

Advanced Text Analytics Sentiment in News Media Impact

on Stock Market Performance

Stephen Teresi MBA Founder & CEO Iapetus Consulting LLC Email:<u>Stephen.teresi@iapetusconsulting.co</u> website: <u>IapetusConsulting.co</u> twitter: <u>@IapetusLLC</u>

Jack Onderisin Statistical Analyst Iapetus Consulting LLC

08/14/2017

ABSTRACT

Quantitative trading is becoming increasingly more popular as we search for new insights with better tools. Iapetus Consulting LLC. is a startup company with a profoundly original behavioral psychology research idea asking the question: Does news impact the volatility in markets and if so, can this be quantified, analyzed, and used real time by investors? The Iapetus real time inventing tool times the market using complex algorithms that trawls news articles and quantifies news media sentiments throughout the day as news cycles change. Our unique training data going back to September 2016 is used to forecast changes in the stock market with Buy/Hold/Sell recommendations. Using only nine different news sources with an array of different ideologies we have been fortunate to find good signal correlating to market movement. Our algorithm scans thousands of articles every day. It searches our unique library of sentiment trigger words which are weighted per our unique scaling design. This makes up the aggregate sentiment quantification of each news source. The goal of Iapetus is to create an investing model that will consistently outperform the stock market. In this paper, we discuss our methodology and design of simulation models that buy and sell funds based on the Iapetus algorithm.

Introduction

Iapetus Consulting LLC. set out to build an investing tool to outperform buy and hold strategies of some of the most common exchange traded funds (ETF's). Iapetus, as the tool is known, uses an algorithm to quantify news media through sentiment analysis (SA). A thought or view based on emotion is called a sentiment. SA or Opinion Mining (OM) is an automatic process of mining opinions, emotions, and attitudes from text or speech by using Natural Language Processing (NLP). OM is an important sub discipline within data mining, NLP, and web mining, which automatically extracts, classifies, and understands the opinions generated by various users (Bing 2012).

The Iapetus interface (Figures 2,3,4) graphs the Psychological Trigger Word News Index (PTWNI) or News Index (NI) for short and overlays stock market performance of the Dow Jones, Standard & Poor's 500, and the Nasdaq. Our Simulation modeling has Iapetus buying and selling the SPDR S&P 500 ETF (SPY) and SPDR Dow Jones Industrial Average ETF (DIA) per calibrated thresholds of the PTWNI. The simulation ran from January 2017 through May 2017. When the nine news sources we follow measure an optimistic news cycle the output is a buying recommendation. When negative news is trending, Iapetus recommends a sell recommendation. In between set thresholds we have a hold recommendation. These thresholds took several months to dial in a high correlation of the PTWNI with market volatility, allowing our artificial intelligence to provide correct recommendations. Buy, hold, and sell times came directly from the software with no human intervention.

Figure 1:	: Examp	ole of early	y version	article s	earch qu	antificatio	on feedbac	ek loop.		
					'score'	: -8.0,	n+1. 01			
					'title'	_word_cou	nt : ZI, ther bank	s follow	ved Wells	Fargo's
scandalo	us lead	?".			circe.		cher bank	3 10000	icu metts	largo s
					'total_	word_coun	t': 795,			
					'url':	'http://w	ww.usatod	ay.com/s	story/mone	ey/
2016/09/	28/have	-other-ba	nks-foll	owed-wel	ls-fargo	s-scandal	ous-lead/	91138118	3/'},	_
			'artic	le_11':	{'averag	e_word_sc	ore': -0.	16666666	6666666666	Ĵ,
					'score	: -3.0,	nt'• 18			
					'title'	: "Deutsc	he Bank r	ecovers	on report	tof
lower se	ttlemen	t, CEO's	reassura	nces",						
					'total_	word_coun	t': 443,			
					'url':	'http://w	ww.usatod	ay.com/s	story/mone	ey/
business	/2016/0	9/30/deut	sche-ban	k-shares	-plunge-	rattling-	markets/9	1311356/	· },	
			artic	te_12*:	'averag	· a a	ore: 0.2	5/14285/	1428571,	
					'scored	word cou	nt': 35.			
					'title'	: 'Britis	h Airways	to end	free food	d in
coach on short flights',										
	(
	2	60.0.1	TEON				Alex	The	The	
	6.2				n p r		guardian	New Hork	Washington	TODAY
	ALIAZEERA	000		NBC		DELITEDS	0	Eimes	post	
	ALJALLINA					REOTERS				
					30 MARCH, 2017					
TITLE					SCORE	WORD COUNT	SCORED WORD	S AVER	RAGE WORD SCORE	TIME

49

694

43

. . . . - 1-1:C . . 1: . . . c 11

1 Fearing voter backlash. Republicans bullish again on ObamaCare repea

11-00 AM

1.14

Through our research we measured that a real time investing tool looking at reliable news only, not social media, needed a frequency of every hour and half to impact trading of news sentiment analytics. In other words, an hour and a half was the sweet spot to get news article turnover as these news sources update content throughout the trading day. Iapetus's algorithm seeks out the most relevant news based on text analytics of the article title. Part of our unique weighting of our PTWNI library is a rolling scale based on the font size of which the word or words are found. For example, Bankruptcy is weighted more negatively based on if the font size in large like a 18 versus a smaller 12. This allows us to weigh our PTWNI library found in article titles at the highest values of the positive and negative spectrum compared to the same word found in the body of the article.

Figure 2. Graphical interface of PTWNI with score of 131 recommending a Buy opportunity with Dow Jones Industrial average overlay.





Figure 3. Graphical interface of PTWNI with score of 120 recommending a Buy opportunity with Standard & Poor's 500 average overlay.

April 18th forecasting S&P 500 to climb based on Positive News Sentiment across 9 sources.



Figure 4. Graphical Interface of PTWNI showing a pessimistic sentiment cluster triggering a sell warning with Dow Jones Industrial average overlay.

April 7th forecasting a dip in the Dow Jones base on Negative News Sentiment Data. Ave. score -129



Data mining focuses on the autonomy of extracting information from structured and unstructured text documents by combining techniques from text mining, Machine Learning (ML), NLP, Information Retrieval (IR), and knowledge management. SA techniques can basically be divided into ML-based approach, lexicon-based approach (Dang et al 2010) and hybrid approach. ML-based approach uses linguistic features with famous algorithms like Support Vector Machine (SVM), Maximum Entropy (ME), and Naive Bayes (NB), while the lexicon-based approaches the sentiment lexicon, which is a predefined sentiment term in the sentiment dictionary. Similarly, lexicon-based approach is divided into dictionary based approach and corpus based approach. (Davi 2017).

The proprietary dictionary core of Iapetus is a unique lexicon-based approach. The Iapetus team thanks Stephen P. Teresi, a life long Lexicographer for Weber's Dictionary and Nancy Truett, a psychology licensed professional for helping to design our psychological trigger word library.

History

Researchers have asked the question, "How precisely do media influence their readers, listeners and viewers?" In the 2009 paper **The social psychology of news influence and the development of Media Framing Analysis,** demonstrated that the field lacks methodology and argued for continued research. In recent years, some psychologists have adopted a more sophisticated approach towards the analysis of media, through advanced study of the media texts (Giles 2009). The Iapetus Tool using the PTWN index has bridged advanced text analytics in news media together with finance. Several researchers over the years have been exploring this relationship of news media psychology with financial markets.

In the Huberman and Regev (2001) study, they evaluated a Sunday New York Times article that discussed improvements of new cancer-curing drugs. Their study showed an impact on rising biotechnology stocks on the following Monday and in the following weeks. Busse and Green (2002) focused on the Morning and Midday Call of segments on CNBC TV. They found that prices responded to reports within seconds of initial mention. Tetlock et al. (2008) found that the more negative words used in news focusing on specific companies predicted low firm earnings. Dougal et al. (2012) showed that financial journalists have the potential to impact investor behavior through the articles they write, at least in a short term.

An example of a training set for news articles can be downloaded by anyone per this site, (https://www.crowdflower.com/data-for- everyone/). Contributors of this data viewed a news article and rated the positivity of the article on a scale 1- 9 with 1 being negative and 9 being positive. This can be further classified in unique ways.

Previous studies have focused on a general relationship between social media and financial market activities (Antweiler & Frank, 2004; Das & Chen, 2007; Kim & Kim, 2014; Leung & Ton, 2015). Iapetus does not look at social media only reliable news sources. Other studies of news media and financial market variables (Tetlock, 2007; Tetlock et al., 2008; Chen et al., 2014) are more closely tied to the Iapetus model. Researchers have explored holding times by enlisting regression with one holding day (Kim & Kim, 2014; Leung & Ton, 2015), and one-day or two-days lead-lag (Antweiler & Frank, 2004; Chen et al., 2014; Leung & Ton, 2015). For

holding days, Tetlock et al. (2008) used one holding day, while others have used longer holding times, such as one month to 36 months holding time (Chen et al. 2014). The Iapetus PTWN index model could have a waiting period of hours to days and weeks depending on how positive or negative news is trending as news cycles evolve.

Sentiment analysis can be based on prior polarity of words. Researchers have used work by Agarwal et al. 2013 using a Dictionary of Affect in Language (DAL) (Whissel, 1989) and extended it using WordNet. This dictionary of about 8000 English language words assigns every word a pleasantness score ($\in \mathbb{R}$) between 1 (Negative) - 3 (Positive). They first normalize the scores by diving each score by the scale (which is equal to 3). Their unique scoring method had a word library with polarity less than 0.5 as negative, higher than 0.8 as positive and the rest as neutral (Agarwal et al 2013). The Iapetus lexicography library has three tiers of positive words and three tiers of negative words where each tier is a uniquely different scale. We then introduce text font analytics as an added layer of weighing.

Simulation Model: Iapetus Model Returns vs. Returns of the Market

Iapetus times the market by quickly analyzing meta data tracking news media trends that may cause upswings or downswings in the market. Monitoring these averages allows us to sell before a downswing, and buy before an upswing. Buying forecasted dips during a bull market is a challenge.

Our simulation modeling was conducted between January 2017 and May 2017. The Iapetus investing model outperformed an S&P ETF (SPY) by 304 basis points or 3.04%. The SPY is a very popular exchange traded fund that corresponds to the yield of the S&P 500.

Trades Throughout the Simulation

Between January 6th and May 25th, 2017, a total of 14 trades occurred. The objective of the simulation was to theoretically invest \$10,000 in SPY, selling when the sentiment average was -15 or below, and buying when the average was 70 or higher. We also included a control group, buying the same amount of SPY at the same time and holding until the end of the simulation. An independent simulation was also conducted, buying DIA instead of SPY, using the same method. DIA is a Dow Jones Industrial Average ETF. Below is the list of trades that were triggered from the sentiment averages of all 9 news sources. Table 1 shows the SPY and Table 2 shows the DIA simulation.

Table 1. SPY simulation

lapetus Model Buy/Sell history	Buy/Sell	Sentiment	Shares of SPY bought/sold	SPY Value	lapetus Return	SPY return
1/6/2017	BUY	72.11	44	\$226.53	0%	0%
2/7/2017	SELL	-21.22	44	\$229.66	1.38%	1.38%
2/7/2017	BUY	75.556	44	\$226.84		
4/5/2017	SELL	-25.556	44	\$236.23	5.39%	4.28%
4/6/2017	BUY	82.556	44	\$235.36		
4/7/2017	SELL	-70.721	44	\$235.07	5.59%	3.77%
4/18/2017	BUY	90.667	45	\$233.85		
4/21/2017	SELL	-15.778	45	\$234.61	5.93%	3.57%
4/24/2017	BUY	80.44	44	\$237.02		
5/11/2017	SELL	-24.44	44	\$239.34	6.96%	5.66%
5/12/2017	BUY	77.889	44	\$238.81		
5/16/2017	SELL	-17.22	44	\$240.63	7.76%	6.22%
5/19/2017	BUY	77.33	45	\$237.32		
5/25/2017	SELL	-17.667	45	\$241.73	9.75%	6.71%

OUTPERFORM 304 basis pts

Table 2. DIA Simulation

lapetus Model Buy/Sell history	Buy/Sell	Sentiment	Shares of DIA bought/sold	DIA Value	lapetus Return	DIA return
1/6/2017	BUY	72.11	50	\$198.73	0%	0%
2/7/2017	SELL	-21.22	50	\$201.30	1.29%	1.29%
2/7/2017	BUY	75.556	50	\$201.00		
4/5/2017	SELL	-25.556	50	\$207.59	4.61%	4.45%
4/6/2017	BUY	82.556	50	\$206.00		
4/7/2017	SELL	-70.721	50	\$206.19	4.70%	3.75%
4/18/2017	BUY	90.667	50	\$205.36		
4/21/2017	SELL	-15.778	50	\$205.25	4.65%	3.28%
4/24/2017	BUY	80.44	50	\$207.33		
5/11/2017	SELL	-24.44	50	\$209.14	5.56%	5.24%
5/12/2017	BUY	77.889	50	\$209.01		
5/16/2017	SELL	-17.22	50	\$210.32	6.22%	5.83%
5/19/2017	BUY	77.33	51	\$206.75		
5/25/2017	SELL	-17.667	51	\$210.58	8.19%	5.96%

OUTPERFORM 223 basis pts

On every occasion except for the sell on 4/7/2017, the Iapetus Model was able to sell at a high, and rebuy the stock at a lower price. This helped us earn a 9.75% return in just 140 days.

The SPY return, buy and hold, during the same period (not accounting for dividends), was 6.71%, 3.04% lower than the Iapetus Model returns. Iapetus outperformed a buy and hold strategy of the SPY by 304 basis points or 3.04%. Figures 5 and 6 are a timeline of returns for both the SPY and DIA respectfully. Each dot represents the point in time which a sell was triggered. The Blue line represents the returns of the Iapetus model, while the Orange represents SPY returns. After the first sell, the Iapetus model consistently outperforms the ETF.



Figure 5: Iapetus model returns vs. SPY buy and hold returns



Figure 6: Iapetus model returns vs. DIA buy and hold returns

There is an important distinction to be made in comparing these results of SPY and DIA simulation. The Iapetus PTWN index isn't customized to a specific market index, these recommendations are for the market in general so our findings that Iapetus is more closely correlated with SPY market movement than DIA is interesting. We understand that the Dow and S&P don't exactly track on top of each other, where one may zip and the other zags. The Wall Street Journal reported on July 10th 2017 that the Dow and S&P are growing apart. The 20-day correlation between the Dow Jones Industrial Average and the S&P 500 fell to 0.47, the lowest since 2003. Data from the last 15 years has the average correlation between the Dow and S&P 500 at 0.96, almost uniform price movement (Dieterich 2017). We'll continue to keep an eye on

this and look to do more targeted research of S&P 500 sector rotations. The S&P 500 interestingly has performed best when technology stocks have dominated.

The Iapetus PTWN index with its better beat of the SPY than the DIA can draw comparisons to the CBOE Volatility Index, known as Wall Street's fear gauge. The CBOE index, called VIX uses the S&P 500 options prices to project expected stock-market swings (Banerji 2017). We'll be overlapping our PTWN index with the VIX and showcase those results in our next paper.

The Business Case

Stock markets drive our world's economy because they provide a large platform for companies to raise more money easily. There are so many factors that go into forecasting the trends of a stock market. This makes it a very difficult and highly complicated initiative. Finding good signal in market forecasting is like mining for a precious metal, when found, its valuable. Some of the factors that make this a complicated task include understanding micro, macro, and behavioral economic conditions, political events, human psychology, investors' sentiments and so on. The stock market indexes are generally dynamic, non-parametric, noisy and chaotic by nature. Finding order in this chaos is why we built the Iapetus platform. The scientific challenges of extracting useful information from this vast source of economic psychological stock market data are great due to its diversity and lack of formal structure (Davi 2017). This is why the Iapetus dataset foundation is built on reliable news that shapes perception and ultimately impacts behavior. Although the scientific challenges are many, our Iapetus model is cutting through the meta data noise offering clear and concise trading recommendations. We are looking forward to our continued research efforts making real time tools that work.

Our current business model is disruptive and ambitious. For free anyone can access our website and see two news sources updating real time. For a few dollars a month an individual can get full access to all nine news sources updating real time and offering Buy/Hold/Sell recommendations based on the psyche of the current news cycle. We offer all this intelligence on our website. This disruptive model makes available cutting edge big data analytics to the masses. Anyone and everyone can have access to this complex algorithm to help them trade the market. At the end of the day it's about having better information and better tools. As quantitative trading gets even more popular and algorithms step in to make buy and sell decisions, financial conglomerates will shift to strategies like meta data sentiment analysis to seek better returns. The Iapetus tool is ready to step in and make an immediate impact.

Statistical Data

Focusing on the individual news sources, we wanted to see the correlations between the news source's sentiment, and the movement in the market. We asked the question: "Are there news sources that correlate better to market volatility?" Using data gathered through the month of April, 2017, we took each sentiment reading, 95 in total for each news source, and compared that number with the market change since the last sentiment reading (Figure 7).

Quadrant one represents positive sentiment and a positive change in the market, Quadrant 2 represents a positive sentiment and negative change in the market, quadrant 3 represents a negative sentiment with a negative change in the market, while quadrant 4 represents a negative

sentiment with a positive change in the market. We see clustering of data points in quadrants 1 and 3 which is a good indicator that signal is strong with meaningful correlation. All of these individual dots add up to the bigger picture, which is seeing changes in the stock market based on sentiments in news media.



Figure 7: S&P 500 Quadrant correlation to USA Today



Figure 8: News sources quadrant correlation to S&P 500

Table 3: Regression Statistics for S&P 500 short study

	Multiple R	R square	Adjusted R	ANOVA			
			Square	significance F			
ABC	0.985725	0.971654	0.63832	0.009585			
Al Jazeera	0.529069	0.279914	-0.05342	0.393105			
NY times	0.991785	0.983637	0.65030	0.005499			
NBC	0.999998	0.999997	0.66666	7.73507E-07			
The Guardian	0.621568	0.386347	0.05301	0.303081			
USA Today	0.952442	0.907145	0.57381	0.032467			
FOX News	0.992810	0.985671	0.65234	0.004811			
Washington Post	0.936071	0.876229	0.54289	0.044000			
SUMMARY	0.986949	0.974069	0.64073	0.008757			

The above regression is a snap shot of a small data set in the month of March 2017. We are continuing to track this information and compare the results across news sources that we are following. Our future papers will have more detailed statistics showing correlation of our PTWN index and market volatility. This short study of a small data set had a high correlation summary.

Conclusion

Since Iapetus' inception, the goal has been to beat the market. After the results of these simulations, outperforming the SPY by 304 basis points, we are confident that our model can forecast market volatility. We found that our algorithm, unique trigger word library, weighing scale, and interface can forecast the dips in the market with an equal precision as forecasting the market gains, which is not easy to do.

As we were working on this paper, again our Psychological Trigger Word News Index (PTWNI) Iapetus model predicted a buy on 7/26/2017, just before a 200 point gain in the DOW. Then forecasted the market fall on 8/09/2017.

SA techniques help to enhance the value of the existing information resources in many ways that can be useful in decision making, which is affected by the opinions formed by leaders, journalist, and other people (Davi 2017).

Forecasting changes in the stock market is about intelligence gathering and swift analytics to trade on the information. Iapetus is a tool that using nine different news sources with an array of different ideologies. The algorithm scans thousands of articles every day. It searches for trigger words, which will help to make up the aggregate sentiment quantification of each news source. After several months of calibration we landed on an average sentiment score of 70 will trigger a BUY, while an average of -15 will trigger a SELL. As we look to improve our model we'll be looking at trending analysis of gradual increases and declines that don't meet our current set thresholds. We want to better understand the movement inside the Hold band, correlating the smaller movements in the PTWN index with market swings.

Our goal in building Iapetus was to create an investing tool using News as our dataset that will outperform the stock market. We have done that on an aggressive time scale where the inception of this model began in September 2016 with some notes on a napkin. We look forward to future creative insights and research opportunities as we create interfaces for more narrowed sectors of the market and for VIX overlays to our PTWN index. Iapetus is a cutting edge news sentiment analysis investor tool providing insights in a complex behavioral economic world.

Works Cited

Devi K.N. 2017. "Social Media Aided Sentiment Analysis in Forecasting," *Online Journal of Communication and Media Technologies*, 7(1):163-173.

Bing Liu. 2012. "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers.

Dang, Y, Zhang, Y & Chen, H. 2010. "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews," *IEEE Intelligent Systems*, 25(4): 46–53.

Agarwal Apoorv et al. 2013. "Sentiment Analysis of Twitter Data" *Department of Computer Science Columbia University NY*, 2013:1-9.

Giles, D.C. & Shaw, R.L. (2009). The psychology of news influence and development of Media Framing Analysis. *Social and Personality Psychology Compass*, 3(4), 375-393.

Whissel C.M. 1989. "The dictionary of Affect in Lan-guage," Emotion: theory research and experience, Acad press London.

Dieterich, Chris. 2017. "Dow, S&P Grow Apart," *The Wall Street Journa,1* Monday July10, 2017/ B9.

Banerji, Gunjan. 2017. "Fear Gauge Of Stocks Hits A High for 2017" *The Wall Street Journal*, Friday August 11, 2017/ B11.

Hu, Tianyou and Tripathi, Arvind, 2016. "The Effect of Social and News Media Sentiments on Financial Markets," Available at *SSRN*. (December 1).

Antweiler, W., and Frank, M. Z. 2004. "Is all that talk just noise? The information content of Internet stock message boards," *Journal of Finance*, 59(3):1259–1294, (doi: 10.1111/j.1540-6261.2004.00662.x).

Busse, J. A., and Clifton Green, T. 2002. "Market efficiency in real time," *Journal of Financial Economics*, 65(3): 415–437, (doi: 10.1016/S0304-405X(02)00148-4).

Chen, H., De, P., Hu, Y. Y. (Jeffrey), and Hwang, B.-H. B.-H. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media," *Rev. Financ. Stud.* (27). hhu001–(doi: 10.1093/rfs/hhu001).

Das, S. R., and Chen, M. Y. 2007. "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science* 53(9): INFORMS, 1375–1388 (available at http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1070.0704).

Dougal, C., Engelberg, J., Garcia, D., and Parsons, C. A. 2012. "Journalists and the Stock Market," *Review of Financial Studies* 25(3):639–679 (doi: 10.1093/rfs/hhr133).

Huberman, G., and Regev, T. 2001. "Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar," *The Journal of Finance* 56(1): 387–396 (doi: 10.1111/0022-1082.00330).

Kim, S.-H., and Kim, D. 2014. "Investor sentiment from internet message postings and the predictability of stock returns," *Journal of Economic Behavior & Organization* (107), Elsevier B.V., 708–729 (doi: 10.1016/j.jebo.2014.04.015).

Leung, H., and Ton, T. 2015. "The impact of internet stock message boards on cross-sectional returns of small-capitalization stocks," *Journal of Banking & Finance* 55(July 1998), Elsevier B.V., 37–55 (doi: 10.1016/j.jbankfin.2015.01.009).

Tetlock, P. C. 2007. "Giving content to investor sentiment: The role of media in the stock market," *Journal of Finance* 62(3):1139–1168 (doi: 10.1111/j.1540-6261.2007.01232.x).

Tetlock, P. C., Saar-Tsechansky, M., and MacSkassy, S. 2008. "More than words: Quantifying language to measure firms' fundamentals," *Journal of Finance* 63(3): 1437–1467 (doi: 10.1111/j.1540-6261.2008.01362.x).