Project Report

- ## Project 1: Volume Synchronized Probability of Informed Trading Indicator (VPIN)

1. Introduction

One of the most significant volatility changes during the last 10 years is the Flash Crash on May 6, 2010. The one-day point of Dow Jones Industrial Average decline 9% to 998.5 points and then bounce back to 1010.14 points in half an hour. Similar phenomena happen on futures market (S&P 500). Both the regulators and investors want to know the reasons behind these fluctuations, and they attach great importance to proposing some signals, or indicators which could help them detect and avoid the risk in the market.

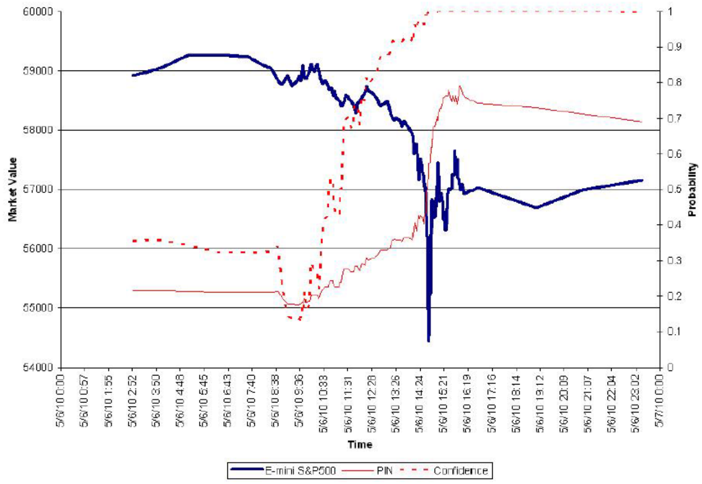
It could be observed that these sudden changes in prices, or high volatility in the market, is usually accompanied with large trading volumes. Usually, high-frequency data is a common choice for dealing with sensitivity-related issues, especially for such special pair anomalies happening in a short time. Thus, VPIN is proposed by several scholars as a market risk indicator.

Volume synchronized Probability of Informed trading (VPIN) could be viewed as a volume-related indicator, which reflect the liquidity-induced volatility for financial products (such as stocks, futures, options and so on). Based on the notation used in another paper by Wu, Bethel, Gu (2013), the parameters needed for calculation are:

- Nominal price of a bar π
- Parameter for the Bulk Volume Classification (BVC) ν
- Buckets per day (BPD) β
- Threshold for VPIN τ
- Support window σ
- Event horizon η

In order to give a more intuitive perception of the relations between VPIN and time, Figure below shows VPIN values of E-mini S&P 500 on May 6th 2010:

Figure 1 E-mini S&P 500's VPIN Metric on May 6th (López de Prado 2011)



For this project, firstly, I need to read the academic paper thoroughly to understand the logic of this method and the mathematical formula. Then connect IB API with Python to receive high-frequency data for different asset classes and back-test the validation of the model. After that, the calculation algorithms and free parameters need to be optimized to reduce the computing time. Finally, I used Plotly to visualize the VPIN value for all historical data and update the plots at fixed intervals and deploy and display them online as trading reference.

2. Resources

The high-frequency price and volume data from different asset classes (stocks, foreign exchange, fixed income, etc.) is extracted from a financial service platform, Interactive Brokers (IB) API, and processed on Python.

When the calculation process is completed, all VPIN values will be stored in MongoDB for future usage, such as visualization and statistical analysis.

3. Methods

Key formula for VPIN calculation:

$$V_j^b = V_j Z\left(\frac{\delta_j}{\zeta}\right)$$

Where V_j denotes the total volume of bar j and V_j^b denotes buy volume. Z denotes the cumulative distribution function of standard normal distribution, ζ is the standard deviation of price difference sequence $\{\delta_j\}$ ($\delta_j = P_j - P_{j-1}$). Also, the sell volume of bar j is $V_j^s = V_j - V_j^b$.

Only a few most recent buckets of volumes are needed for the computation, this is defined as σ , a fraction of one-day's total buckets. Formula for VPIN:

$$\mathrm{VPIN} = \frac{\Sigma\left|\left|V_j^b - V_j^s\right|\right|}{\Sigma \mathrm{V_j}}$$

Normalization of VPIN values:

$$\phi(\mathrm{x}) = \frac{1}{2}[1 + \mathrm{erf}(\frac{x-\mu}{\sqrt{2}\sigma})]\,\Sigma\left|\left|V_j^b - V_j^s\right|\right|$$

where error function (erf) is measured by the normal distribution, μ is the mean of the VPIN values and σ is the standard deviation.

The paper defined 0.99 as a threshold, if at certain time the value exceeds this threshold, we say a VPIN event happens. After we have detected one such event, we need to determine it is a true positive or false positive event. The notation α is defined as the false positive rate (FPR):

$$\alpha = \frac{\#\text{ of False Positive Events}}{\#\text{ of VPIN Events}}$$

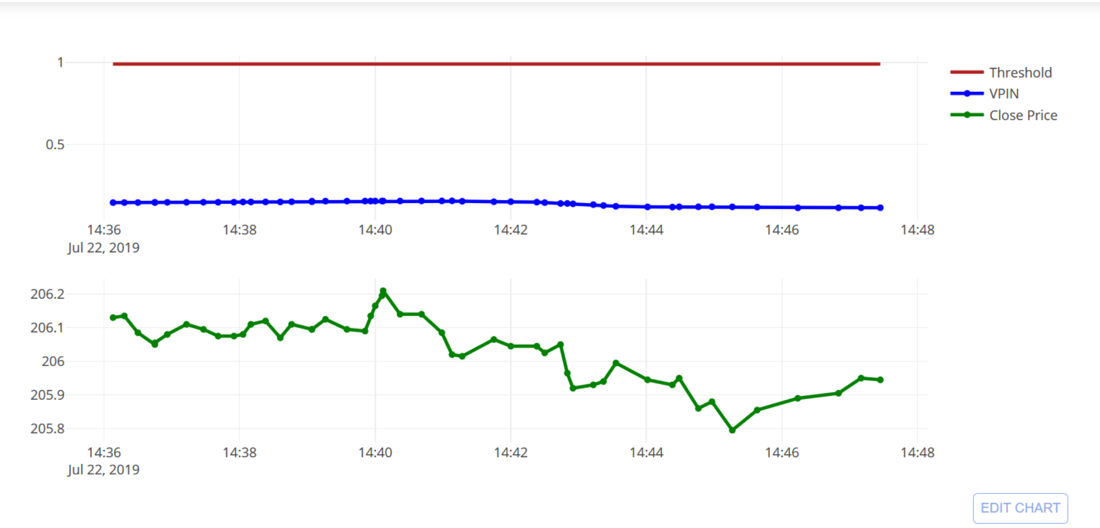
We need to test different parameters to obtain the lowest value of α, then the optimal free parameters will be chosen.

4. Results

After the optimization method, we choose the best parameters with values:

π	β	σ	η	ν	τ
Weighted median	1784	0.0756	0.0093	49.358	0.9936

And the plot is as the following:



5. Conclusion

Through the VPIN project, I learned that trading volume has direct influence to the market volatility, and VPIN could be viewed as a volatility indicator with directions because it separates buy and sell volumes. The regulators could change the threshold according to their perception on how fragile the market is.

Visualization is also a challenging part as it requires live plots updated automatically as new information output from IB. By this means VPIN could serve as a reference for constructing dynamic trading strategies, with components changing between different asset classes.

The validity of VPIN is reflected in its successful prediction on Flash Crash of 2010 and some other obvious event-driven anomalies. However, there are also some debates on the effectiveness of VPIN because it has more assumptions on the correlations between volume and market moving directions, the distribution and so on. Sometimes it conflicts with the real situations, and the prediction ability will be weakened, then the parameters and assumptions must be changed.

- **Project 2: Financial Turbulence, Absorption Ratio and Risk Management**

1. Introduction

The greatest financial catastrophe in 2008 has been witnessed by all government and market practitioners, thus more and more importance is attached to developing effective tools to monitor and quantify all types of risk. Under this circumstance, many risk measures, simple or complex, micro or macro, have been proposed by scholars, trying to grasp the secret of how our financial market is operating.

Two of such risk measures are Financial Turbulence and Absorption Ration, both of which do not require disgusting mathematical proof. I have conducted research on the application of these methods in different asset markets and validated the model in detecting extreme phenomenon during the last 20 years with intraday prices.

2. Resources

All data are received from the financial service platform Quandl. Considering the dividends for stocks, adjusted closing price is used for calculation weekly log returns. If data is not available (NaN), the filling method is backfilled, which means the missing ones are refilled by the last existing value.

3. Methods

Based on the paper Kritzman, M., & Li, Y. (2010), derived from "Mahalanobis distance", which is a distance first used in normalizing the differences of skull characteristics in archeology

$$d = \sqrt{(y - \mu)\Sigma^{-1}(y - \mu)'} .$$

The statistical term "Turbulence Index" is defined as

$$D_t = (y_t - \mu)\Sigma^{-1}(y_t - \mu)'$$

Where
D_t = Turbulence for particular time t
y_t = vector of asset returns for period t
μ = sample average vector of historical returns
Σ = sample variance matrix of historical returns

One direct interpretation of financial turbulence is that it could reflect the current performance of a group of assets compared to its average historical pattern. Turbulence could be used as a variability measure of any group of n return series between point t and its mean value over a fixed period, with the covariance of the same time assets during the same time period.

Another method is called "Absorption Ratio" (AR), which measures the systemic risk through principal components analysis (PCA). It calculates the implied systemic risk as a fraction of the total market from a set of asset returns, according to top eigenvalues of the covariance matrix. It captures the uniformity of the market among different assets, serving as an early warning signal of market stress.

The definition of absorption ratio is the total variance of an asset group explained by a fixed number of eigenvectors:

$$AR = \frac{\sum_{i=1}^{n} \sigma_{Ei}^2}{\sum_{j=1}^{N} \sigma_{Aj}^2}$$

Where,

AR: Absorption ratio

N: number of assets

n: number of eigenvectors used to calculate absorption ratio

σ_{Ei}^2: variance of the i-th eigenvector, sometimes called eigenportfolio

σ_{Aj}^2: variance of the j-th asset

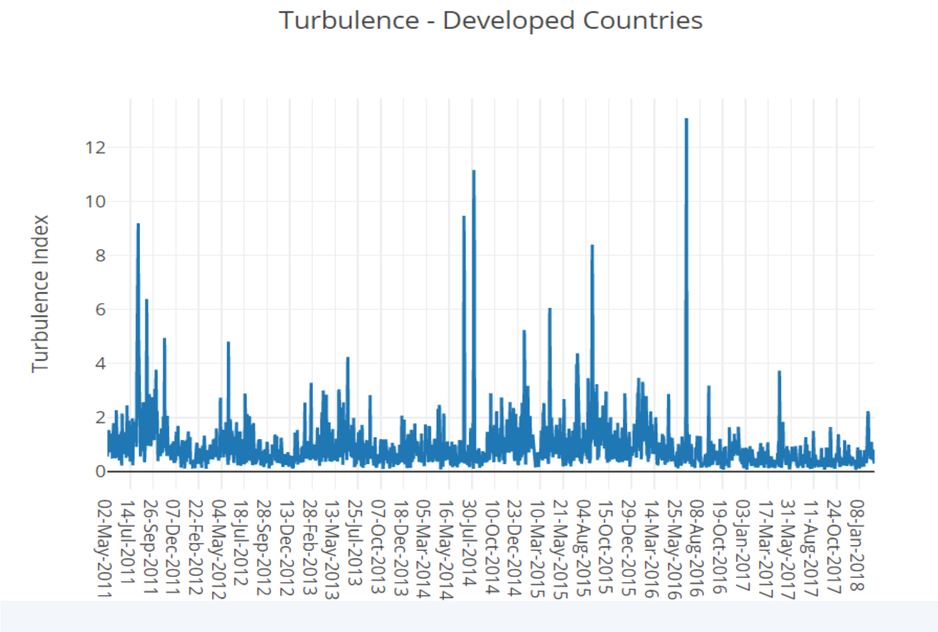
The calculation period is every 500 trading days, and the number of eigenvectors is 1/5 of the total asset number. Also, there could be many choices on the construction of portfolio, AR could be used as a measure of the overall group’s risk compactness. We could change the weights of different assets dynamically in order to avoid the risk and improve the return.

Finally, in order to create scores on Turbulence, I do simple regression of quantile of return based on historical value on the value of Turbulence to construct the “Turbulence Score” which could reflect the prediction of returns for current turbulence. Update the plot for trading reference.

4. Results

The test portfolio classes in Turbulence include: General Market, Developed Stock Market, Emerging Stock Market, U.S. Bond Market, Global Foreign Exchange, etc.

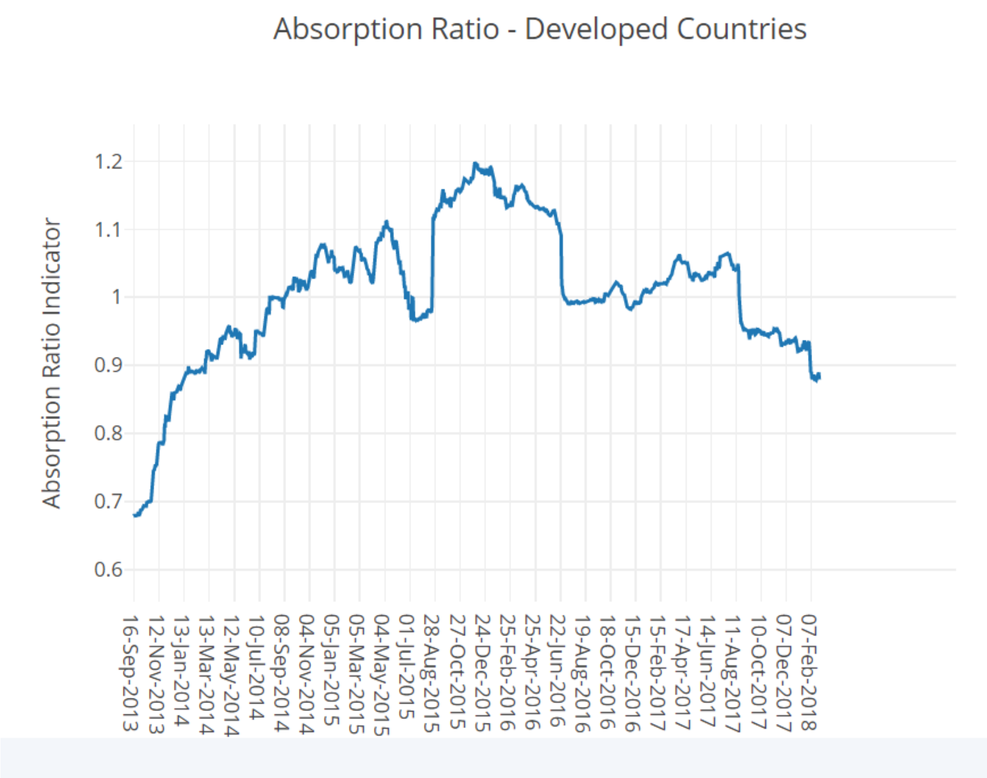
For instance, the plot of Developed Stock Market:



Turbulence has two obvious advantages, the first one is it could be estimated for any asset classes regardless of its liquidity (even with low liquidity, volume and prices still exist, while the plots are not that smooth), and it captures interactions among assets in one group in addition to their individual values (through covariance matrix).

For Absorption Ratio (AR), it is a more applicable method as it may also correlate with other markets, such as the real estate market. It is a leading indicator which could predict the market's future performance. Generally, when the market goes down, there will be a significant increase in the absorption ratio, while when the market goes up, the absorption ratio will decrease.

The Absorption Ratio of Developed market is plotted as the following:



5. Conclusion

As we could view Turbulence as a volatility index, which incorporated both the individual volatility and their correlations, it's explicit and simple. However, summarizing all information in a single index value may sacrifice if the data have high dimensions. It also means two groups of assets may vary greatly with each other even if they show the same turbulence.

For Absorption ratio, it is a good measure of market uniformity. But it also needs careful choice of the calculation period, and the number of eigenvectors chosen in numerator. The longer period for covariance matrix, the more information it includes but also at the risk of changed structure and noisy data.

- **References**

1. Kritzman, M., & Li, Y. (2010). Skulls, financial turbulence, and risk management. Financial Analysts Journal, 66(5), 30-41.
2. Kritzman, M., Li, Y., Page, S., & Rigobon, R. (2011). Principal components as a measure of systemic risk. The Journal of Portfolio Management, 37(4), 112-126.
3. Song, J. H., Wu, K., & Simon, H. D. (2015). Parameter Analysis of the VPIN (Volume synchronized Probability of Informed Trading) Metric. Quantitative Financial Risk Management, 337.
4. Wu, K., Bethel, E., Gu, M., Leinweber, D., & Rübel, O. (2013). A big data approach to analyzing market volatility. Algorithmic Finance, 2(3-4), 241-267.