

LSEG StarMine

# StarMine MarketPsych Media Sentiment Model



**LSEG** DATA & ANALYTICS

The StarMine MarketPsych Media Sentiment Model (MMS) marks the first StarMine equity returns model based on news and social media sentiment. MMS shows significant top-bottom decile spreads as well as low correlations to traditional equity and StarMine factors.

The StarMine MarketPsych Media Sentiment (MMS) model is a stock ranking system that provides a 1 to 100 daily percentile ranking for over 16,000 global stocks. MMS complements the StarMine suite of equity models and follows a similar methodology in research and implementation. The model is derived from LSEG MarketPsych Analytics, a market leader in financial media sentiment data. The output includes an overall score, as well as specific Equity, Business and Management scores. The MMS scores are designed to forecast the next month's relative share price returns, with higher ranked stocks outperforming lower. Historical evaluation demonstrates significant outperformance of higher deciles versus lower ones, with the top-bottom decile spread for global stocks averaging 10.4% annually from 2006 to October 2020, including 12.3% in the out-of-sample period. The MMS scores are uncorrelated with traditional market factors and complement fundamental models.

#### Underlying dataset: MarketPsych Analytics from LSEG

The LSEG MarketPsych Analytics (RMA) are a market leader in aggregated financial media sentiment. The LSEG MarketPsych Analytics provide sentiment and thematic scores for 16,000+ global companies, as well as stock indexes, commodities, currencies, sovereign bonds, countries and cryptocurrencies. The LSEG MarketPsych Analytics represent aggregate scores from 2 million financial articles per day, ingested in real-time from thousands of news feeds, blogs and comments.

The LSEG MarketPsych Analytics contain granular sentiments with coverage including fundamental, earnings, analyst, management, and price sentiment scores. For equities, the LSEG MarketPsych Analytics deliver 34 distinct sentiment scores including emotions such as fear, trust, and surprise as well as themes such as earnings forecast, price forecast, fundamental strength, company innovation, and management change. More details about the LSEG MarketPsych Analytics are available in the MarketPsych Analytics white paper and [user guide](#) and in this [product brochure](#).

#### Backed by external research

Both academic and industry research on non-random share price behavior emphasises two information-related patterns: overreaction (mean-reversion) and underreaction (trending). The terms under- and overreaction refer not only to the price movement, but also to investors' reactions to company-relevant news. The type of reaction occurring depends on news topic; earnings, management, and mergers-related news each produce different price effects. Characteristics of the news, such as media sentiment, audience, vividness, visibility, and anticipation, each can modulate the impact.

#### Constructing the media sentiment model

The StarMine MarketPsych Media Sentiment model was designed to capture both under- and overreaction patterns by optimally weighing these pertinent themes and sentiments. Aside from considering the impacts themselves, MMS also accounts for the persistence of their effects on price, varying by news and social media.

The model construction process was carefully chosen to:

- Identify which LSEG MarketPsych Analytics sentiment scores perform best over time
- Optimise the duration of prediction effect of each LSEG MarketPsych Analytics
- Develop consistent handling of sparse and noisy data
- Create a logical combination of LSEG MarketPsych Analytics that allows for further study

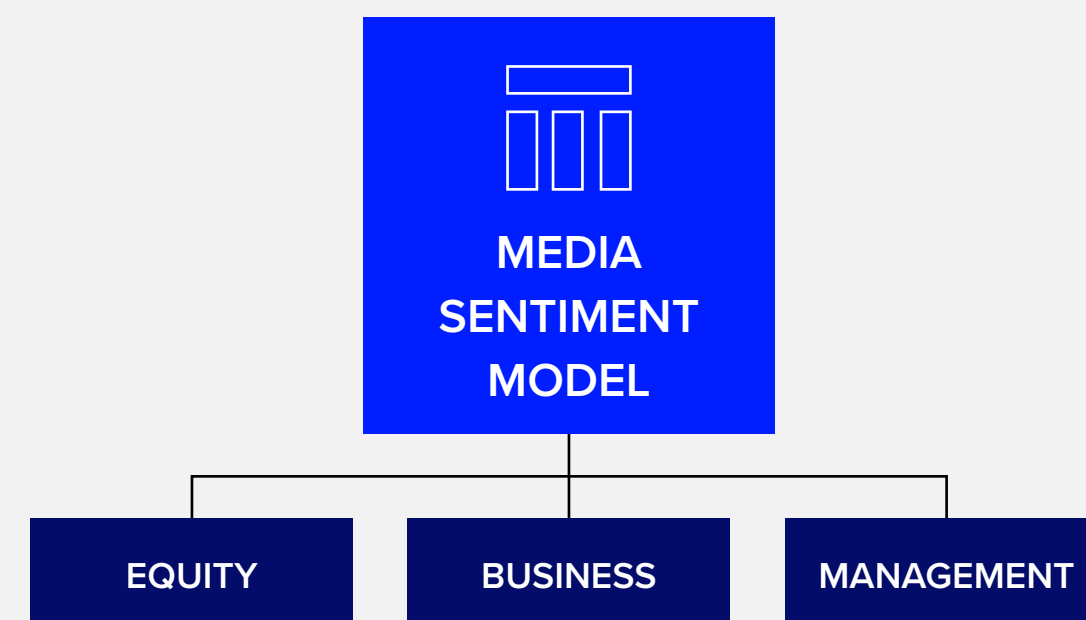
MMS model construction used sentiment scores that fall into three categories. Each group expresses a unique aspect of media sentiment about the company.

- **Equity:** company and its equity price action
- **Business:** fundamentals, earnings, and analyst reports
- **Management:** management and corporate events

**Table 1. Approximate company coverage**

Model Region	No. of assets
United States	6,000
Developed Europe (United Kingdom, Norway, Italy, Spain, etc.)	2,500
Developed Asia (Australia, Hong Kong, Singapore, etc.)	2,000
Emerging Markets (remaining countries)	4,000
Japan	1,000
Canada	600

Figure 1. Overall model composed from sentiments from three categories.



### Modeling process

MMS model rankings are designed to forecast the next-month returns of equities, relative to other equities within each region of the model. MMS model inputs represent transformations of the LSEG MarketPsych Analytics that account for their varying contributions across durations (significance over time) and sources (news vs. social media). The dataset was prepared with an in-sample period ranging from 2006 to 2018, with every third year used as a holdout for the out-of-sample period. 2006 was chosen as the starting point due to significantly lower volumes of online news and social media prior to 2006. Only the final MMS model was tested on the out-of-sample windows. During model preparation and

development, the MarketPsych team maintained a regular dialogue with the StarMine quantitative research team on best modeling practices.

### StarMine methodology

StarMine models employ rigorous methodologies to avoid data mining and overfitting. Generally, such practices result in strong live performance and robust returns across sectors, market cap groups, and time periods.

Similar rigor was applied to the MMS model. Great care was taken to properly manage training and testing samples to avoid cross-contamination. Working within the validation period, model complexity was reduced to further minimise the chances of overfitting.

The trained model demonstrates similar positive performance in the out-of-sample periods as in the training periods and shows robust performance across sectors, size groups, and time periods (including bull and bear markets) in all samples.

Readers are directed to see the MMS white paper for more information on model construction.

### Rank returns

For each model region, the average next-month equity return for each stock decile, ranked by the overall MMS model, is displayed in Figure 2. For example, Decile 10 represents the equities with scores between 91 to 100 on the last day of the prior month, and is expected to outperform on average. The average monthly decile returns show an ascending relationship from lowest to highest rank. The decile averages include both the in-sample and out-of-sample periods.

Figure 3 displays the cumulative performance of the top and bottom deciles and their spread, for each model region. The average monthly returns of stocks with MMS scores in the highest decile, lowest decile, and decile spread are depicted over time. No transaction costs are included. The in-sample time periods are shaded with a gray background and the out-of-sample periods have no background shading. For all stocks globally, the spread averages 10.4% annually from 2006 to October 2020, including 12.3% in the out-of-sample period.

Figure 2. Average monthly return of MMS Deciles – 2006-October 2020

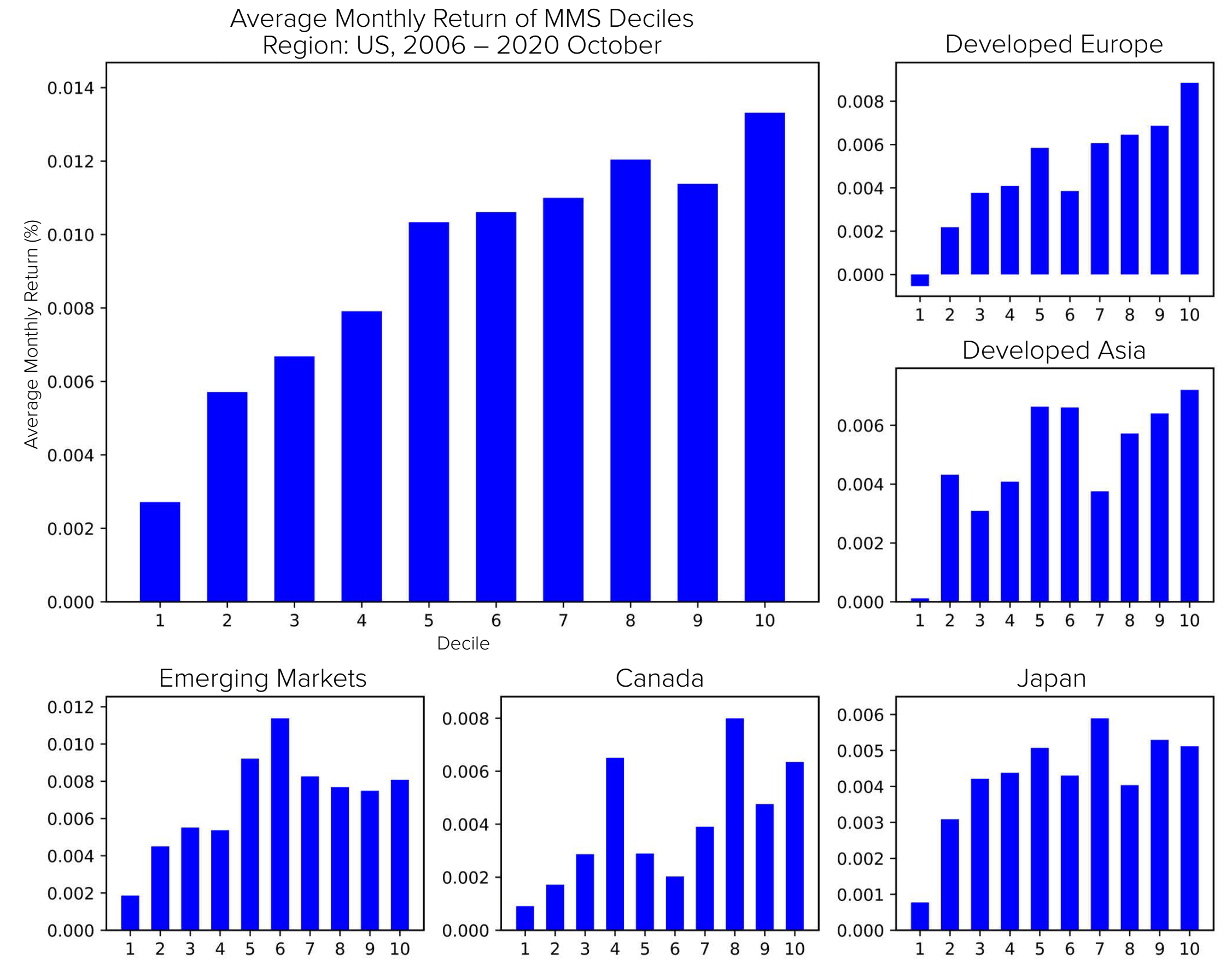
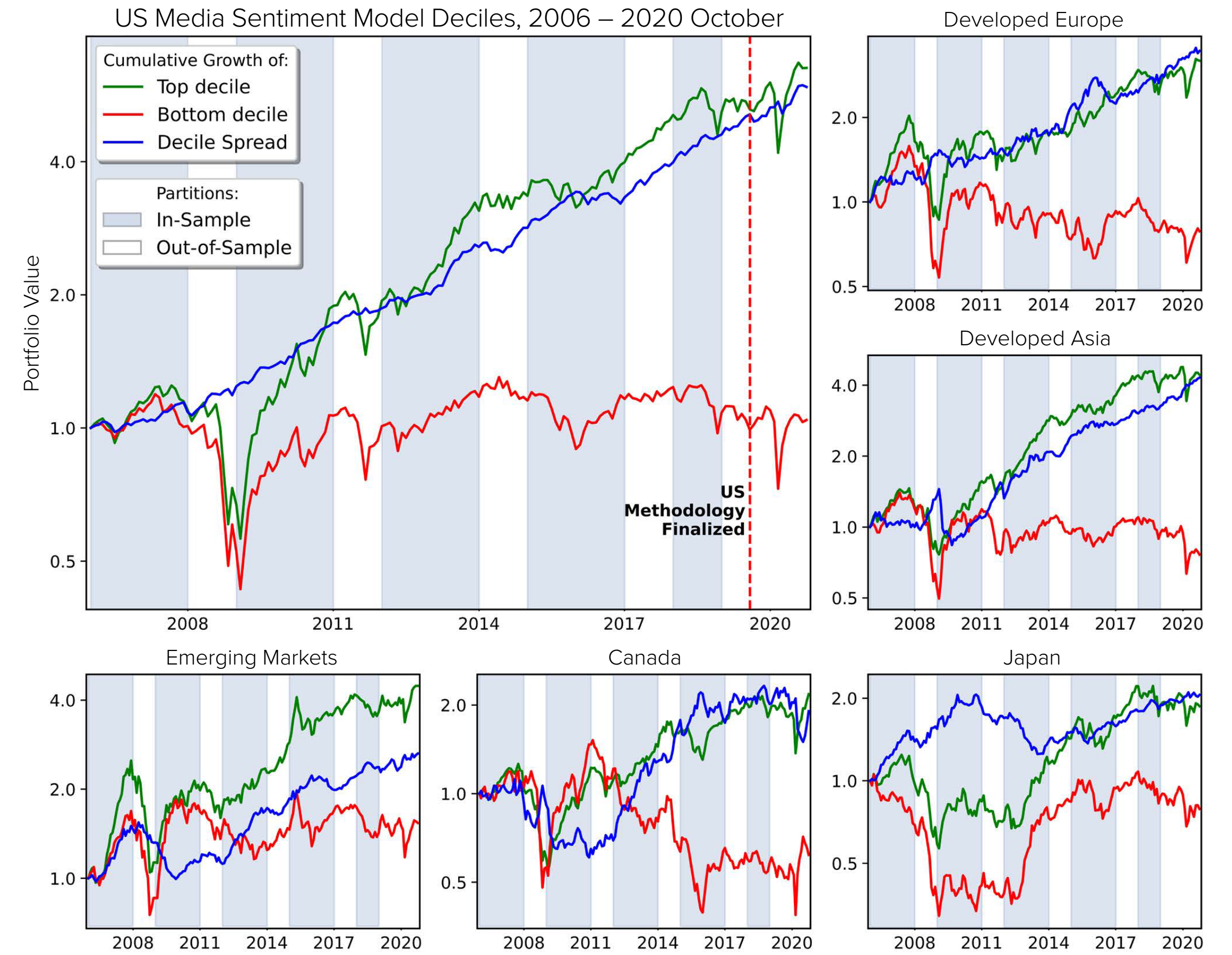


Figure 3. Extreme decile and spread performance from February 1998 through October 2020 for the overall MMS model



### Unique alpha

MMS ranks have a low correlation not only with traditional fundamental and price factors, such as market capitalization, past returns, and volatility, but also with current StarMine models, as can be seen in Table 2 below. This finding suggests that the MMS model captures a uniquely valuable aspect of market information influencing returns independent of past price action and fundamentals.

**Table 2. Correlations between a stock's MMS rank and fundamental or StarMine factors.**

Correlation Coefficient	Market Cap	Past Month Return Rank	Next Month Return Rank	Analyst Revisions	Earnings Quality	Value-Momentum	Smart Holdings
MMS	0.058	0.058	0.026	0.165	0.118	0.058	0.154

Furthermore, the ranks from the three sentiment component categories – Equity, Business, and Management – have low correlations with one another, each providing distinct value in prediction. Each component rank allows the user to leverage these specific themes in their research. Please see the product white paper for more information.

### Regional Statistics

Table 3 below displays the performance statistics for each model region, from 2006 through October 2020. Each model is constructed with both a regional weighting - to account for local behavioral differences - and a shared global model weighting. The first three columns in Table 3 show regional average annualized monthly returns for the Top Decile, Bottom Decile, and the Average Decile Spread, respectively. Information Coefficient refers to the correlation between the company's MMS rank and its next month's return rank. Sharpe of Spread calculates the sharpe ratio on the extreme deciles spread. Average Asset Count is the number of companies with scores, on each of the month ends in the sample. Average Turnover represents the percentage of assets leaving the top and bottom deciles from one month to the next.

**Table 3. Returns Statistics**

Name	Top Decile	Bottom Decile	Average Spread	Information Coefficient	Sharpe of Spread	Average Asset Count	Average Turnover
United States	17.32%	3.23%	13.19%	0.029	1.837	2,900	33.65%
Developed Europe	11.58%	-0.24%	11.36%	0.043	0.829	856	35.37%
Developed Asia	9.14%	0.45%	9.26%	0.038	0.849	786	34.78%
Emerging Markets	10.23%	2.43%	7.24%	0.025	0.735	1,011	45.46%
Japan	6.64%	1.40%	5.46%	0.02	0.527	397	25.57%
Canada	7.38%	1.67%	5.60%	0.035	0.293	244	40.90%

## Conclusion

Drawing from decades of experience in sentiment research, the MarketPsych team created the StarMine MarketPsych Media Sentiment model. The model condenses the predictive power of media topics, themes, and sentiments into a simple 1 to 100 ranking, targeted to forecast the next month's relative stock price return.

The MMS and its component ranks are suitable for use in factor models, buy/sell recommendations, or sentiment profiling. As a pure sentiment factor, MMS rankings were found to have low correlation with traditional market factors and fundamental models, as well as with other StarMine models.

## Questions?

For more information, including delivery options, a detailed white paper, or historical files for backtesting, please visit [website](#) or contact your LSEG representative or the StarMine specialists at [starmine.quantconsulting@lseg.com](mailto:starmine.quantconsulting@lseg.com)