Project summary - Arjun Kalsi & Arthur Boissavy

Investment Idea :

We chose to predict the price of oil using three types of data : the inflation rate, USD value, and inventory levels. Our first supposition is that the price of WTI oil is a function of these three components:

price of oil = F(inflation, USD, inventory)

It was noted in the lectures that inflation and oil prices are highly correlated, and in our final lecture we covered the interconnectedness of inflation rates, exchange rates and inventory levels. Thus we decided to use these 3 factors in order to predict the WTI price as we believed they would form an accurate representation of it. To do this we decided to do a rolling regression where we regress on the previous n days in order to predict the oil price today. The window of the regression would be the parameter we are looking to optimize.

Thus, the idea is simple: we buy when the market price is below the model price (undervalued) and we sell when the market price is over the model price (overvalued). In this way it is more or less a carry strategy where we are looking at the spread of two prices. We will optimize a threshold epsilon above and below which we will sell or buy. We will optimize the threshold in order to maximize the P&L and minimize the cost of transaction.

$$\pi = \begin{cases} 1 & \text{if } predict \geq price \text{ (underpriced)} \\ -1 & \text{if } predict < price \text{ (overerpriced)} \end{cases}$$

We believe this strategy may work if our model is good enough. For that we will optimize all the parameters involved such as the window of the rolling regression. Also we decided to add an epsilon parameter just like the carry strategy that we will also optimize :

$$\pi = \begin{cases} 1 & \text{if } predict - price > \epsilon \text{ (underpriced)} \\ -1 & \text{if } predict - price < -\epsilon \text{ (overerpriced)} \\ 0 & \text{if } |predict - price| \leq \epsilon \end{cases}$$

We use epsilon to avoid whipsaw and thus avoid a too important transaction cost.

We believe this strategy may work because for us it makes sense the that the three biggest factors which significantly change oil price are: inflation, USD and inventory.

Bonus idea: *geopolitical risk reduction*

We then tried to ameliorate our model by minimizing the maximum drawdown and reduce downside risk as much as possible. To do so we realized that the oil market was extremely sensitive to geopolitical issues in some regions of the world such. For example, major conflicts in the Middle East in 2014-2015 affected oil prices as well as volatility, and this was our main motivation for our secondary idea. In order to reduce downside risk we investigated some databases regarding conflict and terrorism in the world. For instance we worked on the global terrorism database from Kaggle (https://www.kaggle.com/START-UMD/gtd) and we identified all the attacks which targeted utilities (oil pipeline etc…) but we had some difficulty to use and integrate in the data in an effective way. We finally decided to use the rate of death due to terrorism in the middle east as an indicator of violence. This is because despite its simplicity, it is incredibly easy to understand and is probably a decent representation of conflict within a region. Note that this data was annual, thus we had to use interpolation in order to make the data daily. Looking at the data we set a threshold death rate for which we halt trading because the situation is too risky according to the data. By doing so we were able to slightly ameliorate our Sharpe ratio and to reduce our maximum drawdown.

Data used

We used the three data sets :
- Inflation rate *(https://fred.stlouisfed.org/series/T10YIE)*
- USD *(https://fred.stlouisfed.org/series/DTWEXBGS)*
- Inventory *(https://www.eia.gov/dnav/pet/pet_stoc_wstk_dcu_nus_w.htm)*

Inflation rate data and USD data are coming from the same website FRED (https://fred.stlouisfed.org/).
These are daily data.

For the inventory, as we are working on WTI we used the US energy data from (https://www.eia.gov/). We decided to focus on the US because
This dataset is made of weekly data. Thus to get daily data we decided to proceed to an interpolation.

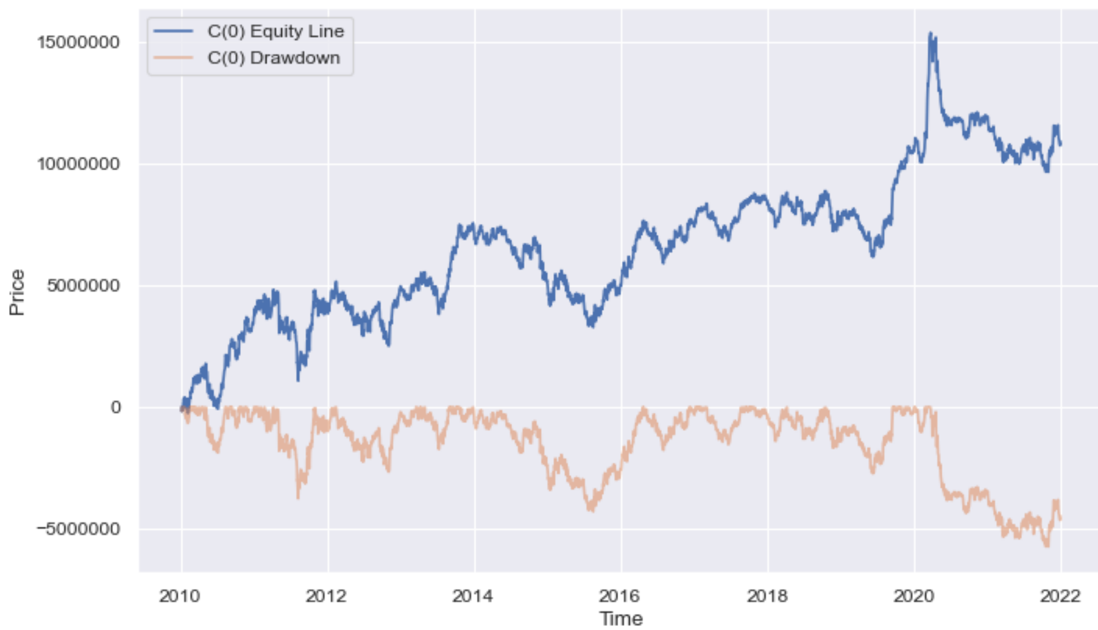
Before to use the USD which is an index we used the exchange rate between USD and Euro because we believed that this exchange rate is a good indictor however it is not fully representative of the value of the American currency. Thus we used the 'nominal broad USD index' which measures the value of the US dollar relative to other world currencies (not only euro). By doing so our model went from a Sharpe ratio of ≈0.20 to a Sharpe ratio of 0.42.

For our geopolitical risk reduction we first started to look at the global terrorism database from Kaggle (https://www.kaggle.com/START-UMD/gtd) and we identified all the attacks which targeted utilities (oil pipeline etc…) But we had some difficulty to use the data.
We decided to use the data of the deaths from conflict and terrorism per 100,000 in the Middle East & North Africa: (https://ourworldindata.org/terrorism). We believe that the conflicts in this region may have a strong impact on the price oil. As the data is yearly we interpolated daily.

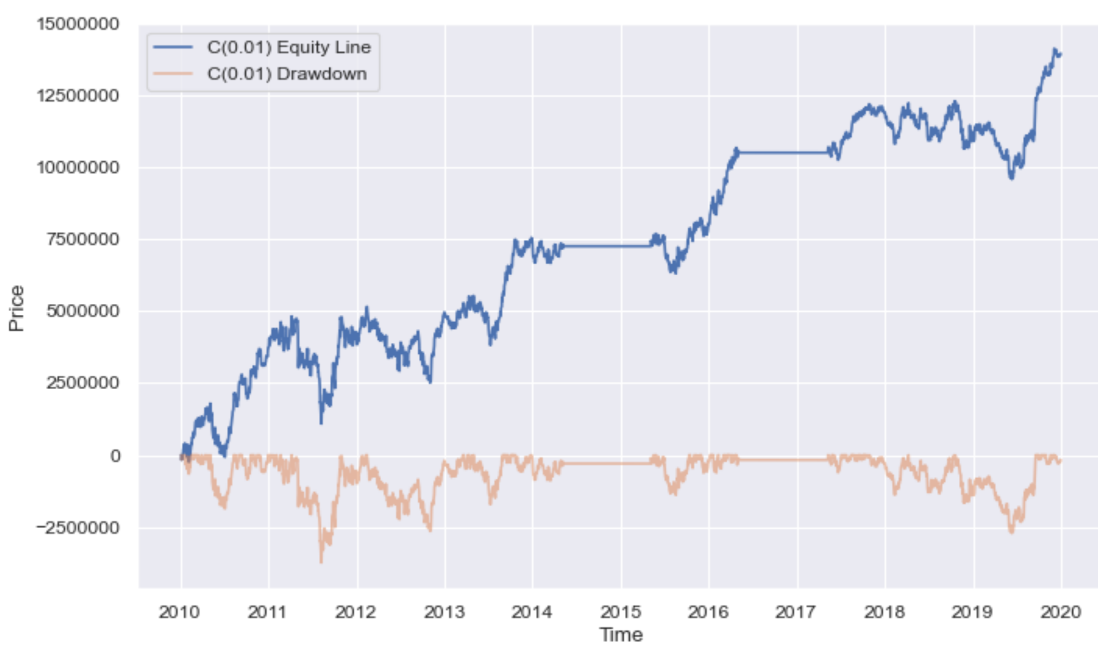
Results summary:

1) Performance Metrics

   a. *Rolling regression strategy equity line and rolling drawdown*
   - without geopolitical risk reduction:



- with geopolitical risk reduction (note that the strategy only goes up to 2020 due to a lack of data, however we can observe some beneficial outcomes of the strategy):



b. <u>Annalized P&L and Maximum Drawdown in $$</u>

- without geopolitical risk reduction:

  * C(0) Average Annual P&L: 909333.3333333308
  * C(0) Annualized Sharpe Ratio: 0.42060292596024096
  * C(0) Maximum Drawdown: 5717000.000000048 at 2021-11-01 00:00:00

- with geopolitical risk reduction:

  * C(0) Average Annual P&L: 1164000.0000000016
  * C(0) Annualized Sharpe Ratio: 0.6115587914972023
  * C(0) Maximum Drawdown: 3742999.999999998 at 2011-08-09 00:00:00

c. <u>Sharpe Ratio (annualized)</u>

- without geopolitical risk reduction:
The annualized Sharpe ratio is 0.42060292596024096 when we optimized the two parameters (epsilon for the carry and the window of the rolling)

- with geopolitical risk reduction:
The annualized Sharpe ratio is 0.6115587914972023 when we optimized the three parameters (epsilon for the carry, the window of the rolling and the threshold)

2) Stability of our output with respect to the choice of model parameters

Regarding of the parameter of the carry (epsilon) we computed different Sharpe ratio using different epsilon within {0. , 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1 ,0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2 , 0.21,0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.3 , 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4 , 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.5 , 0.51, 0.52, 0.53, 0.54,0.55, 0.56, 0.57, 0.58, 0.59, 0.6 , 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.7 , 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.8 , 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.9 , 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99, 1., 10, 9, 8, 7, 6, 5, 4, 3, 2}

We found that epsilon = 0 was the optimal parameter as you can see in this screenshot :
Epsilon equals to 0 kind of make sense for us because as we believe in our model as soon as there is a difference between our model and the ‘spot’ price we should enter a position.

Regarding the window parameter for the rolling regression we tested different values: {11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69}
By doing so we found that the most accurate window for the rolling was 50 (days) as you can see in the screenshot:

```
   0.0 :

48 :

C(0) Average Annual P&L: 573499.9999999992
C(0) Annualised Sharpe Ratio: 0.2652215701008763
C(0) Maximum Drawdown: 6971000.000000034 at 2021-11-01 00:00:00
49 :

C(0) Average Annual P&L: 491666.6666666649
C(0) Annualised Sharpe Ratio: 0.22735156658164488
C(0) Maximum Drawdown: 6941000.000000034 at 2021-11-01 00:00:00
50 :

C(0) Average Annual P&L: 909333.3333333308
C(0) Annualised Sharpe Ratio: 0.42060292596024096
C(0) Maximum Drawdown: 5717000.000000048 at 2021-11-01 00:00:00
51 :

C(0) Average Annual P&L: 857833.3333333299
C(0) Annualised Sharpe Ratio: 0.39677636875247707
C(0) Maximum Drawdown: 5874000.00000005 at 2021-03-23 00:00:00
52 :

C(0) Average Annual P&L: 837833.3333333284
C(0) Annualised Sharpe Ratio: 0.3875004923603048
C(0) Maximum Drawdown: 6140000.00000005 at 2021-03-23 00:00:00
53 :

C(0) Average Annual P&L: 847666.6666666627
C(0) Annualised Sharpe Ratio: 0.39204909639759267
C(0) Maximum Drawdown: 5322000.00000005 at 2021-03-23 00:00:00
```

Here we believe that the window of 50 makes sense because a too short window of days will not be able to explain the price correctly due to a lack of data, and a too long window may lead to incoherent results as the market tends to evolve a lot. 50 days represent a little bit less than two months and this looks like a fair window to us.

Here is the plot of our regression model when we use the optimal window of 50. One can observe that the rolling regression predicts the market price reasonably well, and they seem to revert to one another after major fluctuations:
Regarding the threshold for the geopolitical war reduction we tested different threshold for which we maximize the Sharpe ratio and try to reduce the drawdown. Our optimal threshold is 21 (when looking at the number of death from conflict and terrorism per 100,000 in the Middle East & North Africa.

3) Comments about strategy performance under different market regimes and the risks of the strategy

We observed that the maximum drawdown occurred in 2021 at the end of the covid crisis but did pretty good in 2020 at the beginning of the crisis. Thus we may think that our model tends to work better in a contango situation than during backwardation.

We tried to reduce the risk of our strategy by reducing the geopolitical risk correlated with oil, by using a war and sentiment. The problem with this idea is the lack of data. We searched different dataset but finally we decided to use the number of death from conflict and terrorism per 100,000 in the Middle East & North Africa because we think it is a reasonable indicator of geopolitical instability in the main region where OPEC countries are located. Overall, however, we believe we should use a better indicator because we use yearly data which is way too large for a daily trading strategy, and as a result we were forced to interpolate which is always worse than obtaining real daily data.

The major risks in this strategy lie in reproducibility. We don't have extremely valuable data to use to express war sentiment and its impact on oil prices, and as a result there may be situations in the future where deaths are localised and high, but this simply has no impact on oil price. To add, rolling regressions predict results quite well but in the case of extreme and sustained fluctuations, or even strict government regimes, it may have trouble predicting values and might behave in a 'sticky' manner where predicted price stays low/high until one data point in the window is dropped out of the regression.

Overall, we thought this strategy was fairly effective. The rolling regression worked surprisingly well, and we believe when combined with the appropriate data in order to reduce downside risk, this strategy may actually have some real potential with a strong P&L and Sharpe ratio.