Backtesting Sentiment: A r/wallstreetbets Trading Strategy

Jonas Paul Schmitten, MA Quantitative Economics and Finance
Noah Julian Angara, MA Quantitative Economics and Finance

Thursday, 20th May 2021, Big Data Analytics *(8,272,1.00)*

Dataset

- **Uncleaned 11 GB csv file** with all posts and comments from r/wsb since 2012
  - Variables include author name, date and time, comment/post, etc.
- **Stock prices** from Yahoo Finance (*tidyquant*), **Stock tickers** from NASDAQ (+7,000)
- Data sometimes spotty due to Pushshift API down (comment volume affected)

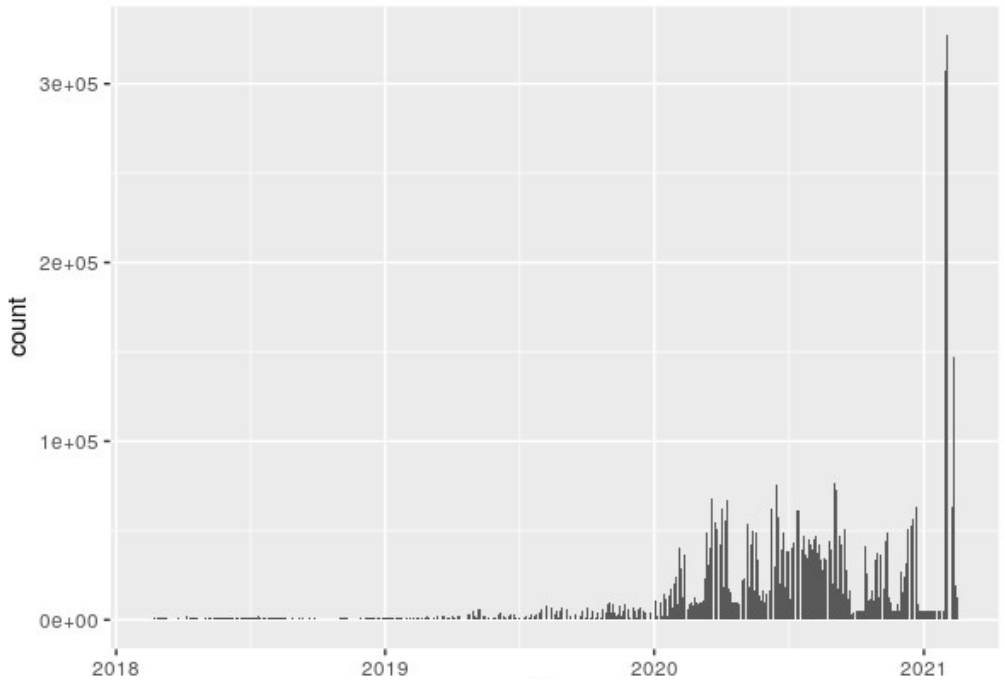
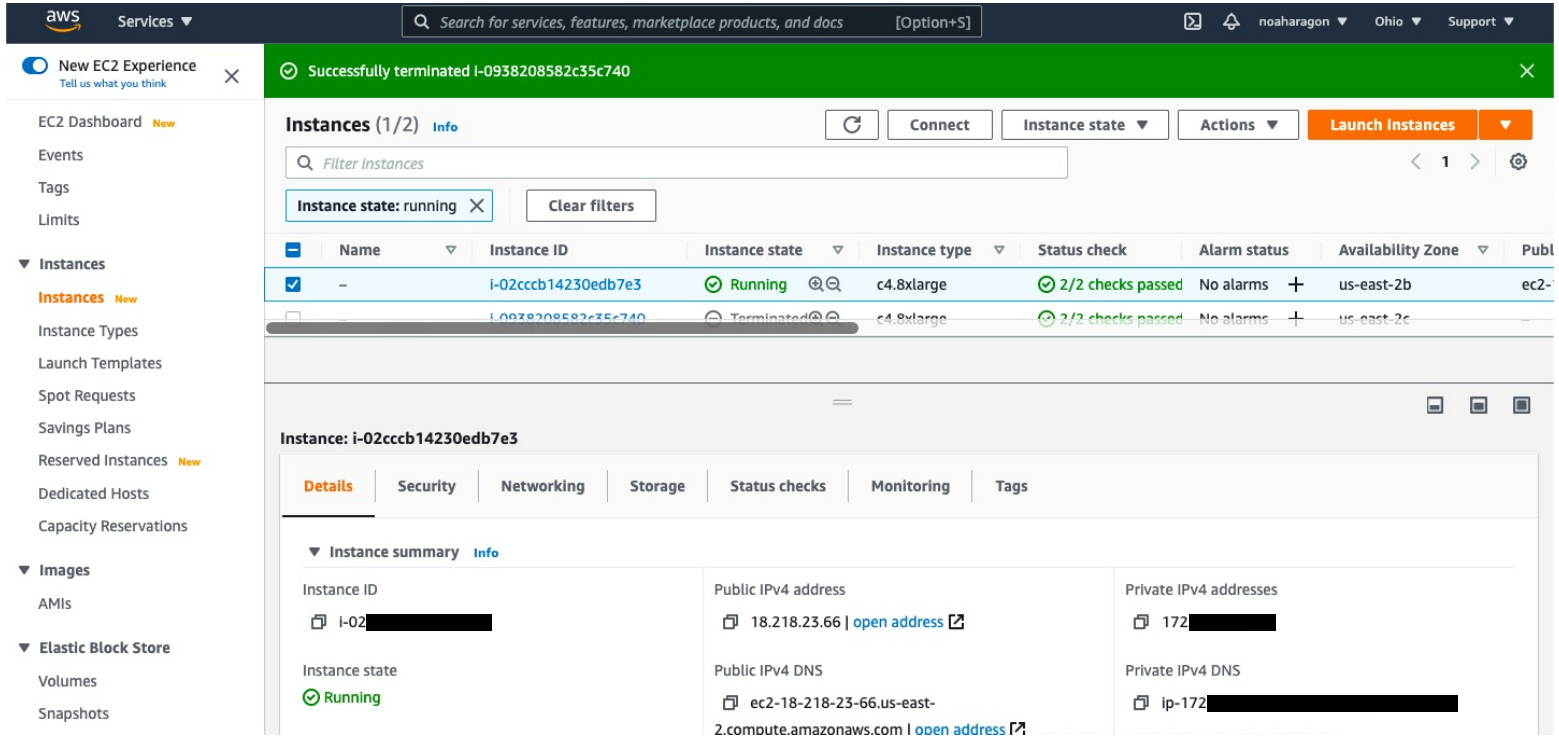
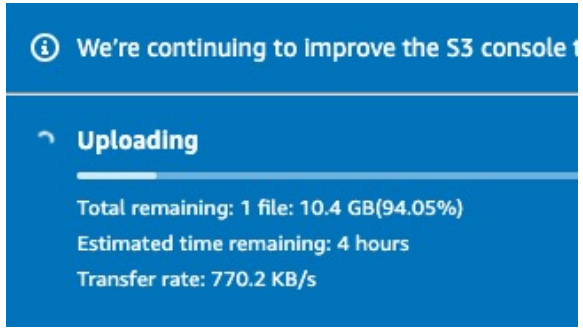


Figure: daily comment volume

2020-04-30	20:11:23	Easy money according to WSB NFLX calls DPZ calls HOG puts AMZN calls what else should ...
2020-04-30	20:11:24	They have been cooking the books forever
2020-04-30	20:11:24	Do you even know how to read? It's a put credit spread. He bet it is going up.
2020-04-30	20:11:25	Sold close to the top at EOD. So happy
2020-04-30	20:11:25	The whole market now lies on Timmy
2020-04-30	20:11:26	If we have a huge down day tomorrow on the one day I switch to calls, I'm going to be sad ...
2020-04-30	20:11:26	Fuck me, THAT HEADLINE!
2020-04-30	20:11:27	[deleted]
2020-04-30	20:11:27	Bet AMZN will be back up 4% tomorrow lmao
2020-04-30	20:11:27	I stayed out bc the day was so red, and TSLA's bizarre plunge after posts. I'm also becomi...
2020-04-30	20:11:28	Will he? considering feud between Trump n Bezos lol . . Sorry I made it sound like real hous...
2020-04-30	20:11:28	Oh shit did you just lose $260?
2020-04-30	20:11:29	Why isn't SPY tanking tho? Isn't amazon like 5% of it lol
2020-04-30	20:11:29	At least my SPY calls are 6/19 lol I don't understand MSFT exceeds expectations market go...
2020-04-30	20:11:29	Omg lol
2020-04-30	20:11:29	i'm holding spy 289C overnight, lets see if everyone notices that AWS will make a loads of ...
2020-04-30	20:11:29	AMZN putssssss

AWS Cloud Computing





- Tested code **locally** with partial dataset (*fread, pryr*) before scaling
- Stored data in Amazon **S3** bucket and then used **EC2**
- Louis Aslett Community **AMI** with RStudio (had to adapt Linux partition)
- 60 GB memory and 32 VCPUs (used **Garbage Collector** in between steps)

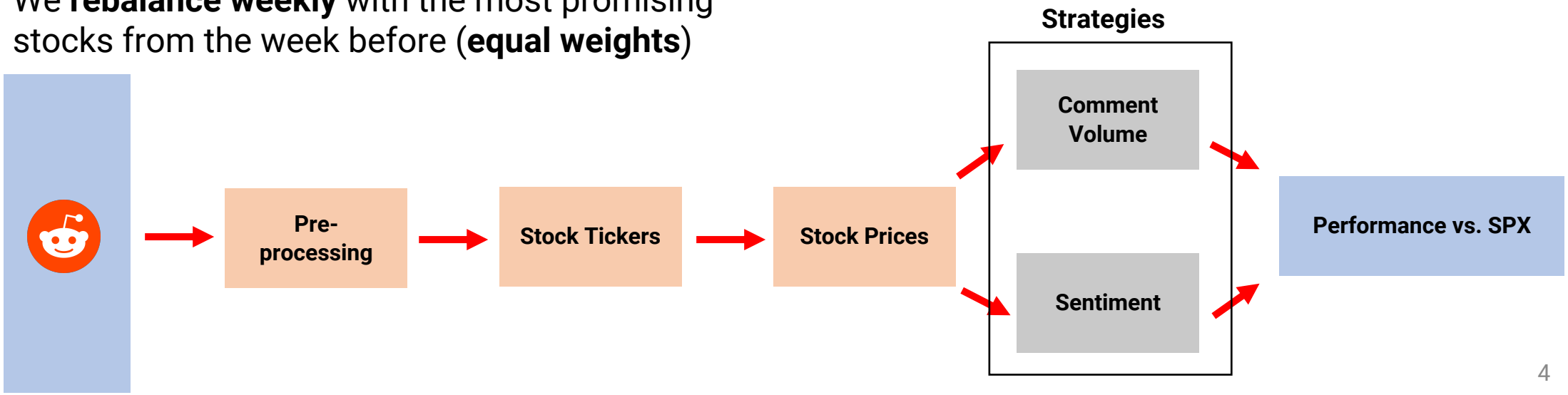
Procedure

- Defined our **own dictionary** to capture r/wallstreetbets specific terms
  - Changed existing CRAN **vader** sentiment

```
c("retard", "retarded", "fuck", "fucking", "autist", "fag", "faggot", "gay", "stonk", "porn",
  "degenerate", "boomer", "ape", "gorilla", "shit"), V2 = 0, V3 = 0.5), # neutral
c("bull", "bullish", "tendie", "tendies", "call", "long", "buy", "moon", "hold",# positive
  "diamond", "hands", "yolo", "yoloed", "free", "btfd", "rocket", "elon", "gain",
  "420", "calls", "longs", "sky", "space", "roof", "squeeze", "balls", "JPOW", "printer",
  "brrr", "HODL", "daddy", "BTFD", "squoze", "full moon", "full moon face", "ox", "astronaut",
  "man astronaut", "gem stone", "money bag", "green", "profit"), V2 = 3, V3 = 0.5),
c("bear", "sell", "put", "short", "shorts", "puts", "bagholder", "wife", "boyfriend",# negative
  "shorting", "citron", "hedge", "fake", "virgin", "cuck", "guh", "paper", "SEC", "drilling",
  "bear face", "briefcase", "roll of paper", "red"), V2 = -1.5, V3 = 0.5))
```

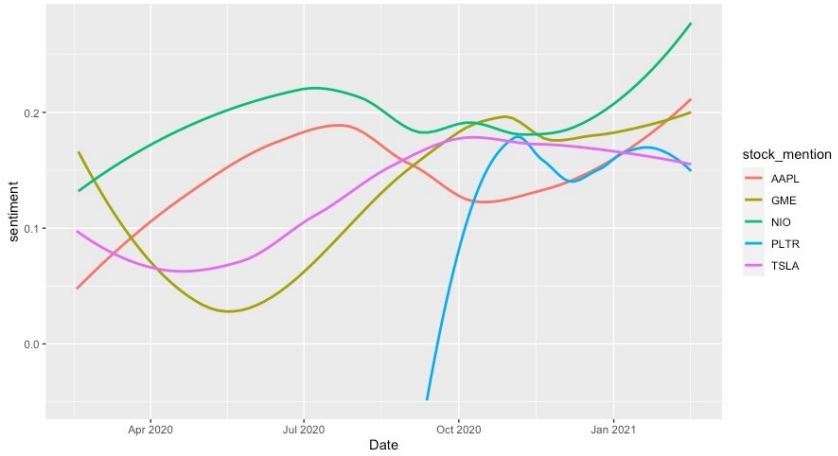
- We **exclude nano and micro** cap companies
- Issues included, among others, SPACs change ticker after take-over and ticker selection
  - Excluded tickers such as **$DD, $RH,** and **$A**

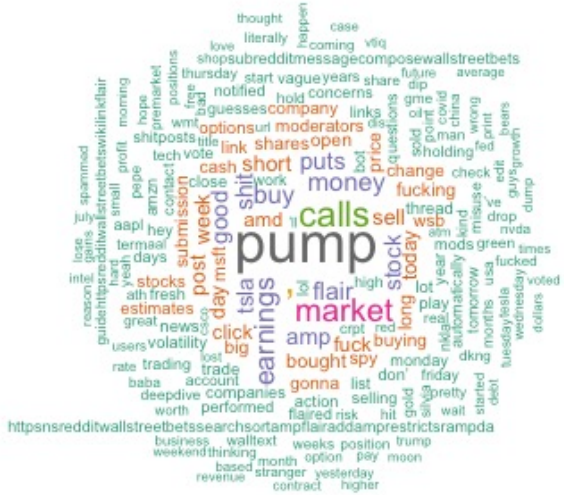
- We **rebalance weekly** with the most promising stocks from the week before (**equal weights**)



Results and Performance

- Sentiment calculations turned out to need a lot of computing power
- Therefore, focus of trading strategy on **comment volume**
  - Our model seems to **find stock tickers well** in comments (few false positives)



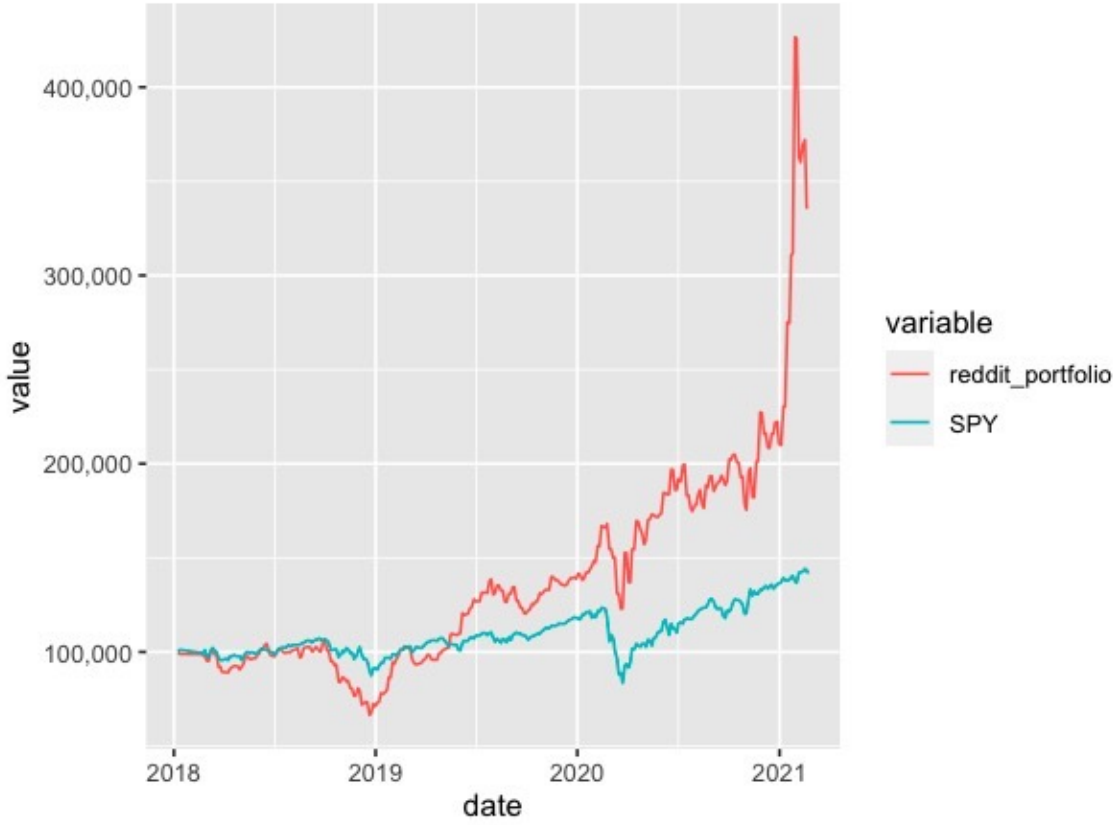


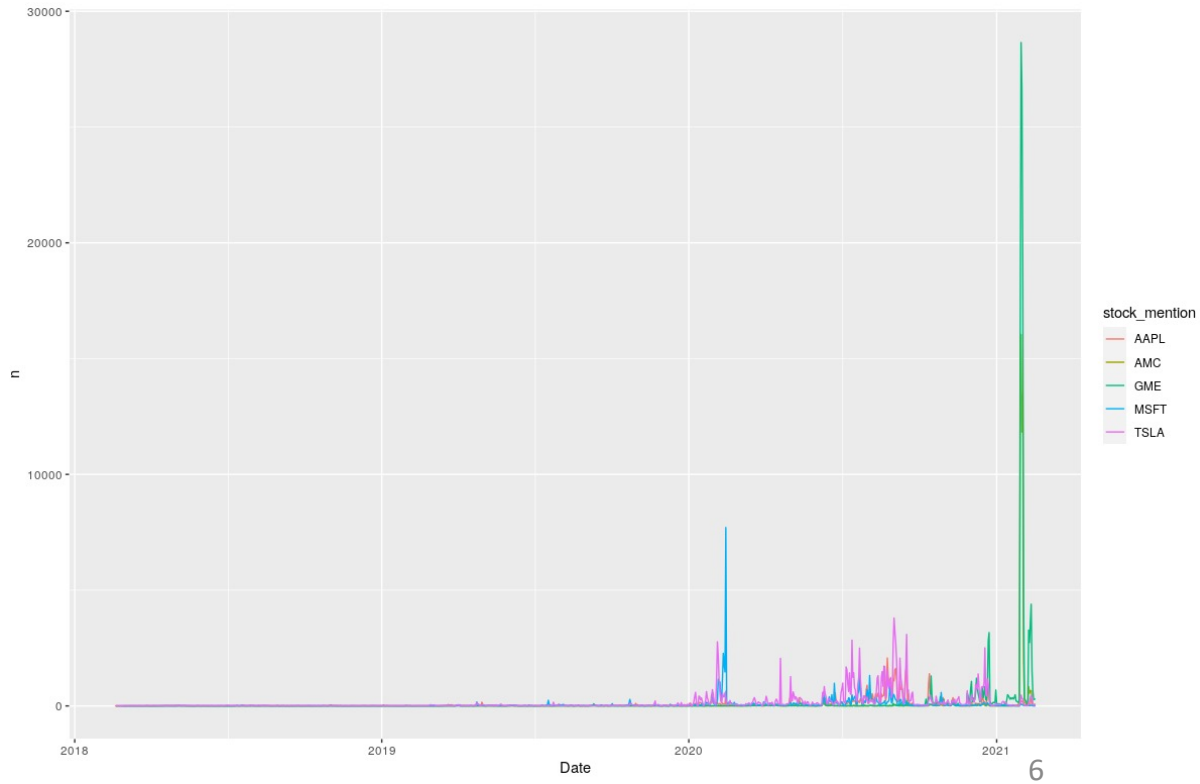
2018-03-12	MU	AMD	GE	SQ	AAPL	NVDA	CRON	AMZN	NFLX	JD
2018-03-19	MU	NVDA	AMD	GE	AMZN	SNAP	BABA	JD	SQ	AAPL
2018-03-26	FB	MU	AMD	AMZN	NVDA	SNAP	MSFT	AAPL	OLED	GE
2018-04-02	FB	TSLA	MU	AMD	FOR	NVDA	AMZN	GE	AAPL	MSFT
2021-01-04	TSLA	NIO	PLTR	GME	MT	BABA	AAPL	PLUG	QS	PSTH
2021-01-11	GME	BB	PLTR	TSLA	NIO	PLUG	AMD	AAPL	CCIV	FB
2021-01-18	GME	BB	PLTR	TSLA	AMD	AAPL	BABA	NIO	NOK	CAN

Results and Performance

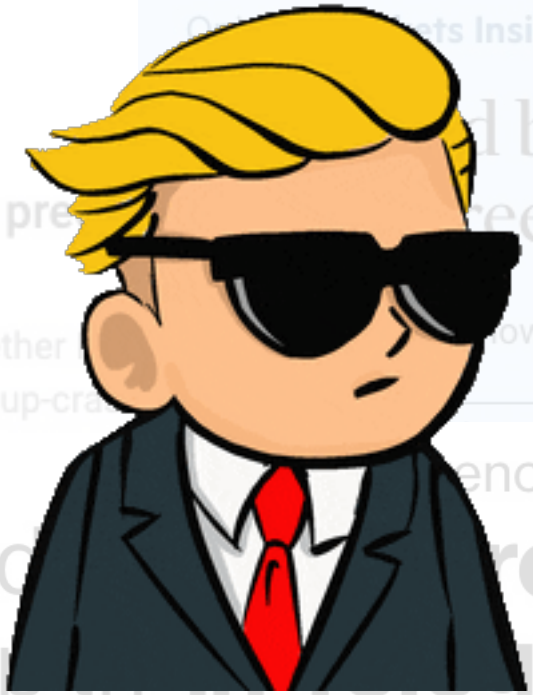
- **Significant outperformance** our - portfolio relative to SPY (SP500 ETF)
  - $100,000 invested in our fund results in close to **$400,000 vs. less than $150,000**
- Performance is correlated with comment volume/ subscriber numbers

- **GME** could have been spotted in **October 2020** already





Thank you



💎🙌