FE 582 Proposal

Topic: Reveal possible features that influences trading volume of stocks

Background

Trading volume represents the total number of stock shares, bonds or commodities futures contracts traded during a certain period of time. Next to price, it is one of the most closely watched indicators. For companies, low trading volume may stop investors from investing in the center stock shares which results in difficulty of raise money to develop companies. For investors, trading volume is a warning mark to sell or buy the shares of stock. Therefore, there is great importance in analyzing the factors that influence the trading volume in the stock market.

There are some possible factors that influences the trading volume of stocks:

(a) size of the company[2] Barron (1995) provides empirical evidence that the trading volume is negatively related to the firm size.

(b)stock life time[3]

(c) stock beta[1]

(d) cash flow position of a firm[2] According to Huberman's theory, high cash flow implies high firm value. Thus, we expect that high cash flow lead to high level of normal trading volume.

(e) leverage risk[2] investors would prefer a low leveraged firm than a high leveraged firm if the investors were risk averse. Thus, we expect that high leveraged firms would have a lower level of normal trading volume.

(f) liquidity ratio[2] Huberman (I984)'s theory of liquidity implies that if the liquidity of the firm is good, then there will be no need for the firm to borrow money which is deemed to be an expensive source of financing.

(g) return on asset[2] Models from Modiglianai and Miller (1958) relate equity risk to the risk of underlying assets of the firm.

Reference:

- Ma, G. (2007). Three essays on trading volume (Order No. 3269266). Available from ABI/INFORM Collection. (304888396). Retrieved from http://ezproxy.stevens.edu/login?url=https://search.proquest.com/docview/304888396?accountid=14052
- Chan, S. F. (2000). Essays on stock liquidity (Order No. 9985148). Available from ABI/INFORM Collection. (304668385). Retrieved from http://ezproxy.stevens.edu/login?url=https://search.proquest.com/docview/304668385?accountid=14052
- Li, W., Wang, F., Havlin, S., & Stanley, H. E. (2011). Financial factor influence on scaling and memory of trading volume in stock market. St. Louis: Federal Reserve Bank of St Louis. Retrieved from http://ezproxy.stevens.edu/login?url=https://search.proquest.com/docview/1697577038?accountid=14052
- Wang, X. (2008). Three essays in insider trading (Order No. 3328988). Available from ProQuest Central. (304577039). Retrieved from http://ezproxy.stevens.edu/login?url=https://search.proquest.com/docview/304577039?accountid=14052
- Gunawan, H., & Lina, E. O. (2015). Mandatory and voluntary disclosure of annual report on investor reaction. International Journal of Economics and Financial Issues, 5 Retrieved from http://ezproxy.stevens.edu/login?url=https://search.proquest.com/docview/1701253290?accountid=14052

Planned Methodology:

- Binarization: use numeric data mining algorithms on categorical data.
- Data-centric Methods: the statistical behavior of the data is used to detect outliers.
- Time Series Analysis: The stock data is in the form of a time series, which means the data generated by continuous measurement over time and then visualizing time-series data to make inference about important components.
- Dynamic Time Warping: stretched the series along the time axis in a varying way over different portions to enable more effective matching to measure time-series similarity.
- Principle Components Analysis: remove irrelevant features or transform the current set of features to a new data space.
- K-means Clustering: divide the stocks into distinct groups based on returns and volatility
- GARCH: model a change in variance in a time series that is time dependent, such as increasing or decreasing volatility.

- ARIMA: exploratory analysis to obtain AIC, which is a measurement of quality used across various model to find the best fit.
- Autocorrelation Function: check each for characteristics of stationarity by looking at ACF, expecting the ACF to go to zero for each time lag.

Planned Implementation:

- Collecting data from websites, like Kaggle.
- Extracting valued and relative features from selected datasets.
- Cleaning data by using R.
- Exploratory Data Analysis.
- Making more analysis by using planed methodologies, including visualization.
- Drawing a conclusion

Sources of data:

Kaggle, Bloomberg and other public datasets.