
Theme Retrieval for Equities

Quantitative Approach for Equity Thematic Exposure Generation

Morningstar Quantitative Research

8 November 2022

Version 1.0

Contents

1	Executive Summary
1	Introduction
3	Data Building Blocks
6	Tenets of Equity Thematic Methodology
11	Evaluation of Theme Basket Outcomes
15	Conclusion

Prashant Srivastava
Senior Quantitative Analyst
prashant.srivastava@morningstar.com

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

Executive Summary

Thematic investing is embodiment of secular growth themes focused on capturing transformational macroeconomic, technological, and ecological trends that have tremendous capability to enhance portfolio outcomes. Given the underlying potential of these themes, there has been a strong interest from institutional and retail investors. This has led to the launch of a range of equity-linked products that have populated the markets and facilitated ways for investors to gain exposure to themes. The information available to investors for evaluating themes is vastly unstructured and accessible in disparate formats from a variety of sources, such as, corporate filings, news outlets, management investment disclosures, and so on. As data on equities continues to expand, it is becoming more challenging for investors to process this growing body of evidence to objectively determine company exposure to relevant themes.

In this paper we discuss a novel approach for identifying equity thematic exposures using emerging technologies to augment equity thematic research. We further determine the economically significant connection between each company and theme exposure measured through our confidence score metric. Our findings indicate that quant-based techniques significantly boosts the efficiency of process for identification of Equity Theme Exposure over vast universes.

Introduction

In recent years, the global menu of equity thematic offerings has expanded in number and breadth like never before. There is an attempt to harness secular growth themes ranging from artificial intelligence to Generation Z. While honing exposure to themes, investors focus on accepting and understanding future return streams that are driven by disruptions and structural changes around technological innovation, consumer habits, business models, economies, markets, regulations, and societal norms, among other things. Thematic intensive strategies are also less constrained by geographic or traditionally rigid approaches of business or industry classification of companies. Also, thematic investing differs from factor investing. Factor investing is based on identifying proven past patterns for which sufficient historical data is available. The thematic investing approach is more forward-looking, arising from new business models, disruptive technologies, or changing consumer tastes and behaviors.

Most disruptive technologies also have an S-curved adoption, and maximum value can be generated through early exposures to theme. There are, however, several challenges to determine thematic exposures for equities during the early phases. First, themes are forward-looking, and it is difficult to identify exposures within historical-structured financial datasets. In most cases there is no direct text

information available within the corporate filings. Instead, we must rely on nonfinancial datasets, such as news and media, which are indicative of an unquantifiable association of a company to a theme. Most of this data may be unstructured, and human-intensive research efforts may not be scalable. Second, it is difficult to identify investable themes. The outcomes of analyst-driven processes are composed of smaller companies that may have tighter liquidity constraints to trade, hindering substantial positions by institutional investors. Underlying companies may also have regional bias, hindering diversification goals and incorporation within existing client portfolios. Third, there may be divergence in theme definitions, limiting exposure to companies directly associated with activities supporting the theme versus in-direct activities, which act as enablers for the theme in the future. Fourth, there may be challenges in evaluating thematic strategies given limited data for back-testing and performance evaluation. An inline weighting mechanism for underlying strategies may pose a challenge to some investors as it may lead to bias toward size-factor exposures. Lastly, regulatory headwinds also prohibit adoption of themes. Lately the SEC has been wary about usage of emerging technologies, like text mining as a stand-alone method for assigning companies to themes. Also, there is a requirement for more disclosure and objectivity with usage of revenue or asset-based tests to determine how linked a company's fortunes are tied to a given theme.

Given the challenges, we have developed a quant methodology (TReE Model) that encompasses emerging technologies like natural language processing to generate thematic exposures alongside bringing objectivity with revenue linkages. Our methodology initiates with a theme definition that is composed of a set of key words that define the building blocks of the theme alongside the supporting pillars. Since these keywords are critical to theme development, they are validated by quant/equity research analysts. Next, we search for the association of these keywords within the relevant text data assets pertaining to companies, such as company business description, revenue activity codes, industry associations, and so on. Once we infer association of the key words, we next determine the exposure objectivity through revenue contribution of identified segments. Once the outcomes are ready, the quant analysts work on theme exposure validation. Analysts check that the outcomes and associated linkages are determined accurately by algorithm. They also check for the occurrence of identified companies within common fund ownership data that tracks similar theme objectives. To further account for information pertaining to new or emerging themes whereby only limited information is available, analysts look at associated news and social media mentions of the company through their respective key words. Then analysts check for volume of relevant news pertaining to partnerships, research and development efforts, patents, and capital expenditures related to the theme. The outcomes of these steps are confidence scores, which are either high, indicative of multiple evidence conforming to the theme exposure, or low scores, indicative of less evidence found to support the theme. Our methodology is transparent, as we can trace back the theme exposure to multiple data points as information contained within corporate filings, news and social media, and revenue attribution.

Morningstar is in a unique position to build this methodology, given rich equity and common ownership data, years of rigorous intellectual property development, and adoption of emerging artificial intelligence and cloud technologies. The newly developed methodology helps improve the efficiency of the theme - generation process.

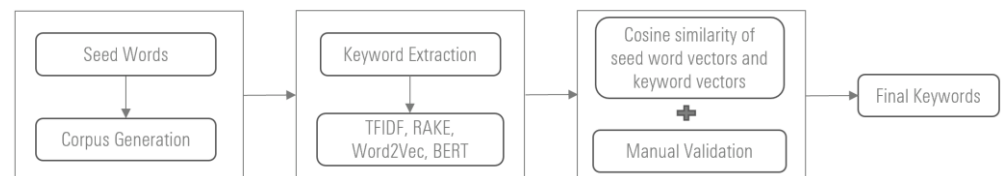
Data Building Blocks

Below are some of the key data points used in building our methodology.

Keyword Dictionary

We begin by articulating the theme objective to reflect the chosen theme. This objective should include high-level business areas associated with the theme while outlining the key sub-ideas that comprise it. We first investigate a theme and capture the seed words that describe the theme. This is enabled either by equity and manager research analysts who have subject matter expertise in the area. Next, we translate and expand a thematic objective ("seed words") into a set of relevant words or phrases ("key words") that are representative of the theme objective. This step is essential, as words can exist in multiple contexts and natural forms. Leveraging natural language-processing techniques, we expand seed words to key words. Word embedding is a natural language processing, or NLP, technique that helps us find relevant words while looking at large word corpus (such as Wikipedia) available globally. These keywords are then used to shortlist the stocks related to the theme of interest within our methodology. Exhibit 1 below describes the underlying workings of keyword-based NLP approach.

Exhibit 1 Expanding Seed Words to Keywords Using Natural Language Processing Techniques



We next try to understand unique subthemes within each theme. To do this, we execute another analytical technique, which is topic modeling, used for discovering the abstract "topics" that occur in a collection of documents. This helps us uncover hidden semantic structures (subthemes) in a text body. The rationale here is that themes have several exposures. As an example, luxury theme could span across multiple subtopics like alcohol, apparel and accessories, beauty consumer products, and so on. Some investors may also have interest to understand and invest in subthemes only. Some companies do not have mentions of keywords defining the themes but may operate in related business areas that are indicated by subthemes. This is indicative of a low theme exposure, which may expand to a high exposure in the future.

Data From Corporate Filings

We leverage data in the business description section of the corporate filings. This is text data and describes the business of the company, such as company beliefs, descriptions of subsidiary business lines, and an overview of operating markets. Morningstar has good coverage of this data for global companies available in English. The section also mentions recent events, competition, regulations, and labor issues. This section provides a very good idea about the key operating units of business for the company. Emerging themes spur out of existing business units of the companies and hence units most related to the keywords are indicative of theme exposures. Exhibit 2 below describes the business description section of corporate filings for LVMH company as it relates to luxury goods theme keywords.

Exhibit 2 Business Description for LVMH

Company Description
LVMH is a global producer and distributor of luxury goods. It operates six segments: fashion and leather goods, its largest and oldest; watches and jewelry ; wines and spirits ; perfumes and cosmetics ; selective retailing (including Sephora and airport duty-free retailer DFS); and other (including publishing). Higher-profile brands include Louis Vuitton, Bulgari, Fendi, Givenchy, Tag Heuer, Hennessy, Moët & Chandon, Glenmorangie, Sephora, and Benefit. LVMH operates more than 5,000 stores around the globe and is headquartered in Paris.

News and Social Media Data

News and social media data play a crucial role in disseminating information and influencing stakeholders, such as investors, and their decision-making processes (Donaldson & Preston, 1995). Bushee and Miller (2012) argue that firms that are less visible to investors have less chance to attract attention among market participants, even if they do disclose information on the firm. These findings can be traced back to the concept of attention-grabbing stocks (Barber & Odean, 2008) and the supposition that investors need to get to know about the firm before making an investment decision regarding it (Merton, 1987). Compared with corporate filings, this data is available in a timely fashion and especially relevant in detecting exposure for new themes. News data has mention of unquantifiable evidence as partnerships, research and development efforts, patent applications, capital expenditure spends, and so on. We use news data from Morningstar News API, which has sound coverage of news events for global companies. Exhibit 3 below describes a sample of news data headlines for companies, as well as a sample of available data for LVMH with exposure to the luxury goods theme.

Exhibit 3 News Headline Data for LVMH

News Headlines
LVMH Shines in a Robust Luxury Field
Louis Vuitton Seen Dominating the Luxury Sphere
LVMH Looks Alluring Amid Luxury Uncertainty
LVMH Has Thrived During the Pandemic. The Luxury -Goods Group's Gains Can Continue

Business Segment Data

Morningstar equity data teams capture revenue attributed to respective business segments of the company. Each segment has a brief description of the operations as described by the parent company. Each business segment is also classified into Morningstar's new GECS, or Global Equity Classification Structure, business activities. Our equity data team relies on sector specialists and a consistent methodology to assign each company's primary line of business to GECS sectors. For theme exposures to be objective it is important to determine if identified segments meaningfully contribute to a major portion (50%) of the company revenue-related activities. A sample of data for LVMH company is described in Exhibit 4 below.

Exhibit 4 Business Segment Revenue Data for LVMH

Segment Name	Segment Description	Morningstar Business Activity Code	Segment Percentage
Wines and Spirits	The wines and spirits segment is divided between the Champagne and Wines segment and the Cognac and Spirits segment. This business group focuses on growth in high-end market segments through a powerful, agile international distribution network. LVMH is the world leader in cognac, with Hennessy, and in champagne , with an outstanding portfolio of brands and complementary product ranges. It also produces high-end still and sparkling wines from around the world.	2051002001	9.2892
Fashion and Leather Goods	The Fashion and Leather Goods segment includes established Maisons with their own unique heritage and more recent brands with strong potential. Whether they are part of haute couture or luxury fashion, LVMH's Maisons have based their success on the quality, authenticity and originality of their designs, created by talented, renowned designers.	1028004001	48.0331
Perfumes and Cosmetics	The Perfumes and Cosmetics segment consists portfolio of world-famous established names as well as younger brands with a promising future. Its Perfumes and Cosmetics business group boasts exceptional momentum, driven by growing and securing the long-term future of its flagship lines as well as boldly developing new products.	2052503002	8.8936
Watches and Jewelry	The watches and jewelry segment operates in the high-end watchmaking , jewelry, and high jewelry sectors. It features some of the most dynamic brands on the market, positioned to complement each other's strengths.	1028004001	13.8163
Selective Retailing	The Selective Retailing business group comprises the Group's own-label retailing activities.	1028004001	18.1891
Other and holding companies	The Other and holding companies business group comprises brands and businesses that are not associated with any of the above-mentioned business groups, particularly the media division, the Dutch luxury yacht maker Royal Van Lent, hotel operations and holding or real estate companies.	3103001001	1.7784

Common Ownership Holdings Data

Morningstar has a rich history of fund holdings and covers over 200,000 funds globally. Further, Morningstar has a methodology for tagging thematic funds based on intentionality of holdings. Our hypothesis of equity exposure to specific themes is based on the existence of equities with high frequency within funds exposed to specific themes. As an example, LVMH stock has a strong exposure to the luxury goods theme and is visible in the majority of fund holdings tagged to luxury goods theme. We refer to this metric as common ownership score. To arrive at common ownership score for equities, we aggregate the holdings of all funds aligned to a theme and then create a frequency score that is indicative of specific equity present across holdings of multiple funds. The higher the frequency of occurrence of the company, the higher the common ownership score. We will describe usage of this metric to reduce the false positives and false negatives as part of our methodology in later sections. The below exhibit describes the process for creating a common ownership score for equities pertaining to the luxury goods theme.

Exhibit 5 Equity Common Ownership Score Example

SinoPac Luxury and Lifestyle Fund	Holding Weight
LVMH	9.60
Hermes International	8.73
Nike	7.89
Apple	6.20

IBK Luxury Lifestyle Master Equity	Holding Weight
LVMH	10.04
Compagnie Financiere	9.09
Tesla	8.11
Hermes International	8.02

Echiquier Luxury I	Holding Weight
LVMH	6.96
Essilorluxottica	6.79
Estee Lauder Cos	5.83
Nike	5.40

Stock Name	Common Ownership Score
LVMH	3
Hermes International	2
Nike	2
Apple	1
Compagnie Financiere	1
Tesla	1
Essilorluxottica	1
Estee Lauder Cos	1

Tenets of Equity Thematic Methodology

We follow a flexible methodology with an automated and efficient way of calculating thematic exposure for equities based on available evidence. Below are some key tenets of our model.

Rule-Based Constituent Selection: We identify long-term investable theme exposures based on evidence of related keywords detected in Corporate Filings data. The constituent selection is done by an algorithm that avoids bias in exposure detection. With a monitored set of keyword-matching rules, we aim to balance the desire to include companies directly and indirectly linked to the theme.

Alternative Datasets: Themes may be nascent, in which case it is difficult to find direct linkages in company financial statements. We therefore leverage alternative data assets pertaining to business description information from companies' filings, news and social media data, business segment names, industry associations assigned to those business segments, business segment revenues, and common ownership scores as part of our methodology. The data is result of years of Morningstar research and intellectual property.

Human-in-the-Loop Review: The methodology outcomes are reviewed by a team of quant analysts. We have decided to take this approach because thematic classification is not always a clear-cut exercise. There are also data errors and hence analysts review multiple data points of identified constituents before confirming the allocation to the theme.

Global Coverage: Consistent with the philosophy of thematic investing, we review a global (as opposed to regions or sectors) coverage universe of companies for our thematic baskets. Our starting universe is much larger with Morningstar Total Market Exposure Index, or TME, which alleviates concerns pertaining to liquidity and invisibility of themes. There is no strict minimum or maximum number of companies in a basket; rather each basket is built to reflect the theme's breadth.

Scalable and Flexible: Unavailability of limited data about companies or themes can prohibit identification of exposure for global equity markets. Our methodology is flexible, and it relies on multiple

inputs to detect a diverse range of themes exposures. The methodology is highly customizable and can be quickly adapted to calculate custom equity baskets for both new and existing themes.

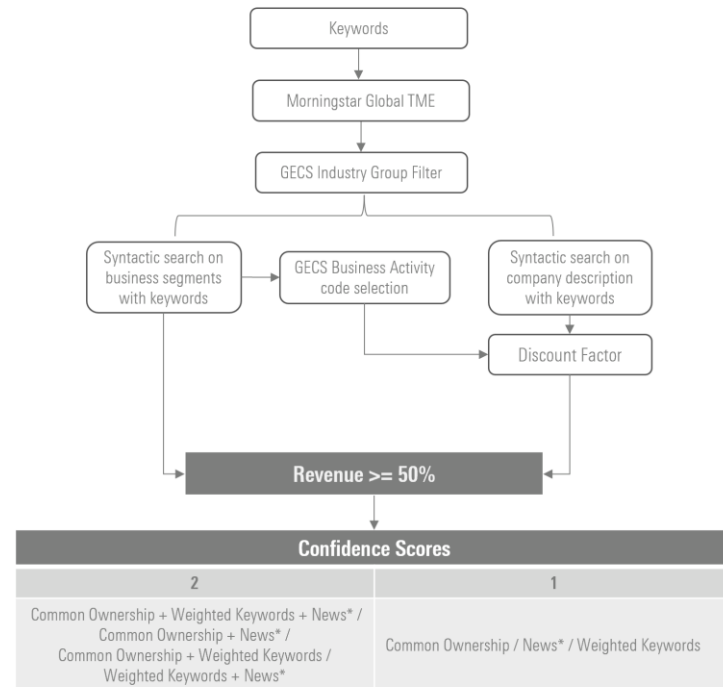
Leveraging Emerging Technologies: The methodology uses machine-learning techniques to build a dictionary of relevant words associated with the theme and can thereby quickly translate a theme objective into an equity basket. This helps reduce the reliance on subject matter experts for each theme and brings a lot of efficiency in the theme exposure detection process.

Transparency and Model Explainability: We want to build a methodology that provides sufficient evidence around inclusion of equities on themes based on multiple data points, such as corporate filings, business segment revenue breakdown, news, and common ownership scores. We believe that this may be important from a regulatory standpoint and can provide comfort to clients around inclusion and exclusion of specific equities.

Theme Governance: On an annual basis we monitor the constituents of the universe to determine if any new companies have exposures to the theme. We also monitor the seed words and key words that are representative of the underlying theme for any changes. These companies are reviewed by analysts to ensure that only relevant companies are added to the basket. Companies that have existing exposures detected would continue to stay in the basket for three years or until the quant analysts determine major changes to exposures of the underlying companies based on supporting evidence.

Methodology Overview

We now move to describing the key building blocks of our Thematic Equity Exposure identification methodology. Exhibit 6 below describes the high-level approach and steps that enable the identification of equity exposures.

Exhibit 6 Constituent Selection Process Overview

* News[Volume, M&A, partnership, patents, Capex]

Step 1: Our methodology first describes a theme in terms of seed words. These seed words are curated in collaboration with equity research and manager research analysts. As described earlier, we then expand these seed words into a larger set of keywords using NLP techniques, the description of which was provided in earlier sections. Next, we identify sub-themes, which are building blocks of the theme. This is accomplished using topic modeling. The below exhibit describes the seed words and key words for luxury theme. While theme keywords essentially describe the theme, the key word for sub-themes helps us identify the supporting pillars for the main theme. As an example, for the Metaverse theme to pollinate, we need gaming companies that help in creation of virtual spaces. We also need companies that are actively involved in the creation of augmented reality and virtual reality devices that help create an immersive experience to interact with virtual spaces. The relevant keywords are reviewed at an annual level by the analysts. Exhibit 7 below describes the supported keywords for the luxury goods theme.

Exhibit 7 Keywords Describing Luxury Goods Space and Associated Keywords for Sub-themes

Luxury	Luxury, luxe, high-end, premium, branded, vintage, high quality, leather, exotic, cognac, athleisure, golf, ski, suites, precious metals, gemstones, yachts	Theme Definition
Alcohol	cognac, whisky, wine, champagne	Value Levers
Apparel & Accessories	apparel, athleisure, suit, handbags, timepieces, watches, accessories, footwear, wearables	
Beauty	cosmetics, skin care, hair care, fragrance, perfume	
Consumer Products	leather, luxury retails, boutique, consumer products, value added, luxury franchise, healthy food, premium chocolates	
Home	décor, interior designing, mansions, villa, arts, luxury furnishings, luxury home builders	
Jewelry	gold, platinum, diamond, jewelry, gemstones, precious metals, sterling silver	
Travel & Recreation	leisure, vacation, tour and travels, ski, resorts, casino, golf, theme parks, recreation, cruise	
Vehicles	sports car, high-end vehicles, yachts, superyachts	

Step 2: Universe Selection

We next select the appropriate universe to be used in our analysis. In our case, use the Morningstar TME Index, which is the widest available equity universe capturing 97% of market cap across global markets. This ensures that the underlying constituents are investable and carry enough liquidity for institutional investor use cases. Usage of a large universe also enables efficient portfolio diversification. We, however, have the capability to determine theme exposure for regional markets based on local Morningstar indexes or other custom indexes.

Step 3: Relevant Industry Groups Matching

We conduct a keyword and contextual match to detect relevant Morningstar GECS Industry Groups related to the theme based on associated keywords and text description. Our hypothesis is that themes are tilted toward specific business activities pertaining to GECS Industry Groups. This also controls for false positives as we would like to limit our search to companies tied to specific GECS Industry Groups. Exhibit 8 lists the associated GECS Industry Group for the luxury goods theme.

Exhibit 8 GECS Industry Group Mapping for Luxury Goods Theme

GECS Industry Groups

Beverages - Alcoholic

Consumer Packaged Goods

Hardware

Homebuilding & Construction

Manufacturing - Apparel & Accessories

Retail - Cyclical

Travel & Leisure

Vehicle & Parts

Source: Morningstar

Step 4: Constituent Selection Through Keyword Matching

We next move to searching the relevant keywords information within data assets of universe constituents, such as corporate filings and business segment revenue text. While conducting the syntactic search, we also like to maintain objectivity, so we match the associated revenue impacted by identified theme exposures. Exhibit 9 describes the keyword search exercise conducted for a sample company LVMH.

Exhibit 9 Keyword-Based Search for LVMH in Luxury Theme

Segment Name	Morningstar Business Activity Code	Inclusion via direct route?	Inclusion via alternate route?	Attributable Revenue
Wines and Spirits	2051002001	Yes		9.2892
Fashion and Leather Goods	1028004001	Yes		48.0331
Perfumes and Cosmetics	2052503002	Yes		8.8936
Watches and Jewelry	1028004001	Yes		13.8163
Selective Retailing	1028004001		Yes	12.7323
Other and holding companies	3103001001	Yes		1.7784

Source: Morningstar. Revenue Exposure = 94.54

In order to determine objective theme exposures, we follow the next two steps.

Direct Linkage via business segments: Company-identified business segment description and segment names are compared against the set of theme-related words and phrases. Companies from the universe, which include at least one relevant word in their business segment names or segment descriptions, are included in the eligible universe. For exposure via this route, the entire identified segment relevant to the theme is selected.

Alternative linkage via a company's business summary description: There are cases where we do not have a direct match of keywords in a company's business segment. The company's summary business description is analyzed in such cases using the set of relevant words. Companies from the universe, which include at least one distinct relevant word in their summary description data, are then included in the eligible universe. If the chosen company has any business segments with GECS business activity codes linked to identified GECS sectors determined in Step 2, then that revenue line is also considered linked to the theme. As this is not a direct linkage, we adjust for the uncertainty in this linkage through application of a discount factor to such companies. This discount factor can vary from 50% to 70% based on whether the theme is established, like luxury goods or an emerging theme like metaverse.

The revenue from matched segments is then aggregated to arrive at a cut-off score that helps us determine whether the theme has an objective exposure. In our case, we apply a threshold of 50% (aggregated revenue) to use for inclusion of the constituents in our theme.

Confidence Score Allocation

Once we have a constituent list of objective companies, we then apply analyst intervention to ensure that we could control for false positives and false negatives. Here, analysts look for multiple data points such as news and social media data (keyword match and number of news articles), weighted keywords (frequency of occurrence of keywords), and funds common ownership score as described below. This can be likened to a voting mechanism whereby theme exposure evidence can be detected across all multiple datasets. Accordingly, analysts allocate either 2(High) or 1(Low) Confidence to the constituents based on logic demonstrated in Exhibit 10.

Exhibit 10 Confidence Score Logic

Confidence Scores	
2	1
Common Ownership + Weighted Keywords + News / Common Ownership + News / Common Ownership + Weighted Keywords / Weighted Keywords + News	Common Ownership / News / Weighted Keywords

Further analyst intervention also acts like a feedback loop in the methodology by revising the methodology to account for some missing companies that have high common ownership score. This could be the result of missing seed words or a higher discount factor application. Accordingly, the seed word and the discount factor would then be revised. The analyst review also serves as a feedback loop to the methodology in case insufficient constituents

are detected, and in case there is a new theme that is found. Finally, analysts also check the theme exposures for sub-themes, following a similar methodology discussed above. The result is a condensed sample set of companies that have high exposures to the theme, as seen below in Exhibit 11.

Exhibit 11 Top Constituent List of Companies in Luxury Goods Theme basket

Company	Sub-theme	Country	GECS Industry Group
LVMH	Apparel & Accessories	France	Retail -Cyclical
Hermes International	Apparel & Accessories	France	Retail -Cyclical
Kering	Apparel & Accessories	France	Retail -Cyclical
Pernod Ricard	Alcohol	France	Beverages - Alcoholic
Moncler	Apparel & Accessories	Italy	Manufacturing - Apparel & Accessories
Estee Lauder Cos	Beauty	United States	Consumer Packaged Goods
Compagnie Financiere	Jewelry	Switzerland	Retail -Cyclical
Ferrari	Vehicles	Italy	Vehicles & Parts
Nike	Apparel & Accessories	United States	Manufacturing - Apparel & Accessories
Diageo	Alcohol	United Kingdom	Beverages - Alcoholic

Source: Morningstar.

Evaluation of Theme Basket Outcomes

Consensus Checks for our Thematic Basket

We try to evaluate our theme basket against available structured products. Here, it is preferred that a higher percentage of holdings match across similar thematic products. We refer to this match as a consensus score. The below exhibit describes the consensus score for a luxury goods basket against other available funds. As can be seen, the consensus theme across the board is varied. The key reason is that structured products have been built on varied universes and theme definitions. In the case of our basket, we aim for a high consensus when compared with other thematic products.

Exhibit 12 Consensus of Thematic Basket vs. Funds

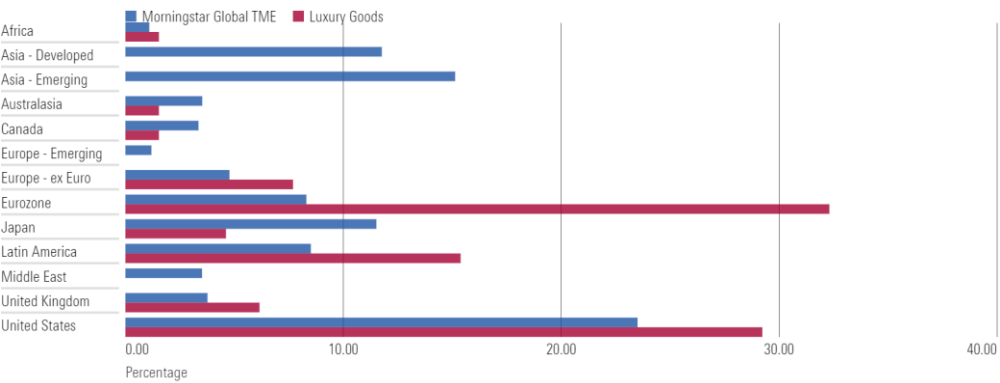
Fund Legal Name	Domicile	Inception Date	Fund Size (USD)	Holdings Count	Consensus (Percentile Rank)
Nomura Global Luxury Brands Fund	Taiwan	21/06/2002	108,189,796	46	0.06
RHB Leisure, Lifestyle & Luxury Fund	Malaysia	22/10/2013	3,463,912	58	0.13
Hua Nan Global Luxury Goods Fund	Taiwan	23/11/2006	13,357,207	33	0.19
Dominion Global Trends - Luxury Consumer Fund	Malta	04/06/2007		26	0.25
JKO Life Style Fund	Taiwan	19/01/2007	6,291,055	29	0.31
Felis Luxury Brands Fund	Liechtenstein	26/09/2019	11,405,526	24	0.38
U.S. Global Investors Global Luxury Goods Fund	United States	17/10/1994	39,511,179	53	0.44
Emles Luxury Goods ETF	United States	24/11/2020	4,606,731	49	0.50
GAM Multistock - Luxury Brands Equity	Luxembourg	31/01/2008	461,258,474	33	0.56
SG Actions Luxe	France	14/10/2002	362,702,949	50	0.63
Diversified Growth Company - Franck Muller Luxury Fund	Luxembourg	10/12/2014	68,909,228	32	0.69
Echiquier Luxury	France	09/05/2007	17,042,172	24	0.75
IBK Luxury Lifestyle Master Equity	South Korea	15/02/2007	57,866,639	28	0.81
SinoPac Luxury and Lifestyle Fund	Taiwan	04/06/2007	7,718,609	27	0.88
E Fund S&P Global Luxury Index Enhanced QDII	China	04/06/2012	26,060,516	26	0.94
Luxury Goods Theme Basket				65	1

Source: Morningstar. Data as of October 31, 2022.

Regional Diversification

We also wish to ensure that our baskets are geographically diversified. In order to do this, we would need to have good representation of holdings across major regions. Exhibit 13 details that our basket has constituents across major regions when measured against the base universe.

Exhibit 13 Regional Diversification of Thematic Basket vs. Benchmark

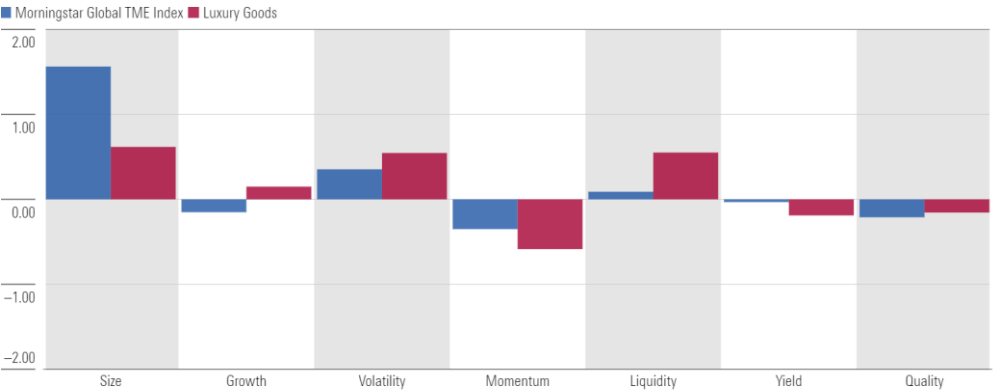


Source: Morningstar

Factor Exposure of Underlying Theme

We also measure the performance attribution of our theme baskets. This is helpful from a performance attribution and risk management perspective. We leverage outcomes from Morningstar Industry Standard Risk Model, which allows us to measure the exposure of the theme basket compared with the benchmarks. This enables users to understand the source of returns and risks for our theme baskets versus the global universe. Exhibit 14 describes the risk factor exposure for the luxury goods basket. As is evident, our theme has higher exposure to growth factor and is composed of smaller companies. The liquidity factor also looks good, indicating that the basket is investable.

Exhibit 14 Risk Model Factor Exposure Theme Basket vs Benchmark

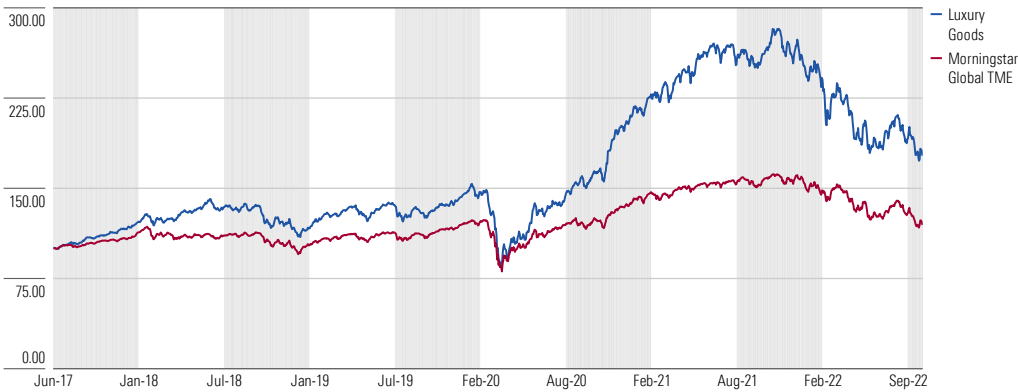


Source: Morningstar. Data as of September 30, 2022.

Theme Basket Performance Back-Testing

Finally, we also compare the back-testing metrics for the theme versus benchmark in a visual below based on a yearly rebalance schedule, using an equal-weighting mechanism.

Exhibit 15 Back-Testing Performance of Theme vs. Benchmark



Source: Morningstar. Data as of October 9, 2022.

Stats	Luxury Goods	Morningstar Global TME
Start	2017-06-30	2017-06-30
End	2022-10-09	2022-10-09
Total Return	76.98%	19.42%
CAGR	11.43%	3.42%
Max Drawdown	-44.78%	-34.63%
Yearly Mean	12.29%	3.45%
Yearly Volatility	32.14%	21.26%
Yearly Sharpe	0.38	0.16

Source: Morningstar. Data as of October 9, 2022.

Conclusion

Thematic investing is an investment approach that has become increasingly popular with both institutional and retail investors, tapping into structured growth themes of the future that are representative of business models, disruptive technologies, or changing consumer tastes and behaviors.

In this paper, we first reviewed the concept of thematic investing. We leveraged Natural Language Processing techniques in order to detect exposure of companies to a secular growth theme. We first built a seed word dictionary dataset that captures the thoughts, products, and services associated with the theme objective. We next expanded these seed words using machine-learning techniques into a wider set of in a keyword dictionary that is representative of the theme. We next assigned objectivity of these associations based on either direct financial linkage to business segments' revenue activity codes or alternatively linking through mentions in the business description text of the company. Further, the outcomes are reviewed by analysts whereby they looked at multiple data points as news (volume, theme associations as it relates to partnerships, mergers and acquisitions, capital expenditures, patents) and common ownership scores to allocate confidence levels to the theme exposures for companies.

The application of the methodology brings a lot of efficiency in our theme-generation process. While the methodology is flexible and can be applied across multiple themes, there may be themes that may need special treatment and therefore the way the exposures are determined would be different for specific companies. As an example, we would need to tweak the methodology for determining exposure to sustainable themes like water or clean ocean by applying sustainability filters on the outcomes of our model. We would also work toward aggregating the theme exposures to fund levels to determine thematic purity scores of the funds. These are subjects that we intend to research and publish papers on in the future.

References

<https://www.sec.gov/rules/proposed/2022/33-11067.pdf>

Donaldson, Thomas, and Lee E. Preston, January 1995. "The Stakeholder Theory of the Corporation: Concepts, Evidence and Implications." *The Academy of Management Review*.
<https://www.jstor.org/stable/258887>

Barber, Brad M. and Terrance Odean. November 2006. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors."

Bushee, Brian J. and Gregory S. Miller. May 2012. "Investor Relations, Firm Visibility, and Investor Following." *The Accounting Review*.
<https://www.jstor.org/stable/23245633>

Merton, Robert C. July 1987. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *The Journal of Finance*.
<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1540-6261.1987.tb04565.x>

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



22 West Washington Street
Chicago, IL 60602 USA

©2022 Morningstar. All Rights Reserved. Unless otherwise provided in a separate agreement, you may use this report only in the country in which its original distributor is based. The information, data, analyses, and opinions presented herein do not constitute investment advice; are provided solely for informational purposes and therefore are not an offer to buy or sell a security; and are not warranted to be correct, complete, or accurate. The opinions expressed are as of the date written and are subject to change without notice. Except as otherwise required by law, Morningstar shall not be responsible for any trading decisions, damages, or other losses resulting from, or related to, the information, data, analyses, or opinions or their use. The information contained herein is the proprietary property of Morningstar and may not be reproduced, in whole or in part, or used in any manner, without the prior written consent of Morningstar. To license the research, call +1 312 696-6869.