Adjusting for the Human Bias in Media Enhances Stock Portfolio Returns

State Street MediaStats' Indicators ® provide a timely measure of media sentiment drawn from hundreds of thousands of curated, unstructured data sources.

Over the past seven years, we have found that this source of data can generate timely insights into the performance of a range of investments.

This real-time pulse of broad media sentiment enables investors to anticipate and evaluate the impact of online chatter, including the impact of unusual media coverage and the degree of disagreement in opinion.

Asset managers strive to make decisions that yield outstanding investment performance for their clients. The strategies informing those decisions have evolved over time, but the underlying tools and analyses have remained largely unchanged. Today, however, new techniques enabled by the proliferation of big unstructured data challenge that status quo and threaten the stability of businesses that adhere strictly to traditional methods. Technologies such as machine learning (ML), artificial intelligence (AI), and natural language processing (NLP) enable investors to sift through the vast amount of information, answer questions and model scenarios that push the boundaries of traditional analytics, delivering differentiated insights with unprecedented precision and speed. In many cases, comprehensive analyses that might have taken weeks in the not-so-distant past are now achievable within hours. Market moving events can be found across potentially hundreds of thousands of news sources, in corporate communications, and on various blogs.

Yet, it is widely recognized that every media outlet offers its own "spin" on news, and this bias has been described in many ways by practitioners and academics alike. For instance, journalists differ markedly in their writing styles such as sentence structure, complexity, article length, and even

pessimism or optimism about market conditions. Additionally, a news article on a given company appearing in a national publication may be more negative compared to another appearing in a local publication. State Street MediaStats helps filter out similar noise by training ML algorithms to account for biases. Our process is trained to weed out redundant articles, and can detect a bias that might exist between Starbucks and a Seattle Times reporter who is frequently and positively covering the company because they are locally-based.

The Media-Based Investment Indicators

Our three stock-level indicators derived from media articles are motivated by economic theory.

Intensity measures the number of media articles that mention a particular stock. The motivation for this measure follows the theoretical work of Merton (1987). His model posits that investors only use those securities they know about in constructing their optimal portfolios and ignore the rest—the "neglected securities". In equilibrium, the neglected securities are under-invested and are therefore associated with a higher expected return compared to better known securities. Using the intensity of a stock as a proxy for the degree of investor recognition of the firm, we find evidence consistent with

these predictions—that is, that high-intensity stocks tend to underperform low-intensity stocks.

Sentiment measures the tone of media coverage. For example, Tetlock (2007) shows that the textual sentiment extracted from a single daily column is correlated with market return. We investigate the predictability of media sentiment across multiple dimensions. The magnitude depends on the type of sources and the investment horizon. More specifically, investors tend to overreact to local coverage compared to distant coverage, and different source types are systematically biased either positively or negatively, as demonstrated in Ozik and Sadka (2013). Overall, the results indicate the ability to outperform by constructing sentiment indicators that consider both investor and source-type biases.

Disagreement considers the dispersion in opinion across media sources for a particular stock. This measure is motivated by the current debate in the literature. On the one hand, the traditional risk-return tradeoff asserts that investors perceive high-disagreement stocks as riskier, and therefore demand a higher expected return. On the other hand, Miller (1977) presents a different model whereby high-disagreement stocks are typically overpriced. The key assumption in Miller's model posits that, under restricted short selling, asset prices typically reflect the views of optimist investors, resulting in over-valuations. Similarly, short-sale constraints in Diamond and Verrecchia (1987) result in slower incorporation of negative views. Diether, Malloy, and Scherbina (2002) study the predictability of analyst forecast dispersion, and find that high-disagreement stocks tend to underperform. Nevertheless, Avramov, Chordia, Jostova, and Philipov (2009) demonstrate that the analyst dispersion anomaly is concentrated only in a small number of the worst-credit-rated firms. The large cross-section of media coverage per security facilitates the unique extraction of disagreement-like indicators. Our results below indicate a significant outperformance of high-disagreement firms among US large-cap stocks, consistent with the risk-return tradeoff.

Indicator construction

The indicators are constructed daily using a 28-calendar-day window. Each indicator is standardized daily in the cross-section of firms to produce firm-level z-scores.

We consider two versions for each indicator: one with an adjustment for a variety of biases and one without an adjustment.

Intensity

• The unadjusted intensity measure represents the simple count of articles in the previous 28 calendar days.

• The adjusted intensity measure considers two types of firm-specific biases: i) information production biases, such as the day-of-week firm-specific effect (e.g., the Wall Street Journal tends to cover General Motors on Wednesdays), and ii) extensive cyclical coverage around corporate events, such as earnings announcements, which causes media spikes. Our model of abnormal intensity is designed to correct for the predictable patterns in media coverage around corporate events. The indicator is constructed daily using a 28-calendar-day rolling exponential moving average.

Sentiment

• The unadjusted sentiment measure represents the simple average stock-level daily media sentiment in the previous 28 days.

• To obtain our adjusted sentiment measure, we implement adjustments for firm-specific biases, as well as source type and geography, in sentiment scoring. The indicator is constructed daily using a 28-calendar-day rolling exponential moving average.

Disagreement

• The unadjusted disagreement measure shows the standard deviation of article sentiment across all articles covering a stock in the 28-day window.

• Our adjusted measure controls for the statistical and behavioral biases in estimating disagreement. Statistical bias refers to the fact that sample standard deviation is not the unbiased estimator of the population standard deviation, and the bias is a function of the number of media articles used in calculating sample standard deviation. Behavioral bias refers to the dependency of media opinion divergence on the amount of coverage. The adjusted disagreement measure considers these biases by accounting for population size. The adjustment procedures have a more meaningful impact on stock disagreement when overall media coverage is low.

Indicators and Future Stock Returns

In this section, we examine whether firm-level media-based indicators can predict cross-sectional stock returns. In particular, we compare the performance of indicators created with and without the adjustments described above. Our sample includes all S&P500 stocks over the time period from May 2011 to December 2017.

Based on each indicator, we form a long/short portfolio using a ladder approach. At the end of each trading day, we sort stocks into ten equal groups—or deciles—incorporating the latest available indicator values. We then construct a zero-cost long/short portfolio by buying the highest decile of stocks and shorting the lowest decile of stocks. The portfolio is then held for the next 20 trading days. Under this ladder approach, every day we hold 20 long/short portfolios, and only one-twentieth of the aggregate portfolio is rebalanced according to the most recent indicator values available. The daily return of the long/short portfolio under our laddered approach is the equal-weighted average of the returns of all 20 portfolios. We plot the cumulative returns of the daily long/short portfolio for Intensity, Sentiment and Disagreement in Figures 1, 2 and 3, respectively. The intensity results represent the low-intensity decile minus the high-intensity decile, whereas the sentiment and disagreement results represent the high decile minus the low decile for each indicator. In Figure 4, we report the average returns (annualized), volatility and
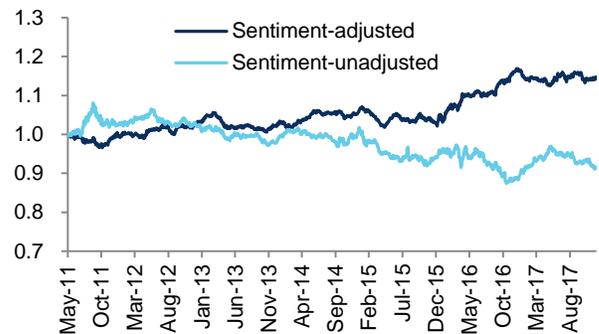
information ratio, as well as associated T-statistics of the daily returns of the long/short portfolios.

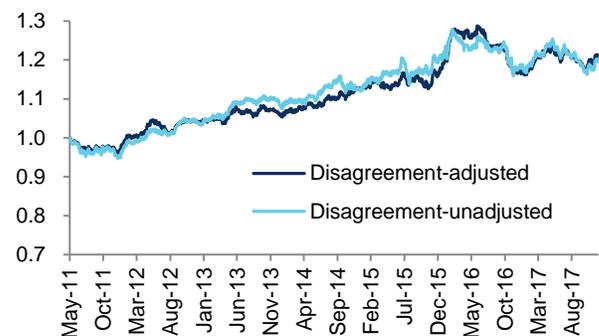**Figure 1: Cumulative Returns, Intensity**
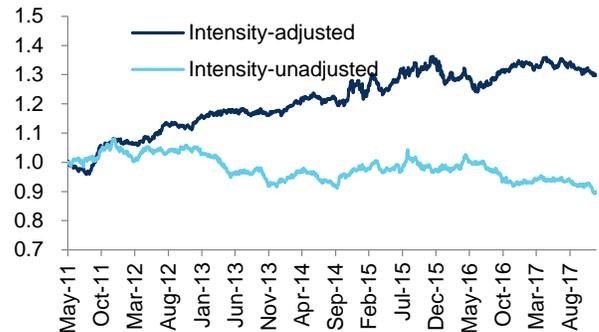
Source: State Street Global Exchange, MKT MediaStats (May 2011-Dec 2017).

Figure 2: Cumulative Returns, Sentiment

Source: State Street Global Exchange, MKT MediaStats (May 2011-Dec 2017).

Figure 3: Cumulative Returns, Disagreement

Source: State Street Global Exchange, MKT MediaStats (May 2011-Dec 2017).

Information Classification: Limited Access

Figure 4: Backtest Summary Statistics

	Intensity	Sentiment	Disagreement
Adjusted			
Average Returns (Ann.)	4.15%	2.10%	2.97%
Volatility (Ann)	5.93%	3.55%	4.58%
IR	0.70	0.59	0.65
T-stat	*1.79*	*1.52*	*1.67*
Unadjusted			
Average Returns (Ann.)	-1.43%	-1.20%	2.83%
Volatility (Ann)	5.76%	5.32%	4.90%
IR	-0.25	-0.23	0.58
T-stat	*-0.64*	*-0.58*	*1.48*
Return Difference |Adjusted-Unadjusted|	5.58%	3.30%	0.14%

Source: State Street Global Exchange, MKT MediaStats (May 2011-Dec 2017).

All three adjusted indicators are associated with future stocks returns. For example, consistent with theory, a lower adjusted intensity predicts higher returns in the following month, suggesting that neglected stocks are underpriced and therefore earn higher returns. More specifically, an equal-weighted portfolio of low adjusted intensity stocks outperforms an equal-weighted portfolio of high adjusted intensity stocks by 4.15% annualized. In contrast, the unadjusted intensity measure is not significantly related to future returns. The average return difference between low unadjusted intensity and high unadjusted intensity is -1.43% with a t-stat of -0.64. The results suggest that the adjustment can improve the performance of a zero-cost trading strategy based on intensity of media coverage by 5.58% per year.

The adjusted sentiment predicts returns positively. The zero-cost long/short portfolio based on adjusted sentiment yields an average return of 2.10%. In contrast, the zero-cost long/short portfolio based on unadjusted sentiment yields an average return of -1.20% although the t-stat is quite low at -0.58. The return difference of the zero-cost portfolios between the adjusted sentiment and unadjusted sentiment is 3.30% per year.

Lastly, our results on disagreement indicate that it positively predicts returns in the cross-section of S&P

500 stocks, consistent with a risk-return relationship. Consistent with the prediction that the adjustment effect is more relevant for stocks with low coverage, the effect of adjustment is small for the S&P500 stocks.

Conclusions

In this paper, we set out to evaluate the performance of State Street MediaStats indicators after accounting for a number of consistent media biases. Our results show that it is possible to extract alpha from media articles. However, there are strong benefits from adjusting for inherent and consistent biases when building predictive models as we achieve cleaner, less volatile return profiles, with a marked improvement in risk-adjusted returns. We demonstrate that media-based indicators can be constructed to predict the returns of the S&P 500 stocks. Intensity predicts returns negatively, while sentiment and disagreement predict returns positively.

Finally, we have provided evidence in favor of the statistical significance of our results. By comparing the Information Ratio and annualized returns of the bias-adjusted portfolios against their respective unadjusted counterparts, we demonstrate that the improvement in annualized returns is meaningful when adjusting for bias in media coverage.

Information Classification: Limited Access

For more information,
please contact:

North America:
Jenni Lee
jenni.lee@statestreet.com

Jamie Stewart
jstewart2@statestreet.com

Globally:
mktmediastats@statestreet.com

References

Avramov, Doron, Tarun Chordia, Gergana Jostova and Alexander Philipov, 2009, Credit ratings and the cross-section of stock returns, Journal of Financial Markets 12, 469—499.

Diamond, Douglas W., and Robert E. Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, Journal of Financial Economics 18, 277—311.

Diether, Karl B., Christopher J. Malloy, and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, Journal of Finance 58, 2113—2141.

Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, Journal of Finance 42, 483—510.

Miller, Edward M., 1977, Risk, uncertainty, and divergence of opinion, Journal of Finance 32, 1151—1168.

Ozik, Gideon, and Ronnie Sadka, 2013, Media coverage and hedge-fund returns, Financial Analysts Journal 69, 57—75.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, Journal of Finance 62, 1139—1168.

STATE STREET. | **MKT MEDIASTATS**

Disclaimers and Important Risk Information

Important Legal Information – Please Read