NATURAL LANGUAGE PROCESSING ON THE FEDERAL RESERVE SYSTEM

Group members:
Stephen Kusrianto
Stanley Wong
Sam Lim
Jiang Jin

AGENDA

Introduction

Literature Review

Topic Modelling

Sentiment Analysis

Conclusion

INTRODUCTION

By Jiang Jin

LITERATURE REVIEW

1.Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent dirichlet allocation - Journal of Machine Learning Research*. Journal of Machine Learning Research. Retrieved December 14, 2022, from https://jmlr.org/papers/volume3/blei03a/blei03a.pdf

2.Blinder, A. S., Ehrmann, M., Fratzscher, M., de Haan, J., & Jansen, D.-J. (2008). Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence. *Journal of Economic Literature*, *46*(4), 910–945. http://www.jstor.org/stable/27647085

3.Hubert, P., & Labondance, F. (2021). The signaling effects of Central Bank tone. *European Economic Review*, *133*, 103684. https://doi.org/10.1016/j.euroecorev.2021.103684

4.Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models - Cornell University. Cornell Bowers Information Science. Retrieved December 14, 2022, from https://mimno.infosci.cornell.edu/papers/mimno-semantic-emnlp.pdf

5.Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013, September 7). *Efficient estimation of word representations in vector space*. arXiv.org. Retrieved December 14, 2022, from https://arxiv.org/abs/1301.3781

6.Hyndman, R. J. (2014). *Measuring forecast accuracy - Rob J. Hyndman*. Rob J Hyndman. Retrieved December 15, 2022, from https://robjhyndman.com/papers/forecast-accuracy.pdf

7.Řehůřek, R., & Sojka, P. (2010) *Software Framework for Topic Modelling with Large Corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. (pp. 45-50). ELRA. http://is.muni.cz/publication/884893/en

8.Dalio, R. (2011, August). *Engineering targeted returns and risks* . Retrieved December 14, 2022, from https://bridgewater.brightspotcdn.com/fa/e3/d09e72bd401a8414c5c0bdaf88bb/bridgewater-associates-engineering-targeted-returns-and-risks-aug-2011.pdf

9. Selenium. (n.d.). *The Selenium Browser Automation Project.* https://www.selenium.dev/documentation/webdriver/

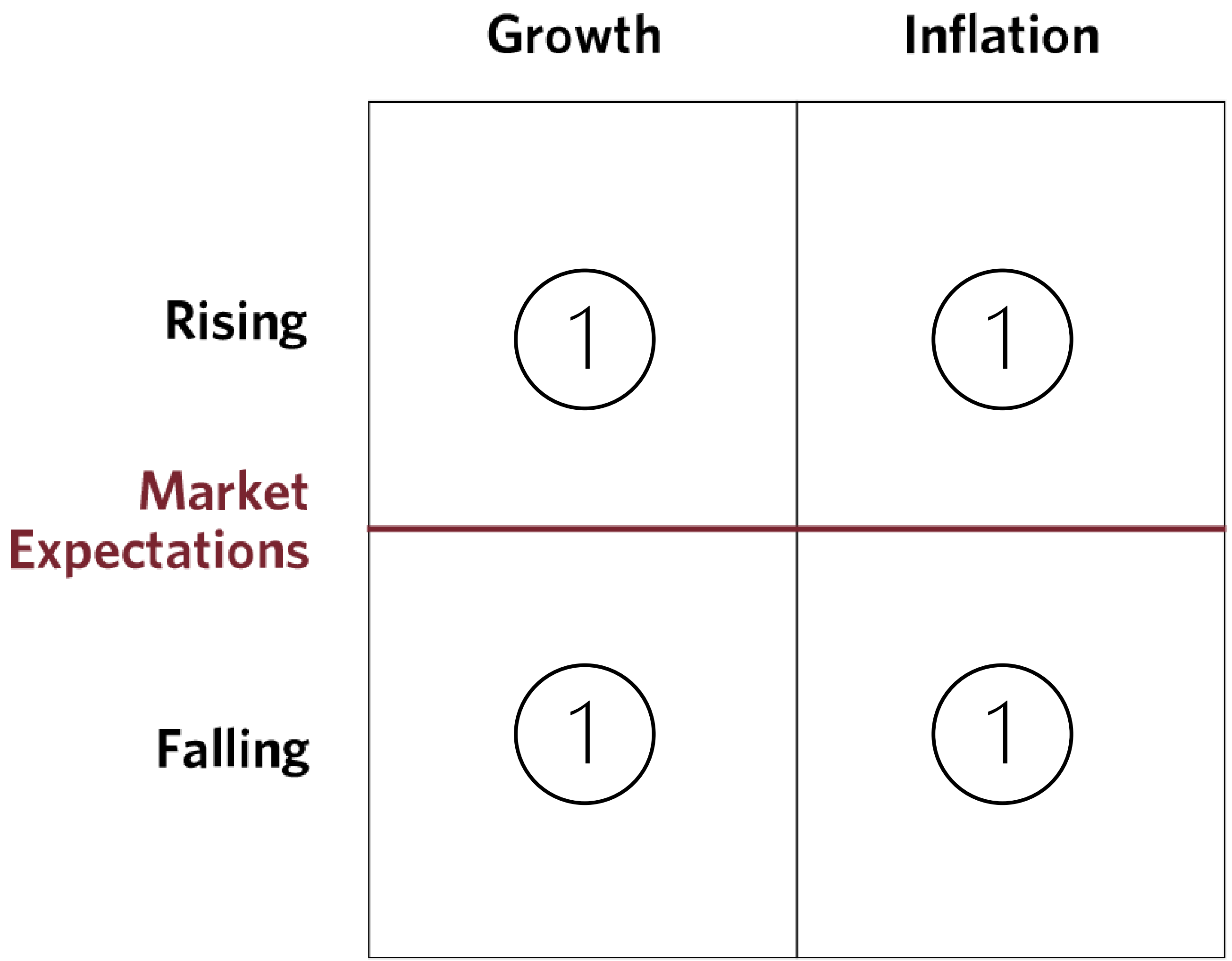
10.PYPI. (n.d.). *Fedtools 0.0.7.* Fedtools · PyPI

All weather portfolio

Market Expectations	Growth	Inflation
Rising	25% of risk Equities Commodities Corporate Credit EM Credit	25% of risk IL Bonds Commodities EM Credit
Falling	25% of risk Nominal Bonds IL Bonds	25% of risk Equities Nominal Bonds

Source: The All Weather Story. (2012, Jan). *Bridgewater Associates.*
https://www.bridgewater.com/research-and-insights/the-all-weather-story

Weather forecast portfolio



Asset Class	ETF
Equity	SPY
Commodities	DBC
Corporate Credit	LQD
Emerging Market(EM) credit	EMB
Nominal bonds	BND
Inflation Linked (IL) bonds	TIP

TOPIC MODELLING WITH WORD2VEC & LATENT DIRICHLET ALLOCATION

By Stanley

PROCESS FLOW

TEXT PREPROCESSING

- Data Sourcing
- Text Cleaning [regular expression]
- Lemmatization
- Filter stopwords
- Generate corpus

→

TOPIC MODELLING

- Word embedding with Word2Vec
- Topic modelling with LDA

WORD2VEC

METHODOLOGIES

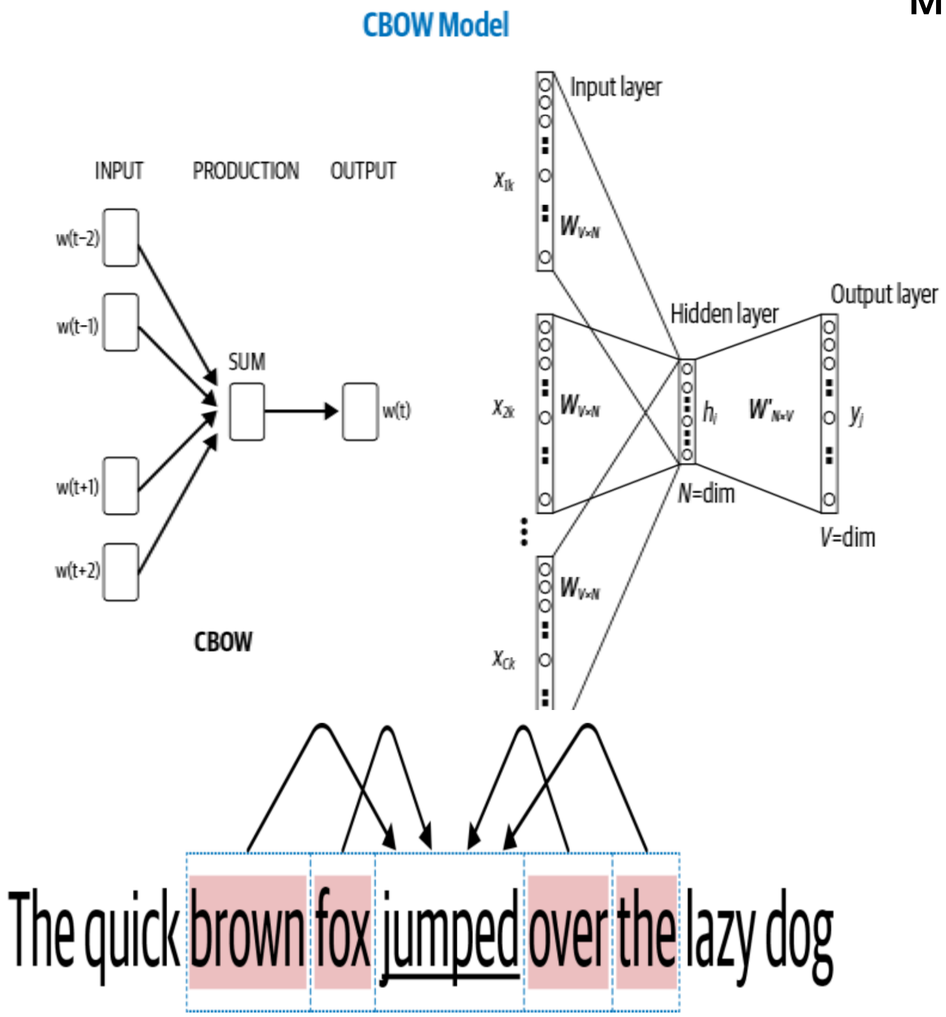
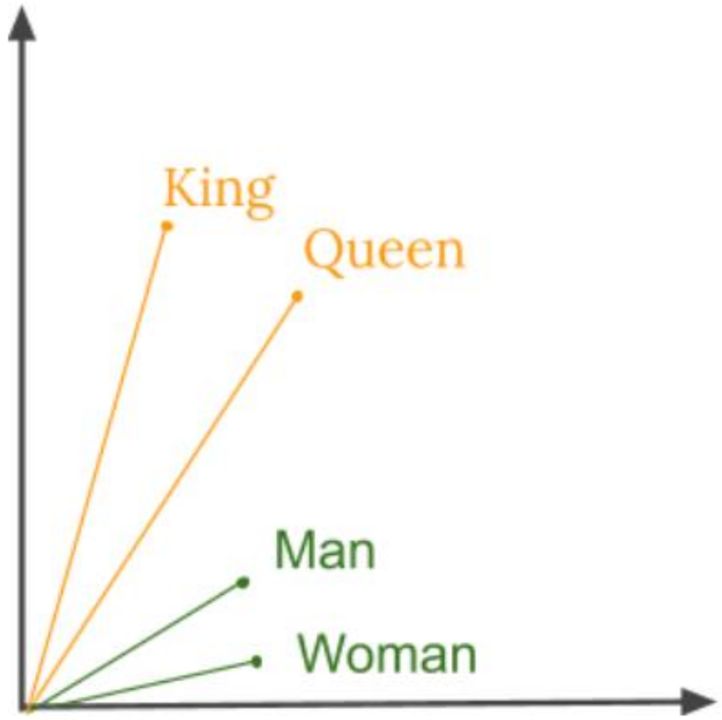


Figure 3-7. CBOW: given the context words, predict the center word

Cosine Similarity



Similar words are closely placed in vector space

Source: **[1]** Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). Chapter 1. NLP: A Primer. *Practical natural language processing: a comprehensive guide to building real-world NLP systems (1^{st} Eds)*. (pp. 9). O'Reilly Media, Inc.
[2] Great Learning Team. (2020, July 20). *What is Word Embedding | Word2Vec | GloVe.* Great Learning. [What is Word Embedding | Word2Vec | GloVe (mygreatlearning.com)]

WORD2VEC

LABEL OUR DTM BASED ON COSINE SIMILARITY RESULTS

INFLATION

Out[55]:

	words	similarity (%)
0	policy	0.889025
1	price	0.888703
2	july	0.875804
3	staff	0.863304
4	telecommunication	0.860242
5	difficult	0.854228
6	period	0.852166
7	did	0.822580
8	range	0.817453
9	really	0.815422
10	puzzle	0.811708
11	trend	0.811264
12	somewhat	0.806327
13	august	0.805652
14	second	0.805585
15	employment	0.801149
16	economy	0.800695
17	ought	0.789595
18	basic	0.788608
19	circumstance	0.785263
20	growth	0.782292

GROWTH

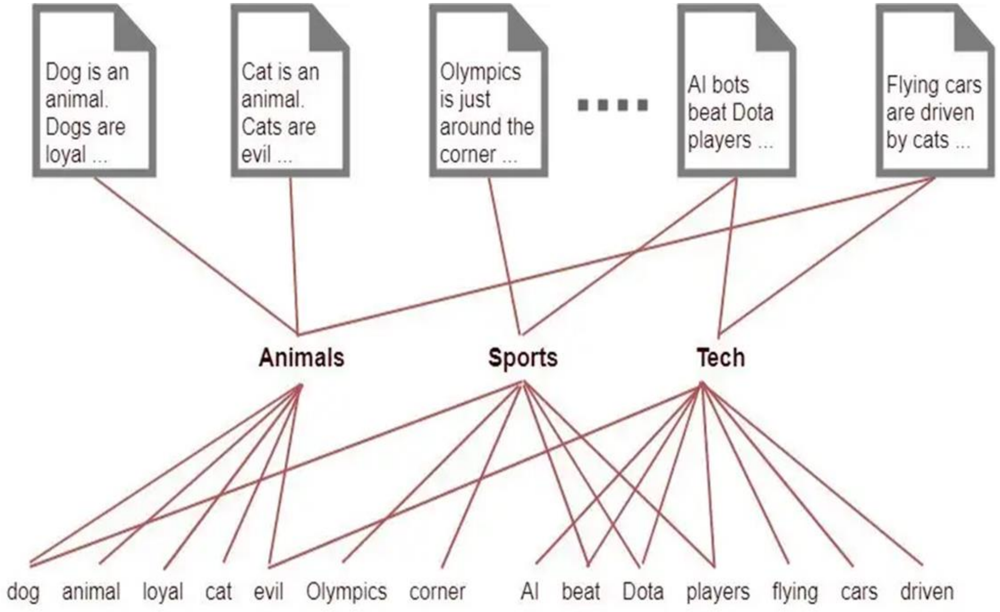
Out[49]:

	words	similarity (%)
0	gdp	0.903379
1	price	0.893671
2	deliver	0.885176
3	background	0.882438
4	recent	0.881655
5	stability	0.881156
6	weighted	0.876470
7	digit	0.872268
8	embedded	0.870827
9	everyday	0.869752
10	labor	0.857519
11	chip	0.851141
12	generate	0.847399
13	belief	0.847308
14	device	0.842703
15	economy	0.841281
16	foreseeable	0.838745
17	demand	0.834329
18	talk	0.831786
19	feel	0.830964
20	language	0.823181

LDA & ELDA

LATENT DIRICHLET ALLOCATION & ENSEMBLE LATENT DIRICHLET ALLOCATION

LDA



ELDA

- Solve the issue of topics reproducibility from LDA model.
- Addresses the issue by training multiple topic models and discard topics that do not occur across ensembles. (More reliable)

Source: [1] "Light on Math Machine Learning: Intuitive Guide to Latent Dirichlet Allocation", Ganegedara T., (2018, Aug 23). *In Medium.* https://towardsdatascience.com/light-on-math-machine-learning-intuitive-guide-to-latent-dirichlet-allocation-437c81220158
[2] Ensemble Latent Dirichlet Allocation. (n.d.). GENSIM

LIMITATIONS OF WORD2VEC & LDA

FUTURE PROJECT IDEA

WORD2VEC

- Requires a large corpus with quality inputs to generate meaningful results.
- Hubness problem, common word vectors share closer distance that may generate noise to cause bias to the evaluation.
- Can't handle out-of-vocabulary (OOV) words and morphologically similar words

LDA

- Doesn’t consider sentence structuring and language flow to derive semantic meaning
- Performs poorly if the dataset is small or length of document is too short
- Computationally intensive training on sparse vector representation

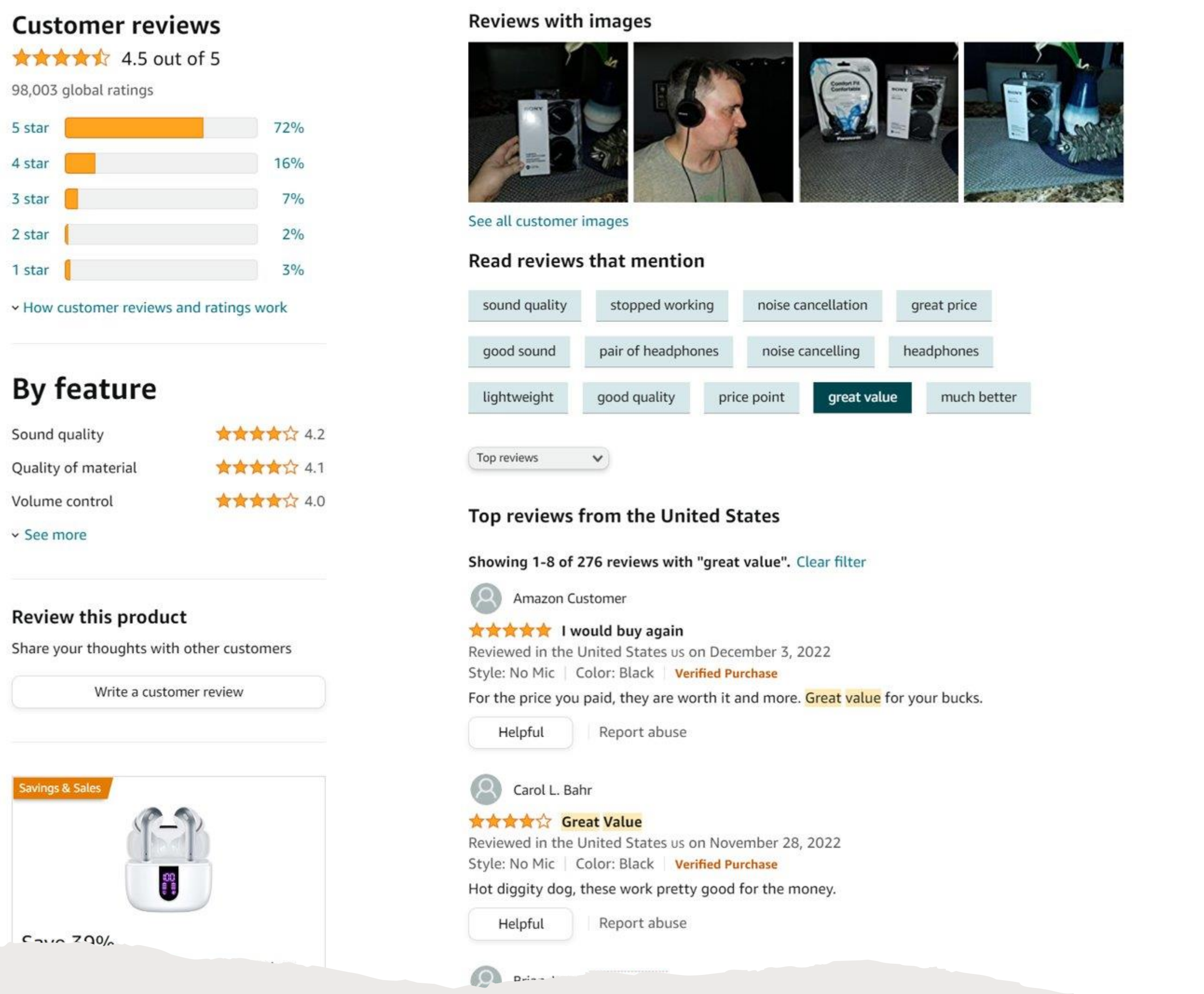
TOP2VEC

- Doesn't require manual preprocessing. Preprocessing has been in-built into function
- Combines document and word embedding vectors to search for semantic meaning to generate topics.
- Reduced high dimensionality sparsity issue using UMAP.

Source: Angelov, D. (2020). Top2Vec: Distributed Representations of Topics. ArXiv, abs/2008.09470. https://arxiv.org/abs/2008.09470v1
Naushan, H. (2020, Dec 3). Topic Modeling with Latent Dirichlet Allocation. In *Medium*. https://towardsdatascience.com/topic-modeling-with-latent-dirichlet-allocation-e7ff75290f8

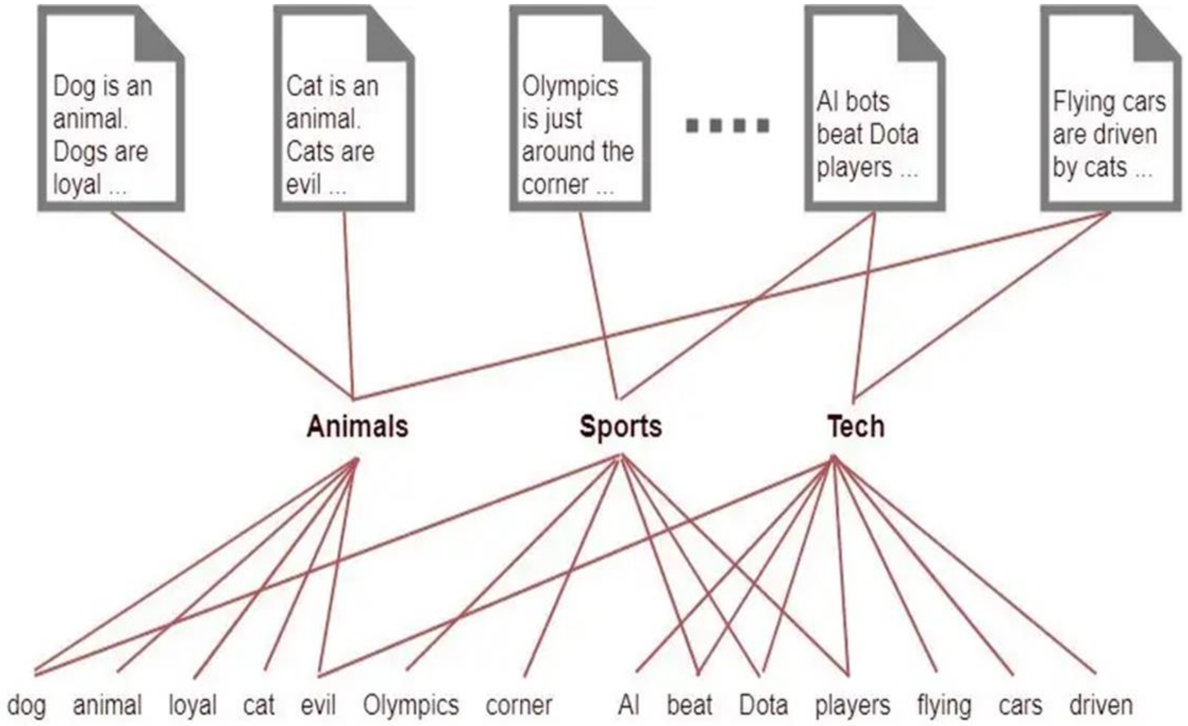
FINDINGS, INSIGHTS, AND TOPIC MODEL SELECTION

By Stephen



LATENT DIRI...WHAT?

TOPIC MODELS




```
[(0,
  '0.028*"district" + 0.026*"price" + 0.024*"contact" + 0.022*"sale" + 0.022*"activity" + 0.020*"demand" + 0.020*"year" + 0.014
*"report" + 0.013*"growth" + 0.013*"firm"'),
 (1,
  '0.023*"committee" + 0.023*"market" + 0.022*"growth" + 0.021*"price" + 0.019*"rate" + 0.018*"member" + 0.015*"inflation" + 0.
014*"policy" + 0.013*"wa" + 0.013*"year"'),
 (2,
  '0.030*"rate" + 0.028*"inflation" + 0.028*"policy" + 0.017*"market" + 0.015*"price" + 0.014*"economy" + 0.011*"term" + 0.011
*"year" + 0.010*"reserve" + 0.010*"percent"'),
 (3,
  '0.031*"bank" + 0.021*"risk" + 0.018*"market" + 0.012*"capital" + 0.010*"reserve" + 0.010*"credit" + 0.010*"institution" + 0.
008*"banking" + 0.008*"community" + 0.008*"ha"')]
```

INSIGHTS

1. Topic Models from Beige Books do not provide meaningful interpretations toward economic regimes:

```
[(0,
  '0.079*"district" + 0.022*"city" + 0.022*"activity" + 0.021*"sale" + 0.020*"dallas" + 0.020*"price" + 0.019*"richmond" + 0.
019*"report" + 0.018*"chicago" + 0.018*"cleveland"'),
 (1,
  '0.029*"contact" + 0.026*"price" + 0.023*"activity" + 0.023*"sale" + 0.021*"demand" + 0.021*"year" + 0.020*"district" + 0.0
14*"firm" + 0.014*"growth" + 0.013*"report"')]
```

2. On the flip side, Fed Minutes provide highly relevant Topic Models for our use case:

```
[(0,
  '0.026*"market" + 0.023*"inflation" + 0.022*"rate" + 0.019*"committee" + 0.018*"participant" + 0.014*"policy" + 0.014*"price"
+ 0.013*"board" + 0.013*"bank" + 0.012*"reserve"'),
 (1,
  '0.022*"growth" + 0.022*"committee" + 0.022*"market" + 0.020*"price" + 0.020*"member" + 0.018*"rate" + 0.015*"inflation" + 0.
014*"policy" + 0.014*"wa" + 0.013*"year"')]
```

We can map these topic models back to every transcripts and allocate assets based on regulator views.

SHORTLISTED ENSEMBLE LDA MODELS

Model Name	K Topics Set	Actual K Topics	Num Models	Num Passes	Perplexity Score	Umass Coherence	Runtime
Full Fed	4	4	25	80	-6.2335	-0.3327	135 min 26 sec
Full Fed W2V	4	4	10	50	-6.2813	-0.3591	33 min 45 sec
Fed Minutes	4	2	25	100	-5.9194	-0.0177	48 min 48 sec
Fed Statements W2V	4	3	25	50	-5.0525	-0.5673	8 min 22 sec
Fed Speech W2V	4	4	25	50	-6.9291	-0.3639	163 min 34 sec

Source:
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent dirichlet allocation - Journal of Machine Learning Research*. Journal of Machine Learning Research. Retrieved December 14, 2022, from https://jmlr.org/papers/volume3/blei03a/blei03a.pdf

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models - Cornell University. Cornell Bowers Information Science. Retrieved December 14, 2022, from https://mimno.infosci.cornell.edu/papers/mimno-semantic-emnlp.pdf

MODEL LIMITATIONS & FUTURE WORK

1. LDA, or rather topic modelling techniques as a classifier is a relatively weak solution in general because they are a generative model, whereas classification is a discriminative problem.
   We used Word2Vec preprocessing to serve as a semi-supervised discriminatory criterion.
2. Accuracy of signal inference on regulator views are highly sensitive to data source, its preprocessing and the interpreter's ability to capture such insights.

1. Consider other NLP techniques (I.e., BERTopic for larger corpus or NMF for smaller ones)
2. Consider topic modelling as an input feature for supervised machine learning models, leveraging other Macroeconomic indicators to achieve greater prediction accuracy

Source: Sbalchiero, S., & Eder, M. (2020). Topic modeling, long texts and the best number of topics. Some Problems and solutions. *Quality & Quantity, 54*, 1095-1108.

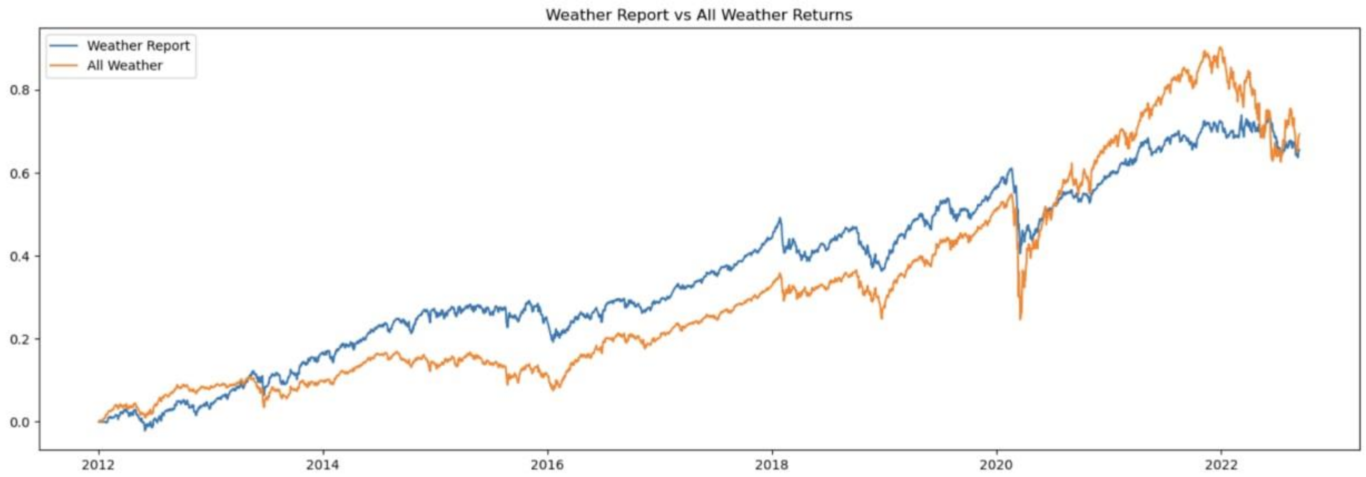
STRATEGY PERFORMANCE

By Jiang Jin

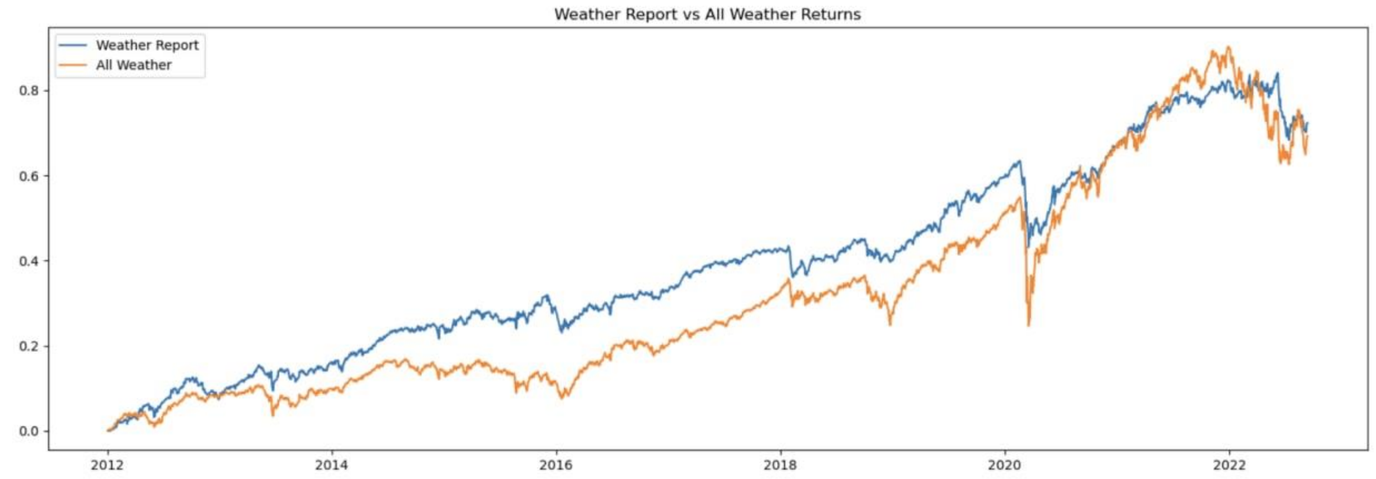
BACKTESTING RESULTS

Model Name	CAGR	Annualized Sharpe	Annualized Vol	Maximum Drawdown Period	Maximum Drawdown (%)
Full Fed	4.06%	0.5893	6.53%	526 Days	32.44%
Full Fed W2V	4.21%	0.6078	6.57%	531 Days	32.90%
Fed Minutes	4.82%	0.8086	5.76%	350 Days	20.54%
Fed Statements W2V	4.56%	0.7608	5.78%	349 Days	20.43%
Fed Speech W2V	5.21%	0.9574	5.03%	279 Days	20.28%

Model: Fed Min



Model:
Fed Speech W2V



STRATEGY LIMITATIONS & FUTURE WORK

1. Assumes no rebalancing for our benchmark (all-weather portfolio)
2. Simplistic inference from using VIX index as proxy for market expectation

1. Social Media data to derive market expectation using LDA
2. Comparison of central bank topic models versus retail topic models
3. Looking at different central bank's statements

SENTIMENT ANALYSIS

By Sam

DATA METHODOLOGIES

DATA SOURCE

Text Data (Fed)

Beige Book

Statements

Minutes

Speeches

Non-Text Data (ETFs)

SPY (Equity)

BND (Bond)

TLT (LT Treasury)

SHY (ST Treasury)

BIL (Cash)

TIP (Inflation Hedge)

GLD (Inflation Hedge)

VNQ (Real Estate)

DBC (Commodity)

Beige Book — Economic conditions before meetings

Statements — Policy decision immediately after meetings; 8 times a year

Minutes — Detailed record of meeting 3 weeks after statements

Speeches — By fed officials in between meetings

TEXT DATA

NON-TEXT DATA

Equity

SPY: SPDR S&P 500 ETF Trust

Bonds

BND: Vanguard Total Bond Market ETF

TLT: iShares 20+ Year Treasury Bond ETF

SHY: iShares 1-3 Year Treasury Bond ETF

BIL: SPDR Bloomberg 1-3 Month T-Bill ETF

TIP: iShares TIPS Bond ETF

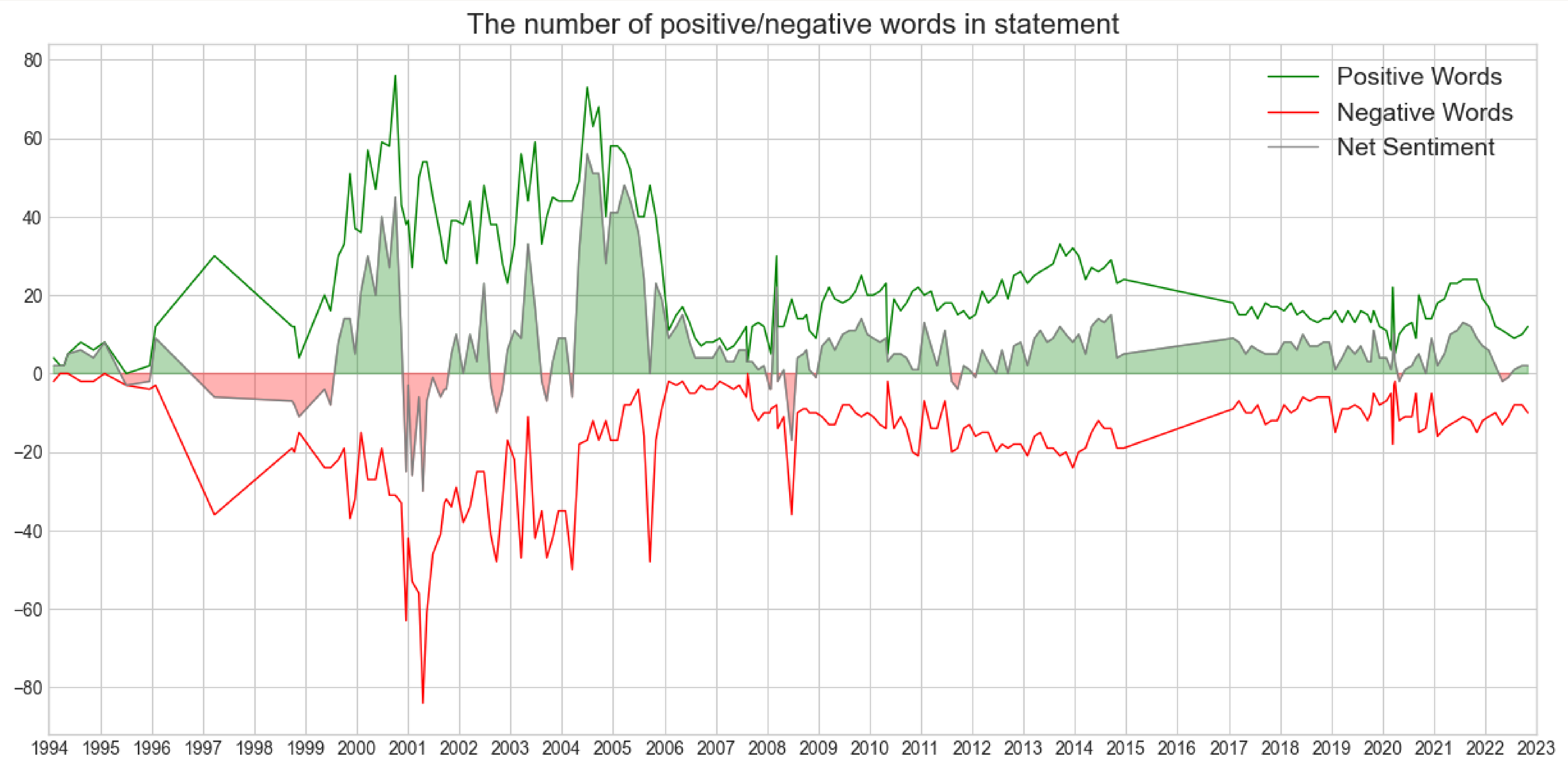
Commodities & Others

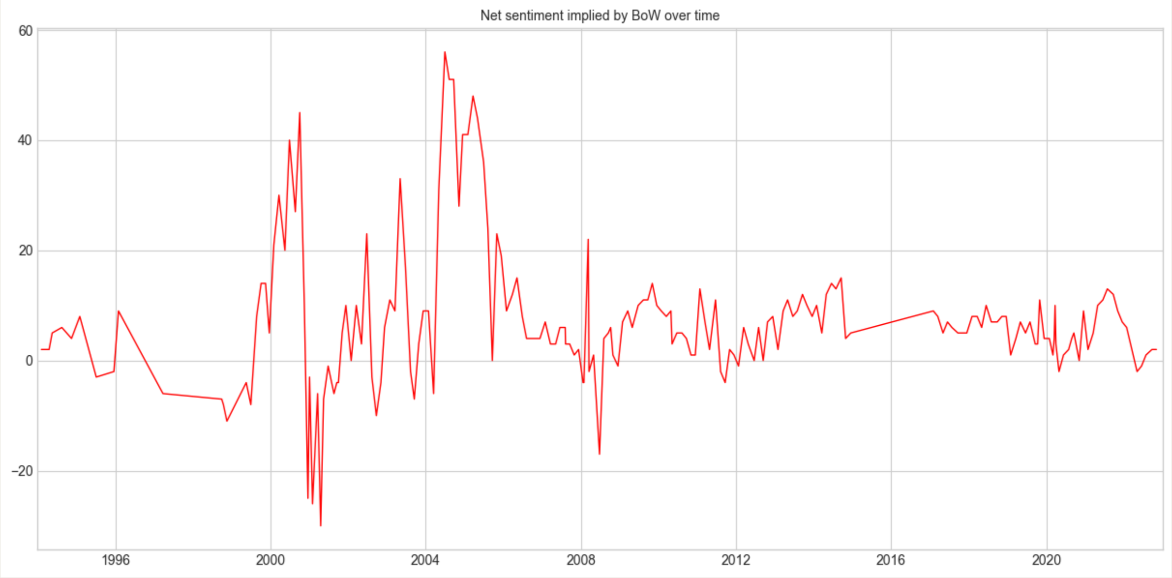
GLD: SPDR Gold Shares

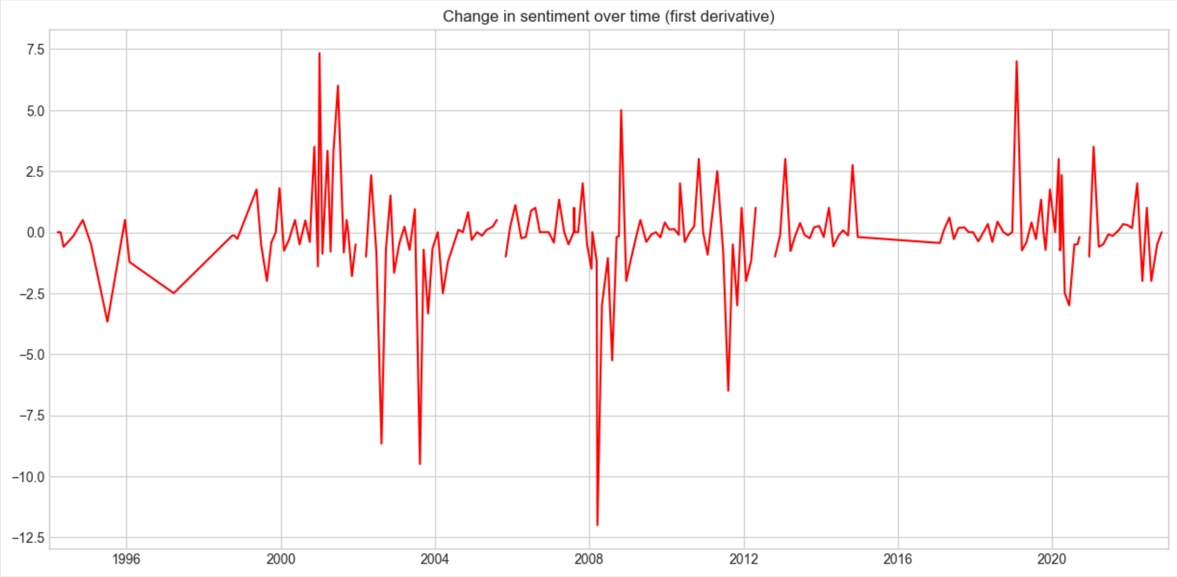
DBC: Invesco DB Commodity Index Tracking Fund

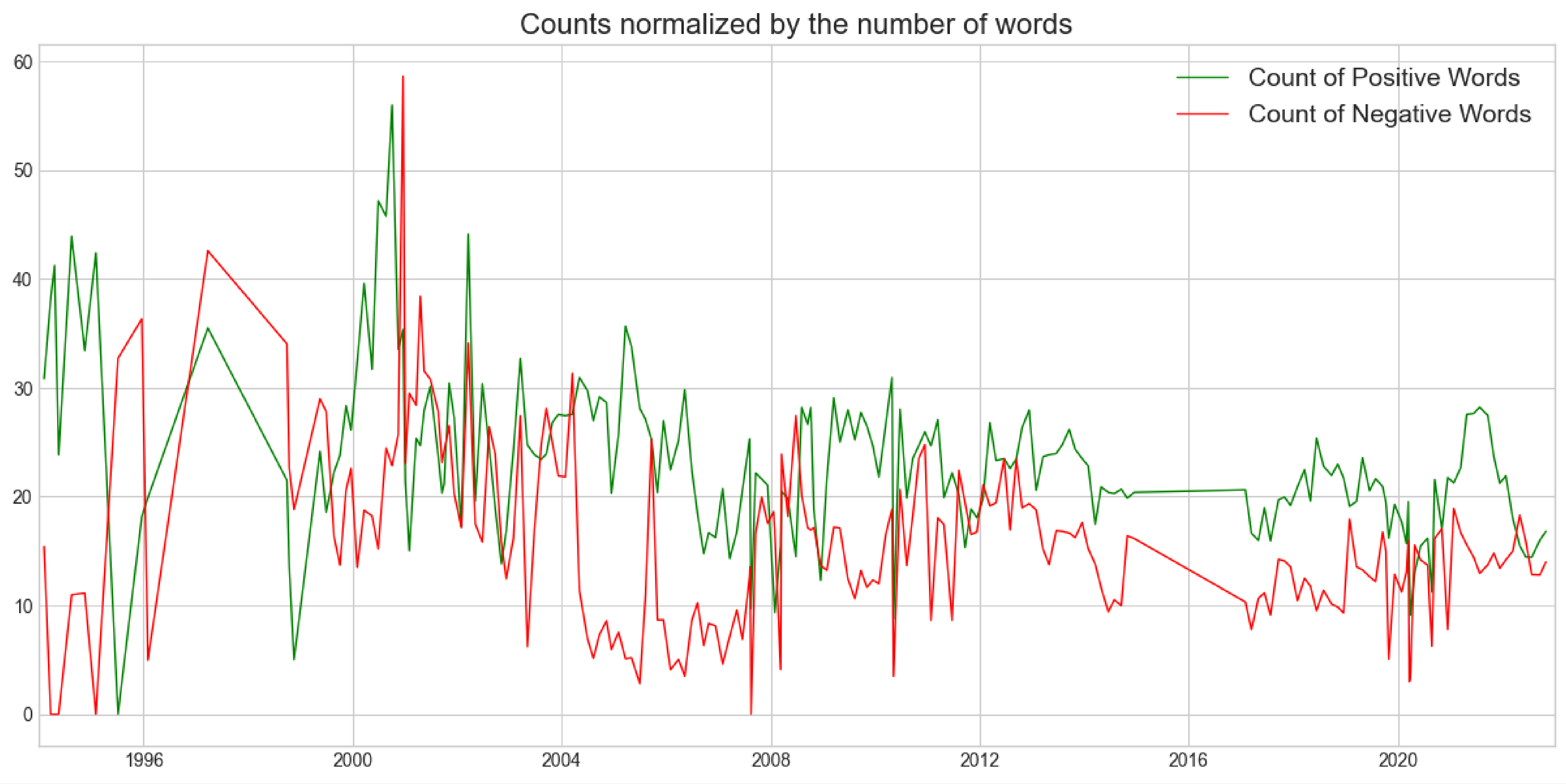
VNQ: Vanguard Real Estate ETF

EXPLORATORY DATA ANALYSIS (EDA)

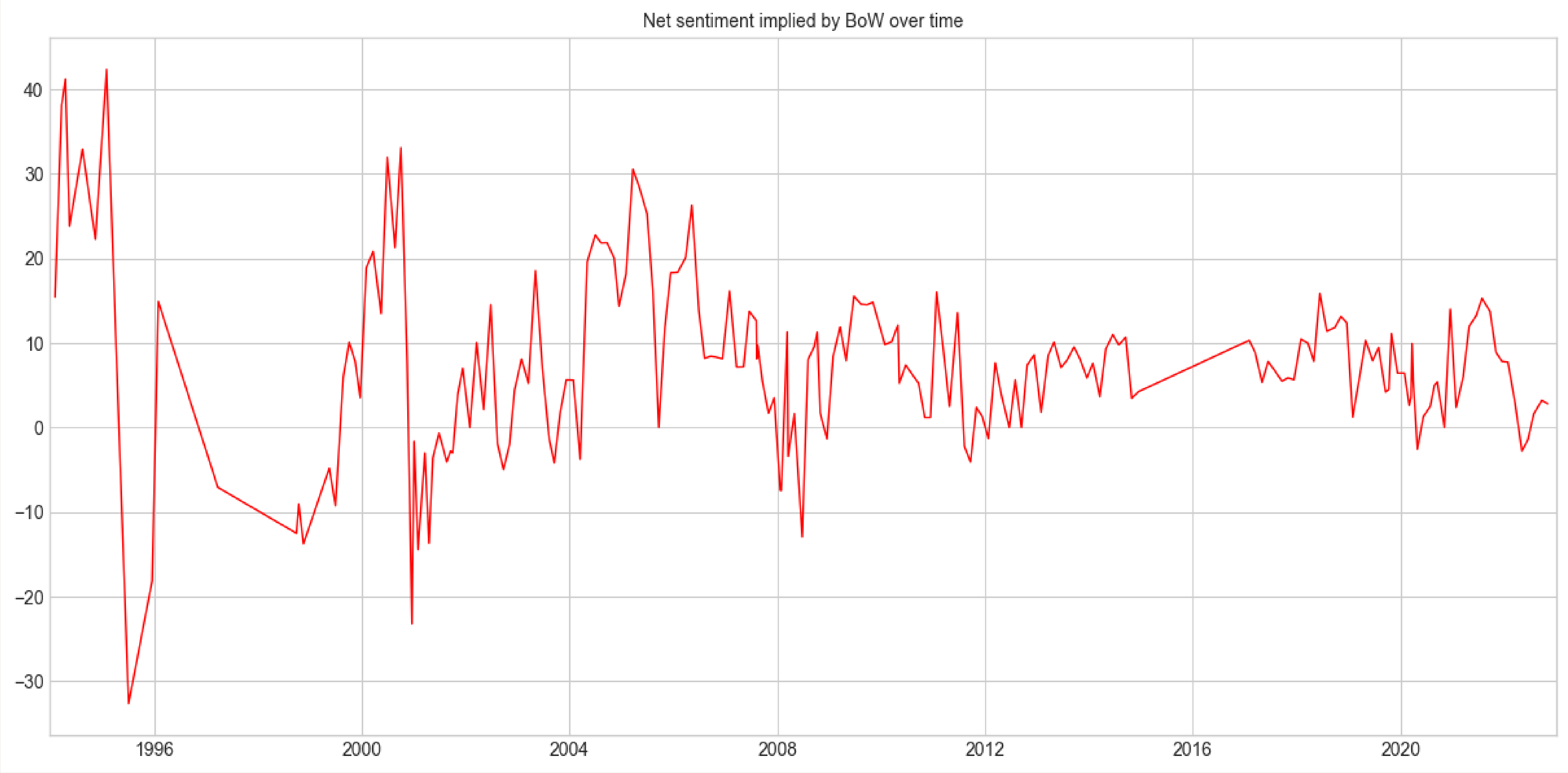

The number of positive/negative words in statement
Positive Words
Negative Words
Net Sentiment
80
60
40
20
0
−20
−40
−60
−80
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023

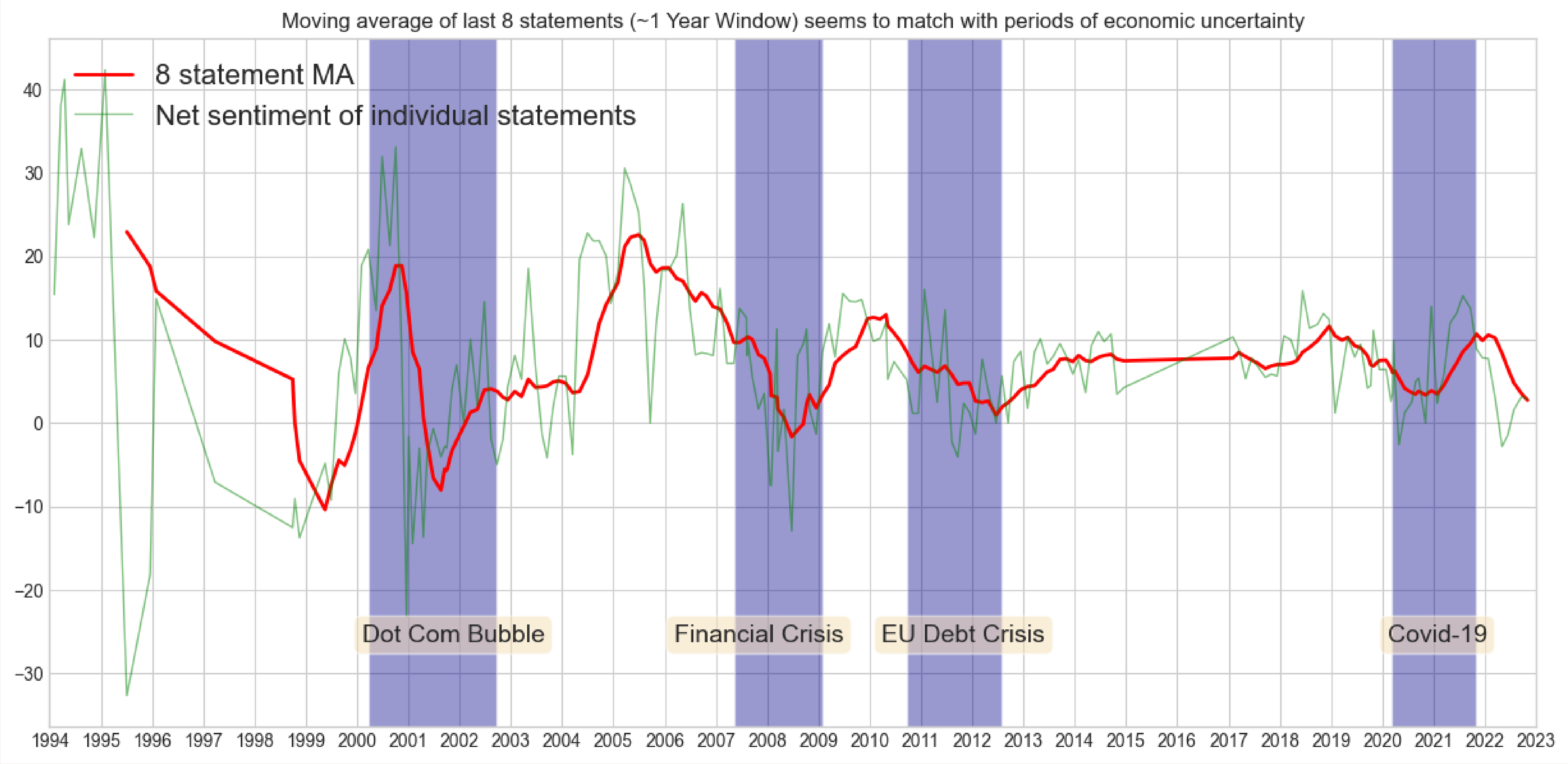

Net sentiment implied by BoW over time
60
40
20
0
−20
1996
2000
2004
2008
2012
2016
2020


Change in sentiment over time (first derivative)
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−12.5
1996
2000
2004
2008
2012
2016
2020



Counts normalized by the number of words
Count of Positive Words
Count of Negative Words
60
50
40
30
20
10
0
1996
2000
2004
2008
2012
2016
2020


Net sentiment implied by BoW over time
40
30
20
10
0
−10
−20
−30
1996
2000
2004
2008
2012
2016
2020



Moving average of last 8 statements (~1 Year Window) seems to match with periods of economic uncertainty
8 statement MA
Net sentiment of individual statements
40
30
20
10
0
−10
−20
−30
Dot Com Bubble
Financial Crisis
EU Debt Crisis
Covid-19
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023

INTERPRETATION OF RESULTS

EVENT RETURNS

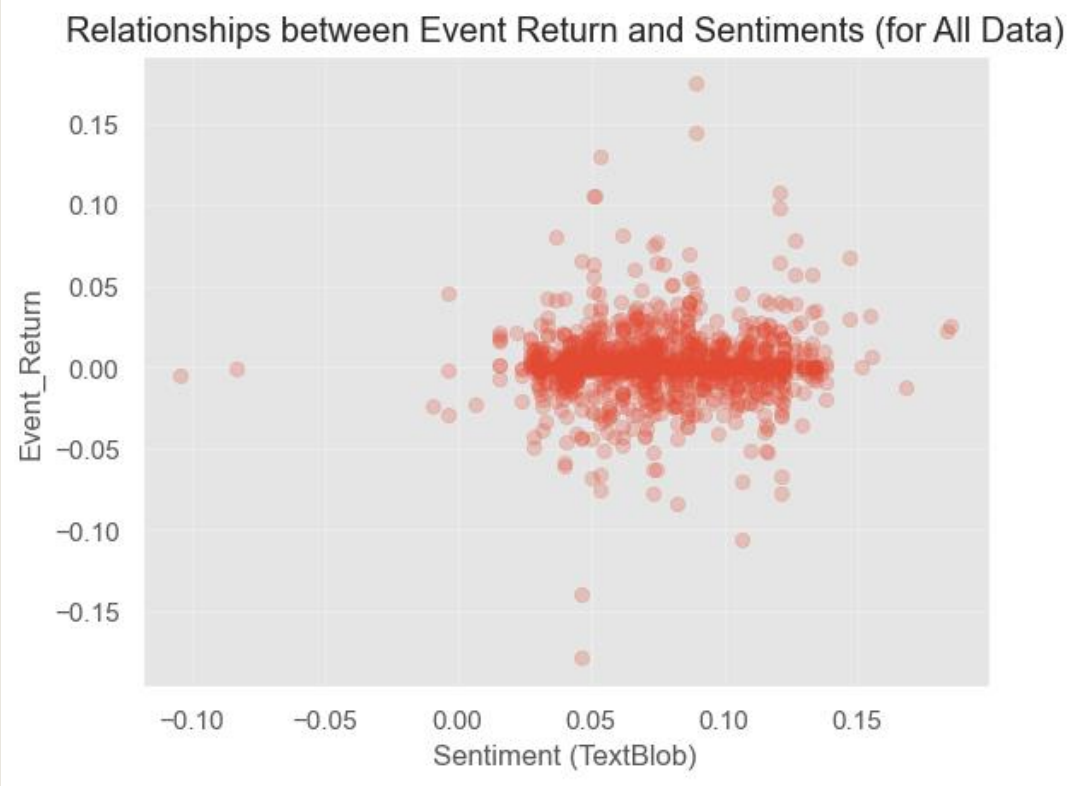
$$R_{t-1} + R_t + R_{t+1}$$

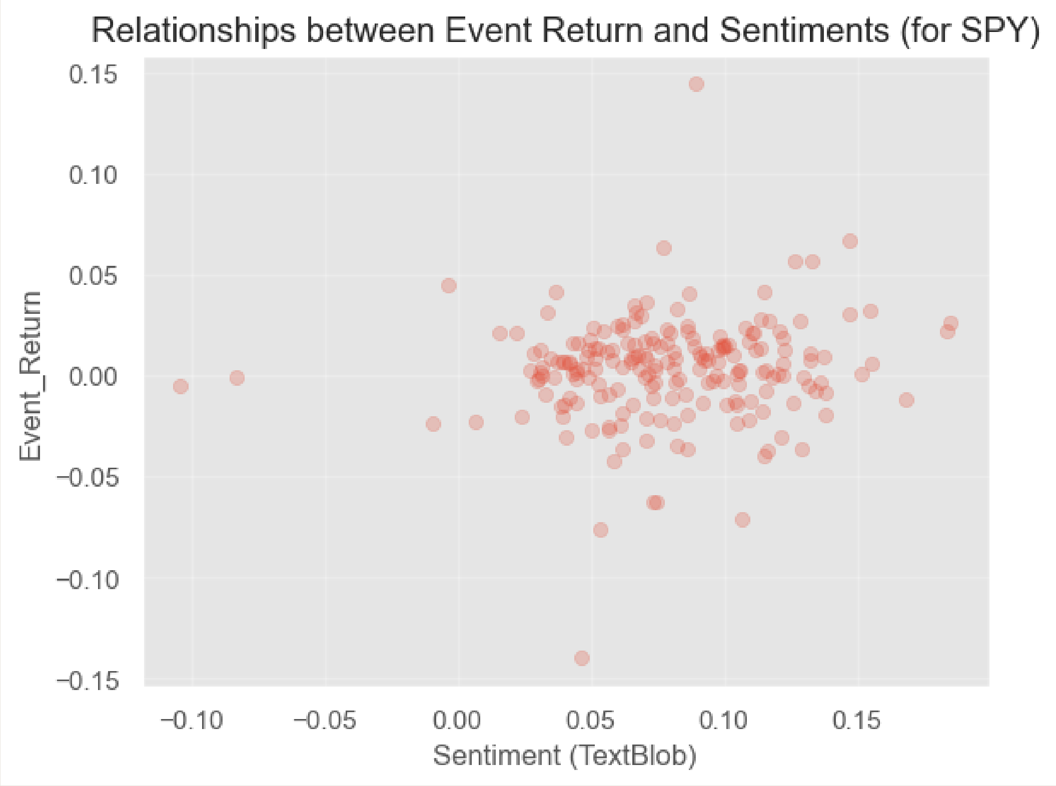
R_{t-1}	is return one day before policy decision
R_t	is is return on the day of policy decision
R_{t+1}	is return one day after policy decision

TEXTBLOB OVERVIEW

1. Pre-trained model based on the Naïve-Bayes classification algorithm
2. Polarity score: convert sentences into a numerical value of sentiment between -1 to +1 (positive-negative)
3. Subjectivity score: convert sentences into a numerical value of subjectivity between 0 to 1 (objective-subjective)

NLP TECHNIQUE 1: TEXTBLOB

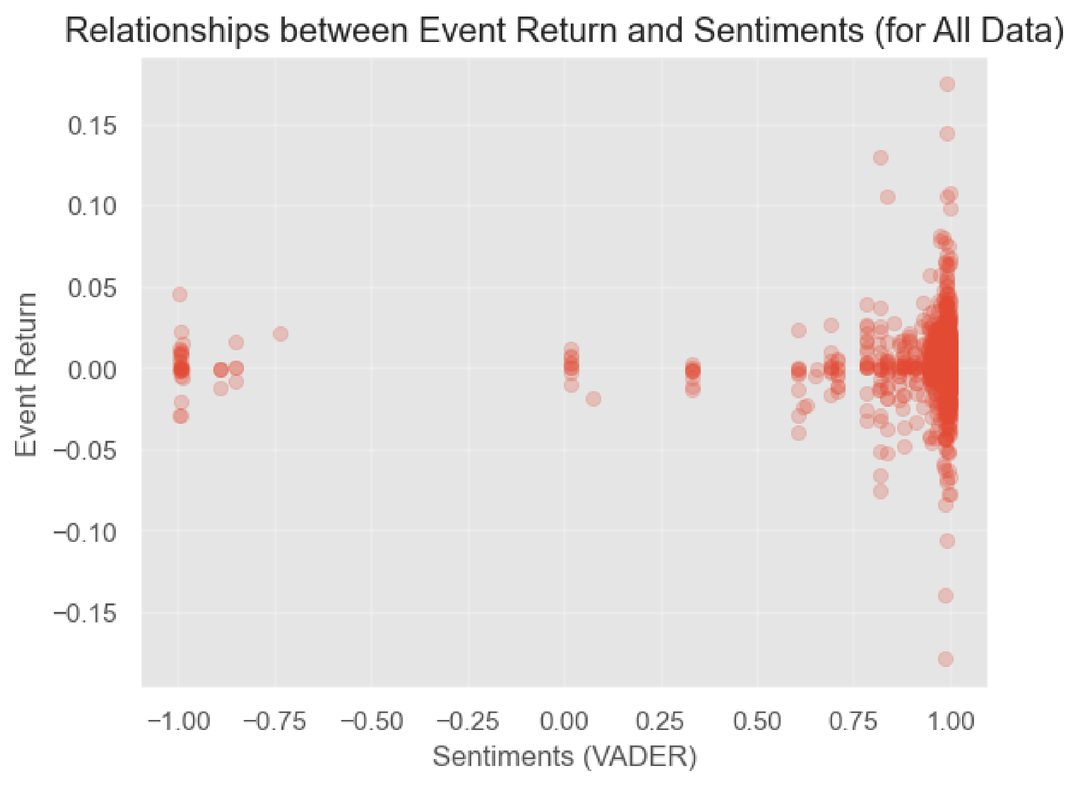


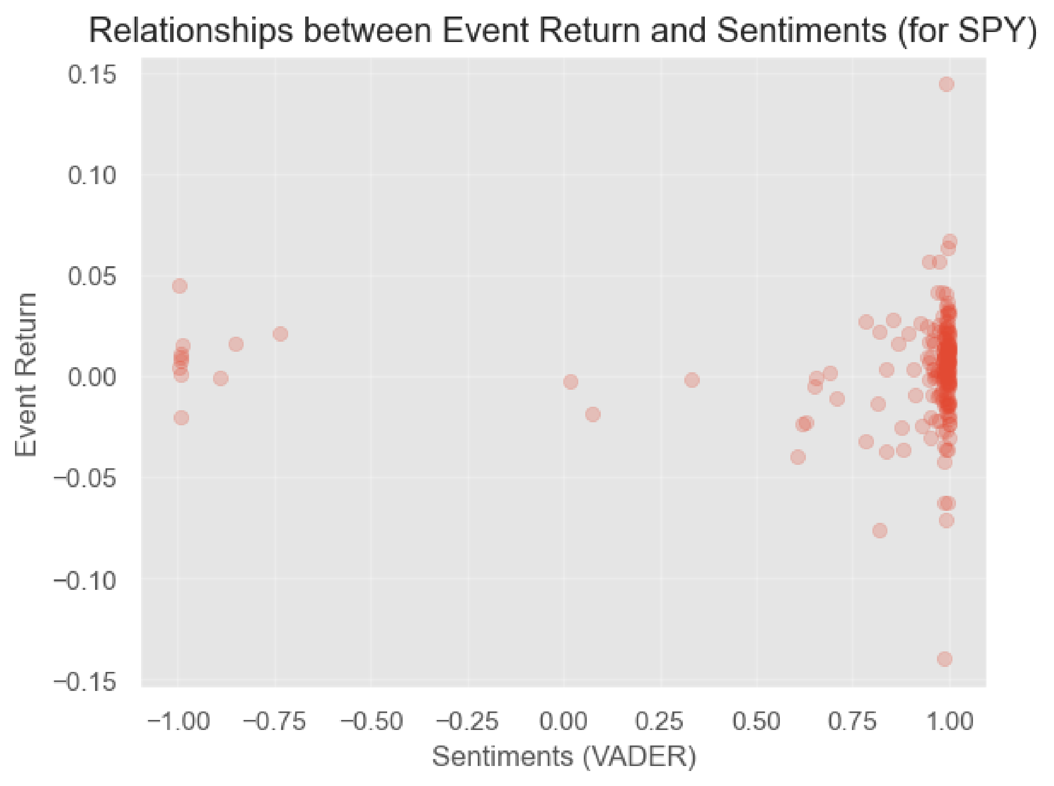


VALENCE AWARE DICTIONARY FOR SENTIMENT REASONING (VADER) OVERVIEW

1. Pre-built sentiment analysis model included in the NLTK package
2. Lexicons are special dictionary or vocabularies created for analyzing sentiments
3. Trained model based on financial lexicon (~20k words) and applying it to fed statements

NLP TECHNIQUE 2: VADER

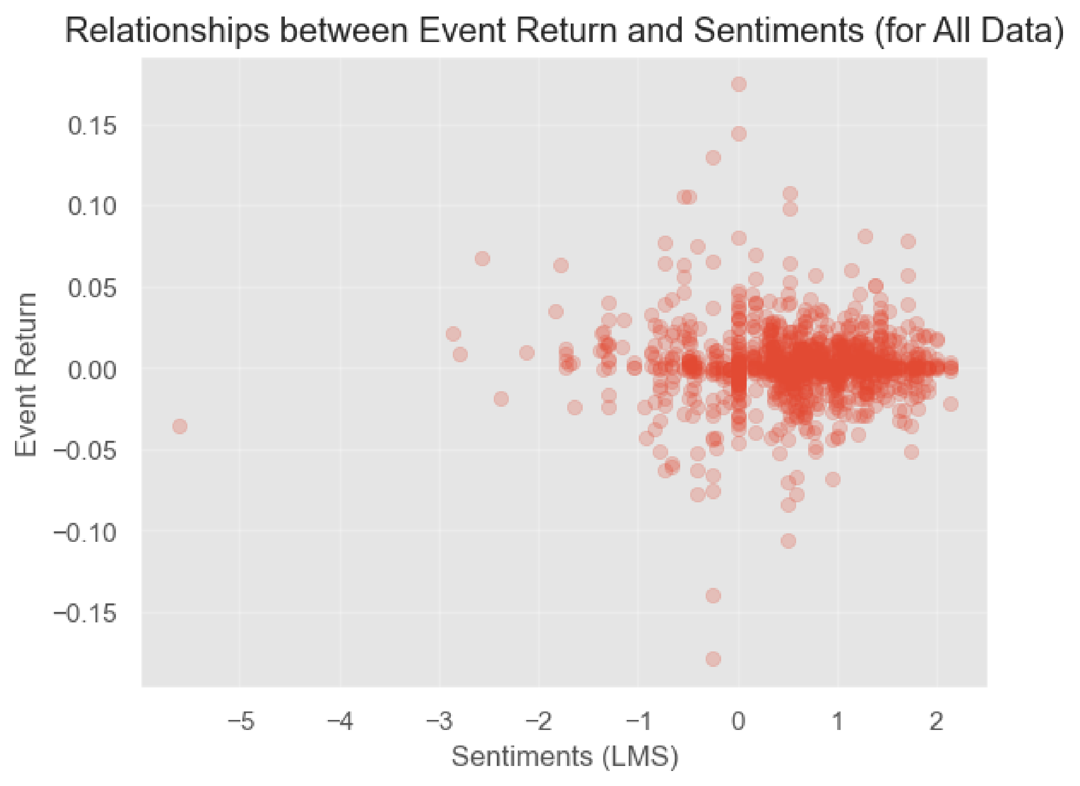


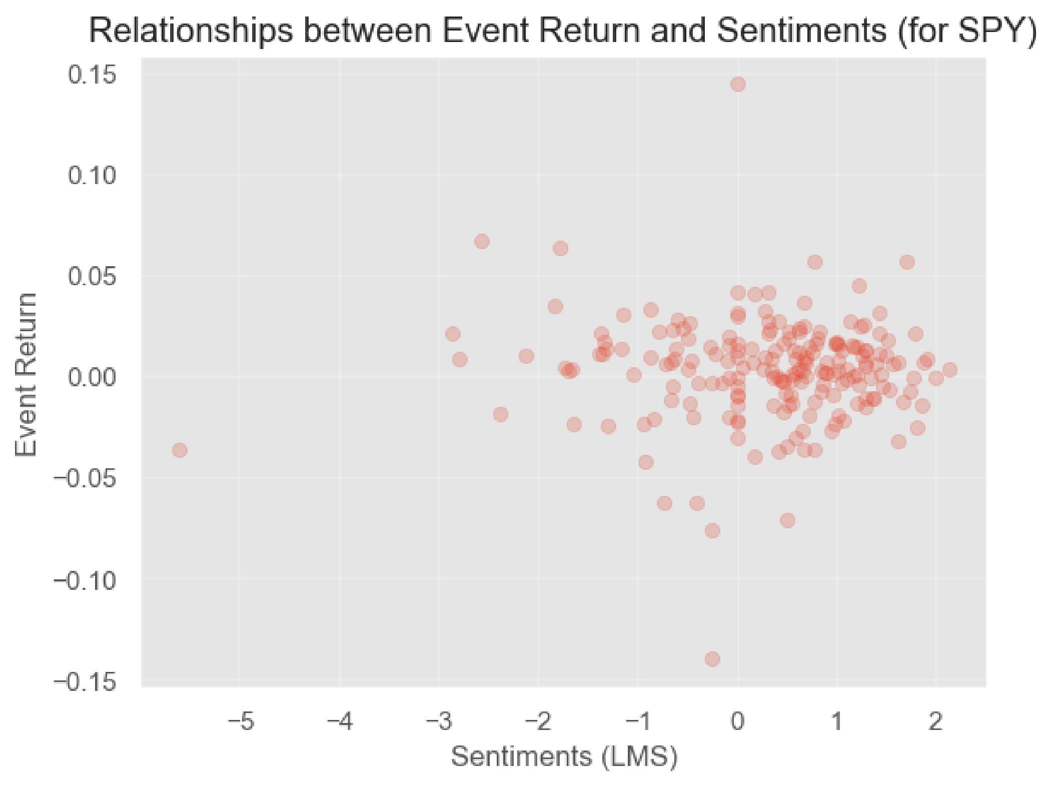


LOUGHRAN-MCDONALD SENTIMENT (LMS) OVERVIEW

1. Gold standard that examines tokens from all 10-K type filings for the full EDGAR 10-K archive and earnings calls from CapIQ
2. Sentiment categories include negative, positive, uncertainty, litigious, strong modal, weak modal, and constraining
3. Trained model based on LMS lexicon (~4k words) and applying it to fed statements

NLP TECHNIQUE 3: LMS





Positive fed = Increasing interest rates

Negative fed = Decreasing interest rates

Increasing interest rates is bad for all asset classes, except commodities and real estate

We shall use the LMS model for our backtesting strategy

BACKTESTING STRATEGY

SPY TRADING LOG (EXTRACT)

SPY
Starting Value of Our Portfolio: 100000.00
1996-01-31, Previous Sentiment -2.38, New Sentiment -0.65 BUY CREATE, 39.53
1996-02-01, BUY DONE, Price: 39.49, Cost: 3948.95, Comm 0.00
1997-03-25, Previous Sentiment -0.65, New Sentiment -2.79 SELL CREATE, 50.06
1997-03-26, SELL DONE, Price: 50.25, Cost: 3948.95, Comm 0.00
1997-03-26, Profit of the operation, GROSS 1076.43, NET 1076.43
...
2021-01-28, SELL DONE, Price: 367.16, Cost: 36130.61, Comm 0.00
2021-01-28, Profit of the operation, GROSS 585.21, NET 585.21
2021-04-28, Previous Sentiment 0.78, New Sentiment 1.57 BUY CREATE, 408.53
2021-04-29, BUY DONE, Price: 411.39, Cost: 41138.64, Comm 0.00
2022-11-29, (MA Period 15) Ending Value 118006.87

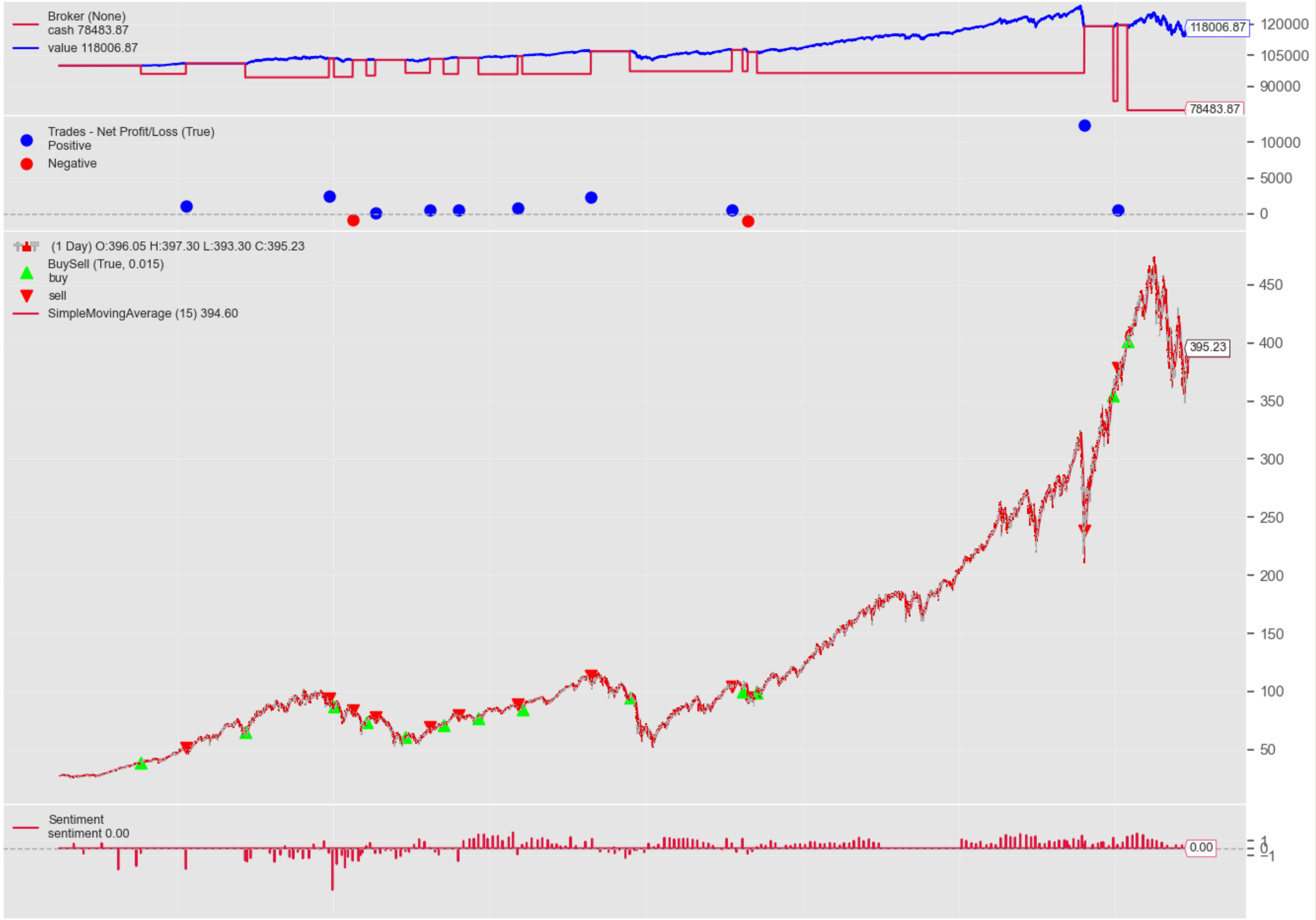
Starting Value of Our Portfolio: 100000.00
Final Value of Our Portfolio: 118006.87
Profit: 18006.87

Cash Value Observer

Trade Observer

Buy/Sell Observer

Sentiment Observer



PORTFOLIO OF ETFS TRADING RESULTS

Performance of Strategy

	SPY	BND	TLT	SHY	BIL	TIP	GLD	VNQ	DBC
Per Unit Start Price	27.54	46.73	41.34	58.12	82.90	55.84	44.38	22.74	22.05
Strategy Profit	18,006.86	199.33	2,075.47	2,125.70	465.73	6,372.85	14,335.00	1,918.22	389.39

Sensitivity Analysis (for SPY)

SPY		Position size		
		90	100	110
Change in sentiment	0.55	+2.10%	+13.44%	+24.78%
	0.5	-10.00%	$18,007	+10.00%
	0.45	-31.63%	-24.04%	-16.44%

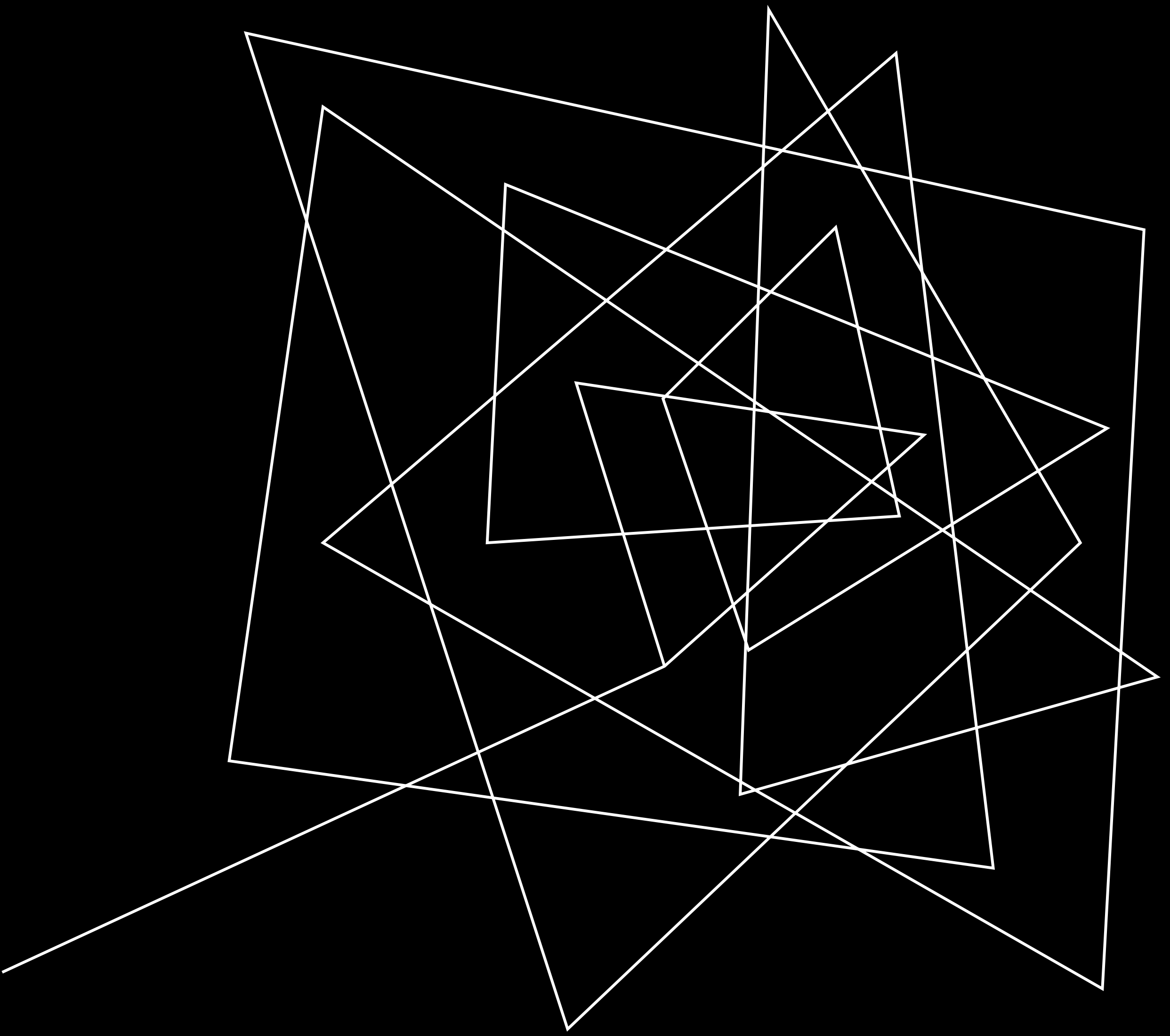
LIMITATIONS & FUTURE WORK

Limitations

1. Lack of training data (8 statements per year)
2. Data trade-offs (statements are timely but lack context compared to minutes; minutes gives more insights but lagged by 3 weeks giving little applicability)
3. Data quality (irrelevant and long paragraphs, ML good at learning ~500 words, splitting words by overlapping 200/50 but lose context, etc.)

Future Work

1. TFIDF
2. LSTM/RNN
3. BERT



CLOSING THOUGHTS

By Stanley

reddit

DISCORD

Selenium

eToro

/(reg)ex/

python

Backtrader