Predicting PNL Using Covariance:

A sum of parts vs the whole analysis

Project by Reid Davis

Applied Data Analysis

Executive Summary:

The goal of this project was to try to predict the PNL of trading company. My major hypothesis was that a group of strategies each with a risk calculation derived from its positions and a covariance matrix, did not accurately model the real risk of the group as a whole. I theorized that this mismodeling of risk could be seen by comparing the sum of the standard deviations vs the standard deviation of all the strategies as one, and I could predict the PNL as a function of this difference.

In order to test my hypothesis, I gathered trading positions from the start of day from January 2022 to March 2023 for 7 trading strategies that all trade the same Crude oil and derivatives of crude oil Futures. I calculated the standard deviation using a covariance matrix that was derived at the start of each day from empirical data of the last 25 trading days. I wrote python scripts to pull position data and covariance matrix transforms, normalize delta positions of contracts with different multipliers, and calculate and export standard deviations as a CVS file .

As additional predictors, I used the opening, and change on day of the CBOE crude oil volatility index, as well as the volume, open interest, percent change on day, and range for the WTI crude oil continuous contract.

I used both various linear and logistic regression method from daily 2022 data to try to predict the daily PNLs of the first five months of 2023. The predictors were unable to accurately model the PNL observations, both disproving my hypothesis on a daily time frame, never reaching a model at exceeded an R-squared value of above 0.30.

Although modeling on a daily time frame was never accomplished much was learned, and future investigation need to be evaluated on smaller time frames.

Introduction:

Currently, I work at a propriety trading firm. My group has many automated trading strategies, that trade the same groups of products. Risk for each product is measured by the std which is calculated based on open positions of the strategy, and a covariance matrix which is generated daily based on historical data of a specific length of time. PNL is tracked for each product and the group as a whole, but no measure of risk is generated by the positions of all the strategies.

I would like to calculate this measure of risk. Then I would like to sum the individual standard deviations and see if I can use the difference between the sum of stds and the std of the whole group to predict the PNL, possibly finding the optimal difference to maximize profits. In addition.

Our goal as a group is to maximize the groups profits.

The goal of most of the strategies is to take on positions we determine to have edge, or positive expected value. Strategies aim to add these positive expected value positions and find exit positions to offset the increase in standard deviation where the sum of the expected value of the entry and exit remains positive and the standard deviation remains unchanged. Most trades have some residual positions where the strategy keeps the positions. Sometimes an incomplete exit position is found, and the strategy holds the position. Over time these residual positions tend to have negative drift. This means that if left for a long period of time, we have found the residual position will have a negative PNL over these longer time frames.

We do not track the standard deviation of all the strategies combined. We currently manage risk by strategy. When I strategy is over a predetermined standard deviation, we actively work to reduce the standard deviation.

I theorize that often we reduce the risk of one strategy, but we are actually increasing the risk of the group as a whole. As an example:

strategy A is has a long 500 position in contract X

strategy B has a short position of 500 in contract Y,

X and Y are 99 % correlated

We might choose to reduce the std of strategy A by selling out some of contract X. This would reduce the std of strategy A, but increases the risk to the group as whole. If we can identify and reduce the risk to the group, we will be more effectively managing risk, reduce the positional drift of the group, and optimize profits.

Accurately Predicting when PNL’s will be have a high probability of becoming negative would be significant in optimizing the profits of the trading group; we can proactively reduce both the current risk positions of the group as well as change the risk parameter our trading strategies.

Detailed Description of Related Work

Although here are no direct papers that focus on my direct hypothesis of summing standard deviation vs standard deviation of the whole as one, There is existing literature on the performance of covariance trading during high volatility events. The following are just two examples

Increased correlation among asset classes: Are volatility or jumps to blame, or both? By Yacine Aït-Sahalia , Dacheng Xiub

Description of related work:

During the 2008 financial crisis, covariation between assets changed dramatically, often increasing. Using high frequency data, this paper looked to evaluate the covariation changes between two time segments: co-movement of the continuous part of the price processes, or an increase in the co-movement of their jumps. Both situations will result in different hedging of risk using portfolio optimization principles. The evidence of the paper supports the former as a larger contributor to the increased co-variation, finding that risk-on, risk-off scenarios lead to large scale portfolio changes over a more continuous increases in hours or days rather than instantaneous changes due to jumps.

Data collection

In order to evaluate assets over a continuous 24 hour period, this paper used data from four futures contracts: S&P 500 eMini futures, the 10-year Treasury futures, crude oil futures, and the euro/dollar currency futures. High-frequency front-month futures contract prices were obtained from Tick Data Inc. from January 1 2007 to July 31 2012 (continuous contracts were used) . Only electronic Globex data was used and volume from pit trading was excluded. 1 minute returns were sampled from 6 PM to 5PM (when Globex is open). Jump events were defined as the 21 most common known news releases during the month

Model used: The paper assumes that asset log-prices form a d-dimensional Itô semimartingale, defined on a filtered space. Time periods are group into continuous and jump events on the basis of the presents of a news event. Correlations are computed for the two groups based on the one minute data between the four continuous contracts and analyzed to find correlations between the contracts. A simulation is run to determine the quadratic estimators using a monte Carlo simulation, to determine the mean and standard deviation under various volume by jump percent conditions. If I understand the paper correctly, the large number of sample points allows the assumption of a normal curve distribution, which is verified by the simulation results.

Results: This paper makes many conclusions from the analysis:

1. The changes in correlations between products, happen slowly (change over time) due to changing news related events. These seem to be changed, not by instantaneous events on news releases, but large-scale macro related trends. The crisis resulted in commonality of asset returns over time, more than nearly instantaneous co-jumps from news events, or event surprises.
2. Although the covariation did show an increase during the crisis time period; it did not lead to a significant change in the breakdown between their respective continuous and discontinuous (jump) contributions, with both moving consistently with one another.
3. Jumps and co-jumps rarely occur overnight despite the relative illiquidity of the market. Most jumps are driven by the surprise component in macroeconomic news that are typically announced prior to the opening of US markets.

From this paper, I like the idea of dividing the data into segments, as returns in a high standard deviation environment might e very different than a low standard deviation environment. It is likely that predictions might be more accurate if the sample are group and evaluated separately will look at dividing my data into quartiles.

Research paper 2: Correlation Breakdowns, Spread Positions, and CCP Margin Models by David Li, Fernando Cerezetti, Roy Cheruvelil

Description of related work: The default of a participant at Nasdaq Clearing in 2018 and the recent COVID-19 events brought to the attention of risk managers to the need to re-evaluate the risk associated with correlation breakdowns. (Although this paper focuses on the CCP margin risk, this risk metric is generally representative of the real value at risk of a portfolio). A case study of the spread positions in energy markets is used to show that correlation breakdown happens more often then originally thought and it causes large variability in margin [and as I look at it, variability in risk]. Spreading one futures contract vs another contract in the same product is often the major method of hedging risk and reducing margins. Basing this correlation off past data, results in excess risk when correlation breakdowns occur especially observed during shocks. In some of these cases the highly correlation of two contracts can become a negative correlation, making a hedged position even more risky. This paper aims to point out the impact of the general volatility and specific volatility of each asset can have a large effect on how spreads should be modeled. Then an “enhanced” Var model is used to try to more accurately the risk of correlation breakdowns.

Data collection:

Spread prices we recorded of the ICE Endex German Power Financial Based Future and the ICE Endex Nortic Power Financial Based Future, covering the first 12 monthly expirations from June 6, 2016 to June 1st 2020. This includes the one week period in September 2018 that spread prices spiked to more than 3X previously observed levels, as well as the Covid shock period. It was not implicitly stated but high frequency data was used from the ICE.

Model used:

This paper Filtered historical simulation Var and a joint MLE method to calibrate multivariate dynamic copula model parameters, alpha, beta and vc. From my reading outside of this paper, this dynamic copula model is a commonly used way to fit asset returns using high frequency data. Correlation breakdown is considered to be when dislocations are outside the predicted 95% confidence interval based on long term averages.

Joint maximum likelihood (JML) estimation, seems to be similar to a regression model when parameters are unknown, but data is used to fit observations by tuning multiple parameters at once in order to minimize error.

This model statistically significant for all terms above 90% significance.

Results:

1. There is no discernable pattern to how correlations break down occurred with respect to the 12 month curve. Correlation breakdowns were observed more frequently than expected from both increasing and decreasing correlations.
2. Despite the lack of pattern, correlation breakdowns in 2018 and 2020 both resolved ultimately in the same way, a quick reversion to the long term average.
3. All models with varying parameter suffered shortfalls due to correlation breakdowns although the models that were more accurately modeled the data up until the breakdown suffered small shortfalls.
4. For margin purposes a margin floor needs to be used, as well as a volatility floor, because these correlation breakdowns ultimately result in losses that are difficult to model.

This paper suggests that correlation breakdowns occur in an unpredictable manner, but they seem to be observed during high volatility of the underlying. It might be possible to add an additional factor into my modeling, realized volatility of the underlying. In addition, it might also be useful to predict PNL in two groups large and small ranges for the daily change in the underlying

When looking at the linear regression plot of Total standard deviation it is evident the truly randomness of the scatterplot

A picture containing text, screenshot, diagram, line

Description automatically generated

Detailed description of my solution and related work:

1. Creating the daily standard deviations:

I created a script to query the trade data base. I aggregated this trade data by strategy, and underlying contract. I normalized each contract standardizing all contracts relative to WTI Crude Oil (CL) deltas. I mapped the position deltas into a vector format matching the daily covariance matrix that I was able to pull from the database. The variance was then found as the dot product of the position vector transposed the covariance matrix and the position matrix. Then the square root of the variance is the standard deviation.

This was done for each strategy daily along with all the positions cumulatively, and finally entered into a data frame.

1. Downloaded historical data on the Crude Oil VIX index from barcharts.com, and WTI crude prices, volumes and open interest from StockCharts.com
2. Saved all data to a CVS and combined into one Dataframe in Google Colab

I tried to fit the data with linear regression, logistic regression, K-nearest neighbors, decision trees, and best subset selection. None fit my data well. I believe that daily timeframe needs to be in a smaller 10 to 60 minute intervals to get better results.

What worked:

The data collection took a lot of time and effort, but in the end the script I wrote were able to pull the data need and produce standard deviations that I could verify were accurate when compared to gui outputs on our current systems.

What didn’t work out and why

Overall, I was unable to find a model that would accurately predict the PNL. There are many reasons for my failure to accurately model the PNL:

1. More predictors: It is very likely that I just didn’t have the correct predictors to predict the pnl. From the papers research, I believed volatility would have a large effect on the PNL. For this reason, the additional predictors I chose focused on movement and volatility measures. If I spend more time gathering predictors instead of focusing on deriving standard deviations, I could have created data on inventory levels, crack spreads, or switch values. This might have been more valuable in hindsight.
2. Time Frame: From the papers I researched, I believed covariance breakdown would occur and this would result in significant similarities in positions between products. This would result in a convergence between the sum of the standard deviations and the standard deviation of positions grouped together. The second research paper , Correlation Breakdowns, Spread Positions, and CCP Margin Models, suggested these break downs occur for multiple days then revert back to the expected covariance relationship. I believed that a daily timeframe would be sufficient to display these finding and support my hypothesis. I now believe that the choice of a daily time frame is may have been too large and these correlation breakdowns may only be visible in the Crude complex in minute or hour time frames. Looking at too large a time frame might result in a poor model no matter how many predictors were chosen.

-Detailed presentation of the results obtained and analysis.

**Linear regression –** I started with 16 predictors: (Project 4 notebook)

[PNL, STD\_A, STD\_B, STD\_C, STD\_D, STD\_E, STD\_F, STD\_G, STD\_Sum, STD\_Total, STD\_Diff, Open\_Vol, Perc\_Change\_Vol, Perc\_Change\_Price, Volume, Open\_Int, Range]

I removed any predictors with a correlation to PNL less than 0.2. I then compared the predictors for collinearity, and I made sure that no pair of predictors had a correlation above 0.75. I was left with STD\_E, STD\_F, Open volatility, percent change in price and Range as my predictors. I ran a linear regression and saw that STD\_E was not significant, so I removed it, and re-ran a linear regression. The regression model only has a R-squared value of 0.25. From this I can conclude the linear regression model does not accurately model the observations.

Just to validate my conclusion I used the model to predict the PNL for the 5 months of daily test PNL data. I found the RMSE of over 251,000, which is massively large considering most of the PNLs were between -250,000 and 250,000.

**Best Subset Selection (**Project 4 notebook)

I then ran best subset selection. I again found that the lowest RMSE was over 200,000 and had a R-Squared value of only 0.251, which again verified the model does not accurately fit the data.

**Logistic Regression. (**Project 5 notebook)

Since the linear regression yielded a poor model, I next changed the PNL data to a categorical variable. I found the mean PNL and labeled all daily PNLs above the mean as High(1) and all PNL’s at or below the mean as Low (0) . I trained the model of the daily PNL values from 2022 and tested the model on the 5 months from 2023.

A picture containing text, screenshot, diagram, colorfulness

Description automatically generated

I found a correct prediction rate of 0.513, a precision score of 0.478, and a recall score of 0.647. This yet again shows the poor performance of this model.

**K-nearest Neighbors(project 6 notebook)**

I ran a k-nearest neighbors with k = 1,2,3,5,10 on the categorical variable of above and below the mean so see if we could predict the whether a PNL would be above or below the mean. I found that the best prediction rate was with k=3 and only had a precision rate of 0.527, with most of the test having a test precision below 0.5. This again does not model the data well.

**Decision Trees (project 7 notebook)**

I divided the PNL in above and below the mean. Trained the data on the PNL from 2022 and tested on the data from 2023.

precision recall f1-score support

0 0.42 0.47 0.45 40

1 0.28 0.24 0.25 34

accuracy 0.36 74

macro avg 0.35 0.36 0.35 74

weighted avg 0.35 0.36 0.36 74

With a precision and recall scores all well under 0.5, this model too doesn’t come close to accurately modeling the data.

Suggestions for future work along the same lines:

1. In the future, I would recommend prioritizing pulling data on as many predictors that can be easily pulled from other source over spending time deriving predictors (too time intensive).
2. I would also recommend that when data is collected in time intervals that you try to get the shortest time frame available. Time frames can always be combined, and numbers averaged to create larger time frames, but data collection on intervals that are too large, just leaves you with a poor model. I don’t know if I could have gained PNL access for smaller time frames, but it certainly would have been better than the current project and results.
3. In future work ,I think that creating a measure of comparison between the covariance matrix used for trading (using empirical data from weeks of data) and creating a rolling covariance matrix over a smaller time frame might be a valuable predictor of PNL. This would be very time intensive, but it might be an idea to investigate further.

Description of efforts including relative efforts on different activities (estimates) :

* 60% (maybe more ) of my time and effort was spent deriving the standard deviations. I started working on this part during week 3.
  + 10% spend learning how to make database calls , and scripting
  + 20% normalizing deltas from different products (CL to HO) and different contracts of the same product that had different sizes (Cl to MCL to UCL)
  + 10% formatting data, consolidating positions, and mapping position vectors to fixed covariance matrix mappings.
  + 20% working through various problems on why my standard deviations didn’t match the outputs of our current displays. As it turns out, some strategies alter covariance diagonals to alter risk, and other strategies decide to run covariance matrix calculations off different timeframes. This was only found after hours of code inspection and frustration ultimately resolved in meetings with software engineers.
  + 5% finding other predictors from data sources with downloadable data
  + 5% data preprocessing, and combining data sets
  + 20% Analyzing data – After I realized how poor my models were at predicting PNL, I tried to divide data into high and low segments to see a different view would yield better models
  + 10% writing report and making the presentation.

What was learned:

Despite the poor results, I did learn a myriad of things:

1. I learned how standard deviations are calculated from position data and covariance matrices. As well as how to interact with my current databases.
2. I learned the importance of time intervals in data collection.,
3. I learned that in data analysis, it is always valuable to have as many predictors as possible.

Skills already processed --- fill in later

Already stated previously, but I spend more time deriving standard deviations as predictors, than fine tuning my predictors or reading research papers.