Contents

1 Results

1.1 Text Analysis

Firstly, we begin with a purely semantic analysis. Using the whole sample of more than 438 thousand r/wallstreetbets unique submissions, word tokens, Porter's stems and lemmas are extracted. The overwhelming positive sentiment is evident at first sight as shown by the Fig. [1], depicting the most frequent word tokens and bigrams. The prominency of words *rocket, buy, hold, moon* or *let* tokens indicates a large bullish sentiment. Broader context is presented by the bigram model extracting phrases such as *rocket rocket, let us, moon rocket, us go* or *hold line*[2]. Fig. **??** shows the activity on the subreddit throughout the year corresponding to most volatile periods of GME.

Moreover, highly frequent *short-squeeze, shares* or *short interest* terms manifest the relevancy of the short-selling phenomena in the trading frenzy, it arguably being one of its triggering factors. The relevance of the selected sample of companies, as shown by Fig. **??**, demonstrates the large association users made between GameStop and other tickers. The key to identifying ticker mentions is based on a simple convention that ticker must be preceded by a $ sign and contain up to six letters.

Sentiment Analysis

Moving beyond a simple textual analysis, we assign sentiment scores to the scraped submissions. Owing to the specific lexicon predominantly used in the r/wallstreetbets subreddit, we believe that slight modifications of the lexicon are necessary to better reflect bearish and bullish signals.[3] Therefore, we construct two sentiment scores, before and after applying the lexicon adjustments, the overview of which is given in Appendix **??**. Our analysis primarily employs the latter.

Fig. **??** further confirm the two distinct periods in which sentiment surrounding GameStop peaked on the subreddit. It is shown that positive sentiment dominated throughout the year, regardless of the performance and volatility of GME. Interestingly, aggregating the scores intraday, we detect large deviations and periods of strong negative sentiment, which almost disappear in aggregations of frequencies larger than 6 hours.[4] This may suggest a fickle and reactive behaviour on the subreddit as negative mood is quickly suppressed and drown out by the trend of overwhelming positive sentiment.[5].

Comparing the news and Reddit sentiments, we observe very different patterns how-

[1] number

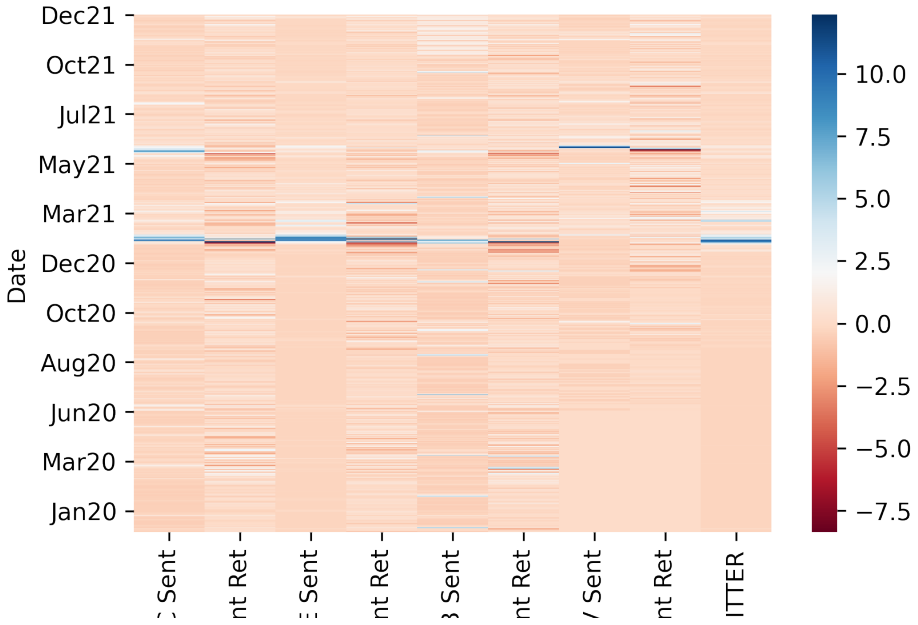
[2] translation of emojis may distort the most used phrases

[3] overview of some of the changes will be referred here, in appendix

[4] either figure included here or comparison in appendix

[5] echo chamber, repetition?

Figure 1: Heatmaps



ever, note the large discrepancy in the respective datasets' sizes. While the former shows significant peaks and troughs, Reddit users' sentiment seems to be unwaveringly positive throughout the period. On the other hand, comparing the two social media sentiment measures from Twitter and Reddit, we observe an even more stark contrast. The Twitter sentiment displays more or less consistently opposite signals compared to the Reddit sentiment, being largely negative especially during the most active periods.

Nevertheless, all sentiment datasets pick on the increased activity and heightened volatility since the beginning of 2021 corresponding to the GameStop trading frenzy. Likewise, the news sentiment surrounding the other relevant tickers displays similar patterns, as shown by Fig. **??**.

1.2 Wavelet Coherence

Applying the wavelet framework to analyse the effect of Reddit investors' sentiment during the GameStop retail trading frenzy, we discover interesting dynamics. Figure [6] shows wavelet coherence plots between the scraped sentiment and intraday 30-min returns, historical and implied volatility, and lastly volume.[7]. The colour scale on the right side corresponds to the level of wavelet coherence where warmer (colder) signify higher (lower) coherence. The arrows represent the relative phase relationship between the two series

[6] numref

[7] figure note that some series are sampled at different frequencies

Figure 2: Word Frequency

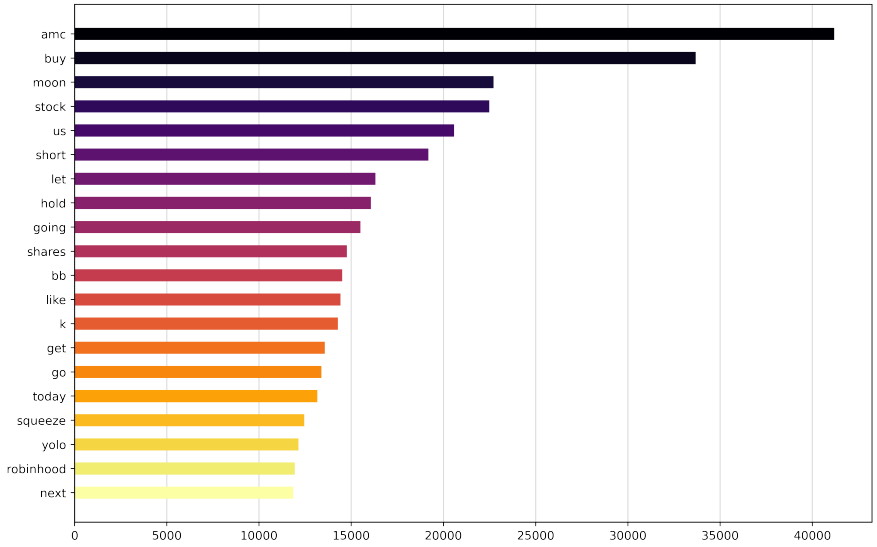


Figure 3: Bigrams Frequency

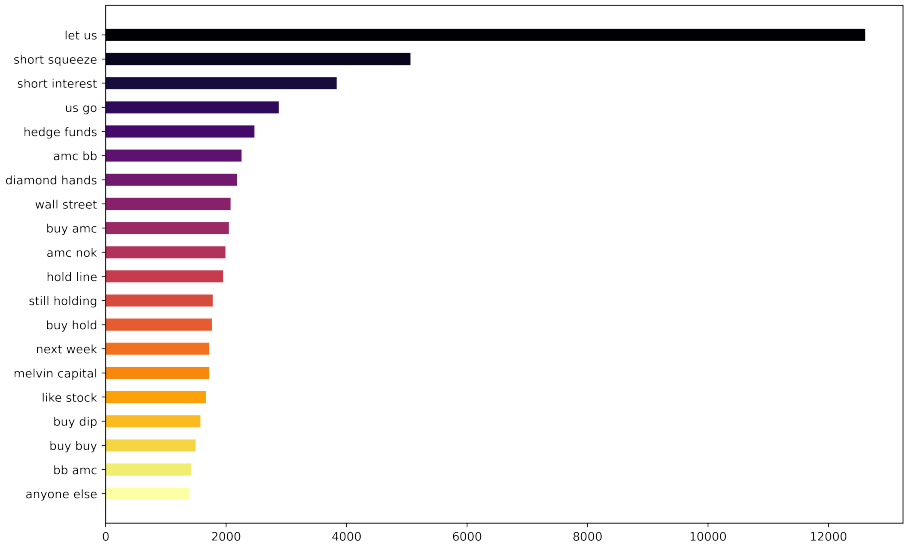
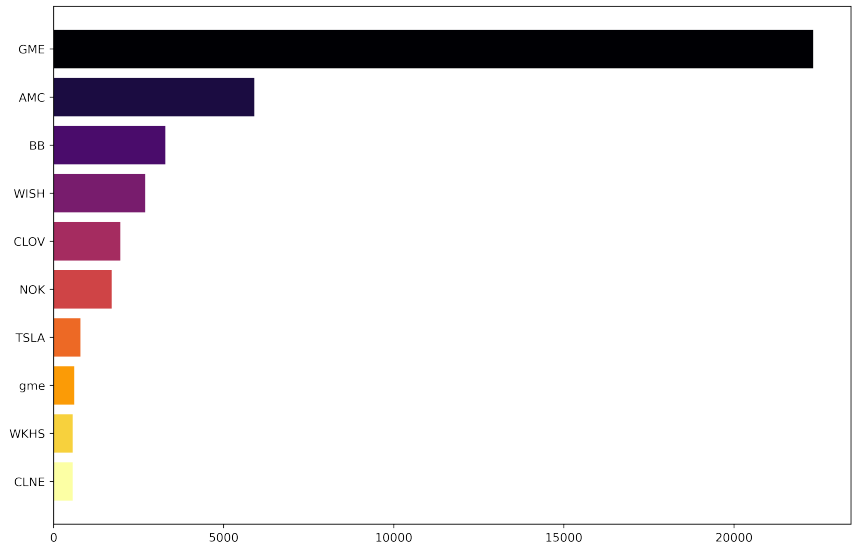


Figure 4: Most Mentioned Tickers



which can be indicative of causality. The arrow notation follows Torrence and Campo's (1998) [1]. In-phase (anti-phase) behaviour is denoted as ↑ (↓). In-phase and anti-phase relationships can be alternatively described as positive and negative correlation. The phase difference is then indicated by the horizontal plane of the arrows' notation where right → (left ←) pointing arrow suggest that first (second) series leads the second (first) by 90°. For instance, in Panel A ↗ implies that sentiment and GME returns are in-phase with the former leading the latter by 45°, i.e. sentiment positively influences GME returns. Horizontal axis corresponds to the time dimension while vertical axis shows the frequency, respectively period, dimension. The scale (period) ranges from 2 up to 128 intervals which depend on the sampling frequency, i.e. 1 hour to 64 hours[8] or 1 day to 128 days. White contours designate areas where wavelet coherence is strongly statistically significant at 5% level given by Monte Carlo simulations as described above. Finally, cone of influence is given by the cross-hatched region wherein the edge effects can be effectively ignored.

Panel A shows number of short-lived areas of significant coherence at higher frequencies between the sentiment and GME return series, which can however, occur by chance. More telling are the large areas of coherence between the 8-128 scales the timing of which corresponds to the most active periods in the subreddit. The two substantial periods of significant coherence occurring in Jan-Mar and May-Jun periods show two series being

[8] remember 30-min intervals

largely in phase with sentiment leading, suggesting that in scales from 8 hours up to 64 hours, corresponding to approx. 1 to 8 trading days, sentiment positively influenced GME returns. Further, beginning August there are also larger however shorter-lived areas of significant coherence at high frequencies, notably in the 8-32 frequency bands. Although, we observe a phase switch as in these time-scale regions both series are interestingly anti-phase with no distinct phase difference. That is, increased activity larger sentiment had no effect on the stock's returns and moreover, larger sentiment was simply met with lower returns which is in contrast to previous periods of heightened activity on the subreddit.

Sentiment and GME returns series shows number of significant coherence areas at higher frequencies which are however, very short in duration. Large areas of coherence on the 8-128 scales can be found during the most active periods in the GameStop saga namely, Jan-Mar and Aug-Oct. We can observe that the the until July the series were largely in phase with sentiment largely leading, suggesting that retail investors' sentiment in these periods were positively influencing GME returns. Perhaps even more interesting finding, further highlighting the usefulness of the utilised wavelets framework, is the changing relationship in the later part of the year. The significant periods of coherence from August onwards show two series out-of-phase with a mixed directional impact.

Contrary to our expectations that sentiment is better predictor of volatility rather than series [9], we observe fewer and smaller areas of significant wavelet coherence between the sentiment and volatility series. Again, during the most active periods larger areas of coherence are identified in the 2-32 hour frequency bands[10], where the two series are mostly in-phase with ambiguous directional relationship. Notably, in the later half of the year there are periods where sentiment and volatility are out of phase with the latter leading.

We also employed daily data on traded volume and implied volatility [11]. IMVOL PLOT shows large areas of coherence even at higher frequencies, where the time-periods coincide with the large volatility of GME. Moreover, the phase arrows indicate that sentiment is positively influenced by volatility, sustaining the argument that greater volatility simply spurs more discussion and possibly feeds the growing sentiment on the subreddit. At lower frequencies in the 16-32 days band there is consistent large coherence between the two series up until August when the relationship seemingly disappears.

Similar results are shown by the volume plot, comparing the trading activity and the activity on the subreddit, respectively sentiment surrounding it. Large areas of coherence, which however lasted only shortly, can be found in the 2-4 days bands. Longer significant

[9] atkins et al paper

[10] check

[11] check period

Figure 5: PANEL A

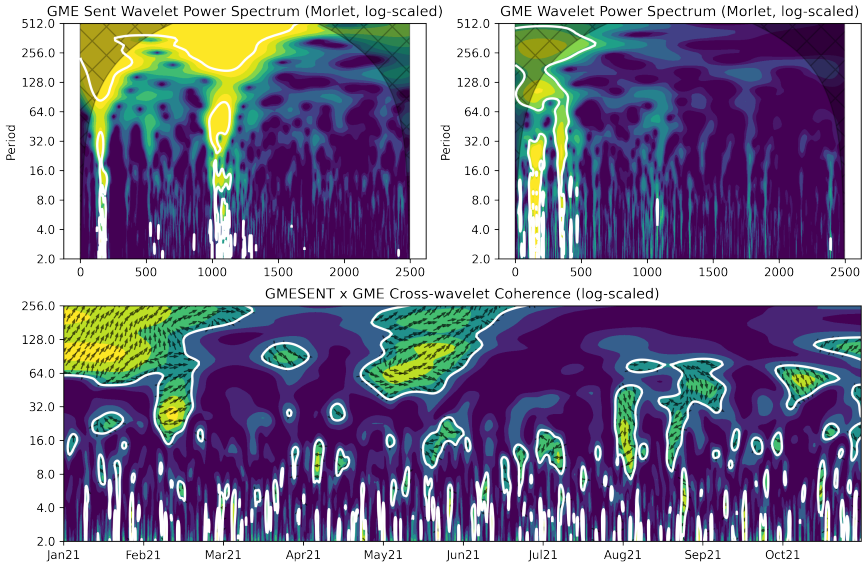
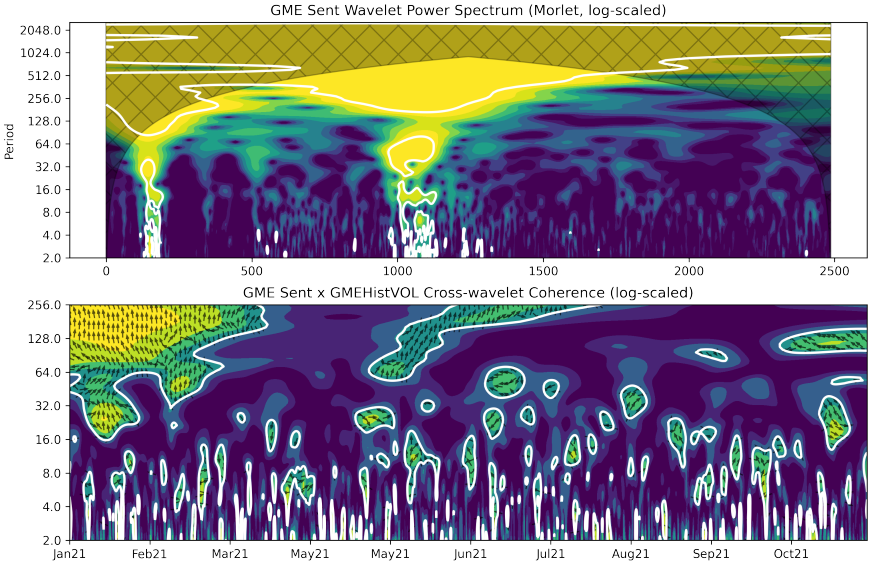


Figure 6: Panel B



time-scale periods can be found up from the 16-days band. The phase arrows indicate largely in-phase relationship with the volume series predominantly leading, except for the June/July period where both series presumably do not necessarily influence each other given the near-zero phase. These results suggest that simply heightened activity on the market led to higher activity on the subreddit by fuelling the sentiment.

News Sentiment

The news dataset compared to the Reddit scraped sentiment extends back to December 2019 offers unique findings in addition. Similarly, there are large time-scale areas across all frequencies with significant wavelet coherence, especially from September 2020 onwards. What is more interesting, is the changing phase relationship at the turn of 2021, as news sentiment begins to take over and lead the both series. Overall, in both cases, before and after the structural break, news sentiment and GME returns are largely in-phase in the relevant frequency bands (2-64 days).

In addition for a more direct comparison we sample the Reddit dataset on a daily basis and re-run the wavelet analysis, the results of which are included in plot [12]. The large areas of significant coherence do not change and correspond to the higher frequency dataset we used initially. The phase arrows however, indicate an out-of-phase relationship at higher frequencies with GME returns largely leading the series. Thus, confirming our suspicion that GME returns simply reinforce and drive the sentiment on the subreddit. Although, we also find a possible difference in the news and social media sentiment, as news sentiment seems to be positively correlated with GME returns while social media sentiment exhibits out-of-phase relationship.

Lastly, the coherence plot between the two datasets show large coherence until August even at higher frequencies up to 16-days. The news sentiment positively influences the Reddit sentiment apart from the period at the beginning of the year, where the two are out-of-phase.

1.3 Comovements

The predominantly warm coloured areas in PANEL A show extremely high levels of coherence. The comovement between GME and AMC returns is by far the largest out of the selected sample. The wavelet coherence occurs consistently throughout the year across all frequencies moreover, coherence becomes arguably stronger from September onwards. As expected considering the results of textual analysis, both series are largely in-phase however, with unclear directional relationship.

[12] numberref

The other tickers do not display the same level of comovement with GME as AMC. Nevertheless, there are large significant areas of comovement even at higher frequencies which coincide with the most active periods in the GME saga. Similarly to the sentiment series, we observe large short-lived comovements in the higher frequesncy bands when retail investors' sentiment has peaked. All series largely exhibit in-phase relationship however, there is no unequivocal evidence as to whether GME directly caused/influenced the other stocks. Therefore, apart from AMC the contagion to other relevant tickers is limited to the periods when the discussion surrounding GME peaked. These periods moreover, coincide with the largest volatility and largest price spikes within the saga, suggesting that retail investors impact also spilled over to other stocks affecting their returns and volatility.

Notes

- Sentiment
  - GMESent x GME

    While showing large areas of high coherence, the areas are predominantly at lower frequencies, which might suggest that sentiment might be much more persistent compared to GME returns. Although, few notable areas of significant coherence are found at higher frequency bands, in the Jan-Feb and Sep-Oct period. The two series are, as expected, largely in phase and more importantly, the significant areas show that it is the sentiment series largely leading the return series. In other words it could be that the social media sentiment is the primer driver of the stock's returns in these periods, with both being in-phase, i.e. larger sentiment leads to larger returns.
  - GMESent x AMC (due to extremely large coherence)
  - daily GMESent x GME
  - News Sentiment x GME

    Switch of phase relationship at the beginning of 2021. Instead of GME returns leading the both series, at the turn of 2021 the news sentiment begins to take over and lead. In both cases before and after the structural break, the two series are in largely in phase at higher frequencies.
  - Twitter Sentiment x GME
- Volatility/Volume
  - GME Sent x Historical vs Implied Vol (daily)

Using the daily dataset and extending the sample to include November 2020 reveals an interesting development before the GameStop saga hit its peak in January 2021. There is a large and significant in-phase relationship between the sentiment and GME volatility at 4-16 days frequency band, the relationship peaks at the end of February where it extends across all frequencies. Moreover, phase arrows indicate that volatility series is leading, sustaining the suspicion that greater volatility simply led more discussion and further fed the growing sentiment on the subreddit.

- GME Sent x HistVol using intraday data (30min)

  Interestingly, the wavelet coherence is significantly lower compared to the return series. While there are discernible large areas of coherence, they mostly occur at lower frequencies and sporadically. The time periods of significant coherence are significantly smaller compared to the return series. At the start of the year, when GME stock showcased extremely large volatility, the coherence plot shows mostly in-phase relationship with however, ambiguous lead/lag nature.

- GME Sent x Volume?

  Small periods of strong coherence consistent with the timing of the GME saga peaks and the sentiment and return series coherence. Although, there is weaker coherence at the beginning of the year at the 2-8 days bands. Larger significant period of coherence also exhibited during May-June period however, on a more than weekly basis only.

- Comovements

  As opposed to the sentiment series, we expect larger and more consistent areas of comovement even at higher frequencies. Figure xx [13] shows the coherence plots between the individual pairs with GME. All panels exhibit again exhibit large areas of significance coherence even across high frequency bands (1-4 hours). The time periods at which comovements appear largely coincide together and with the periods where sentiment surrounding GME peaked, where we revealed significant in-phase relationship between sentiment, GME returns and volatility.

  - GME x AMC

    GME and AMC exhibit the very strong comovement across all time-scales. At lower frequencies there is relatively consistent in-phase relationship with the directional impact rather ambiguous. Apart from a small window in June,

[13] numberref

there is large and significant comovement in the 2-8 hours frequency bands, which gets notably stronger from September onwards. As expected from the relevancy of AMC as shown by the textual analysis, we find large and significant in-phase comovement with lead/lag relationship switching, presumably as both series feed off each other.

- GME x BB
- GME x CLOV
- GME x NOK
- GME x RKT
- GME x Short Interest Indices

  Very strong coherence across most frequencies during the peak periods (Jan-Mar and July-Sep). Even though, the beginning of the year volatility is largely associated with large short positions in GME and the following short-squeeze phenomena, the coherence plot indicates a out-of phase relationship with GME leading albeit only by a little degree.-
- ??? GME x S&P 500

  Finally, Panel XX [14] shows large significant comovement even before the GameStop peak in January 202. The phase relationship completely switches in January with the two series being out-of-phase. In the later half of the year we can also large scale comovement up to a month frequency.

2 Discussion

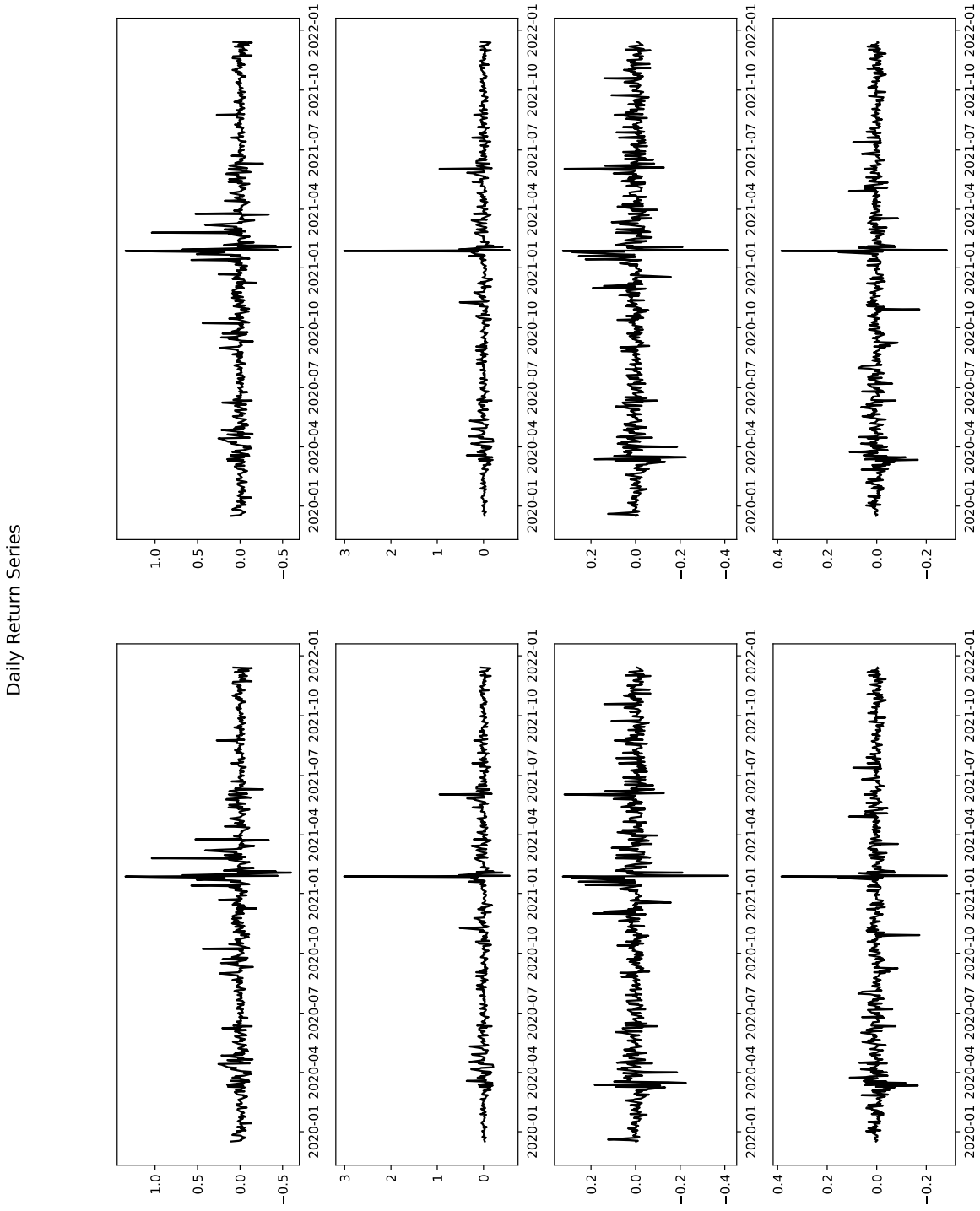
3 Conclusion

[14] numberef

References

[1] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological Society*, vol. 79, no. 1, pp. 61 – 78, 1998.

Appendix A


Daily Return Series
2020-01
2020-04
2020-07
2020-10
2021-01
2021-04
2021-07
2021-10
2022-01

Appendix B

If you need more then one section...

@misctime, author=, title=, year = , month = , howpublished=, note =