Topic: Evaluating the Role of Volatility in Predicting SET Index and Exploring Volatility Behavior From SET Index to Individual Stocks

Abstract:

This study evaluates the role of technical indicators in four categories (trend, momentum, volatility, volume) in predicting changes in SET index movement in multi-class classification framework. Models conducted in this study are logistic regression and decision tree. This predictive model uses training data from 2008 to 2017, validating data in 2018 and testing data in 2019. Specifically, the role of volatility in prediction model is evaluated by using feature importances from decision tree model, and its associated interpretation is done using the results from logistic regression. To start exploring the behavior of this variable, 14-day standard deviation on SET index is chosen as representing volatility under the linear regression framework. However, the result from using SET index does not always generalize to individual stocks due to their heterogeneity. Therefore, the snapshot of volatility behavior in individual stocks for the past few years (2015-2019) is examined by considering historical volume traded as one aspect of heterogeneity. By sorting the stocks on historical volume traded and splitting stocks into ten deciles, linear regression was employed in each group for each year. The exploratory data analysis conducted in this study could explain intuition behind volatility clustering and response to shocks found in SET Index and among individual stocks. This result may give implications on applying technical indicators as trading strategies.

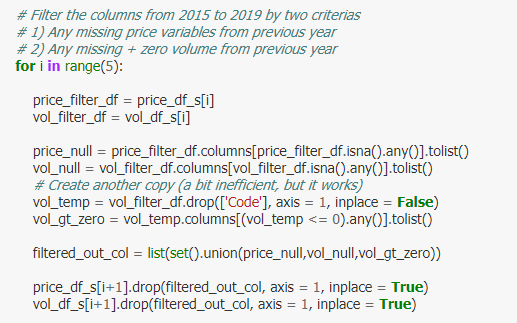
// Prev Version

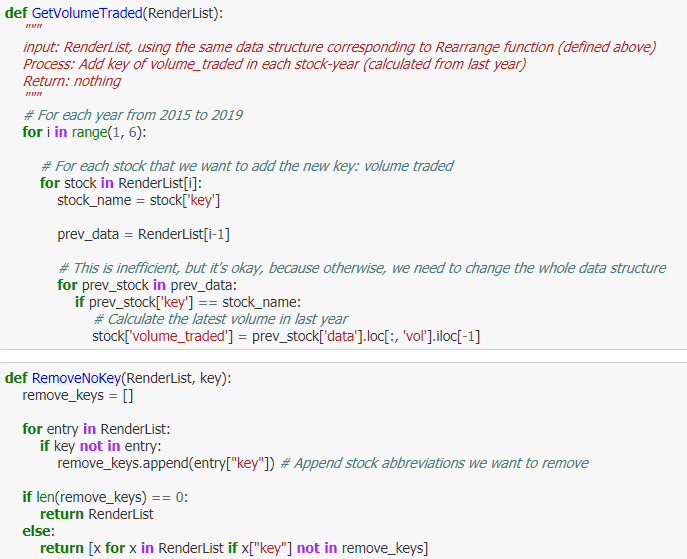
This study evaluates the role of technical indicators in predicting change in SET index direction in classification framework. This predictive model uses training data from 2008 to 2017, validating data in 2018 and testing data in 2019. To investigate explanatory power of technical indicators in each category, feature importances derived from decision tree model are used. Specifically, the feature importance of volatility proxy (n-day standard deviation) is evaluated under different windows of forecast. To interpret the role of each technical indicator in predicting SET index movement, coefficients from logistic regression are used. Further, this study explains behavioural differences of the volatility proxy on individual stocks listed in SET. The behaviour is illustrated by sorting on historical volume traded and dividing stocks into ten groups. In each group, linear regression was employed in each group for each year. The model is either run across all stocks in a particular decile or run on individual stock and the results are aggregated then. The period of study is from 2015 to 2019. The findings give implications on applying technical indicators as trading strategies.

Note on study period (reasons):

- In time-series ML framework, there is trade-off between choosing longer and shorter time horizon. The longer time horizon, the more data points we get, but too far past data points can actually hurt model performance on validating and test sets. Hence, 2008-2017 (10 years) is one possible choice.

- For individual stock, there is also trade-off between choosing longer and shorter time horizon even though it is EDA. Any missing data on prices, volumes, or zero volume on year t will cause the same individual from year t+1 not being considered. For two consecutive years (year t, t+1), only individuals where available information can be retrieved from year t are included in investigation for year t+1. By introducing longer time horizon, the number of observations in later year will be reduced, and this did not happen for SET index where there is only one individual to be considered. Hence, to retain a number of individuals in stock market, shorter time horizon should be considered. (Merit of having longer time horizon is the variable of interest is averaged over time, resulting in more reliable result)





Literatures:

Part 1

Title: Predicting Stock Prices Using Technical Analysis and Machine Learning

Content: Modification of MACD to capture trend (short-run – long-run)

Volume Agent as the second opinion

Title: Stock Selection and Trading Based on Cluster Analysis of Trend and Momentum Indicators

Content: Momentum indicator as rate of change

By using clustering method on five attributes (trend + momentum), and the strategy of choosing best cluster works well except during major sell-offs. (TH)

Title: Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques

Content: Intuition of MACD as following the stock trend

The idea of decision tree being efficient (the paper uses random forest)

We have to tune hyperparameters such as depth, # of features to prevent overfitting

Stock market data is non-stationary. They converted using deterministic data-set.

Plugging in continuous values such as trend deprives model (no info about the transition).

Title: Relative Movement to Support Stock Forecasting of the Thai Market using the Neural Network

Content: This paper uses data processing by relative movement to improve stock forecasting

By using technical indicator in ANN, it outperforms traditional model, random walk and buy and hold strategy.

Traditional model uses exact value as inputs and the model predicts future price.

The paper uses only three fundamentally strong bank stocks because they claimed that in other sectors there might be other factors affecting price movement.

One indicator from each category is used, because otherwise we need to do PCA to decorrelate the input data and remove unimportant redundant factor in the model.

The benchmark models they used include 1) buy and hold and 2) random walk.

The evaluation method they used is k-fold cross validation and t-test.

The drawback of this model is that it does not generalize to other companies with substantial impact. (Trade-off)

Title: Profitability of Simple Technical Trading Rules in the Thai Stock Market

Content: The paper tests technical trading rule from April 30, 1975 to June 28, 2013 using daily

observation of closing prices.

The paper compares proposed method and buy-and-hold strategy, and proposed method provides higher risk-adjusted return.

The paper uses MACD to generate buy and sell trading rule.

The shorter time periods used in calculations, the more sensitive the average is to small price changes. The higher sensitivity level implies higher transaction cost.

They evaluated using risk-adjusted return.

By using MACD to SET Index, it can avoid large drawdowns during market crashes.

Title: Performance of technical trading rules: evidence from Southeast Asian stock markets

Content: Transaction costs of 0.5% of transaction value for Thai stock market.

They used RSI, STOCH, MACD, DMI, OBV … as features.

Technical indicator does not help so much in terms of market timing, but helps in terms of behavioural biases (disposition effect).

Basically, traders cannot expect to buy at a relative low price and sell at a relative high price by just using technical trading rules.

Even profitable strategies could not reliably predict subsequent market directions. They make money from having higher average profit from profitable trades than an average loss from unprofitable ones.

Part 2

Title: GARCH Modeling of Individual Stock Data: The Impact of Censoring, Firm Size and Trading Volume

Content: The study runs on 1014 Australian companies using ARCH model.

The result shows that ARCH model estimation is impacted by the degree of censoring, firm size and trading volume.

Low trading volume, small firm size and high censoring leads to higher persistence of ARCH effects in estimated model.

A potential implication in moving from index to individual stock data is the possibility of thin trading in some stocks. Further, this could induce the censoring.

Considering the GARCH, the two smallest trading volume categories have high alpha1 and beta1. The persistence in other eight categories is fairly similar.

Title: Modelling Volatility in the Stock Markets using GARCH Models: European Emerging Economies and Turkey

Content: Financial data have leptokurtosis, volatility clustering, volatility pooling and leverage effects.

GJR-GARCH allows conditional variance having different response to past negative and positive innovations.

From the study of European markets, volatility shocks are quite persistent and impact of old news on volatility is significant.

Title: Asymmetric volatility of the Thai stock market: evidence from high-frequency data

Content: The study uses data from SET index to test leverage and volatility feedback effects from 2005 to 2013. The result shows both of those.

The result shows large positive and negative shocks positively affect conditional volatility, but the impact of negative return shocks is much stronger.

The volatility negatively causes return, which supports the volatility feedback effect.

The study uses one-day return.

Title: Trading Volume and Returns Relationship in SET50 Index Futures

Content: By using modified version of GARCH, trading volume has significant positive effect on returns volatility. (contemporaneous effect)

By using system of equation by GMM, it suggests that past info of trading volume and volatility can be used to explain current trading volume and volatility.

Title: Test of Information-Based Variance Model: Evidence from the SET

Content: The reliance on trading volume as a proxy of information arrival shows that persistent in volatility diminishes when trading volume is incorporated in GARCH and EGARCH.

This supports information-based variance concept.

The rate of information flow, measured by trading volume and its lags, has an explanatory power on variance of stock returns to a fairly acceptable level.

**Difference between before2020 and 2020.**

EDA:

1) the number of observations indicating downtrend increases compared to before 2020, and this difference is evident when we move to 14 and 20-day window.

*-> Does shifting sand occurs across individual stocks?*

2) When using cross-tabs, the number of observations belong to the same class increase slightly in all of three pairs (1 vs 10, 3 vs 14, 5 vs 20)

Model:

1) The model accuracy generally worsens (look at the range on y-axis)

2) Logistic regression in many horizons performs worse than benchmark!

3) Tree model is the clear winner in most horizons.

4) Looking at classification report on tree model, the recall on +1 class is high while both recall and precision on -1 class are still low (same as previous result).

5) Before 2020, the accuracy follows U-shape along forecast horizon. However, for 2020, the accuracy follows inverted-U-shape along forecast horizon. The intuition before crisis may not carry over. (Refer to v0\_part3\_revised)