VRIJE UNIVERSITEIT VAN AMSTERDAM

Research Project

The Effects and Trading Applications of News Sentiment in Crude Oil Futures Markets

Authors:
Sander Schleeper
Maurits van dan Oever
Connor Stevens

Supervisor:
Svetlana Borovkova

Group:
9

Contact probability & partners:
Svetlana Borovkova
Sergiy Ladokhin
Ying Wu

Student numbers:
2622915
2613642
2702708

31-01-2021

1 Introduction

Market sentiment has long been a recognised phenomenon, but its promise as a tool for forecasting security prices has recently grown as a result of the democratisation of the internet, social media and the digital nature of how people interact with news coverage. Across various areas of study and practice throughout the finance industry, several developments have been made due to the advent of ‘Big Data’. Few areas of financial study are better positioned to benefit from this than market sentiment.

The seminal paper by Baker and Wurgler (2007) defines investor sentiment as beliefs about future cash flows and risks that are not justified by the facts at hand. We argue that this may be an outdated interpretation of 'market sentiment' and does not fit well with how the term in used in the industry and news coverage. Perhaps a more contemporary concept of market sentiment is to consider it a distillation of the general views of market participants in terms of the directions they expect the market to move in the future and their feelings about where it is currently. There is no doubt that while market participants rely on numerous sources of data and information to form their views, the information being fed to them via the various forms of news media play a significant role. One such example of this is the 'Cramer Bounce', which refers to a sudden overnight increase in a stock's price after being favourably mention by 'Mad Money' television show host Jim Cramer on news channel CNBC. Bolster, Trahan and Venkateswaran (2012) found that buy recommendations on 'Mad Money' saw short-lived increases in the prices of recommended stocks while sell recommendations saw far more persistent negative effects. This example highlights the influence that news coverage can have on the decisions of market participants and suggests that news sentiment may have potential as an explanatory variable in forecasting returns and volatility.

While papers like that of Schleifer and Vishny (1997) and Baker and Wurgler suggest that investors acting on sentiment are typically retail investors, we argue that this may not be entirely true today. In the time since those papers were written, the speed at which information is spread and consumed has increased significantly. We propose that the influence of news coverage and social media discussion is not limited to retail investors alone. Ultimately, it is of little consequence exactly which segments of the financial markets are responsible for any possible causal relationship. This paper's goal is to explore if any such causal relationship exists and whether or not it can be exploited with the use of exchange traded options to reliably establish profitable positions.

The notion of using sentiment metrics as explanatory variables in forecasting is not a new idea and generally takes one of two flavours; metrics based on market proxies and metrics based on alternative data such as distillations of news coverage, Google Trends and social media data. Both of these flavours of sentiment metrics have been found to have explanatory power and improve forecasting accuracy to various degrees. This paper will be focusing on news sentiment metrics. News sentiment metrics are useful and give us a good idea of whether the news coverage at a given moment is positive or negative, but falls short in that we do not know how many people were exposed to it, and hence potentially influenced by it. This paper's focus on practicality means that some potentially promising alternative data sources such as Twitter data have to be excluded on the basis of limited access to data and others on the basis of an inability to reliably obtain daily data. Having established the explanatory variable to be tested and the method to be implemented, all that remains is to choose an underlying.

The commodity futures market has seen a remarkable increase in activity in the years following 2006, often described as the 'financialisation' of the commodities markets. Market activity, as measured by average daily open interest, in crude oil futures has more than doubled since 2006 (eia.gov, 2021). Figure 1 below, illustrates the sheer scale of the increase in crude oil futures market activity.

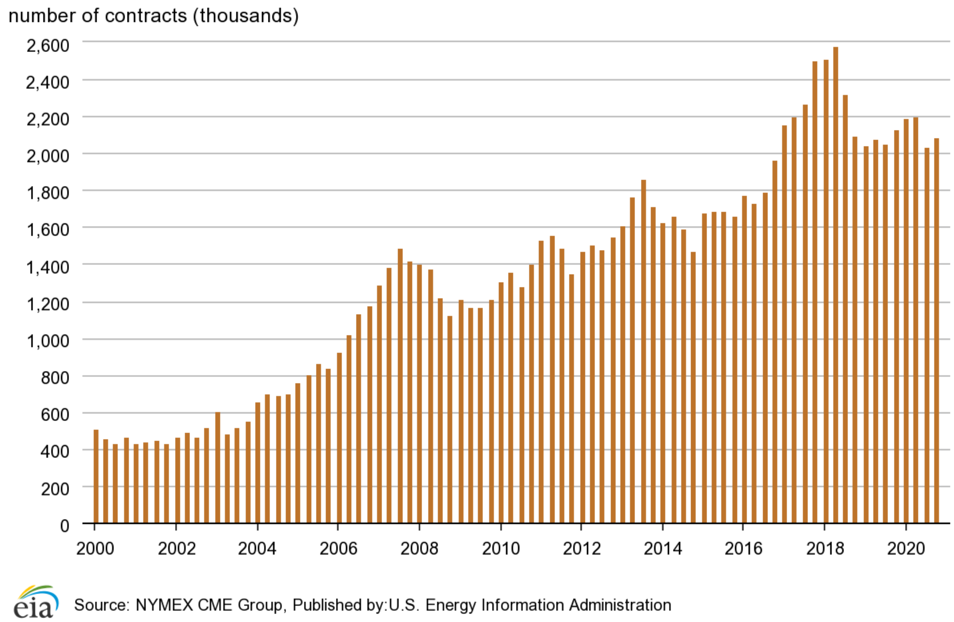


Figure 1: Daily Average Open Interest in Crude Oil Futures Market

While oil companies, storage facilities, airlines and other industry merchants remain active in the market, a large proportion of the new activity comes from financial institutions. The key difference to note here is that these two market segments' demand for these derivatives has very different drivers. While the demand of merchants and producers is rooted in consumption, the demand of financial institutions is rooted in the search for profitable investments. The problem that arises is the apparent disagreement between these two market segments. With the exception of a short period during 2008, money managers in the United States have been net-long the oil futures markets while producers and merchants have been net short, with a few exceptions. This clear and sustained disagreement between the the industry that produces the oil and the financial institutions which speculate on its' movements makes it a perfect underlying for our exploration into the use of sentiment to forecast asset prices. The fact that the market has seen such an influx of participants and a resulting increase in liquidity only supports our choice of crude oil futures as underlying. Figure 2 below shows a plot of the realised volatility of the West Texas Intermediate (WTI) futures price as well as the news sentiment metrics for crude oil. It is clear in figure two that volatility in the WTI futures market and news sentiment move somewhat together. We do not even attempt to suggest some sort of lead-lag relationship yet, but rather present this figure to the reader to indulge curiosity and highlight the potential of news sentiment as an explanatory variable.

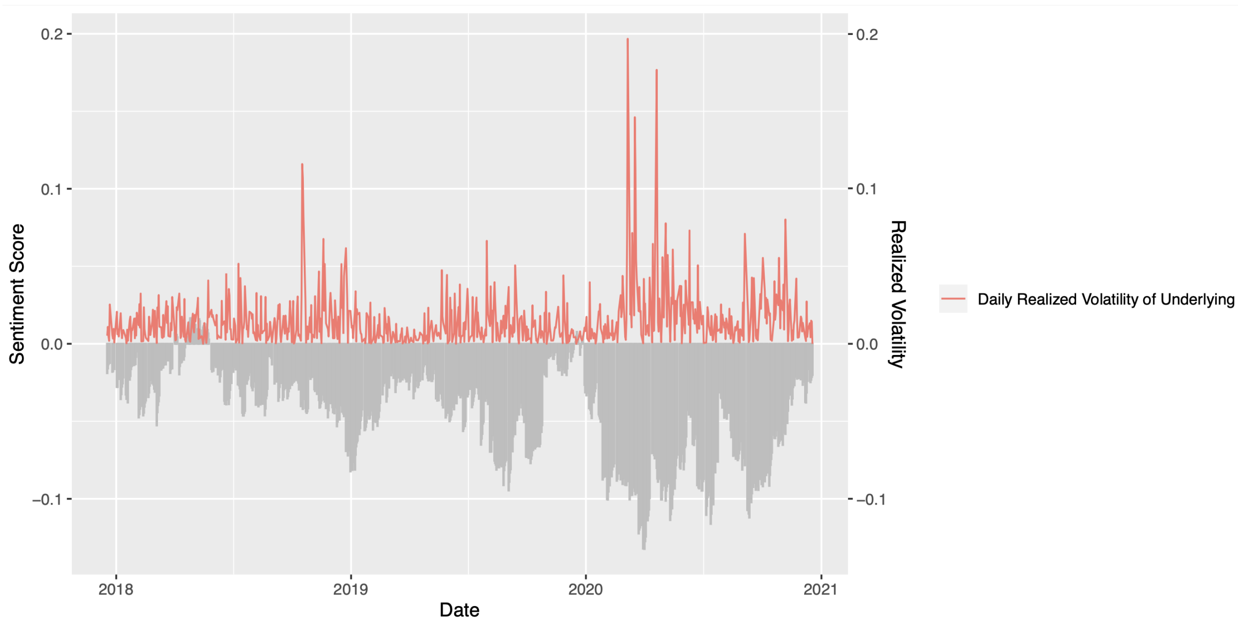


Figure 2: Daily Realised Volatility Alongside Daily News Sentiment Metrics.

2 Literature Review

The root of this paper is market sentiment and the effect it has on the investment decisions of its' participants. The earlier-mentioned Baker and Wurgler used a 'top down' approach to examine the effects of investor sentiment on individual stocks and the market in general. This paper breaks away from the definition of sentiment as stated in Delong, Shleifer, Summers and Waldmann (1990) and instead defines sentiment as optimism or pessimism about stocks in general. The authors construct their own sentiment index using market data such as share turnover, initial public offering (IPO) volume, IPO first day returns, equity share in new issues, closed-end fund discount (CEFD) and dividend premium. These market proxies for positive and negative investor views are then aggregated into a single sentiment index which the authors use in their analysis. The sentiment index constructed is specified to be an index of individual investor sentiment and the authors link this to arbitrage opportunities created by the influence of sentiment on investor decisions.The authors find that stocks that are difficult to arbitrage or value experience the most notable effects of investor sentiment.

The next step towards the topic of this paper is the connection between commodity futures and sentiment, made by Gao and Süss (2015). The authors take direct inspiration from Baker and Wurgler (2006), adapting their methodology and applying it to the commodity futures market in the wake of the 'financialisation' of the commodities market in the years following 2006. Again, a sentiment index is constructed along the lines of Baker and Wurgler, but curiously the authors explicitly use equity market proxies for their market sentiment index. Option implied volatility, option implied skewness, CEFD, first day returns of IPOs, NYSE share turnover and dividend premium is used in composing their sentiment index. The authors justify their decision to use equity market proxies with the liquidity of the equity market and its tendency to reflect market sentiment in general as well as the ability to generate daily sentiment data, which was not possible using the some of the proxies of Baker and Wurgler. This paper is the first of many to find the presence of an asymmetric sentiment effect in commodity futures; a larger decrease in commodity returns following a decrease in sentiment than increase in commodity returns following an equal increase in sentiment. Rather importantly, the authors discover a strong impact of investor sentiment on commodity prices and find investor sentiment has predictive power for commodity futures returns using daily time-steps, but they do note that energy futures are the exception to this finding. However, Deeney et al. (2015) found that an oil-industry specific sentiment metric had explanatory power in the crude oil markets , using a similar approach to Gao and Süss.

Borovkova (2011) attempts to measure sentiment more directly by way of a distillation of the aggregated news data product offered by Thomson Reuters News Analytics Engine. The service produces probability values indicating the likelihood of news items being either positive, negative or neutral about a given topic, stock or industry. The author then distills this data into a single news sentiment value for each day for crude oil and uses it to test how crude oil futures markets respond to positive and negative sentiment in the news. Although the sentiment metric used here is totally different to that used in the previously discussed papers, Borovkova also finds an asymmetric sentiment effect. Negative sentiment shocks resulted in larger negative returns than the positive returns resulting from positive sentiment shocks. These asymmetric effects were pronounced in futures with nearby maturities and for markets in a state of contango. On the topic of the forward curve, negative sentiment shocks saw increases in the slope in both backwardation and contango markets with contango deepening and backwardation flattening. Borovkova and Mahakena (2015) conduct a similar analysis on natural gas futures and additionally examine the forecasting ability of both generalized autoregressive conditional heteroskedasticity (GARCH) and high-frequency-based volatility (HEAVY) models when news sentiment data is included in the model. They find that HEAVY models outperform GARCH models significantly in improving predictive power when including the news sentiment data and that GARCH models with news sentiment performed at least as well as those without. Again, the asymmetric sentiment effect was observed and the natural gas futures markets exhibited very similar responses to negative and positive sentiment shocks with more persistent negative responses to negative shocks and smaller, mean reverting responses to positive sentiment shocks.

In line with this paper's goal of determining the whether sentiment data can be used to profitably trade oil futures options, we endeavour to forecast both futures returns and, perhaps more realistically, volatility as a source of information from which to establish option positions. Our exploration of volatility models begins with a paper by Kang, Kang and Yoon (2009) who compared the ability of GARCH, component-GARCH (CGARCH), and fractionally integrated GARCH (FIGARCH) to model the persistence in the volatility of crude oil prices. The authors then examine which model best forecasts oil price volatility and finds that for Brent and Dubai, FIGARCH performed best while CGARCH performed best for WTI volatility forecasting. This is of particular importance for this paper since our focus is on WTI oil futures and the options that trade on them. Making the connection between volatility forecasting of asset returns and sentiment is not a novel idea and sentiment has been successfully integrated in some form or another into volatility forecasting models by several authors. Kumari and Mahakud (2015) used sentiment in the context of GARCH and vector autogregressive-GARCH (VAR-GARCH) and found that investor sentiment had a significant effect on stock market volatility in India. The authors also observed the asymmetric response of the markets to negative sentiment, as was seen in the commodity markets. Similar findings are made using sentiment in the context of GARCH, threshold GARCH (TGARCH) and exponential GARCH (EGARCH) by Uygur and Taş (2012) across a variety of the major market indices across the globe.

Based on the papers that have come before, some clear trends and repeated results occur. First, investor sentiment, measured by both market proxies and news metrics shows explanatory power to varying degrees in some form or another. Second, there is a clear asymmetry in how the market responds to positive or negative news. Both sentiment metrics saw larger responses to negative sentiment than to positive sentiment, which is not surprising based on what we know about human attitudes towards risk. Prospect Theory of Kahneman and Tversky (1979) tells us that human beings are risk averse and will choose risk aversion over possible gains. We believe that this explains the asymmetric response to positive and negative sentiment days seen across the financial markets and it is precisely this behaviour we hope to exploit.

3 Data

3.1 WTI Futures Prices

WTI crude front month futures prices for contracts trading on The New York Mercantile Exchange (NYMEX) were obtained for our period of 01-01-2010 to 31-12-2020 and standard returns were calculated. We believe that ten years of data is more than sufficient for our purposes and starting the sample further back would only see results be less representative of contemporary market behaviour due to the digital nature of news proliferation and consumption. We chose 2010 as a starting year to avoid starting in the middle of the financial crisis and subsequent global recession. In order to test our forecasting methods, we define an in-sample and out-of-sample period with in sample beginning at the start of our data set and ending at 01-04-2106 from where the data is considered out-of sample. Descriptive statistics are shown below. In general, figures are as expected with returns centering around zero with the exceptions coming from the period of negative prices in middle of 2020. These negative prices result in the very large skewness and kurtosis figures for returns seen below.

	Prices	**Returns**
Mean	69.48212	-0.0007995236
Q1	49.4800	-0.0108637724
Median	65.545	0.0003061036
Q3	92.4575	0.0115091035
Std Dev	23.16876	0.05358862
Min	-13.1000	-1.7229580574
Max	113.9300	0.5706401766
Skewness	0.06436861	-22.85409
Kurtosis	-1.216238	745.6663

Table 1: WTI Futures Descriptive Statistics

3.2 Option Data

Exchange-traded WTI crude futures options data is used for a variety of maturities between 5 and 15 days, where slightly differing maturities are used to fill gaps in data and ensure positions can be set up for a given strike. Options data for the same 2010 to 2020 period was used. Where implied volatility was missing it was calculated by interpolation, but this was done by the data provider and we are not sure of the exact method used. There is some missing data which has also been excluded from the calculations. All other expected option data was provided, such as bid and ask prices, open interest, implied volatility and the Greeks. Again, some missing data was present and these observations were removed from the data set. Missing data meant that some options did not have ask prices for certain days. In these cases, settlement was taken as option price if a settlement price existed where an ask was missing.

3.3 News Sentiment Data

Our news sentiment data is a distillation of the Thomson Reuters News Analytics data. The data gives probabilities of each day being a positive, negative or neutral sentiment day. Our distillation of these metrics is the difference between the probability of a given day having a negative news sentiment day and the probability of it being a positive sentiment day. The our data uses the same 2010 to 2020 sample.

$$SentimentMetric = P(+sentiment)_t - P(-sentiment)_t \tag{1}$$

The nature of how our sentiment metric is calculated means that we are not surprised to see that is is mostly negative. This does, however, mean that when days are assigned a positive sentiment metric we are fairly certain they are positive sentiment days with little ambiguity.

	Sentiment Metric
Mean	-0.0404
Q1	0.0619
Median	-0.0424
Q3	-0.0213
Std Dev	0.0478
Min	-0.137
Max	0.0576
Skewness	0.1772
Kurtosis	0.1678

Table 2: Sentiment Metric Descriptive Statistics

4 Methodology

4.1 Event Study

We carry out an event study in order to show a practical basis from which we extend our exploration into the explanatory power of sentiment metrics. This event study is presented as a simple proof of concept and makes clear that our results are not simply data mining.

An event is defined in two ways, either in the form of a very high or low level of our sentiment metric or a very large increase or decrease in our sentiment metric between days. To this end, we split our sentiment data into sentiment level and change in sentiment. Those sub-samples are then further divided into positive sentiment levels and negative sentiment levels as well as increases in sentiment and decreases in sentiment. For each of the positive value sub-samples we calculate the top 20% of values and the bottom 20% of values for negative sub-samples.

For each day in our newly constructed sentiment shock sub-samples, we calculate the average daily returns and cumulative average daily returns for the 20 days before and after the predefined events, where 20 days were chosen based on the findings of Borovkova and Mahakena (. Where a sentiment shock takes place near the start or end of the sample, we include the maximum possible number of days in the calculation while still staying within our 01-01-2010 to 31-12-2020 sample. We then analyse the results of our event study in order to determine the best possible rule based strategy to implement for both our rule-based trading strategy and our directional return bets.

4.1.1 Rule-Based Trading

Based on the results from the event study a rule based trading strategy is adopted. First the trading strategy involves creating trading signals based on the upper and lower quantiles of sentiment levels and changes in sentiment. The quantiles are not taken over the whole sample period but rather over an expanding window, in keeping with our focus on practicality. If today's sentiment was either larger than the upper quantile or smaller than the lower quantile of all sentiment observations, a trading signal of 1 or -1 is created respectively indicating high or low sentiment. The same process is carried out on the changes in sentiment data. When the sentiment is in neither of the quantiles a 0 signal is generated. For this study we examine the lower 30%, 20%, 10%, 5% and upper 70%, 80%, 90%, 95% quantiles. Although we examined only the upper 20% and lower 80% in the event study, we expand the scope here to test the optimal threshold. The signal generation is summarized in equation 2 below. In total 4 time series of signals are produced, with the quantile pairs being 70% & 30%, 80% & 20%, 90% & 10% and 95% & 5% for both the positive and negative levels of sentiment and the positive and negative changes in sentiment.

$$signal_{1,t} \begin{cases} 1, & \text{if } sentiment_t > 70\%, 80\%, 90\%, 95\% \; quantile \\ -1, & \text{if } sentiment_t < 30\%, 20\%, 10\%, 5\% \; quantile \\ 0, & otherwise \end{cases} \tag{2}$$

With the signals created we take up options positions based on high or low sentiment (or large changes in sentiment) and the outcomes from the event study. Finally the profits and losses are calculated according to the payoff profile of the option at maturity and the price of the option when the positions are initiated. Finally the cumulative profit and losses are shown in a graphical representation and interpreted.

4.2 Returns modelling

4.2.1 ARMAX model

The returns of the crude oil futures is modelled using an Auto-Regressive Moving Average and Exogenous (ARMAX) framework, an extension of the well known Auto-Regressive Moving Average (ARMA) model. Introduced by Whittle (1951), it incorporates lag polynomials as a tool to understand and predict time series. One lag polynomial serves the auto-regressive part of the equation, while the other refers to the moving average part, or a linear combination of contemporary and lagged error terms. The ARMAX framework extends on the ARMA framework by incorporating exogenous regressors to the standard ARMA equation. The general ARMAX equation is shown in equation 3.

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{m=1}^{M} \beta_m X_{m,t} \tag{3}$$

In this case, the exogenous regressors are to be the variables S and ΔS, denoting the level of sentiment and the changes in sentiment respectively. Another thing to note is that the error terms ϵ_t are assumed to be student-t distributed, to take into account fat tails of returns, which accommodates our large, negative returns in 2020. The exogenous variables are split into positive and negative observations using indicator functions, so as to better understand the asymmetric effects of the level and changes of the sentiment metric. This also allows us to capture the asymmetric responses of the market to positive and negative sentiment. In total, three models will be compared to each other, a general ARMA model, an ARMAX that incorporates sentiment, and an ARMAX that incorporates changes in sentiment. Sentiment and changes in sentiment is are separated into two different models due to possible (and likely) collinearity between the two variables.

4.2.2 ARMAX specification and estimation

Specifying the correct amount of lags for both the ARMA and exogenous components is necessary to prevent collinearity between lags of the same variable. When this happens, the lags will individually be estimated to have an insignificant effect on the dependent variable. Considered collectively however, the lags together do have a significant effect. The lag order of the ARMAX is determined by calculating Akaike's Information Criterion (AIC). This metric penalizes the inclusion of more parameters, but this penalty can be compensated by the explanatory power they have.

Estimation is done through two methods. First, a conditional sum-of-squares is performed by changing the parameters $\mathbf{\Phi} = (\phi_0, ..., \phi_p, \theta_1, ..., \theta_q, \beta_1, ..., \beta_M)$, to initialize the values of the parameters. Secondly, a maximum likelihood approach is employed to find the final estimates. This is done by maximizing the sum of individual conditional log likelihoods over the in-sample period.

4.3 Volatility modelling

We model the volatility of the underlying WTI front month futures using both GARCH and GARCHX as well as CGARCH and CGARCHX, where the exogenous variables to be included will be the positive and negative sub-samples of both sentiment level and change in sentiment. This comparison is carried out in order to determine if variations of our news sentiment metric can indeed increase the forecasting performance in comparison to 'vanilla' volatility models. GARCH and CGARCH are used as benchmarks for their sentiment-inclusive variants.

4.3.1 GARCH Model

Volatiltiy of WTI futures returns are first modelled using the classic GARCH model, as introduced by Engle (1982). Borovkova and Mahakena showed that GARCHX with news sentiment data performed at least as well as GARCH without news sentiment data and we aim to examine if this is still the case in the five years since that paper has been published.

$$r_t = \mu + \epsilon_t \qquad \epsilon_t \sim T(0, \sigma_t^2, \nu) \tag{4}$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{5}$$

We then conduct fit the following two GARCHX models to include the sentiment level .

$$r_t = \mu + \epsilon_t \qquad \epsilon_t \sim T(0, \sigma_t^2, v) \tag{6}$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \upsilon_1 S_t \mathbb{1}_{\{S_t>0\}} + \upsilon_2 S_t \mathbb{1}_{\{S_t<0\}} \tag{7}$$

As well as the change in the sentiment level.

$$r_t = \mu + \epsilon_t \qquad \epsilon_t \sim T(0, \sigma_t^2, v) \tag{8}$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \upsilon_1 \Delta S_t \mathbb{1}_{\{\Delta S_t>0\}} + \upsilon_2 \Delta S_t \mathbb{1}_{\{\Delta S_t<0\}} \tag{9}$$

4.3.2 CGARCH model

Next, the volatility of the crude oil futures returns is modelled using the CGARCH model, a variation of GARCH. The article of Kang, Kang and Yoon (2009) has shown evidence of the CGARCH outperforming the GARCH, FIGARCH and IGARCH models in terms of MSE and MAE in the out of sample predictions for the WTI crude oil market, confirming that the CGARCH model is best fitted for WTI crude oil.

The CGARCH model as described in Abu and Siriopoulos (2008) is different from the regular GARCH model by incorporating a time varying mean reversion level for the variance instead of a static mean to the unconditional variance. We first examine the standard CGARCH and then add contemporaneous sentiment level and changes in sentiment. The CGARCH model is defined as:

$$r_t = \mu + \epsilon_t \qquad \epsilon_t \sim T(0, \sigma_t^2, v) \tag{10}$$
$$\sigma_t^2 = q_t + \alpha(\epsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) \tag{11}$$
$$q_t = \omega + \rho q_{t-1} + \gamma(\epsilon_{t-1}^2 - \sigma_{t-1}^2) \tag{12}$$

Next, we add contemporaneous sentiment level:

$$r_t = \mu + \epsilon_t \qquad \epsilon_t \sim T(0, \sigma_t^2, v) \tag{13}$$
$$\sigma_t^2 = q_t + \alpha(\epsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) + \upsilon_1 S_t \mathbb{1}_{\{S_t>0\}} + \upsilon_2 S_t \mathbb{1}_{\{S_t<0\}} \tag{14}$$
$$q_t = \omega + \rho q_{t-1} + \gamma(\epsilon_{t-1}^2 - \sigma_{t-1}^2) \tag{15}$$

Finally, we examine CGARCH with contemporaneous changes in sentiment included in the variance equation.

$$r_t = \mu + \epsilon_t \qquad \epsilon_t \sim T(0, \sigma_t^2, v) \tag{16}$$
$$\sigma_t^2 = q_t + \alpha(\epsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) + \upsilon_1 \Delta S_t \mathbb{1}_{\{\Delta S_t>0\}} + \upsilon_2 \Delta S_t \mathbb{1}_{\{\Delta S_t<0\}} \tag{17}$$
$$q_t = \omega + \rho q_{t-1} + \gamma(\epsilon_{t-1}^2 - \sigma_{t-1}^2) \tag{18}$$

In which r_t is the daily return of the WTI Crude Oil futures price and ϵ_t the daily innovations. The variable q_t is the time varying long run variance mainly driven by the term $(\epsilon_{t-1}^2 - \sigma_{t-1}^2)$. Both the sentiment level and changes in sentiment are separated into positive and negative by way of an indicator function. Additionally, the transitory short term component is defined by the term $(\sigma_{t-1}^2 - q_{t-1})$. Due to recent findings as confirmed by Khaki Gunay (2017) about non-normality and fat tallness of energy futures returns, we have chosen to assume the student-t distribution for the innovations in the returns and expect to find enhanced results.

4.3.3 GARCH and CGARCH order specification and estimation

It is essential to correctly specify the lag order of the variance component of the GARCH and CGARCH models. Therefore we have chosen to select the lag order by use of the Akaike information criteria where a balanced trade off is made between likelihood explanatory power and the amount of parameters and added estimation error of the model.

Estimation of the GARCH and CGARCH models is performed by maximum likelihood estimation, using the prediction error decomposition the total log likelihood function is obtained by summing all individual conditional log likelihoods over all in sample time points. Next the total log likelihood is maximized with respect to the parameters $\boldsymbol{\theta} = (\mu, v, \alpha, \beta, \phi_1, \phi_2, w, \rho, \gamma)$ using the (SOLNP) solver in R. The total conditional log likelihood is defined as:

$$L(\theta; F_T) = \sum_{t=1}^{T} log \frac{\Gamma((v+1)/2)}{\Gamma(v/2)} - \frac{1}{2} log(\pi v \sigma_t^2) - \frac{1}{2}(v+1) log(1 + \frac{(r_t - \mu)^2}{v \sigma_t^2}) \tag{19}$$

Where, F_T is all information available until $t = T$.

4.4 Out of sample testing

After the parameters have been estimated on the in sample period, a subsequent one day ahead forecast will be produced for each out of sample time point t. The estimated parameters are refitted after each forecast based on a rolling window with length equal to the in sample period such that the model incorporates the most up to date information. Subsequently, the mean absolute error (MAE) and mean squared error (MSE) as given by equation 20 and 21 are calculated for each out of sample time point and to evaluate significant improvements in MAE and MSE between the models we perform a Diebold Mariano test.

$$MAE = \frac{1}{N} \sum_{t=1}^{N} |\sigma_t^2 - \hat{\sigma}_t^2| \tag{20}$$

$$MSE = \frac{1}{N} \sum_{t=1}^{N} (\sigma_t^2 - \hat{\sigma}_t^2)^2 \tag{21}$$

4.5 Trading strategies

Using the results from the event study as well as the forecasts of both the returns and conditional volatility, different options and option packages positions are taken based on trading signals. The positions are long or short ATM put or call WTI crude oil futures options with the price equal to the prevailing observed market settlement price of the day. All positions time to maturity is equal to 15 trading days (or less depending on availability data availability) and are held till maturity. Subsequently, a profit or loss is calculated by taking the payoff profile of the option at maturity plus or minus the bought or sold settlement price depending on a long or short position. There are gaps in our data where there is no option settlement price available, in such case we do not take any position and our profit or loss is equal to zero.

4.5.1 Directional bet

The directional bet is determined by the forecasted return. If the forecasted return is in the top 20, 10, and 5% in terms of absolute value, an option position is taken based on the forecast. The option positions to be taken will be dependent on the outcome of the event study. If the absolute value of the forecast is within either the top or bottom 20, 10 or 5% , an option position will be taken. These contracts are held to maturity, and are chosen in such a way that maturity is 15 days or less. This threshold is chosen due to availability of settlement and bid/ask data and recommendations from existing literature.

4.5.2 Volatility bet

The volatility bets are made based on the shock in the volatility forecast compared to present realised volatility, as formally shown in equation 22. With the shocks defined, the following buy and sell signals are created as in equation 23. Here 1 indicates to long one straddle, -1 indicates to short one straddle and 0 represent no position taken for trading day t. The straddle positions are composed of one long or short

ATM call option and one long or short ATM put option with the same strike. Additionally, the position is closed out after one trading day by selling the options in the market and the profit or loss on the position is equal to the price difference between the day the straddle package is bought and the price one day later. Although short straddles are far riskier than a butterfly, they are much cheaper and accommodate our expected smaller profit margins. Given that we unwind the position after one day, the risk associated is smaller and we consider this trade-off acceptable.

$$shock_t = \hat{\sigma}_{t+1} - \sigma_t \tag{22}$$

$$signal_t \begin{cases} 1, & \text{if } shock_t > magnitude \\ -1, & \text{if } shock_t < -magnitude \\ 0, & otherwise \end{cases} \tag{23}$$

5 Results

5.1 Event study results

In theory, we believe that a particularly strong sentiment measurement for a given day will result in some sort of market reaction and it is the nature of this reaction we aim to discover and subsequently exploit. Based on prior literature results and behavioural finance theories, we expect to see a persistent drop in cumulative returns following a negative sentiment shock as market participants react in accordance with the findings of Tversky and Kahneman. Similarly, we expect a increase in cumulative returns leading up to a positive sentiment shock followed by a swift correction. Following the findings of Tversky and Kahneman, separating sentiment shocks into positive and negative is logical since we know that investors do not respond symmetrically to good and bad news.

Below is a table of descriptive statistics for the variables examined in the event study. The event study uses four permutations of the news sentiment metric and the descriptive statistics of each are shown for comparison. Going forward in the event study, the focus will be on the graphs of cumulative returns with commentary on the average daily returns provided where necessary.

	Sentiment Metric	**+Sentiment**	**−Sentiment**	**+ΔSentiment**	**−ΔSentiment**
Mean	-0.0433	0.0175	-0.0504	-0.0377	-0.0033
Q1	-0.0649	0.0081	0.0051	0.0011	-0.0044
Median	-0.0444	0.00147	-0.0479	0.00245	-0.0027
Q3	-0.023	0.0261	-0.0298	0.0061	-0.0013
Std Dev	0.0326	0.0123	0.0263	0.0064	0.0033
Min	-0.137	0.00007	-0.1370	-0.000	-0.0447
Max	0.0576	0.0576	-0.0001	0.0426	-0.0000
Skewness	0.1519	0.6799837	-0.4788	2.1071	-3.9547
Kurtosis	0.0546	-0.1003	-0.1057	4.5641	29.2833
N	2764	288	2476	1091	1672

Table 3: Sentiment Metric Sub-Sample Descriptive Statistics

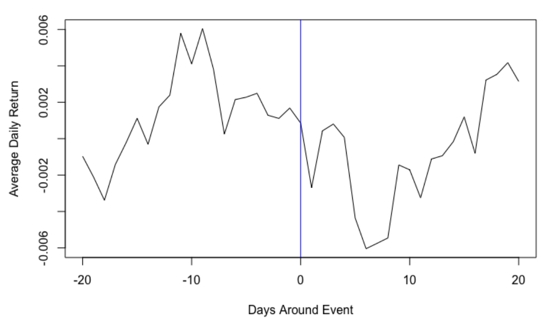
5.1.1 Positive Sentiment Level Shocks

Figure 3 graphs (a) and (b) show the event study for the upper 20% of sentiment level days. The average daily returns in graph (a) are not incredibly informative other than to suggest information leakage and the remedying of an apparent over-correction.

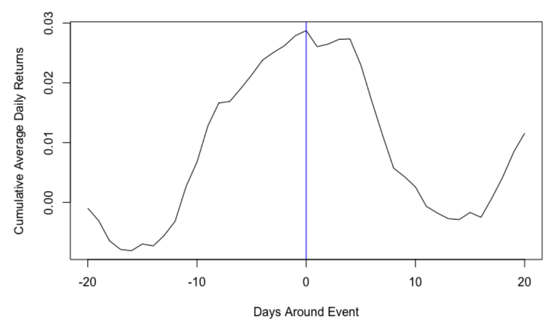
Looking at graph (b) we see a very clear peak at the positive sentiment event with two troughs on either side. Again, the apparent information leakage is seen with returns rising from about 15 days prior to the sentiment event and a subsequent correction in about 15 days following the event. Following the correction, cumulative returns start to climb again, perhaps as a result of an over-correction, but this is just speculation. The cumulative returns graph suggests that a long, at-the-money put option purchased at the end of the day of the event is a good choice for a profitable position, but the maximum possible average gain of around three percent is less than ideal and is reduced further by the cost of the options themselves.

5.1.2 Negative Sentiment Level Shocks

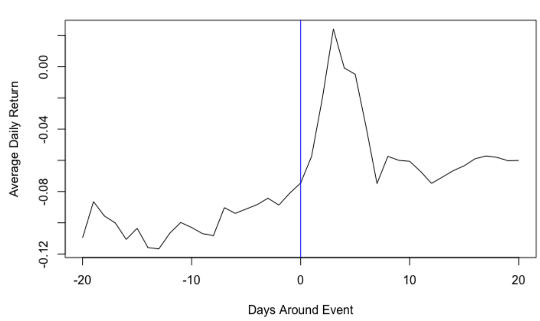
Graphs (c) and (d) show the average daily returns and cumulative returns respectively. The cumulative returns for the negative sentiment shock are very interesting. Again we see severe information leakage and steeply negative cumulative returns. The slight plateau just following the event is peculiar and we cannot be sure as to why it occurs. Perhaps the most interesting feature of this graph is the scale; we see cumulative returns starting at around zero and running down to as low as -300%. While this graph suggest that long, at-the-money puts could be incredibly profitable for high negative sentiment events, we know that this comes as a result of the period of negative WTI futures prices during 2020. This same graph with 2020 removed (available in appendix) is largely of the same shape and shows the same degree of information leakage, but the scale is far more modest, with less than 20% decrease in cumulative returns following the sentiment shock. This is, nevertheless, a significant profit opportunity for long, at-the-money puts.



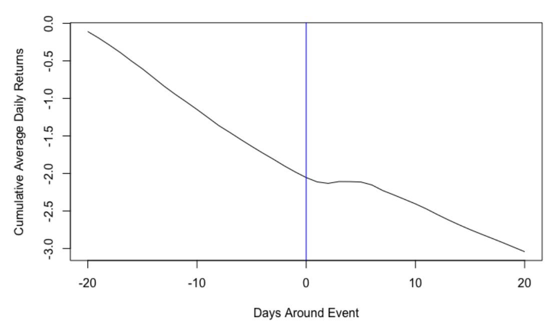
(a) Ave daily returns positive sentiment shocks



(b) Cum daily returns positive sentiment shocks



(c) Ave daily returns negative sentiment shocks



(d) Cum daily returns negative sentiment shocks

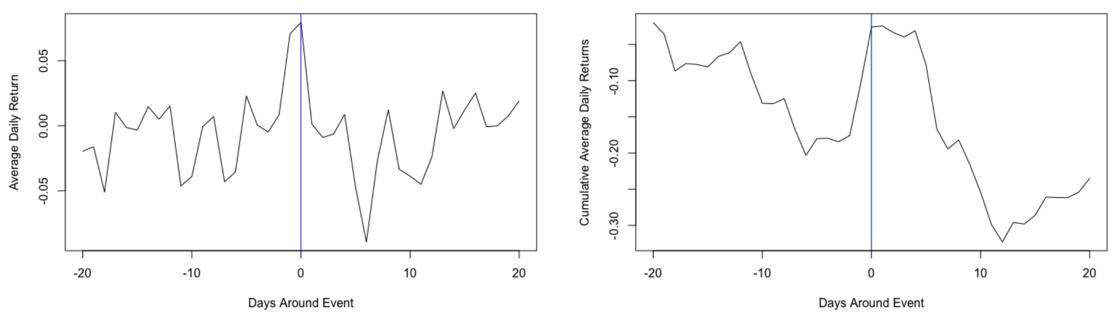
Figure 3: Average daily and cumulative average daily returns for top 20% of positive and negative sentiment shocks.

5.1.3 Positive Change in Sentiment Shocks

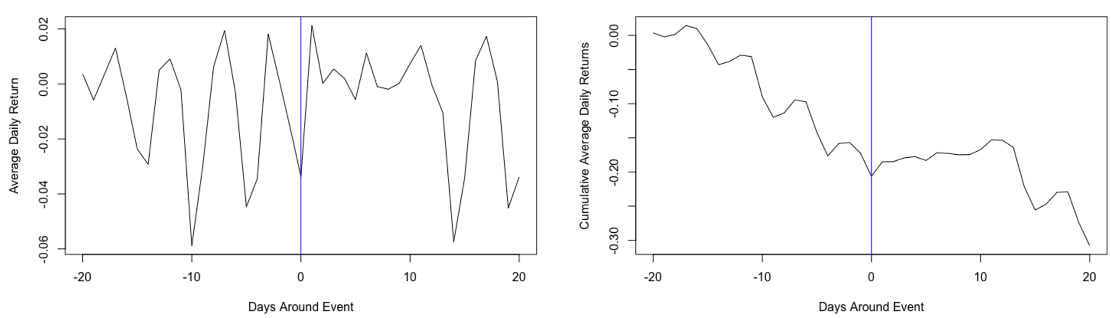
Graphs (a) and (b) of figure 4 now move to examining sentiment shocks in the form of the top 20% of increases in our sentiment metric. Immediately, the average return figure proves to be more promising than before with a significant drop in average daily returns following the sentiment shock followed by a further drop a few days later. This drop appears to be a correction for possible overly zealous market reaction to good news that was known to markets before making its way into news coverage. This is echoed in the cumulative returns in graph (b) with around -25% in cumulative returns 20 days after the event. Again these results are affected by the negative prices in 2020, but this graph maintains its general shape and trend when 2020 is excluded but with significantly smaller negative returns to the tune of 2% to 3%. These results also suggest that long at-the-money puts with around 15 days to maturity could provide consistently profitable options positions.

5.1.4 Negative Change in Sentiment Shocks

Graphs (c) and (d) portray returns data for the 20% most negative decreases in sentiment. The average returns in (c) do not seem to be particularly informative other than to possibly highlight some volatility in the markets leading up to big decreases in news sentiment. The cumulative returns in (d) repeat the recurring theme of information leakage as most of the adjustment in terms of cumulative returns has already taken place before the shock is recognised. The graph does show cumulative returns drop of around 10& following the shock which seems like a good opportunity, but this is again as a result of the anomalous year that was 2020. The same event study conducted for 2020 alone is much flatter, hovering around -4% cumulative return, following the shock with no real obvious option position opportunity. Based only on graph (d) however, we would be inclined to again recommend a long at-the-money put option.



(a) Ave daily returns positive delta sentiment shocks (b) Cum daily returns positive delta sentiment shocks



(c) Ave daily returns negative delta sentiment shocks (d) Cum daily returns positive delta sentiment shock

Figure 4: Average daily and cumulative average daily returns for top 20% of positive and negative delta sentiment shocks.

5.1.5 Summary

The sentiment level shock results provided decent option opportunities with long at-the-money put options for both events appearing to be the most likely profitable opportunity. Although the negative prices of 2020 do obscure the scale of our results somewhat, the shape of graphs and the takeaway form them endures. There appears to be reliable, exploitable responses to high and low levels of news sentiment. One downside of the sentiment level shocks is their tendency to be messy and overlapping; very high positive (or low negative) sentiment days tend to be surrounded by other high (low) days and the event windows can often overlap. This could result in some unreliable information being extracted from this event study.

The changes in sentiment event study results are potentially of more interest and value as they play into behavioural finance doctrine. We expect investors to respond more severely to sharp changes in sentiment, which would be a deviation from previous held beliefs and expectations, than to persistently high or low sentiment which is likely baked into expectations at that point. Large increases in sentiment showed promise both in the full sample and the sample excluding 2020 (although less so), with both suggesting long at-the-money puts could be profitable. The case for large decreases in sentiment seems promising for the full sample but loses all credibility when excluding 2020. Using the entire sample to determine events removes nuance and means that many 'genuine' events in earlier years may be overlooked as a result of later years. We were aware of these shortcomings, but carried out the event study in this manner to mimic previous papers and examine whether their findings still hold with more recent data. This shortcoming is remedied in our rule-based trading where quantiles for each day are calculated using only past data.

Lastly, we look to previous literature and examine whether we witness similar trends and behaviours. Based on sentiment level shocks ,the aforementioned asymmetric sentiment effect is present with far larger drops in cumulative returns following negative sentiment level shocks than increases following positive sentiment level shocks. Curiously, the asymmetric sentiment did not seem to hold during event windows of change in sentiment shocks. The magnitude of the difference between cumulative return highs and lows was about the same for both categories of changes in sentiment shocks. This was not as we expected and in fact disappointed us as we expected this to take advantage of behavioural finance doctrine. More generally, the WTI futures markets seem to have an immense degree of information leakage. We believe this to be as a result of the market participant mix of merchants and producers as well and money managers.

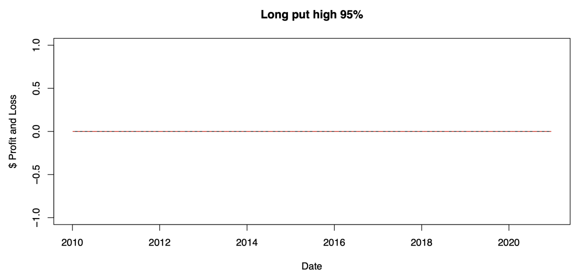
5.2 Rule based trading strategy results

Since we have no solution as to knowing when a sentiment spike event is going to occur, we can only trade based on the observed course of returns after the event, that is the 20 days after the event. In the following section we evaluate the trading strategies first based on the event study results and then for all different combinations of options positions for signals created by both sentiment in levels and differences.

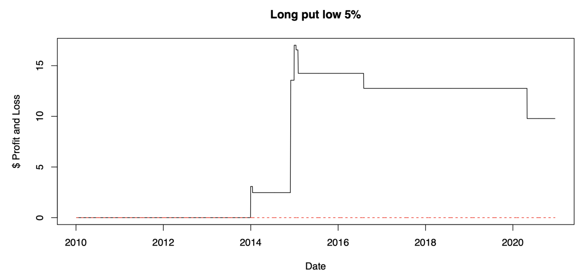
5.2.1 Rule based trading strategy sentiment in levels

As the results from the event study in section 5.1 show, namely figure 3 (b) and 3 (d), both for high values of positive sentiment as well as low values of negative sentiment the cumulative returns follow a decreasing pattern afterwards. This indicates that a long position in put options might be most profitable for both cases.

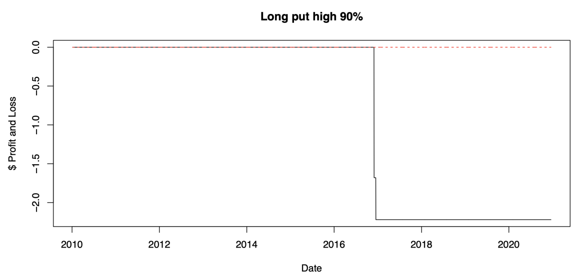
Figure 5 below displays the cumulative profit and losses for the whole sample period following the trading of long puts in high and low sentiment times for different quantiles of sentiment. As observed, taking long positions in puts when sentiment is low results in substantial trading profits although in 2020 it seems to perform less well. As shown when the price of the underlying goes down the profits are substantial while if the price goes up the losses are only equal to the option premium paid. However longing puts in high sentiment times does not confirm our findings from the event study, the trades result in considerable cumulative loses with the upper 95%, 90% and 80% quantiles producing very low and even no trading signals. For the strategy of trading long puts in low sentiment times we find that classifying low sentiment as below the 30% quantile performs best.



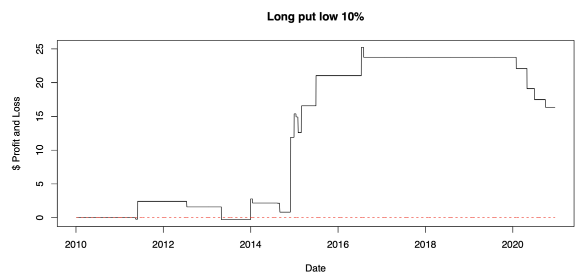
(a) PnL long puts high sentiment using 95%



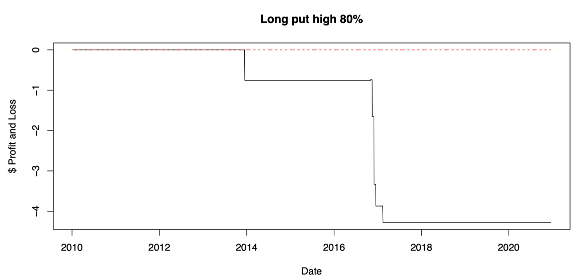
(b) PnL long puts low sentiment using 5%



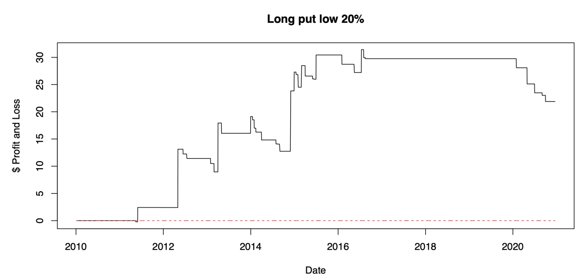
(c) PnL long puts high sentiment using 90%



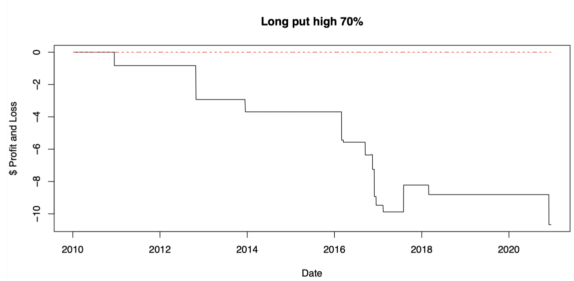
(d) PnL long puts low sentiment using 10%



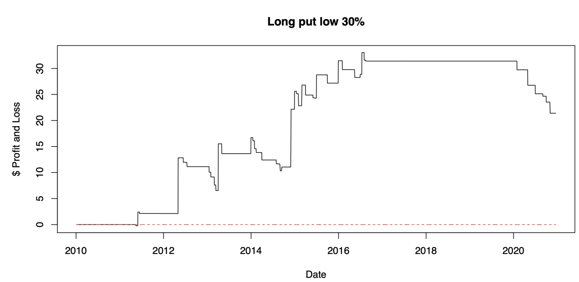
(e) PnL long puts high sentiment using 80%



(f) PnL long puts low sentiment using 20%



(g) PnL long puts high sentiment using 70%



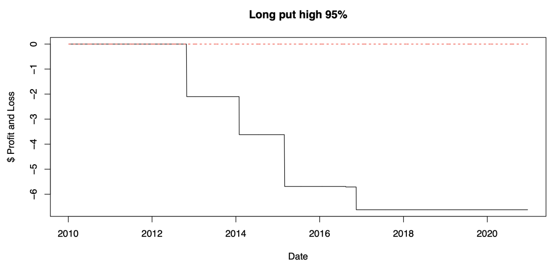
(h) PnL long puts low sentiment using 30%

Figure 5: Cumulative profit and loss (PnL) graphs employing the rule based strategy, focusing on long put positions for high and low levels of sentiment using the quantiles 95%, 90% 80%, 70% and 30%, 20%, 10%, 5%.

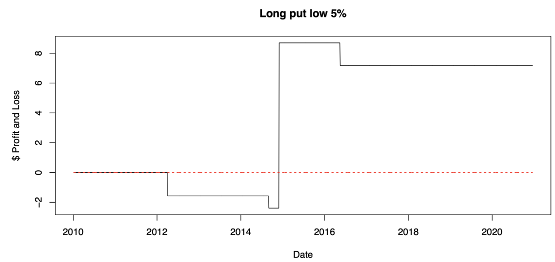
Additionally, appendix 7.2 figure 12 shows the accumulated profits and losses for all possible options combinations of long, short, call, put, high sentiment, low sentiment and all sentiment quantiles. As the results show, the strategy of taking long put positions (d) when sentiment is low appears to be very lucrative and also displays the highest profit of all strategies. Strikingly, also trading long calls (b) in low sentiment times shows substantial profits but only for the 30% lowest quantile. Looking at the short strategies, we find that shorting/writing out put options when sentiment is elevated (g) creates significant profits trough the option premiums collected. On the other hand from a risk management perspective shorting puts and calls in low sentiment times seems to be a losing strategy and could potentially lead to substantial loses as indicated by sub graphs (f) and (h).

5.2.2 Rule based trading strategy sentiment in differences

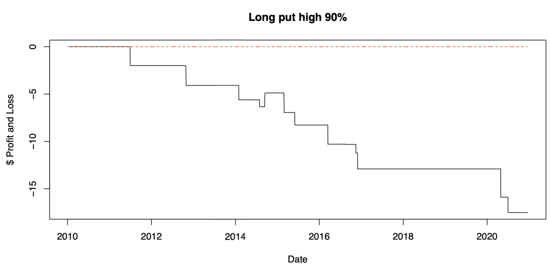
Focusing on the results from the event study it is evident that the cumulative returns as seen in figures 4 (b) and (d) again follow a decreasing pattern in the days following a large positive as well as negative sentiment shock. This indicates that taking long put positions in both cases should result in the best performance. Figure 6 shows the cumulative profit and losses taking long positions in puts based on high or low delta sentiment. The results are very similar to the outcome from the sentiment in levels figure 6. We find that taking long put positions after large positive changes in sentiment leads to considerable loses for all quantiles, while taking long puts after large negative changes in sentiment produces profits mainly at the end of 2014 and for the 30% quantile signal classification.



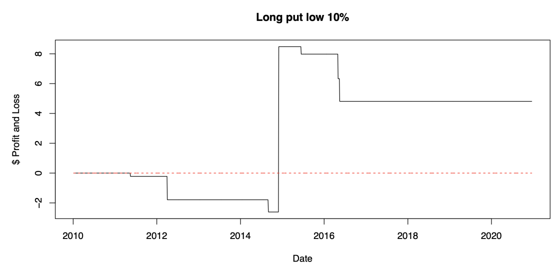
(a) PnL long puts high delta sentiment using 95%



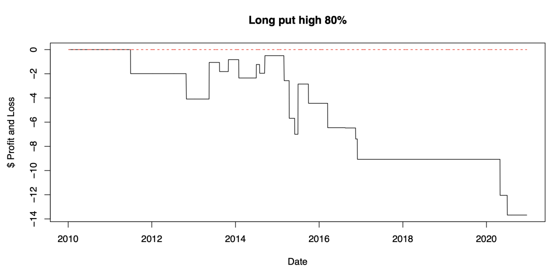
(b) PnL long puts low delta sentiment using 5%



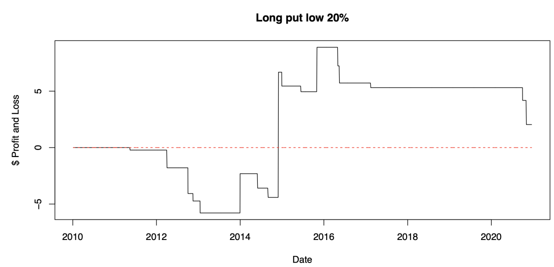
(c) PnL long puts high delta sentiment using 90%



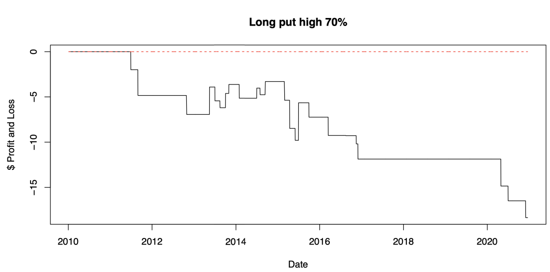
(d) PnL long puts low delta sentiment using 10%



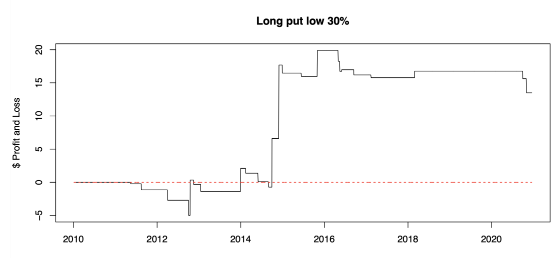
(e) PnL long puts high delta sentiment using 80%



(f) PnL long puts low delta sentiment using 20%



(g) PnL long puts high delta sentiment using 70%



(h) PnL long puts low delta sentiment using 30%

Figure 6: Cumulative profit and loss (PnL) graphs employing the rule based strategy, focusing on long put positions for high and low differences in sentiment using the quantiles 95%, 90%, 80%, 70% and 30%, 20%, 10%, 5%.

Once more we analyse the cumulative profit and losses of all possible combinations of puts, calls, long, short for high and low sentiment and using different quantiles for the buy and sell signals as shown in appendix 7.2 figure 13. The results show that taking long put (d) as well as call (b) positions following large negative changes in sentiment are on average accompanied by profits using the 30% quantile. Also, shorting/writing out puts after positive changes in sentiment (g) leads to consistent profits over time trough the accumulation of option premiums received. Remarkably, shorting calls following negative sentiment shocks (f) leads to tremendous loses compared to the other trading strategies and should be avoided.

5.3 Directional strategy results

According to Akaike's Information Criterion, the optimal ARMA and ARMAX specification was of order (1,1). The optimal amount of lags to include for the exogenous variables resulted in one lag each as well.
In table6, the parameters of model (1) are that of a standard ARMA model, excluding any sentiment related data. As one can see, both the auto-regressive as the moving average components of the model are estimated to be insignificant, meaning there is no clear autocorrelation in returns, and no dependency on lagged innovations. The intercept is not estimated to be significantly away from zero, fitting known characteristics of returns. Furthermore, because the models innovations are assumed to be student-t distributed, a ν estimate is included, and is estimated to be 3.54, with similar estimates in the other models. This confirms our assumption of leptokurtism in futures returns.

Model (2) shows the estimated parameters for the same model, but adjusted so that the level sentiment data are included in the model equation. The auto-regressive and moving average components of the model show similar estimates and significance as model (1). Model (2) does differ from the first model in that the intercept is now estimated to be significantly away from zero, as the level data of sentiment likely captures information otherwise hidden in the intercept. When sentiment in negative on a given day, there is a significant correlation with returns, resulting in lower returns. However, when sentiment is positive, this relationship disappears. Negative sentiment having a stronger effect than positive sentiment fits with conclusions from behavioural finance. Furthermore, because correlation of sentiment on returns is contemporaneous, their influence is likely mutual.

Model (3) includes changes in sentiment as exogenous regressors. When compared to models (1) and (2), the estimates related to r_{t-1} and ϵ_{t-1} are smaller in absolute value, but have similar standard errors. The intercept is now estimated to be significantly negative, albeit still close to zero. The asymmetric effect surrounding the external regressor is still present, albeit that the relationship is flipped. Positive changes in sentiment correlate highly significantly with contemporaneous returns. Negative changes in sentiment however do not seem to have a significant effect.

	ARMA	ARMA+sent	ARMA+Δsent
MAE	0.0199	0.01922	0.01975
MSE	0.00119	0.00121	0.00124

Table 4: Out-of-sample MAE and MSE, comparing the standard ARMA model, the ARMA model including sentiment level data (ARMA+sent) and the ARMA model including changes in sentiment (ARMA+Δsent).

In table 4 the mean absolute error and mean square error are shown, comparing the three ARMA(X) models in terms of out-of-sample prediction accuracy. As one can see, the ARMA model excluding exogenous data performs best in terms of MAE and MSE, meaning that the errors are smallest on average, and also the most consistent. The reason for this is likely due to a lack of autocorrelation in returns. The ARMA+sent and ARMA+Δsent forecasts are influenced by the sentiment data, which pushes the forecast in some way. The ARMA model does not suffer from this problem, meaning the forecasts are more general, leading to smaller forecast errors, especially because of the higher probability density around the zero mean of returns. Because of the inclusion of exogenous data, which is the only data that shows a significant contemporaneous effect, the ARMAX models suffer from mild overfitting.

MAE	ARMA	ARMA+sent	ARMA+Δ sent
ARMA	-	-2.6694***	-4.892***
		[0.007]	[0.000]
ARMA+sent		-	-2.403**
		-	[0.016]
ARMA+Δsent			-
			-

MSE	ARMA	ARMA+sent	ARMA+Δ sent
ARMA	-	-5.5721***	-5.5843***
		[0.000]	[0.000]
ARMA+sent		-	-5.589***
		-	[0.016]
ARMA+Δsent			-
			-

Table 5: Return forecasts Diebold Mariano test results and their respective [p-value]. When the statistic is positive (negative) the row model showcases larger (smaller) errors than the column model.

Table 5 shows the Diebold-Mariano test to compare out-of-sample performance. As described before, the ARMA model performs the best in terms of forecasting. Note that all the DM-tests yielded a significant statistic. The difference between performance of the ARMAX models is smaller than when they are individually compared to the ARMA model. In terms of MSE, similar conclusions can be drawn as when looking at the MAE's. The statistics are somewhat larger in absolute value, pointing towards the fact that the ARMA model is also more consistent than the ARMAX models.

	Dependent variable:		
	r_t	r_t	r_t
	(1)	(2)	(3)
r_{t-1}	−0.518	−0.406	−0.289
	(0.4878)	(0.5673)	(0.546)
ϵ_{t-1}	0.486	0.367	0.244
	(0.480)	(0.577)	(0.552)
intercept	−0.0004	0.002^{*}	-0.002^{***}
	(0.00004)	(0.001)	(0.0006)
$S_t \mathbb{1}_{\{S>0\}}$		−0.024	
		(0.067)	
$S_t \mathbb{1}_{\{S<0\}}$		0.050^{**}	
		(0.019)	
$\Delta S_t \mathbb{1}_{\{\Delta S>0\}}$			1.212^{***}
			(0.185)
$\Delta S_t \mathbb{1}_{\{\Delta S<0\}}$			0.136
			(0.091)
ν	3.54^{***}	3.55^{***}	3.48^{***}
	(0.3541)	(0.426)	(0.348)
Exogenous included	None	Sentiment	Δ Sentiment
Observations	1576	1576	1576
Log Likelihood	3955.715	3959.718	3988.39
σ^2	0.022	0.022	0.022
Akai Inf. Crit.	−5.1580	−5.0161	−5.0525
Note:			*p<0.1; **p<0.05; ***p<0.01

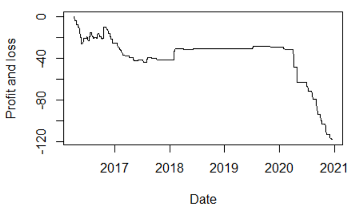
Table 6: Estimated parameters for the ARMA(X), where the first column is the ARMA model that excludes exogenous regressors, and the second column is an ARMAX model, which includes sentiment, while the third column is an ARMAX, which includes sentiment changes.

In table 7 the profit and losses of the directional trading strategy resulting from our return forecasts. The strategy related to the 1% of the most extreme forecasted returns are omitted due to missing data, resulting in uninformative results. Since there is no significant autocorrelation in returns, the strategy yields losses in every case, since one can not consistently predict the price, especially not for the time horizon of our contracts, which usually spans about two to three weeks. ARMA and ARMA+Δsent exhibit similar results, likely due to the fact that Δsentiment is a differenced time series, thereby being closer to zero, and having not such a large effect on the forecast. For most models, trading according to a more conservative metric leads to smaller losses, but this is not the case for the ARMA+sent model, marking the only instance where losses are smaller per trade.

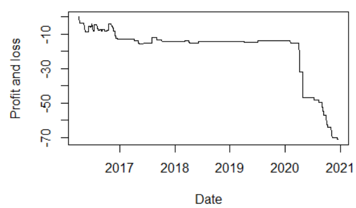
Shown in figure 7 are the cumulative plots for the profits and losses. The strategies show a similar pattern across the years, no doubt driven by the extremities of both returns and sentiment data, as they affect each other contemporaneously. At the end of 2016, all strategies move toward a slight loss, with varying degree. The ARMA+sent performs best in this time-frame, as it manages to produce some positive trades. From 2017 onward there is relatively little trading, caused by a decrease in overall volatility. As the quantiles of return forecasts are calculated over the entire out-of-sample period, a less volatile period will produce less extreme forecasts. By the time 2020 starts, all strategies start making large losses at the same times, as the tail end of the graphs all have a similar shape.

	ARMA	ARMA+sent	ARMA+Δ sent
20%	-118.18	-79.02	-119.61
10%	-71.13	-49.96	-66.03
5%	-62.96	-66.03	-58.85

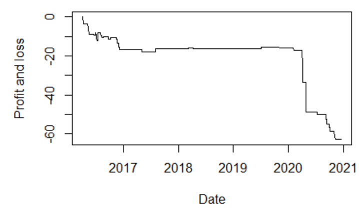
Table 7: Profits and losses in dollars made with directional bets



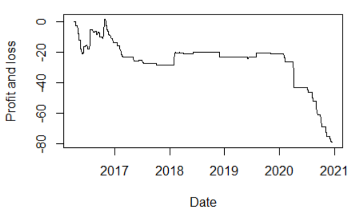
(a) PnL ARMA using 20%



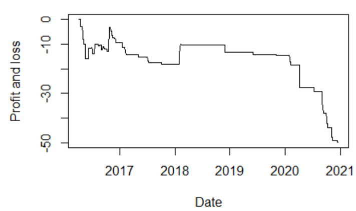
(b) PnL ARMA using 10%



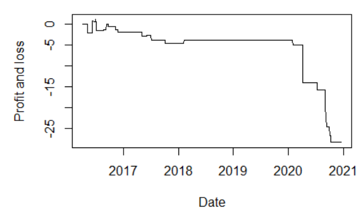
(c) PnL ARMA using 05%



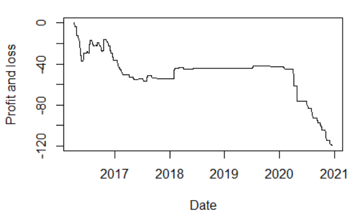
(d) PnL ARMA+sent using 20%



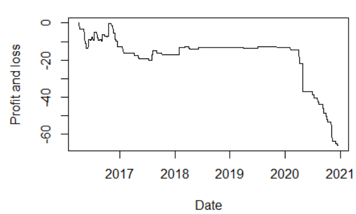
(e) PnL ARMA+sent using 10%



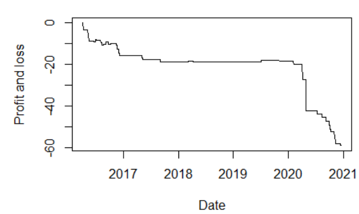
(f) PnL ARMA+sent using 05%



(g) PnL ARMA+Δsent using 20%



(h) PnL ARMA+Δsent using 10%



(i) PnL ARMA+Δsent using 05%

Figure 7: Cumulative profit and loss (PnL) graphs employing the directional strategies of puts only, or puts and calls, using the 80, 90, and 95% quantiles of absolute return forecasts

5.4 Volatility strategy results

Based on the Akaike information criterion we found the optimal lag order to be of (1,1), for both the GARCH and CGARCH models. Accordingly, the estimated parameters for the GARCH models are given in table

8 below. As the results show, the sentiment data does not seem to have a significant effect on conditional volatility, as can be seen when considering the coefficients related to the sentiment variables. In table 9 the estimated parameters are shown of the component-GARCH models. Just like the GARCH models, the CGARCH models do not show significance in their exogenous variables. They do however show significance in the estimates of ρ and γ. This point presents a justification for utilizing the CGARCH models, and reinforces conclusions drawn by previous literature by Kang, Kang Yoon (2009).

	Model		
	$GARCH(1,1)$	$GARCHSent(1,1)$	$GARCH\Delta Sent(1,1)$
μ	-0.000076	-0.000073	-0.000075
	(0.000379)	(0.000379)	(0.000381)
ω	0.000002	0.000002	0.000002
	(0.000002)	(0.000002)	(0.000001)
α	0.060368***	0.060086***	0.060164***
	(0.015011)	(0.014076)	(0.012177)
β	0.937720***	0.936882***	0.937932***
	(0.015894)	(0.016217)	(0.014359)
$S_t \mathbb{1}_{\{S>0\}}$	-	0.000148	-
	-	(0.000131)	-
$S_t \mathbb{1}_{\{S<0\}}$	-	0.000000	-
	-	(0.000017)	-
$\Delta S_t \mathbb{1}_{\{\Delta S>0\}}$	-	-	0.000000
	-	-	(0.000558)
$\Delta S_t \mathbb{1}_{\{\Delta S<0\}}$	-	-	0.000000
	-	-	(0.000485)
Observations	1573	1573	1573
Log Likelihood	4106.79	4199.69	4106.79
Akaike Inf. Crit.	-5.2053	-5.3304	-5.2028
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 8: Estimated parameters for GARCH, GARCHSent and GARCHΔSent. Values in parentheses represent the corresponding parameter standard errors.

	Model		
	$CGARCH(1,1)$	$CGARCHSent(1,1)$	$CGARCH\Delta Sent(1,1)$
μ	-0.000084	-0.000086	-0.000088
	(0.000371)	(0.000374)	(0.000376)
ω	0.000001	0.000001*	0.000001
	(0.000001)	(0.000001)	(0.000001)
α	0.103533***	0.103554***	0.103461***
	(0.034494)	(0.034431)	(0.033952)
β	0.370875	0.371213***	0.372786***
	(0.126749)	(0.107629)	(0.142091)
ρ	0.997044***	0.997042***	0.997037***
	(0.000498)	(0.000399)	(0.000406)
γ	0.044409***	0.044444***	0.044438***
	(0.000805)	(0.001124)	(0.000323)
$S_t \mathbb{1}_{\{S>0\}}$	-	0.000000	-
	-	(0.000077)	-
$S_t \mathbb{1}_{\{S<0\}}$	-	0.000000	-
	-	(-0.000008)	-
$\Delta S_t \mathbb{1}_{\{\Delta S>0\}}$	-	-	0.000000
	-	-	(0.000507)
$\Delta S_t \mathbb{1}_{\{\Delta S<0\}}$	-	-	0.000000
	-	-	(0.000494)
Observations	1573	1573	1573
Log Likelihood	4111.476	4111.476	4111.477
Akaike Inf. Crit.	-5.2087	-5.2062	-5.2062

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Estimated parameters for CGARCH-sentiment model. Values in parentheses represent the corresponding parameter standard errors.

Using the initial estimated parameters the one day ahead out-of-sample volatility predictions are calculated where the parameters are refitted at each iteration on a rolling window. Figure 8 displays the volatility predictions together with the realized volatility for the CGARCH-Sentiment model. From a visual point of view it is noticeable that the CGARCH-sentiment model is able to capture the volatility dynamics with a high degree of accuracy.

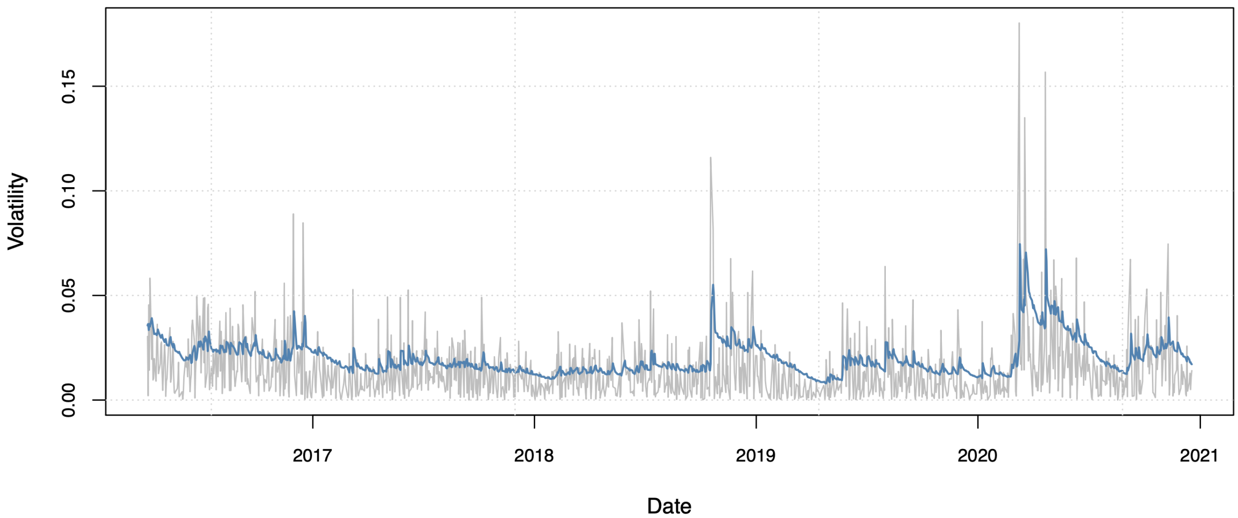


Figure 8: Out-of-sample daily volatility forecast (in blue) and realized volatility (in grey) from 01-04-2016 until 18-12-2020.

Table 10 show the MAE's and MSE's related to the different GARCH versions. CGARCH seems to outperform GARCH, for every version, both in average error and in terms of consistency, as the MAE and MSE are both lower. The models that include sentiment exhibit very similar metrics to their non-sentiment counterpart. This is to be expected since the sentiment estimates are all insignificant and close to zero.

Although the coefficient of sentiment are generally zero and non-significant, there is a difference when considered through the lens of a Diebold-Mariano test, the results of which are shown in table 11. One can see that when looking at the MAE, every version of CGARCH outperforms every version of GARCH. Note that the more general models outperform the less general models, as the models that do not include sentiment have lower errors than the models that do include exogenous data. The only time a non-significant result from a DM-test is seen is when comparing the CGARCH+sentiment with the CGARCH+Δsentiment, meaning they have similar performance when forecasting the realized volatility. Similar conclusions can be drawn when looking at the MSE's. The DM-stats are a somewhat lower, pointing to an increase in consistency of the GARCH models relative to the CGARCH models, as the MSE punishes outliers more. In other words, the CGARCH models are closer on average while the GARCH models are relatively more consistent in how close they are.

	GARCH	GARCH+sent	GARCH+Δsent
MAE	0.01478	0.01478	0.01479
MSE	0.00059	0.00059	0.00059
	CGARCH	CGARCH+sent	CGARCH+Δsent
MAE	0.01396	0.01397	0.01397
MSE	0.00049	0.00049	0.00049

Table 10: Out-of-sample MAE and MSE, comparing the standard GARCH and CGARCH model, the volatility models including sentiment level data (GARCH/CGARCH+sent) and the volatility models including changes in sentiment (GARCH/CGARCH+Δsent).

MAE	GARCH	GARCH+sent	GARCH+Δsent	CGARCH	CGARCH+sent	CGARCH+Δsent
GARCH	-	-9.5969***	-11.601***	6.6706***	6.6483***	6.6699***
	-	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
GARCH+sent		-	-12.5133***	6.6973***	6.6750***	6.6967***
		-	[0.000]	[0.000]	[0.000]	[0.000]
GARCH+Δsent			-	6.7650***	6.7426***	6.7646***
			-	[0.000]	[0.000]	[0.000]
CGARCH				-	-7.5764***	-4.2789***
				-	[0.000]	[0.000]
CGARCH+sent					-	0.2734
					-	[0.7845]
CGARCH+Δsent						-
						-
MSE	GARCH	GARCH+sent	GARCH+Δsent	CGARCH	CGARCH+sent	CGARCH+Δsent
GARCH	-	-6.8465***	-7.8393***	3.2538***	3.2536***	3.2581***
	-	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
GARCH+sent		-	-8.3021***	3.2590***	3.2587***	3.2633***
		-	[0.000]	[0.001]	[0.001]	[0.001]
GARCH+Δsent			-	3.2719***	3.2716***	3.2762***
			-	[0.001]	[0.001]	[0.001]
CGARCH				-	-2.5724**	-1.9267*
				-	[0.013]	[0.054]
CGARCH+sent					-	-1.1379**
					-	[0.026]
CGARCH+Δsent						-
						-

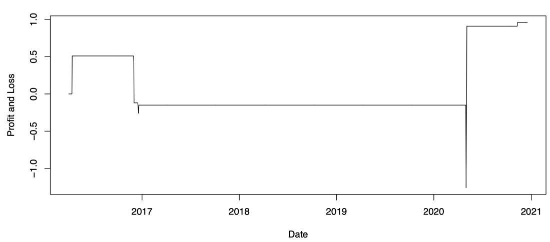
Table 11: GARCH model and variants Diebold-Mariano results, where a positive (negative) DM statistic means the row model showcases larger (smaller) errors than the column model.

We now consider the trading strategy based on the out-of-sample volatility forecasts from the CGARCH(1,1)-Sentiment model. The trading strategy involves taking volatility bets in the form of straddle option packages, a long straddle position is taken if a positive shock in volatility is forecasted and a short position for a negative shock. The straddle position is closed out and sold in the market after one day corresponding to our forecasting horizon. Table 12 shows the trading profit and losses using the mentioned strategy for different magnitudes of forecast shocks. As seen by the results, the trading for signals of magnitude 0.03 and 0.02 show a profit, while 0.01 does not.

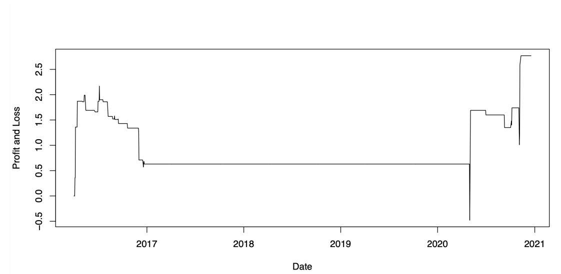
Magnitude of Shock size	P&L
> 0.03 or < -0.03	0.96
> 0.02 or < -0.02	2.77
> 0.01 or < -0.01	-0.54

Table 12: Out-of-sample profit and losses for different magnitudes of shocks ranging from 01-04-2016 until 18-12-2020. A shock is defined as $\hat{\sigma}_{t+1}$ - σ_t, where $\hat{\sigma}_{t+1}$ is the forecasted next days volatility and σ_t is today's realized volatility.

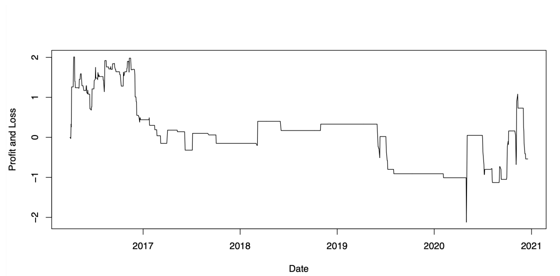
Additionally, the cumulative profit and losses over time are shown in figure 9 below. Clearly the more conservative magnitude of a shock size of 0.03 results in much less signals and trading activity than a shock magnitude of 0.01. It is also noteworthy that for the more conservative magnitudes as seen in (a) and (b) there is no trading activity from 2017 until the beginning of 2020, this is likely due to a lack in extreme volatility during this period as well as gaps in our options prices data, such that no position could be taken because there was no price information.



(a) PnL using shock magnitude of 0.03.



(b) PnL using shock magnitude of 0.02.



(c) PnL using shock magnitude of 0.01.

Figure 9: Cumulative profit and loss (PnL) graphs incorporating the volatility strategy using long and short straddle option packages for different magnitudes of forecast shocks.

6 Conclusion

Overall, it seems that sentiment does indeed show promise in providing trading signals for profitable rule-based strategies but falls short as an explanatory variable. Our event study shows empirically that sentiment shocks can be used to time profitable option positions, but that the scale of these profits is greatly reduced as a result of severe information leakage. The well documented asymmetric sentiment effect is present in our event study in the context of sentiment level shocks, but not in the context of change in sentiment shocks. This was not in line with our expectations based on behavioural finance reasoning.

The rule based trading strategies produced some straightforward insights, unlike what one would expect, trading long put and call options in high sentiment times did not yield profits while some short positions did. Remarkably the trading strategies showed that long call and put positions in low sentiment times leads to the best results, additionally writing puts in high sentiment times also created profits. Clearly the resulting cumulative profit and losses are very promising and lead in some cases to consistent profits indicating that sentiment creates the right trading signals at the right times. The rule based trading strategies could be implement in the real world starting from today, however obviously these results are applicable to the used time horizon and are no guarantee for prospective results in light of future market conditions. For instance some strategies showed great performance overall but seemed to be doing worse during 2020, so the results can be different depending on the date when the trading strategy is first initiated. It would be safe to say that these strategies should be performed and analysed multiple times on different types of oil or even different commodity types before any real world trading is attempted.

Our ARMAX estimates showed significance related to negative sentiment and positive changes in sentiment, pointing towards some explanatory power concerning sentiment and contemporaneous returns. However, due to the serial structure of returns, the general ARMA method outperformed the ARMAX models in terms of out-of-sample predictive power. Moreover, the trading strategy employed led to losses for all models and all trading thresholds, where most losses occurred in 2020.

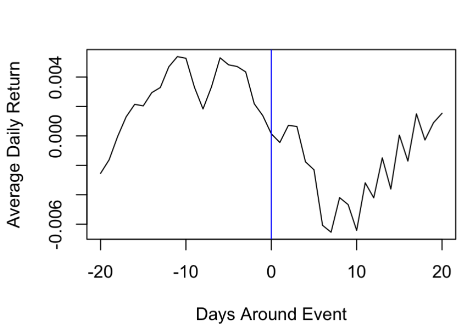
The strategy of incorporating sentiment into volatility CGARCH forecasts did not yield any significant results in terms of coefficients relating to sentiment. It can therefore be concluded that sentiment has no explanatory effect on volatility in this research setting. Additionally, a sentiment Incorporated CGARCH model did produce significantly more accurate forecasts than the regular CGARCH model. This result is to some extent surprising since it would be logical that very low or high sentiment in the market would lead to an increase in volatility due to high expectations of speculators in high sentiment times and hedgers wanting to eliminate risks in low sentiment times. We suspect that the high degrees of information leakage seen in the event study may be partly responsible for this.

Due to time constraints, some recommendations for future research came naturally. These recommendations include exploring the effect of sentiment data on possible calendar spread strategies, as sentiment effects are contemporaneous, and markets adjust to fundamentals quickly afterwards. Another recommendation for future research is to add Google trends data to see if there is a relationship between overall online interest in a market and its returns and volatility.

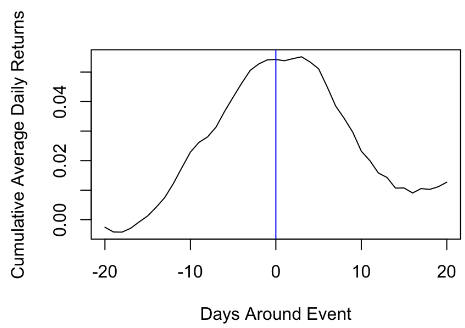
Lastly, we suggest the implementation of machine learning methods as opposed to typical econometric methods for returns and volatility forecasting. The implementation of machine learning techniques have shown promise in this field of study in the past and perhaps its' combination with alternative sources of data such as our news metrics, Google Trends and perhaps even twitter data could yield more fruitful results.

7 Appendix

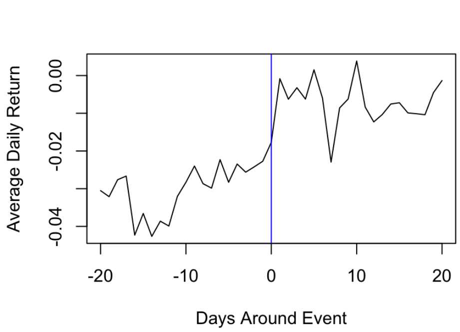
7.1 Event Study Excluding 2020



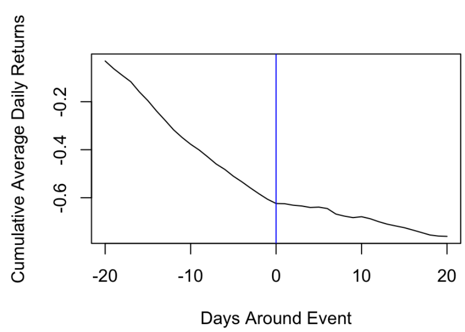
(a) Ave return positive sentiment shocks



(b) Cum return positive sentiment shocks

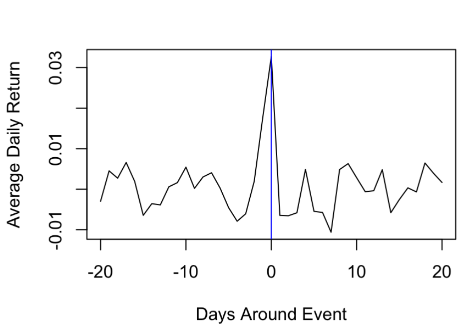


(c) Ave return negative sentiment shocks

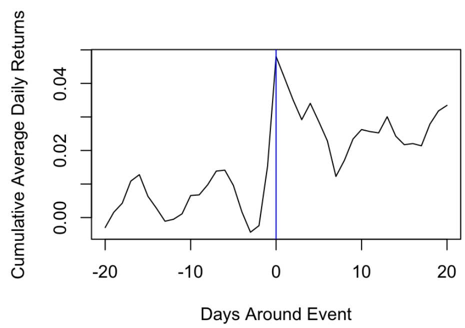


(d) Cum return negative sentiment shocks

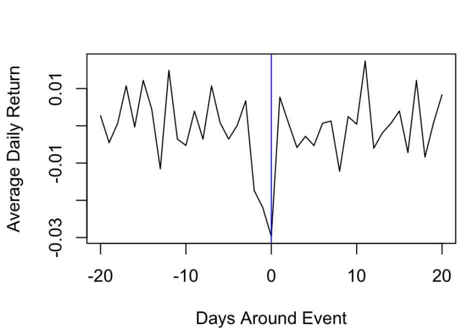
Figure 10: Average daily and cumulative average daily returns for top 20% of positive and negative sentiment level shocks. Excludes 2020



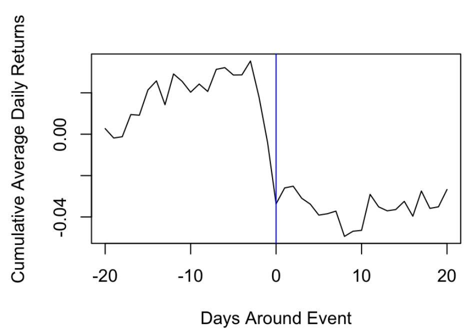
(a) Ave returns positive Δsentiment shocks



(b) Cum returns positive Δsentiment shocks



(c) Ave returns negative Δsentiment shocks

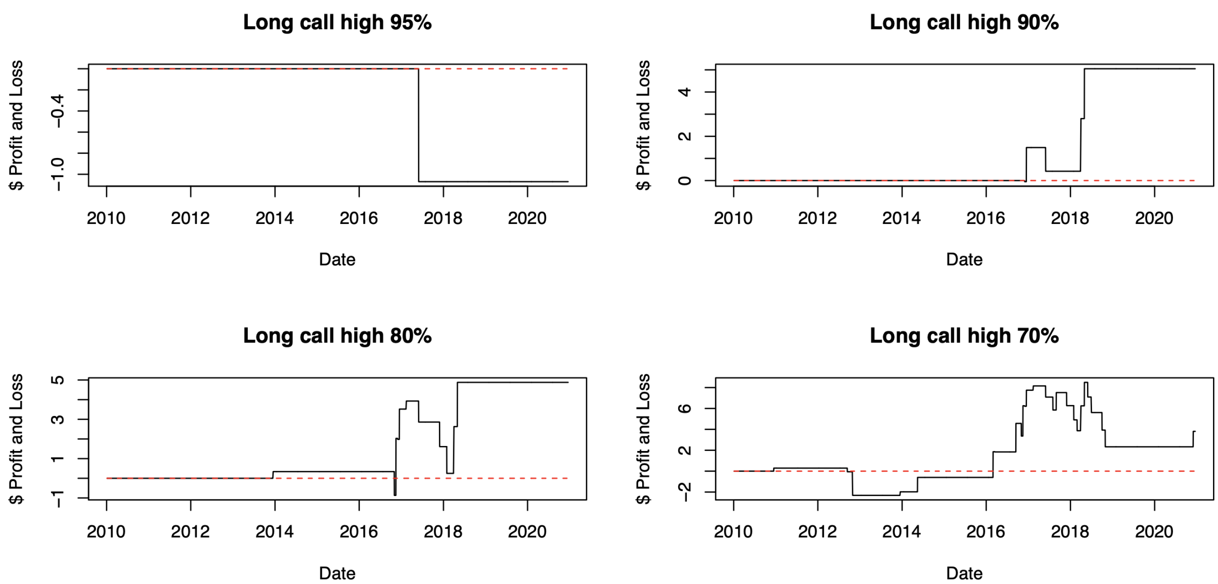


(d) Cum returns positive Δsentiment shock

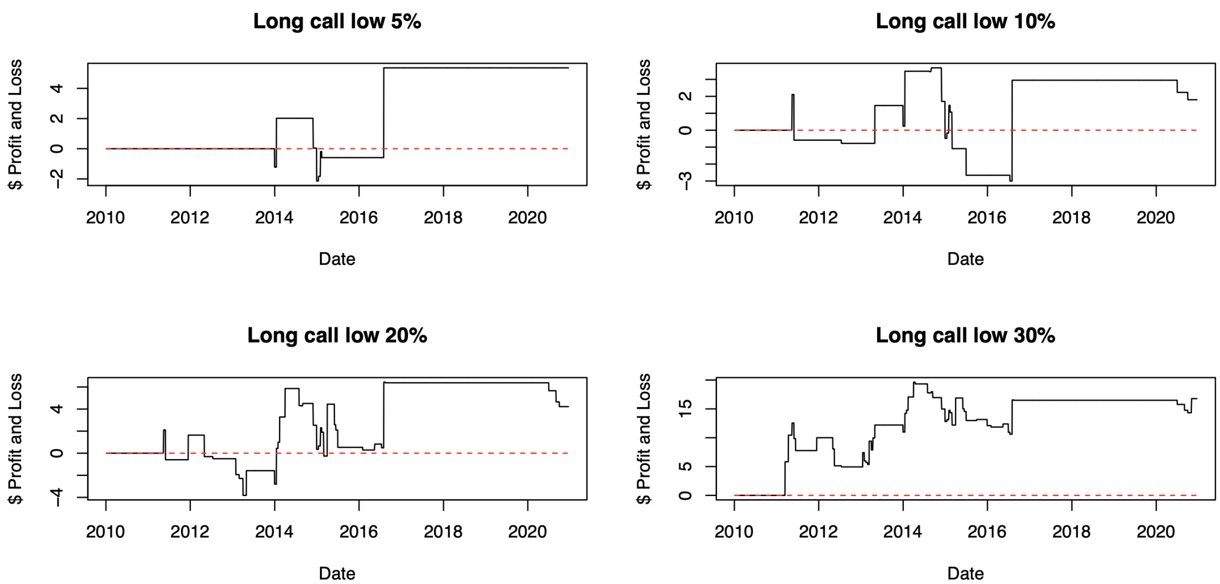
Figure 11: Average daily and cumulative average daily returns for top 20% of positive and negative delta sentiment shocks. Excludes 2020

7.2 Profit and Loss Graphs Rule Based Strategy

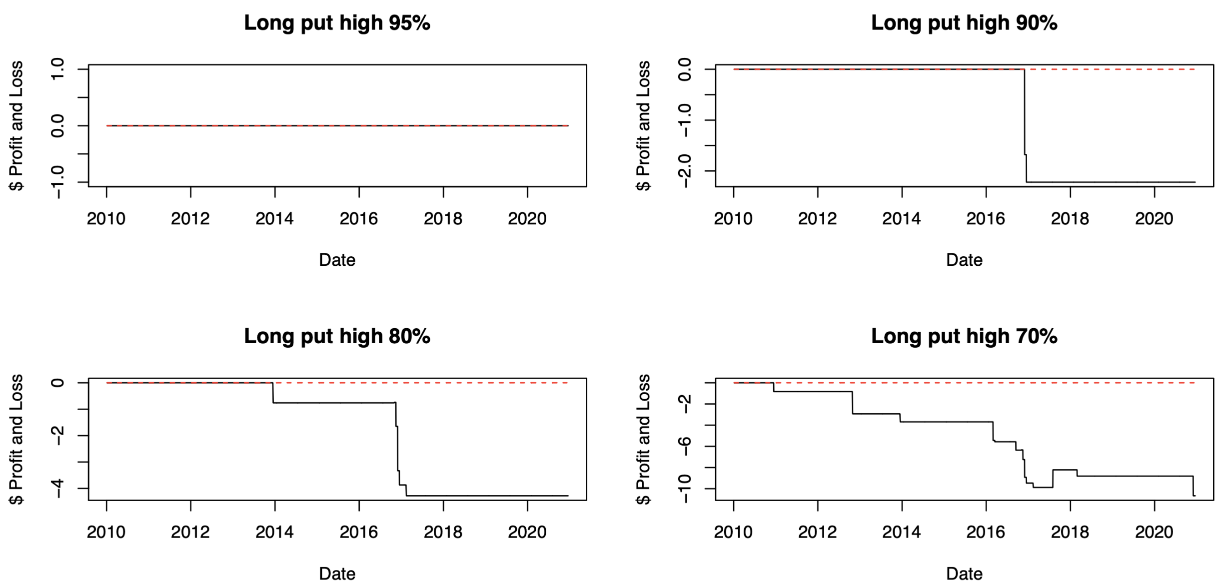
7.2.1 Sentiment in levels



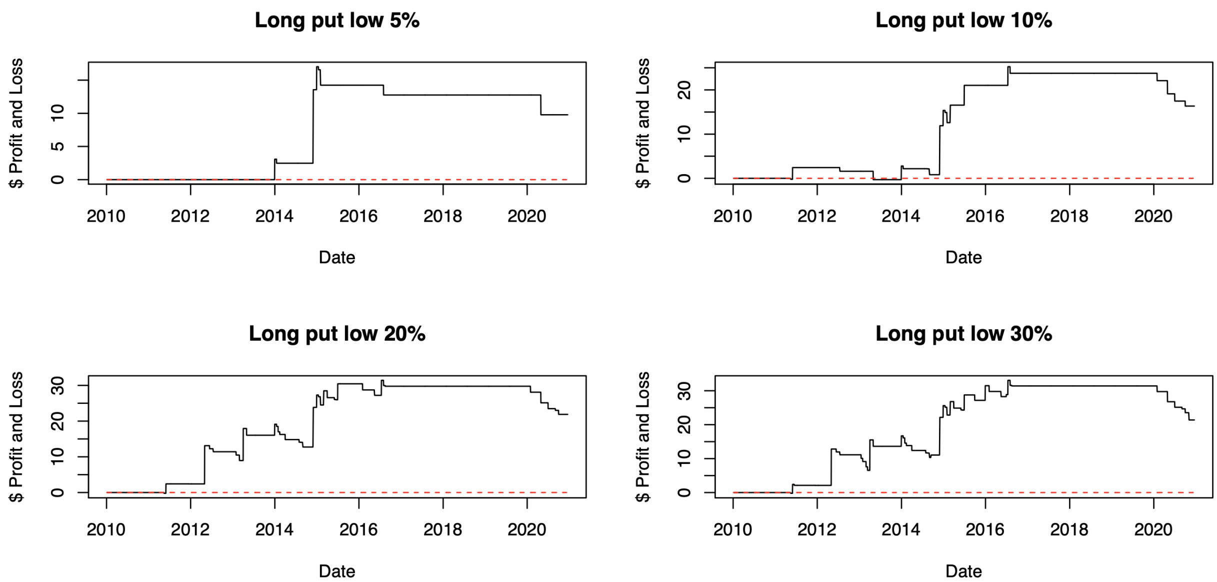
(a) Long calls when sentiment is high



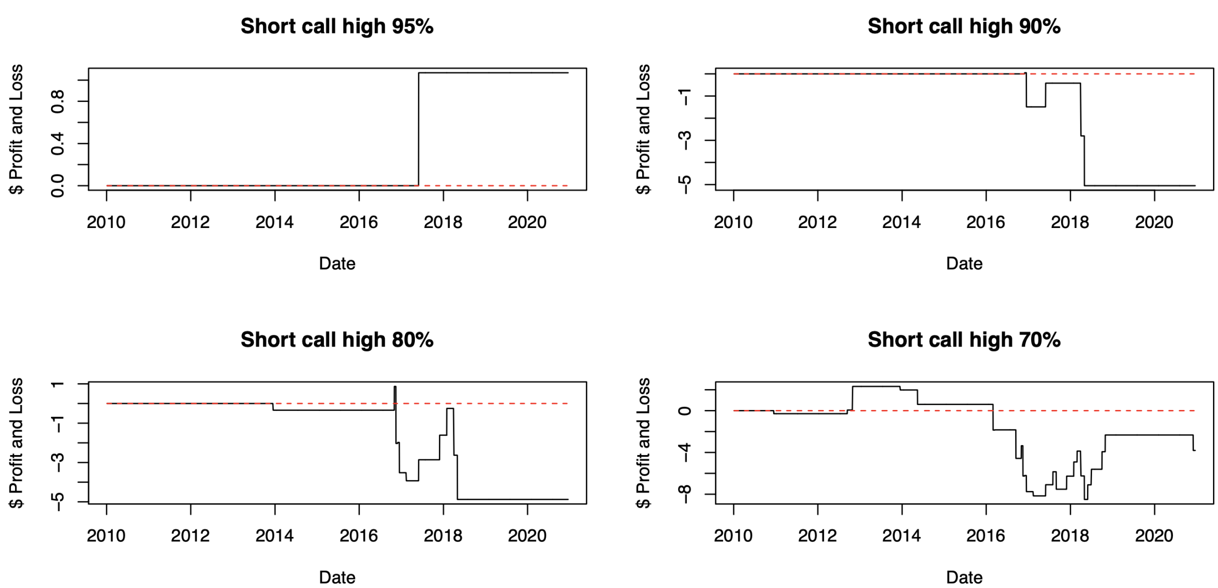
(b) Long calls when sentiment is low



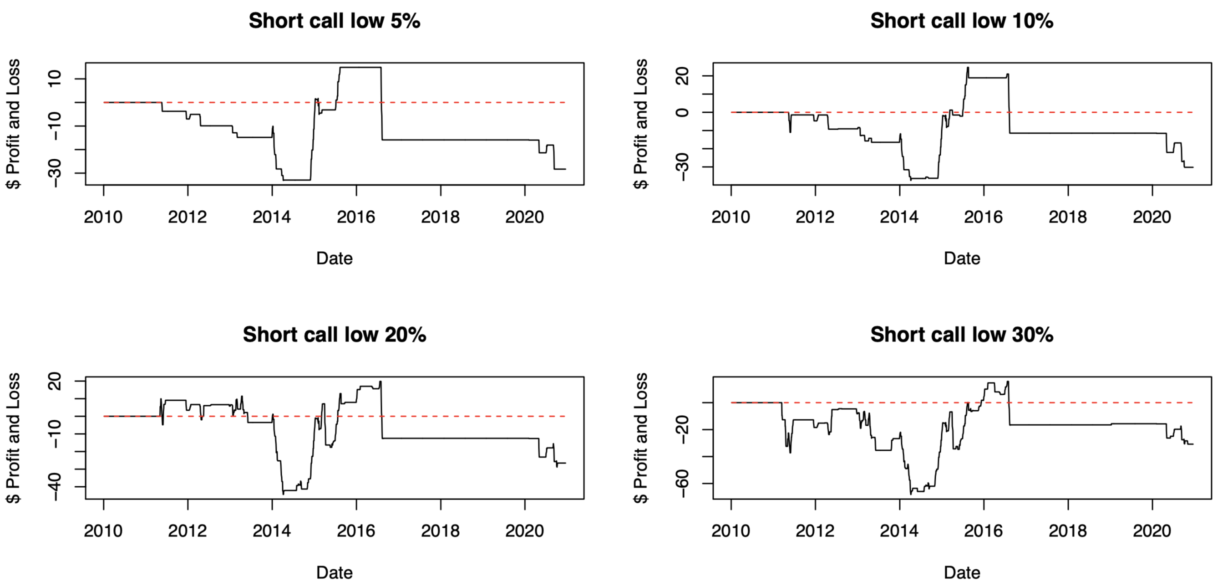
(c) Long puts when sentiment is high



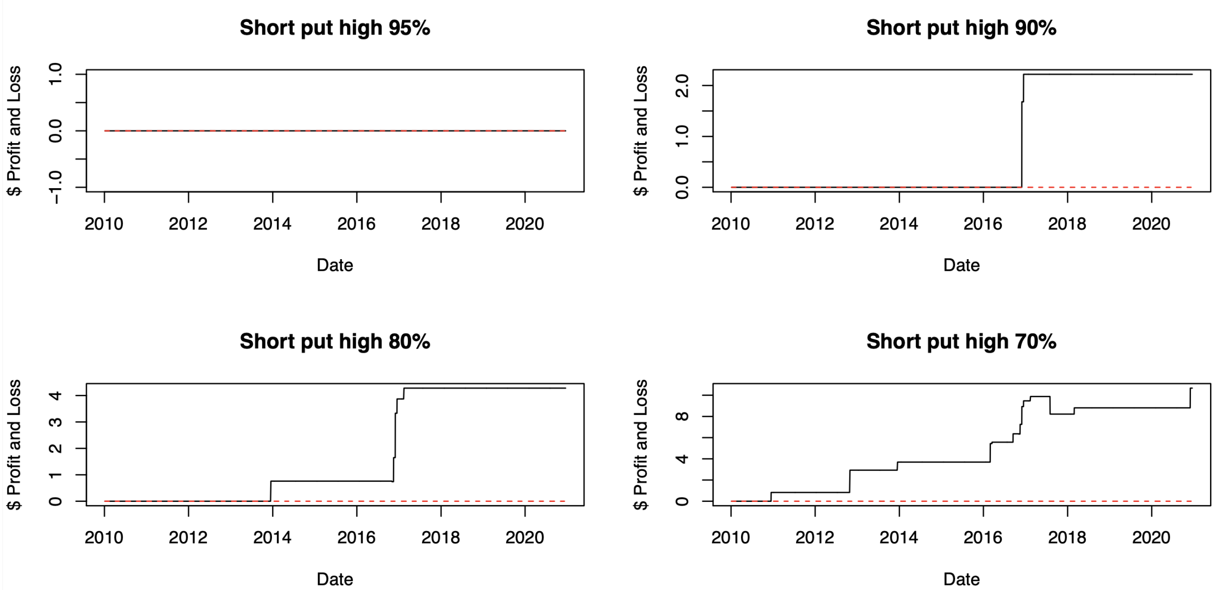
(d) Long puts when sentiment is low



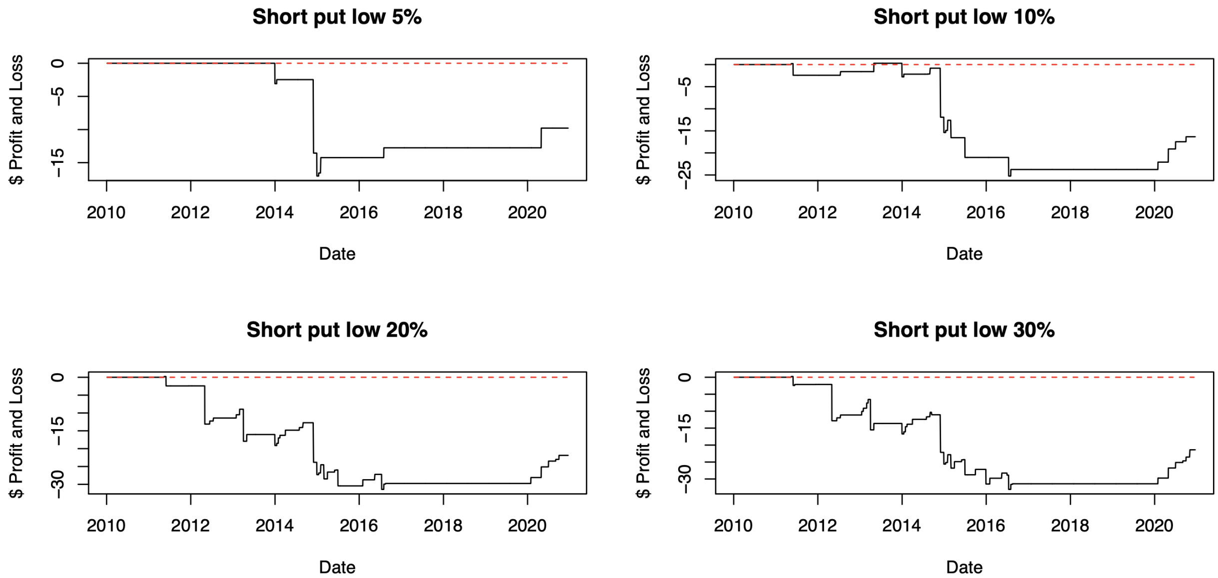
(e) Short calls when sentiment is high



(f) Short calls when sentiment is low



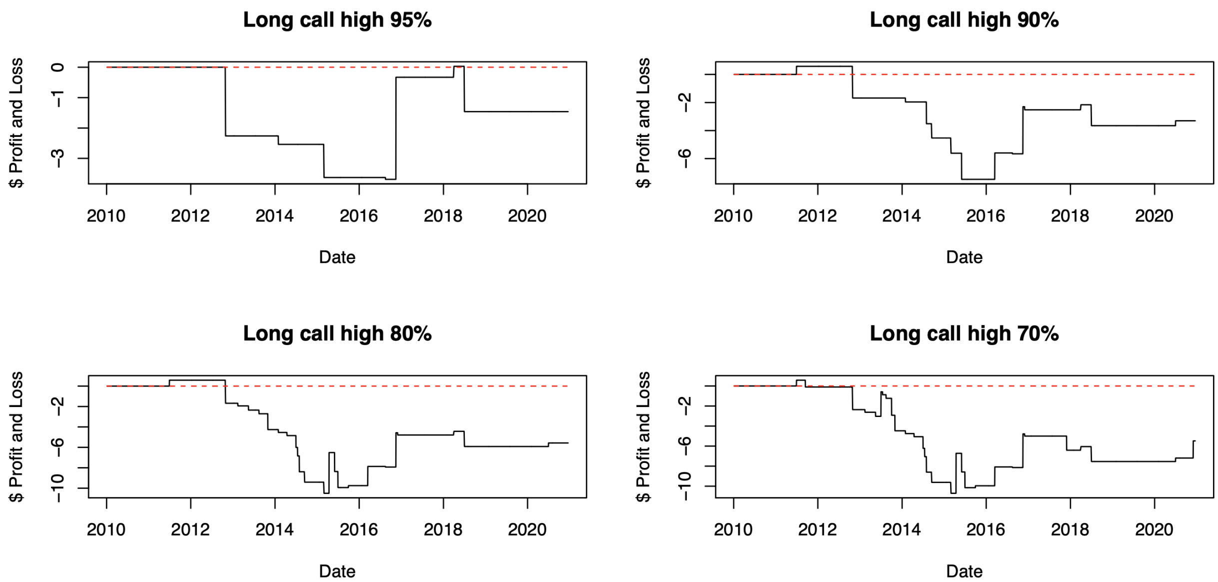
(g) Short puts when sentiment is high



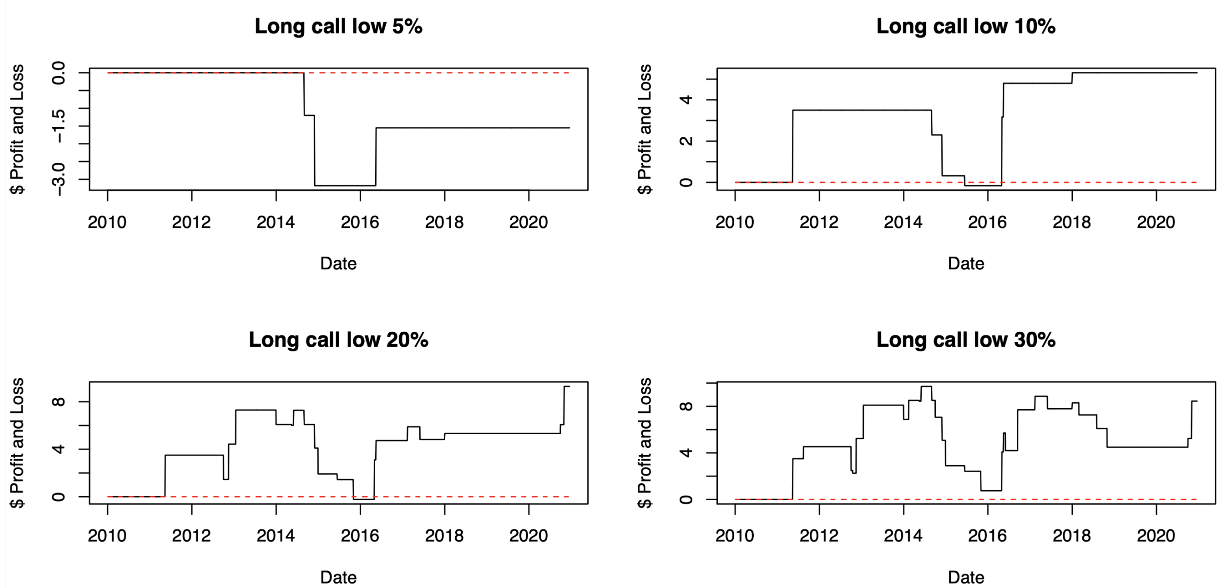
(h) Short puts when sentiment is low

Figure 12: Cumulative profit and loss (PnL) graphs employing the rule based strategy, focusing on all possible combinations for high and low levels of sentiment using the quantiles 95%, 90%, 80% and 20%, 10%, 5%.

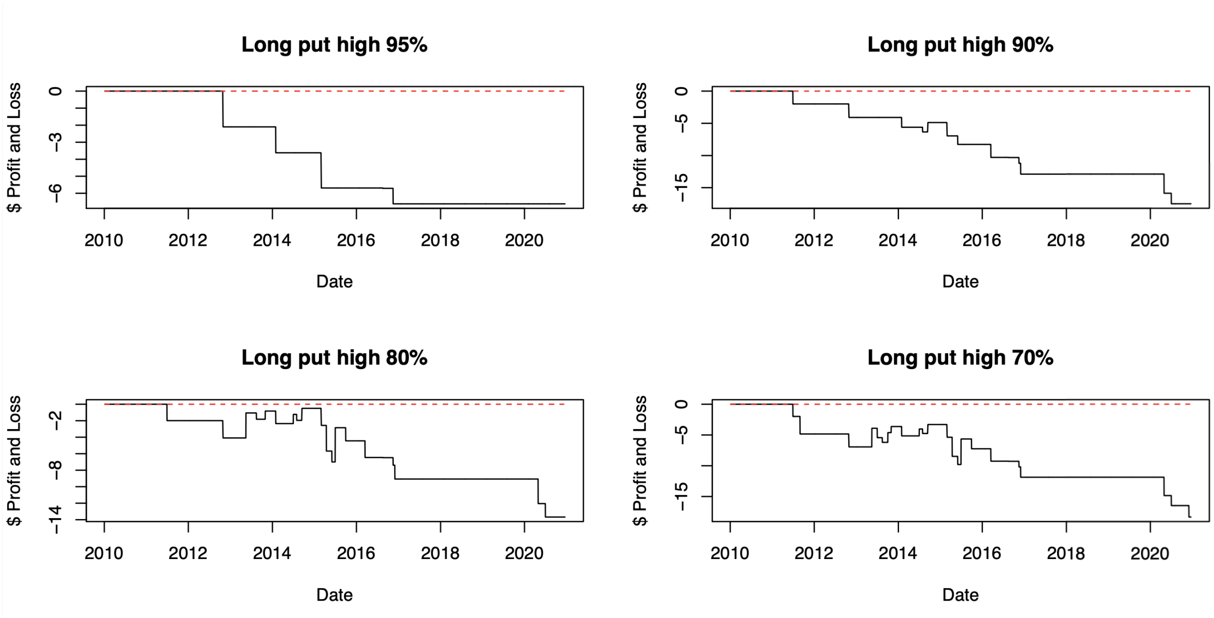
7.2.2 Sentiment in differences



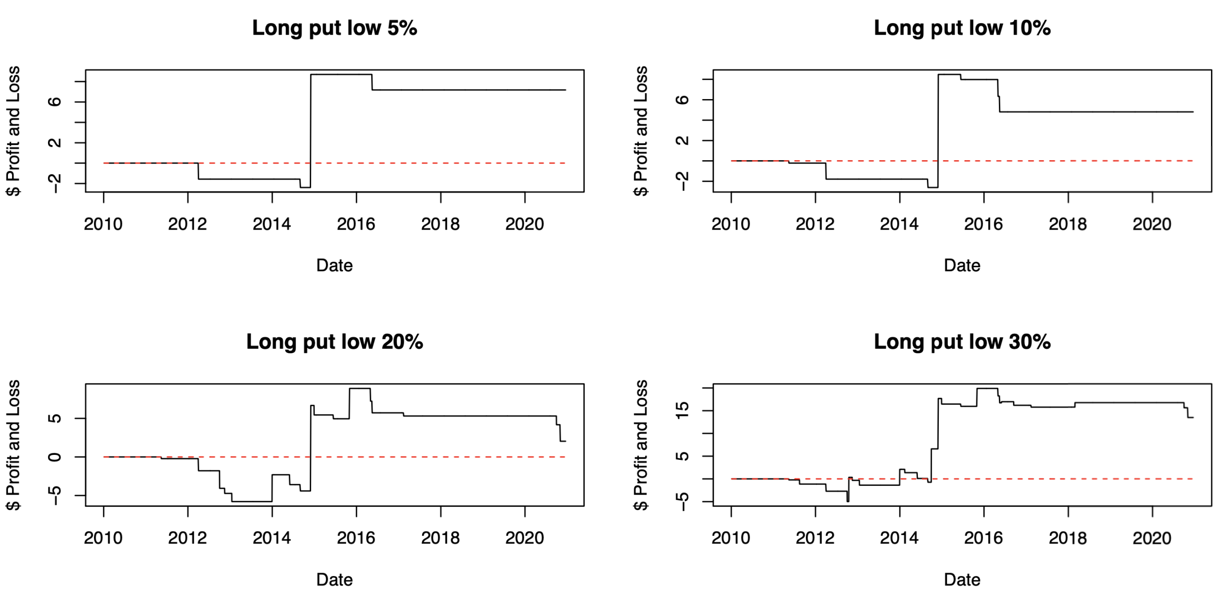
(a) Long calls when delta sentiment is high



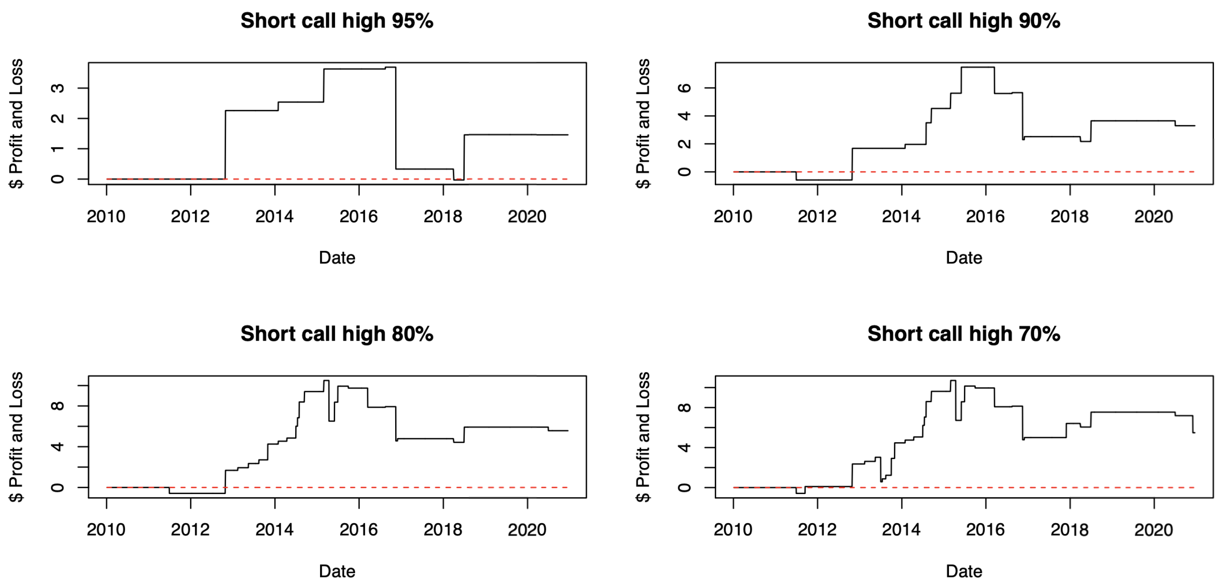
(b) Long calls when delta sentiment is low



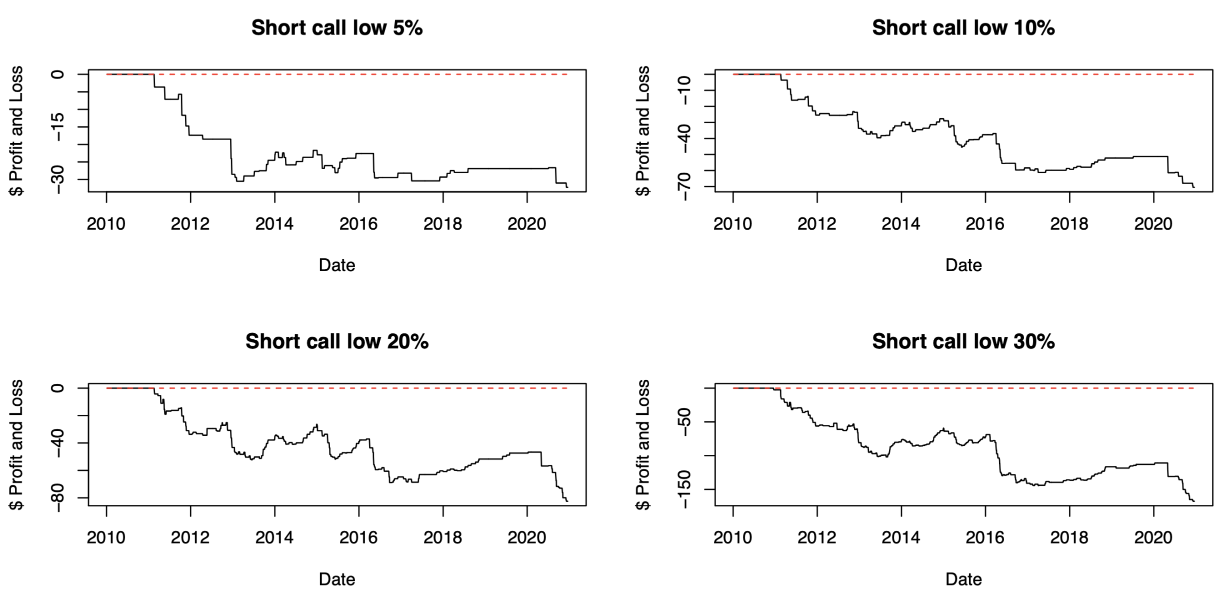
(c) Long puts when delta sentiment is high



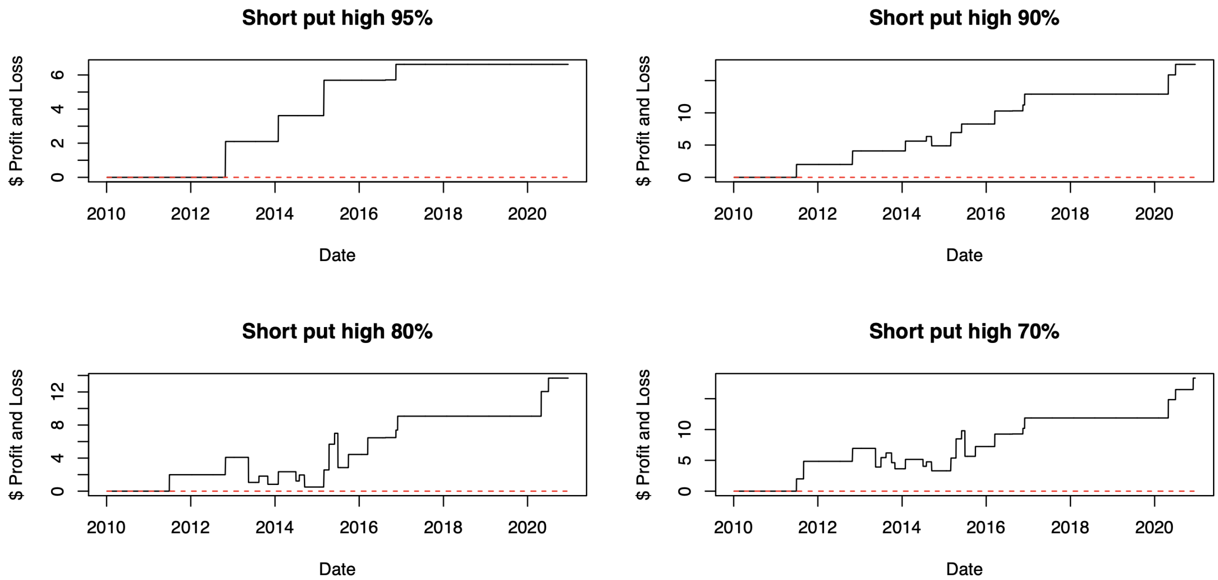
(d) Long puts when delta sentiment is low



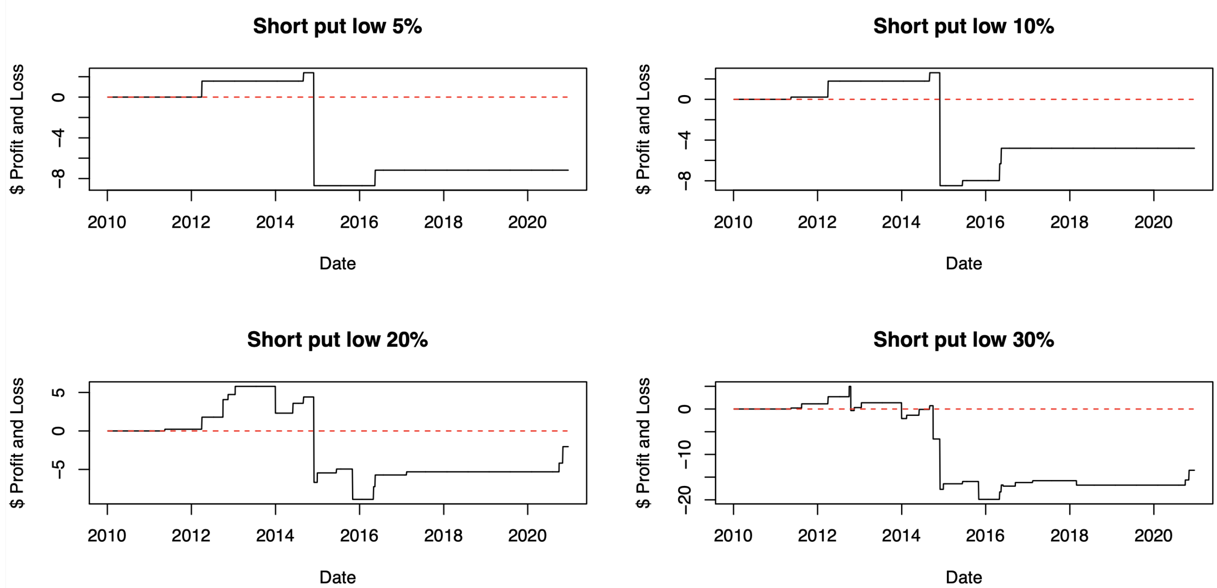
(e) Short calls when delta sentiment is high



(f) Short calls when delta sentiment is low



(g) Short puts when delta sentiment is high



(h) Short puts when delta sentiment is low

Figure 13: Cumulative profit and loss (PnL) graphs employing the rule based strategy, focusing on all possible combinations for high and low differences in sentiment using the quantiles 95%, 90%, 80% and 20%, 10%, 5%.

References

Abu, Z. B., & Siriopoulos, C. (2008). Transitory and permanent volatility components: The case of the middle east stock markets [Publisher: De Gruyter]. *Review of Middle East Economics and Finance*, *4*(2), 80–92. Retrieved January 31, 2021, from https://econpapers.repec.org/article/bpjrmeecf/v_3a4_3ay_3a2008_3ai_3a2_3an_3a3.htm

Barber, B., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors [Publisher: Society for Financial Studies]. *Review of Financial Studies*, *21*(2), 785–818. Retrieved January 13, 2021, from https://econpapers.repec.org/article/ouprfinst/v_3a21_3ay_3a2008_3ai_3a2_3ap_3a785-818.htm

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity [Publisher: North-Holland]. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Bolster, P., Trahan, E., & Venkateswaran, A. (2012). How mad is mad money? jim cramer as a stock picker and portfolio manager [Publisher: Institutional Investor Journals Umbrella Section: Article]. *The Journal of Investing*, *21*(2), 27–39. https://doi.org/10.3905/joi.2012.21.2.027

Borovkova, S. (2011, May 31). *News analytics for energy futures* (SSRN Scholarly Paper ID 1719582). Social Science Research Network. Rochester, NY. https://doi.org/10.2139/ssrn.1719582

Borovkova, S., & Mahakena, D. (2015). News, volatility and jumps: The case of natural gas futures [Publisher: Routledge _eprint: https://doi.org/10.1080/14697688.2014.986513]. *Quantitative Finance*, *15*(7), 1217–1242. https://doi.org/10.1080/14697688.2014.986513

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention [Publisher: American Finance Association]. *Journal of Finance*, *66*(5), 1461–1499. Retrieved January 13, 2021, from https://econpapers.repec.org/article/blajfinan/v_3a66_3ay_3a2011_3ai_3a5_3ap_3a1461-1499.htm

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices [Publisher: Society for Financial Studies]. *Review of Financial Studies*, *28*(1), 1–32. Retrieved January 13, 2021, from https://econpapers.repec.org/article/ouprfinst/v_3a28_3ay_3a2015_3ai_3a1_3ap_3a1-32..htm

De Long, J. B., Shleifer, A., Summers, L., & Waldmann, R. (1990). Noise trader risk in financial markets [Publisher: University of Chicago Press]. *Journal of Political Economy*, *98*(4), 703–38. Retrieved January 13, 2021, from https://econpapers.repec.org/article/ucpjpolec/v_3a98_3ay_3a1990_3ai_3a4_3ap_3a703-38.htm

Deeney, P., Cummins, M., Dowling, M., & Bermingham, A. (2015). Sentiment in oil markets [Publisher: Elsevier]. *International Review of Financial Analysis*, *39*, 179–185. Retrieved January 13, 2021, from https://econpapers.repec.org/article/eeefinana/v_3a39_3ay_3a2015_3ai_3ac_3ap_3a179-185.htm

Energy & financial markets - crudeoil - u.s. energy information administration (EIA). (n.d.). Retrieved January 31, 2021, from https://www.eia.gov/finance/markets/crudeoil/financial_markets.php

Engle, F. R. (1982, July). *Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation*. https://www.jstor.org/stable/1912773

Flair: State-of-the-art natural language processing (NLP). (n.d.). Retrieved January 13, 2021, from https://research.zalando.com/welcome/mission/research-projects/flair-nlp/

Gao, L., & Süss, S. (2015, June 28). *Market sentiment in commodity futures returns* (SSRN Scholarly Paper ID 1934397). Social Science Research Network. Rochester, NY. https://doi.org/10.2139/ssrn.1934397

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk [Publisher: [Wiley, Econometric Society]]. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kang, S. H., Kang, S.-M., & Yoon, S.-M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, *31*(1), 119–125. https://doi.org/10.1016/j.eneco.2008.09.006

Khaki, A., & Gunay, S. (2017, December 1). *Best fitting fat tail distribution for the volatilities of energy futures: GEV, GAT and stable distributions in GARCH and APARCH models* (SSRN Scholarly Paper ID 3085454). Social Science Research Network. Rochester, NY. Retrieved January 31, 2021, from https://papers.ssrn.com/abstract=3085454

Kinyua, J. K., Mutigwe, C., Cushing, D. J., & Poggi, M. (2021). An analysis of the impact of president trump's tweets on the DJIA and s&p 500 using machine learning and sentiment analysis. *Journal of Behavioral and Experimental Finance*, *29*, 100447. https://doi.org/10.1016/j.jbef.2020.100447

Kumari, J., & Mahakud, J. (2015). Does investor sentiment predict the asset volatility? evidence from emerging stock market india [Publisher: Elsevier]. *Journal of Behavioral and Experimental Finance*, *8*, 25–39. Retrieved January 13, 2021, from https://econpapers.repec.org/article/eeebeexfi/v_3a8_3ay_3a2015_3ai_3ac_3ap_3a25-39.htm

Shleifer, A., & Vishny, R. (1997). The limits of arbitrage. *Journal of Finance*, *52*(1), 35–55.

Uygur, U., & taş, O. (2012). Modeling the effects of investor sentiment and conditional volatility in international stock markets. *J. Appl. Financ. Bank.*, *2*, 239–260.

Wurgler, J., & Baker, M. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*, 1645–1680. https://doi.org/10.2139/ssrn.464843