can predict subsequent performance, we analyzed the stock performance for the company the following year. We found that the language complexity of the text increased sequentially in the risk factors section, which translated into lower returns the following year.

The following paragraphs are taken from the risk factors section of the 2015 10-K filing of Bed Bath & Beyond. In contrast to the 2014 report, we observe that this section covers additional risks, such as assessment and implementation of emerging technologies, identification and availability of suitable locations to relocate existing stores or opening of additional stores, and debt obligations. These risks had the potential to raise the cost of doing business and could have a negative impact on the company's equity performance. Exhibit 2 depicts an excerpt from the 2015 10-K filing that elaborates these risks.

Exhibit 2 Excerpt From 2015 10-K Filing of Bed Bath & Beyond

Operational Risks

"The success of the Company is dependent, in part, on its ability to establish and profitably maintain the appropriate mix of virtual and physical presence in the markets it serves. The Company's success depends, in part, on its ability to develop its omnichannel capabilities in conjunction with optimizing its physical store operations and market coverage, while maintaining profitability. The Company's ability to develop its omnichannel capabilities will depend on a number of factors, including its assessment and implementation of emerging technologies."

Debt Obligation Risks

"The Company's business would be adversely affected if the Company is unable to service its debt obligations. The Company has incurred indebtedness under senior unsecured notes and has entered into a senior unsecured revolving credit facility. The Company's ability to pay interest and principal when due, comply with debt covenants or repurchase the senior unsecured notes if a change of control occurs, will depend upon, among other things, sales and cash flow levels and other factors that affect its future financial and operating performance, including prevailing economic conditions and financial and business factors, many of which are beyond the Company's control."

Performance of Bed Bath & Beyond

In Exhibit 3, we illustrate the performance of Bed Bath & Beyond from 2010 to 2020. Following the 10-K filing of 2015, we observed a 37% drop in stock prices in 2016. We further observed decline in the stock price, which may be due to the competition from Internet retailers, given that the company's investment in digital initiatives came later than many of its competitors, along with challenges such as the optimization of the retail network, product differentiation, low switching costs for customers, and increased promotional activities in the industry.

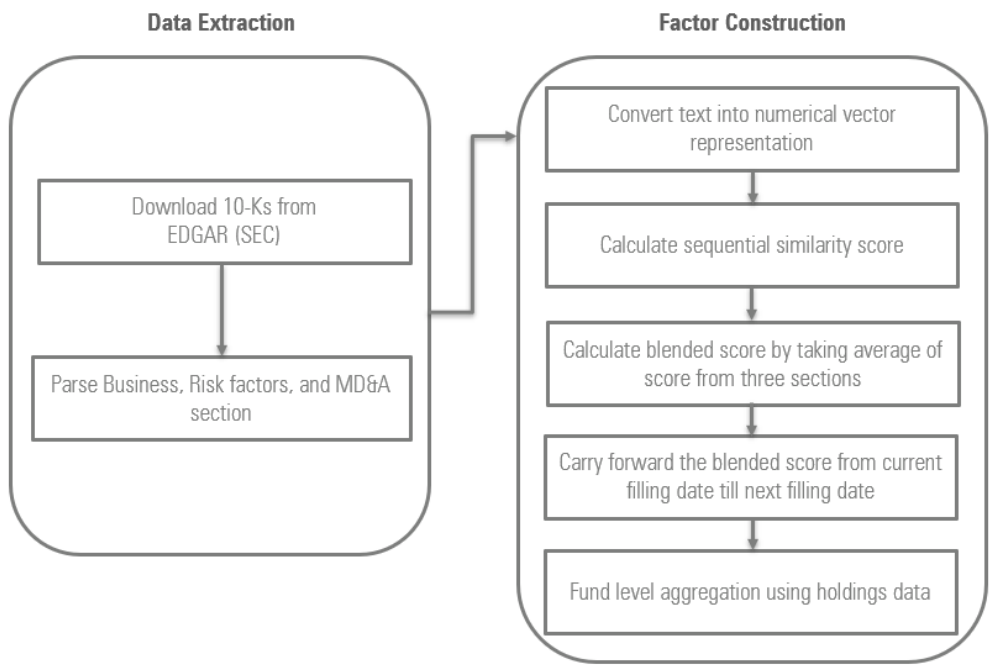
Exhibit 3 Stock Performance (2010-2020)

Source: Morningstar, Inc. Data as of December 31, 2020.

Factor Construction

Exhibit 4 depicts the process we used to build the complexity factor. First, we downloaded the 10-K company filings from the SEC EDGAR website for the Russell 3000 from 2008 to 2021. We then extracted the business, risk factors, and MD&A sections from these filings. After extracting the relevant sections, we obtained the document embeddings and measured the semantic textual similarity of each company across sequential filings. From here on throughout the paper, we will refer to low complexity (high factor values) as favorable and high complexity (low factor values) as unfavorable to investor outcomes.

Exhibit 4 Block Diagram of Framework



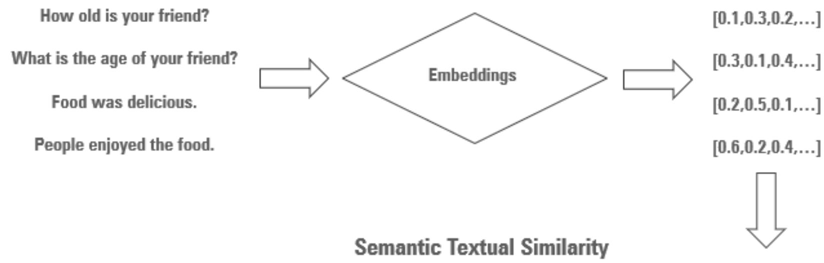
Document Embeddings

Natural Language Processing is the branch of artificial intelligence that deals with training a machine to understand, process, and generate language. Machine-learning models can't read and understand the text in any human sense. Hence, there is a need to convert a text document into numerical vector representation. Embeddings are the main approach to solve this problem, and it is widely used in almost every NLP project. It allows words with similar meaning to have similar representation. To get the document embeddings, we use Google's Universal Sentence Encoder model. The encoding model uses a deep averaging network, or DAN (Iyyer et al., 2015), in which the input embeddings for words and bigrams are initially averaged together before passing through a feedforward deep neural network, or DNN, to produce 512-dimensional embeddings.

Semantic Textual Similarity

Semantic textual similarity helps us determine how the two text documents are close to each other in terms of their context or meaning. In order to measure the degree of similarity, we use the cosine similarity metric. Cosine similarity is used to measure the text similarity between two documents irrespective of their size in NLP. The text documents are represented as vectors in a multidimensional vector space, and the cosine of the angle between two vectors is measured. The closer the documents are by angle, the higher is the cosine similarity (that is, cos theta or cos θ). In this way, the higher the score, the greater the similarity between the documents.

Exhibit 5 Illustration of Semantic Textual Similarity



Semantic Textual Similarity

	How old is your friend?	What is the age of your friend?	Food was delicious.	People enjoyed the food.
How old is your friend?	1.00	0.85	0.15	0.08
What is the age of your friend?	0.85	1.00	0.18	0.10
Food was delicious.	0.15	0.18	1.00	0.80
People enjoyed the food.	0.08	0.10	0.80	1.00

Once we calculate the document embeddings of annual filings for each company, we use the cosine metric to measure the similarity between vectors of sequential annual filings. We calculate the score at the company level for each section over time compared with the last year's filings, and we then create a complexity factor by taking the equi-weighted average of the score from all three sections. We then carry forward the score from the current filing date till the next filing date. In order to calculate the complexity factor at the fund level, we calculate the weighted-average complexity scores based on holdings data as of the given point in time.

Factor Differentiation

Ideally, we want the language complexity factor to be unique and help explain returns through attributes that have not been discovered by earlier factors. Exhibit 6 illustrates how complexity factor is correlated with other popular factors from risk model, including value, size, and momentum from March 2009 to February 2021. We calculate monthly Spearman rank correlation across factors and then take the average. We did not find a strong correlation between these factors, which is less than 0.15. This indicates that the complexity factor is indeed differentiated.

Exhibit 6 Rank Correlation With Popular Factors

	Correlation With Complexity Factor
Size	-0.07
Momentum	0.05
Value Growth	-0.02
Risk	-0.03
Quality	0.11
Volatility	-0.12
Popularity	0.01
Lottery	-0.09
Liquidity	-0.02

Source: Morningstar, Inc. Data as of Feb. 28, 2021.

Factor Efficacy Evaluation

Efficacious factors relate to interpretable and recognizable dimensions of the market. Essentially, these are credible stories that relate to consistent generation of alpha over time. Ideally, a factor should have a high information coefficient, or IC, and rank autocorrelation to be considered credible. IC is a measure of rank correlation between factor and forward returns. We uncover if the factor is investable through Factor Efficacy evaluation.

Exhibit 7 Factor IC and Rank Autocorrelation

	IC Mean	Mean Factor Rank Autocorrelation
1Y	0.039	0.525
2Y	0.054	0.351
3Y	0.061	0.294

We observed positive relationship of 3.9% between factor and forward stock performance, with average factor stability of 52.5% for a one-year horizon. The factor autocorrelation is indicative of portfolio turnover. Given a one-year horizon (indicative of yearly rebalancing), the number seems reasonable.

Factor Incisiveness Evaluation

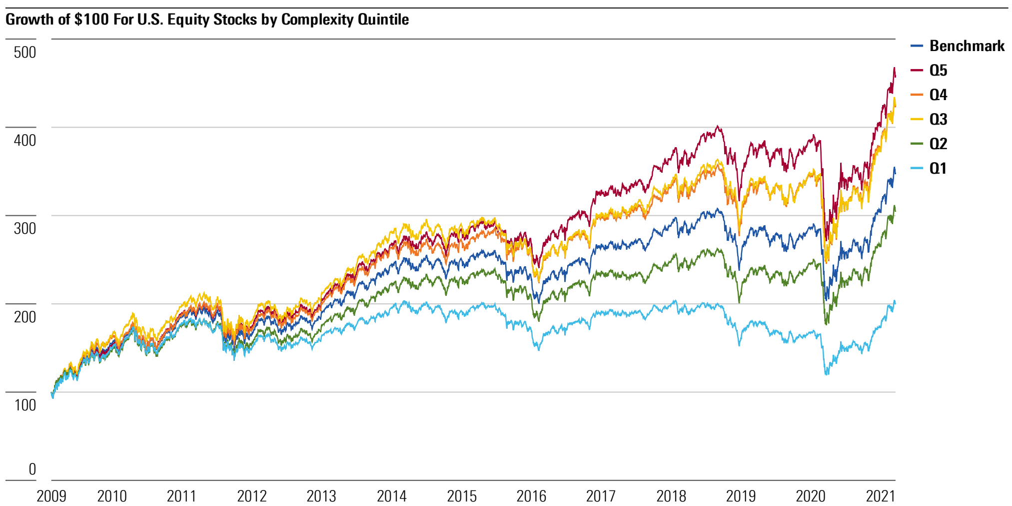
Factor Incisiveness translates to long-short portfolio analysis. In this section, we examine how the portfolio of companies with lower complexity compares with those with higher complexity scores. To perform this analysis, we divide our universe of stocks into five quintiles by sorting them based on the complexity factor in the month of March every year, March 2009 through March 2021. The portfolio strategy is indicative of yearly rebalancing. Because there are delays in filings from January onward, we selected March because most filings are available by this time. Quintile 5 comprises companies with the lowest complexity, while Quintile 1 comprises companies with the highest complexity. Complexity factor scores are calculated using the most recent report published at the time of the annual rebalancing. We then analyze the cumulative growth of each quintile portfolio over the period. Investors may find the portfolio-based tests more intuitive.

In Exhibit 9, we show the performance of the portfolios of stocks and complexity factor quintile. It shows the growth of $100 for U.S. equity stocks by quintile. Here, we observe that the top-quintile portfolio (Q5), having stocks with lowest complexity, outperforms its peers, while the bottom-quintile portfolio (Q1), having stocks with highest complexity, underperforms other portfolios. We also note that the top-quintile portfolio (Q5) has a higher Sharpe ratio of 0.80 compared with 0.59 of the benchmarks (equi-weighted equities in the universe). In contrast, the Sharpe ratio for the bottom-quintile portfolio is 0.25. In early 2020, we observed a dip in the market mainly due to the novel coronavirus crisis. Overall, we observe that the portfolio with companies having lower complexity scores generally perform better than those with higher complexity scores.

Exhibit 8 Portfolio Statistics

Stat	Benchmark	Q5	Q4	Q3	Q2	Q1
Sharpe ratio	0.59	0.80	0.70	0.66	0.53	0.25

Exhibit 9 Growth of $100 for U.S. Equity Stocks by Complexity Quintile



Source: Morningstar, Inc. Data as of March 18, 2021.

Fund Level Interest Evaluation

In this section, we outline the methodology we used to understand the "explainability" of the complexity factor at fund level. To perform the analysis, first we calculate the weighted-average complexity factor at the fund level using holdings data. Based on this fund-level complexity factor, we created quintiles of funds and analyzed their performance from March 2009 to February 2021. In addition, we performed a Fama-MacBeth cross sectional regression analysis at fund level to understand how much of the outperformance of a fund can be explained using the complexity factor.

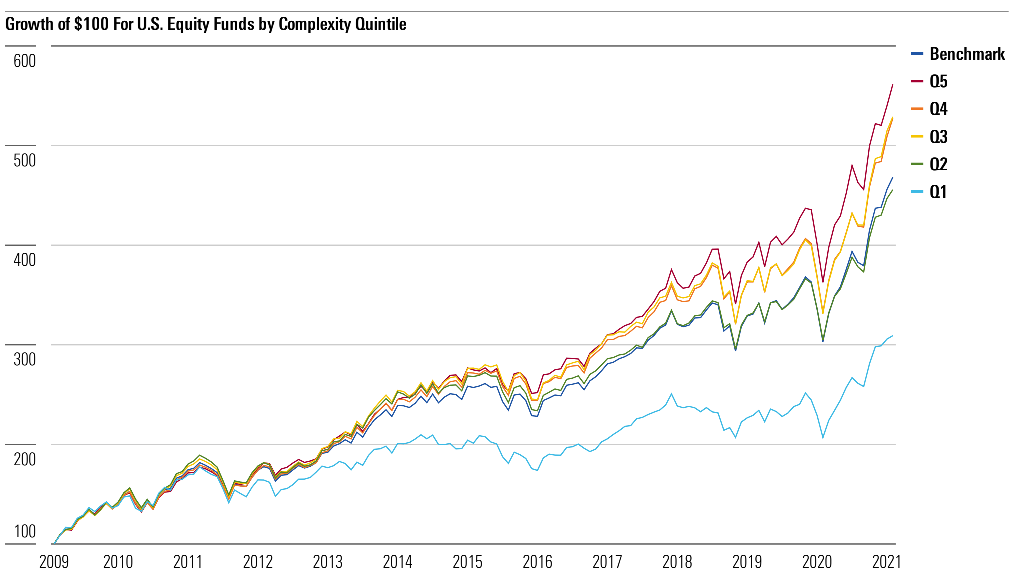
Fund Analysis

In this section, we examine how the portfolio of funds having companies with lower complexity compares with those with higher complexity scores. To perform this analysis, we divide our universe of U.S. equity funds into five quintiles by sorting them monthly (rebalancing frequency) based on the complexity factor from March 2009 to February 2021. Quintile 5 comprises funds having companies with the lowest complexity, while Quintile 1 comprises funds having companies with the highest complexity. In Exhibit 11, we show the growth of $100 for U.S. equity funds by quintile. Here, we observed that the top-quintile portfolio (Q5) of funds outperforms peers while the bottom-quintile portfolio (Q1) of funds underperforms. Also, the top two quintile portfolios outperform than the benchmark (equi-weighted funds in the universe) while bottom two quintile portfolios underperform. We also observed the higher Sharpe ratio for the top-quintile portfolio than the benchmark.

Exhibit 10 Portfolio Statistics

Stat	Benchmark	Q5	Q4	Q3	Q2	Q1
CAGR	13.02%	14.76%	14.14%	14.16%	12.73%	9.18%

Exhibit 11 Growth of $100 for U.S. Equity Funds by Complexity Quintile



Source: Morningstar, Inc. Data as of February 28, 2021.

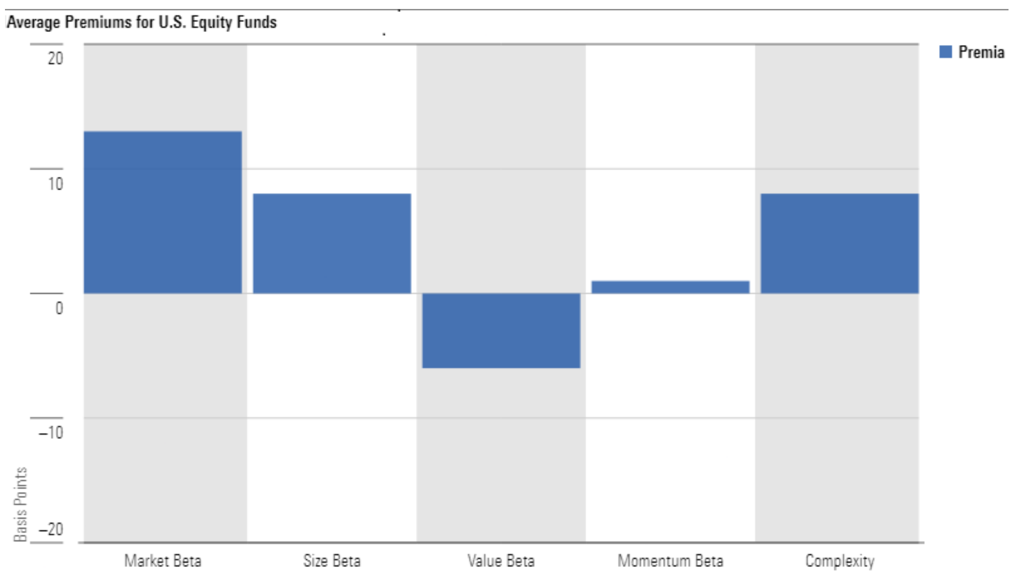
Fama-MacBeth Analysis

We used a Fama-MacBeth cross-sectional regression procedure to test the hypothesis that complexity factor can explain the forward one-month fund net asset value returns. We chose to control for well-known dimensions of returns to better isolate the contribution from complexity factor on performance beyond these drivers of returns. Specifically, we controlled for market, size, value, and momentum. Our universe consists of U.S. equity funds. We conducted monthly cross-sectional regressions using monthly data from March 2009 to January 2021. We estimated the cross-sectional regression monthly and collected all coefficients. After estimation, we applied the Fama-MacBeth procedure to the monthly coefficients to calculate means and standard errors of these coefficients across time. The full description of the model and the variables included is found in the Appendix.

Exhibit 12 shows the average coefficients obtained from the Fama-MacBeth regression for U.S. equity asset class.

We observed that the complexity factor shows to be significant after controlling for 36 months of estimated Carhart factors. The high exposure (that is, less complexity) to complexity factor has a positive impact on fund returns. Additionally, we observed an improvement in the model's accuracy in terms of average adjusted R-squared by 7%, increased from around 23% to 30%.

Exhibit 12 Average Premiums for U.S. Equity Funds



Source: Morningstar, Inc. Data as of January 31, 2021.

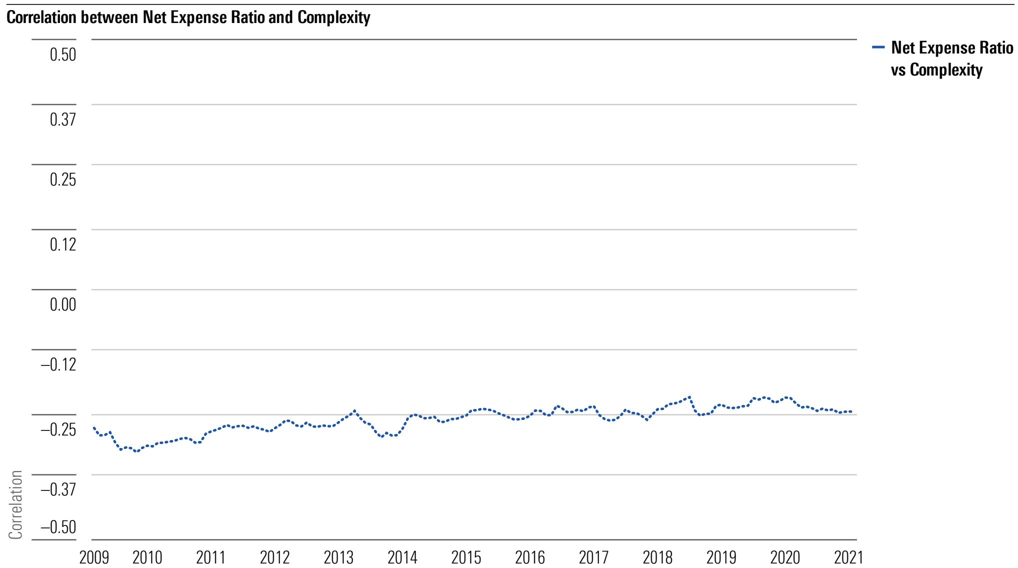
Relationship Between Complexity and Fund Characteristics

To understand the relationship between complexity and fund characteristics, we perform analysis in which we calculate the monthly Spearman rank correlation between the complexity and fund characteristics and plot the values across the time period. Correlation helps us determine how two variables move in relation to each other. A correlation of +1 means that the values of two variables move in the same direction, whereas a perfect negative correlation of -1 means that the values of two variables move in the opposite direction. A correlation of 0 means that the movement of one variable of one investment has no effect on the other variable.

Net Expense Ratio

The net expense ratio of a fund helps us determine the percentage of assets under management directed toward the fund's operating and management costs. We employ an asset-weighted approach to calculate net expense ratio at the fund level. As we can observe below in Exhibit 13, the correlation is consistently negative with average value of negative 0.26. This suggests that funds with companies having low complexity are less expensive and help investors reap more benefits on their investments over time.

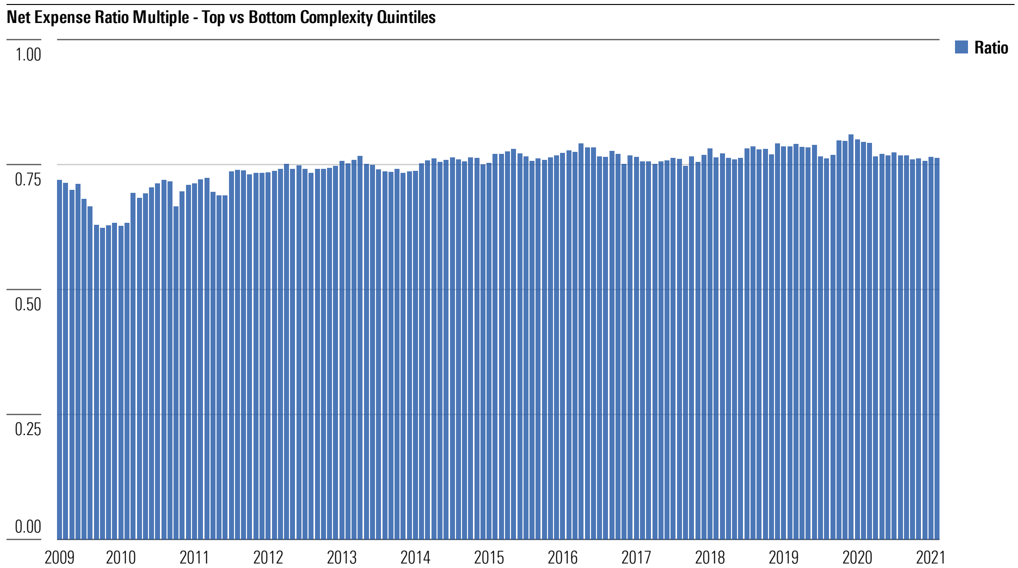
Exhibit 13 Correlation between Net Expense Ratio and Complexity



Source: Morningstar, Inc. Data as of February 28, 2021.

In Exhibit 14, we observe the ratio of average monthly net expense ratio of funds in the top-complexity quintile relative to those in the bottom-complexity quintile, which shows that the funds having companies with low complexity have 0.25 times less expense ratio than the funds having companies with high complexity, on average. The ratio is consistently below 1. Also, we observed average net expense ratio of 0.92% in top-quintile portfolio of funds having companies with low complexity, whereas 1.24% in bottom-quintile portfolio of funds having companies with high complexity. Both analyses suggest that funds that hold companies with lower complexity also tend to have lower expense ratios.

Exhibit 14 Net Expense Ratio Multiple - Top vs. Bottom Complexity Quintiles

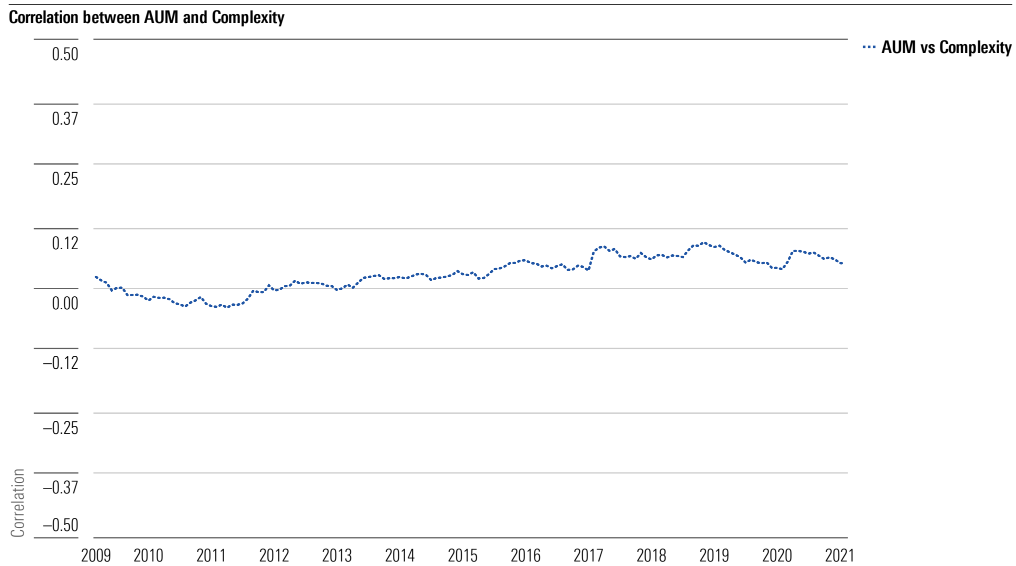


Source: Morningstar, Inc. Data as of February 28, 2021.

Fund Size

The total assets under management of a fund indicates the total market value of all investments that a fund manager manages for her or his clients. It provides investors an indication of the size and popularity of a fund relative to its competitors. In correlation results, as shown in Exhibit 15, we observed that the average value is around 0.03, indicating that funds with companies having low complexity attract inflows. From 2012 onward, we observed continuous positive correlation between complexity and AUM.

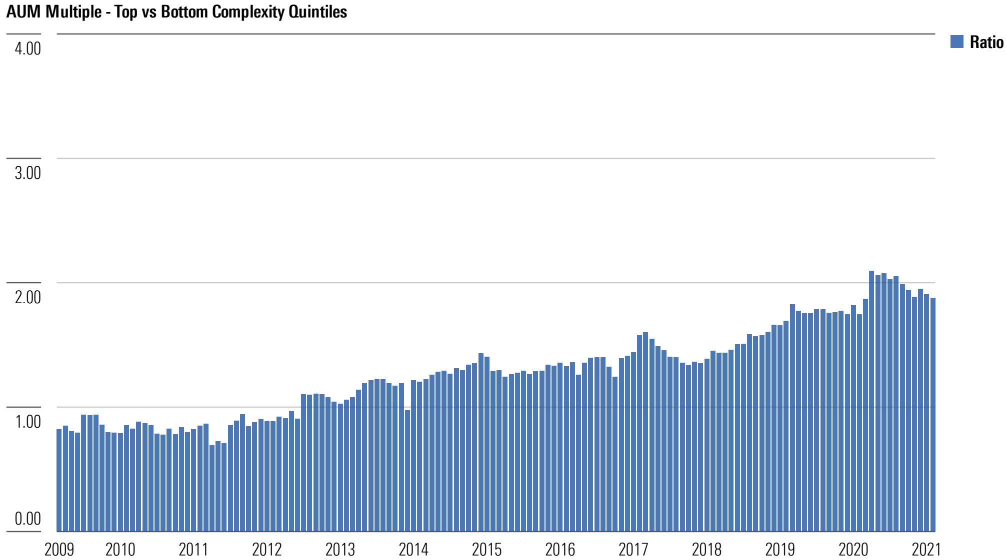
Exhibit 15 Correlation between AUM and Complexity



Source: Morningstar, Inc. Data as of February 28, 2021.

In Exhibit 16, we observe the ratio of average monthly AUM of funds in the top-complexity quintile relative to those in the bottom-complexity quintile, which shows that the funds having companies with low complexity receive 1.3 times more AUM than the funds having companies with high complexity, on an average. This reiterates the fact that investors are drawn more toward funds having companies with low complexity. After 2012, the ratio is consistently above 1. Also, we observed the average AUM of $3.5 billion in the top-quintile portfolio and $2.6 billion in the bottom-quintile portfolio. Both analyses indicate that funds that own stocks with lower complexity also tend to have higher assets. This may be due to the fact that such funds may outperform their counterparts, attracting more flows as a result.

Exhibit 16 AUM Multiple - Top vs Bottom Complexity Quintiles



Source: Morningstar, Inc. Data as of February 28, 2021.

Factor Distribution and Stability

In this section, we discuss the distribution of complexity factor to understand how the asset managers make use of corporate disclosure. In Exhibit 17, we show the overall distribution of complexity factor where a higher score denotes low complexity. We found that the top 25 percentile has lower complexity with a score above 0.79; the median score falls around 0.68. The inclination toward a higher score suggests that asset managers prefer companies with low complexity. This may also mean that they would sell companies that may give weak signals in their corporate disclosures, as measured using our complexity factor.

Exhibit 17 Complexity Factor Distribution

	Complexity Factor Distribution
Mean	0.59
Std. Dev.	0.27
Minimum Value	0.00
25th Percentile	0.46
50th Percentile	0.68
75th Percentile	0.79
Maximum Value	1.00

Factor Stability

In this section, we observe the stability of complexity factor over the period. We create monthly quintiles of complexity factor and then, for each fund and time period, we look at forward 12 months of stability. In Exhibit 18, we show the proportion of stability over a 12-month horizon. We observed that the factor looks consistent. There is slight movement in the quintile buckets; however, it is still in the closer buckets. For instance, the complexity factor stability of Q3 is 0.54, indicating that 54% of the funds remain in Q3 after 12 months. For the same set of funds that were in Q3, we observe that around 16% and 24% moving into Q2 and Q4, respectively. In contrast, the movement of funds to distant quintiles (that is, Q1 and Q5) is very small.

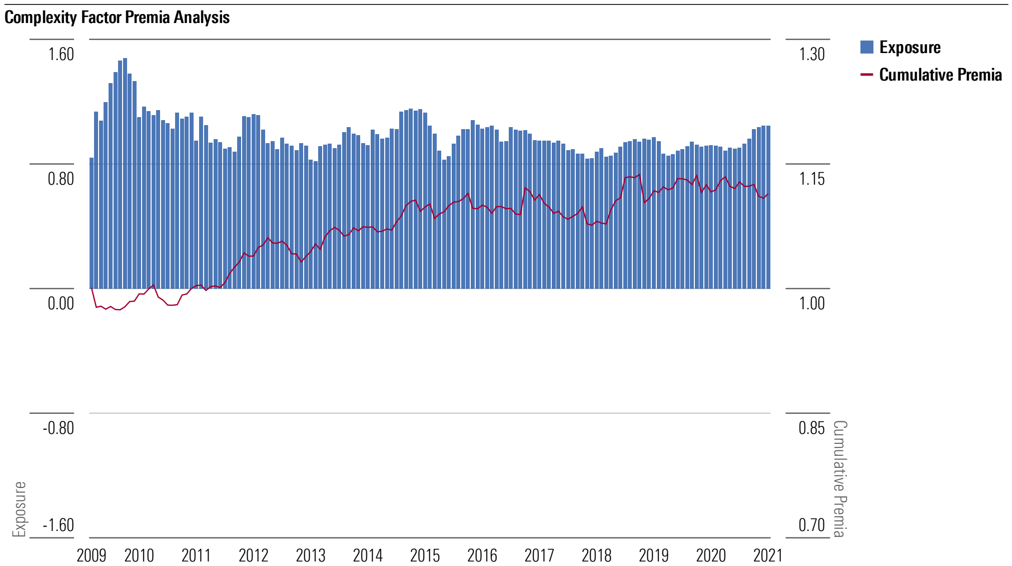
Exhibit 18 Complexity Factor Stability - 12-Month Horizon

	t+12 Months				
Time t	Q1	Q2	Q3	Q4	Q5
Q1	0.89	0.10	0.00	0.00	0.00
Q2	0.05	0.70	0.20	0.04	0.01
Q3	0.00	0.16	0.54	0.24	0.06
Q4	0.00	0.03	0.23	0.50	0.24
Q5	0.00	0.01	0.05	0.23	0.71

Case Study

We would now use the complexity factor for Factor Performance Attribution. Specifically, we evaluated the performance of large-blend funds with AUM more than $1billion and Morningstar Analyst Ratings of Neutral. We observed MFS Blended Research Core Equity (Fund ID: FSUSA000CD). In Exhibit 19, we found that the fund has consistently high exposure (that is, low complexity). Further, we observed the performance of the fund with annualized returns of around 17% for the period.

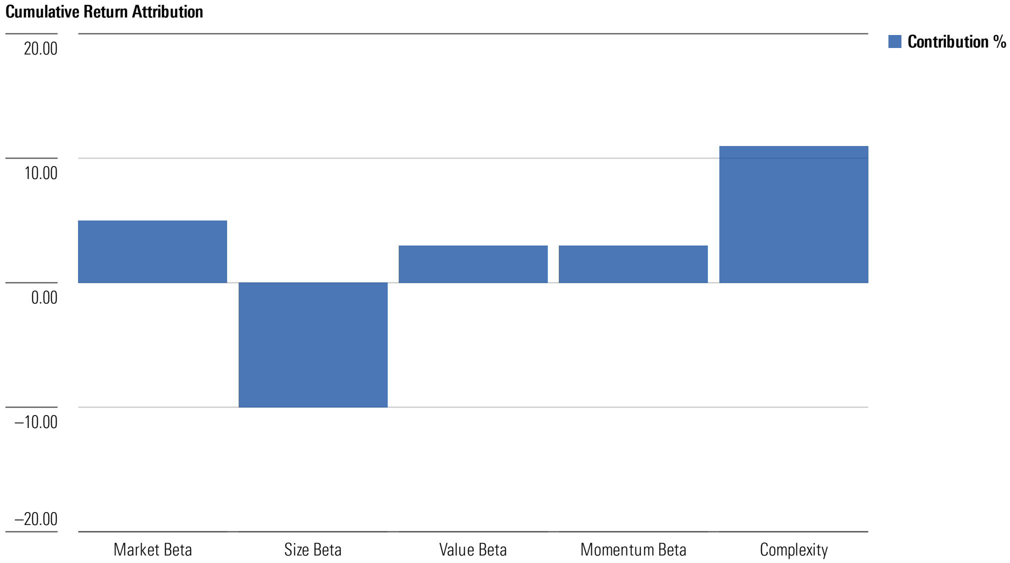
Exhibit 19 Complexity Factor Premia Analysis



Source: Morningstar, Inc. Data as of January 31, 2021.

In Exhibit 20, we found that consistently high exposure (that is, low complexity) to complexity factor attributes toward outperformance, whereas size beta factor attributes toward underperformance for this fund.

Exhibit 20 Cumulative Return Attribution



Source: Morningstar, Inc. Data as of January 31, 2021.

Conclusion

In this paper, we explored the complexity factor to see if it can explain stock/fund performance. Based on our analysis, we can conclude six main findings. First, the language complexity factor is unique and differentiated compared with other popular style factors, such as value, size, and momentum, over analysis period of 2009-20. Second, the complexity factor is investable, having a positive forward relationship with future returns alongside a moderate turnover. Third, investors would generate excess return for funds holding companies with the lowest complexity scores. Fourth, the complexity factor can explain fund returns, whereby the inclusion in the Fama-MacBeth regression improved explainability by 7%. Fifth, funds holding companies with low complexity factor are generally cheaper and attract more inflows. Sixth, in our performance attribution case study, we found that positive exposure to complexity factor can help improve investor outcomes. We do want to highlight the fact that the complexity factor is just one metric that can be obtained using corporate disclosures. There are also different types of disclosures apart from the 10-K (yearly), such as the 10-Q (quarterly) and the 8-K (within four days of unscheduled material events), that are available more frequently. We expect to continue this research and obtain other meaningful NLP factors that can help us dissociate the returns of funds and stocks and make our model more exhaustive. Similarly, we expect to do further research on the relationship of complexity factor with net expense ratio and AUM. In addition, we will continue to enhance the methodology to improve its performance. We will note methodological changes in this document as they are made.

References

Our methodology uses the regression approach pioneered in Fama and MacBeth (1973) to easily calculate standard errors that correct for correlation across assets. Furthermore, using the approach found in Fama and MacBeth (1973), we are able to easily build models in which the independent variables change over time.

Carhart, M.M. 1997. "On Persistence in Mutual Fund Performance." *The Journal of Finance*, Vol. 52, No. 1, P. 57.

Fama, E.F., & French, K.R. 1993. “Common Risk Factors in the Returns on Stock and Bonds.” *Journal of Financial Economics,* Vol. 33, No. 1, P. 3.

Fama, E.F., & French, K.R. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics,* Vol. 116, No.1, P. 1-22.

Cao, S., Jiang, W., Yang, B., & Zhang, A.L. 2020. “How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI.”
https://www.nber.org/papers/w27950

Cera, D., Yang, Y., Kong, S., et al. "Universal Sentence Encoder."
https://arxiv.org/pdf/1803.11175v1.pdf

U.S. Securities and Exchange Commission EDGAR
https://www.sec.gov

Appendix

Fama-MacBeth

Data Tables

In the table below, we show the aggregated monthly cross-sectional regression results. Coefficients are expressed in percentage terms. Below the coefficients, t-statistics are presented in the table. Coefficients can be interpreted as the change in the monthly returns of a fund given a one-standard deviation increase in the factor or, for the cases of dummy variables, when the factor is True.

Exhibit 21 U.S. Equity Asset Class Results

Factors	
Intercept	0.83% 2.17
Market Beta	0.13% 2.43
Size Beta	0.08% 1.22
Value Beta	-0.06% -1.05
Momentum Beta	0.01% 0.30
Complexity	0.08% 1.22

Source: Morningstar, Inc. Data as of Feb. 28, 2021.

Key Sections From 10-K Filings

Item 1 Business requires a description of the company's business, including its main products and services, what subsidiaries it owns, and what markets it operates in. This section may also include information about recent events, competition the company faces, regulations that apply to it, labor issues, special operating costs, or seasonal factors. This is a good place to start to understand how the company operates.

Item 1A Risk Factors includes information about the most significant risks that apply to the company or to its securities. Companies generally list the risk factors in order of their importance. In practice, this section focuses on the risks themselves, not how the company addresses those risks. Some risks may be true for the entire economy, some may apply only to the company's industry sector or geographic region, and some may be unique to the company.

Item 7 Management's Discussion and Analysis of Financial Condition and Results of Operations gives the company's perspective on the business results of the past financial year. This section, known as the MD&A for short, allows company management to tell its story in its own words. The MD&A presents:

1. The company's operations and financial results, including information about the company's liquidity and capital resources and any known trends or uncertainties that could materially affect the company's results. This section may also discuss management's views of key business risks and what it is doing to address them.

2. Material changes in the company's results compared with a prior period.

3. Critical accounting judgments, such as estimates and assumptions. These accounting judgments and any changes from previous years can have a significant impact on the numbers in the financial statements, such as assets, costs, and net income.

About Morningstar Quantitative Research

Morningstar Quantitative Research is dedicated to developing innovative statistical models and data points, including the Morningstar Quantitative Rating, the Quantitative Equity Ratings, and the Global Risk Model.

For More Information

+1 312 244-7541
lee.davidson@morningstar.com

MORNINGSTAR®

22 West Washington Street
Chicago, IL 60602 USA