Building a Robust Low Volatility Indicator for China Public Equities Market with a focus on A-Shares

Team #11: Chi Ziheng, Jason Andika Lie, Lim Xin Hui Kelly, Lucas Ng Yi Leang, Tan Yin Ling

ESD ENGINEERING SYSTEMS & DESIGN

AXA Investment Managers

INTRODUCTION

AXA Investment Managers operates as an investment management company. AXA IM have observed that low volatility can be an effective predictor for future returns in China A-shares market. Taking an active, long-term approach, it is vital that AXA IM constantly find new ways to analyse data to help clients secure a better investment experience.

OBJECTIVE

Our Project focuses on building a **company-level predictive model** for identifying stocks with **low future volatility.**
Our task is to generate a **quantitative signal** that is indicative of a company's future volatility over the **next 12 months**.

METHODOLOGY

#1. Rolling Standard Deviation

1st Methodology

Date	Return	Rolling_Std
Jan 2018	0.5076	
Feb 2018	0.8275	
Mar 2018	0.3782	
Apr 2018	0.4761	0.23626
May 2018	0.9074	0.28156
Jun 2018	0.3025	0.31146
Jul 2018	0.1246	.
.	.	.
.	.	.
.	.	.

To calculate rolling standard deviation (Std), after every data point, the rolling window is slid forward by one sample to process the next data point.

Each standard deviation is calculated over a sliding window of **length 12**.

* Instead of 3 (in diagram), we used a rolling window of 12 (space limitations)

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}}$$

s = sample standard deviation
N = the number of observations
x_i = the observed values of a sample item
$\overline{x}$ = the mean value of the observations

#2. Regression Approach

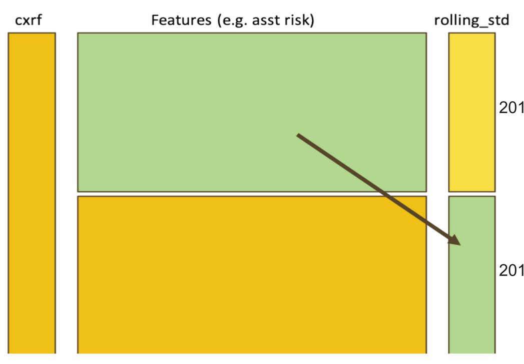
2nd Methodology

Our initial approach is to find the standard deviation of the entire year. However, we recognise that this approach is **not** a feature-by-feature comparison and will lead to inconclusive results.

Limitations of initial approach

1. It is highly possible to miss, e.g. a 3-month or 6-month period where the volatility spiked or plummeted or did both.
2. It is likely for important results to remain undetected with a large sample size.

Thus, we used a **different approach** (regression approach) to understand how volatility has changed over time or behaved in different market conditions.



#3. Time Series Analysis

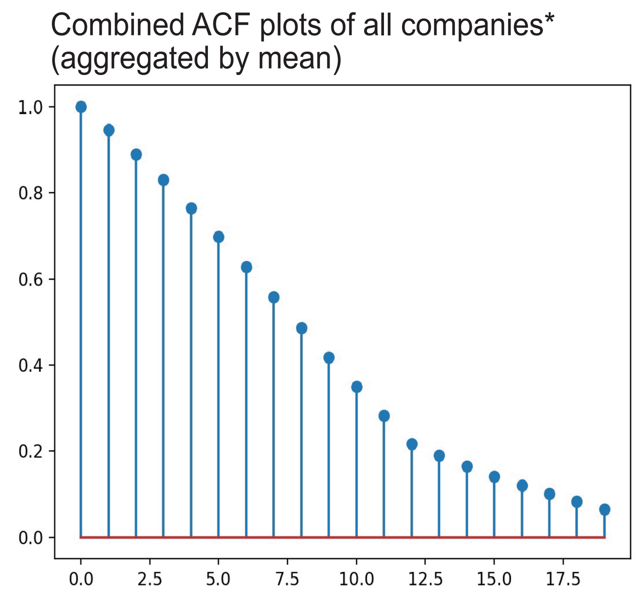
3rd Methodology

ARIMA, 'AutoRegressive Integrated Moving Average', is a **forecasting algorithm**. ARIMA models time series data such that the past values of the time series can alone be used to predict the future values.

✓✓ Pros of ARIMA

ARIMA can be use to account for
1. A pattern of growth/decline in the data --------------------- AutoRegressive
2. The rate of change of the growth/decline in the data --- Integrated
3. The noise between consecutive data ----------------------- Moving Average

Combined ACF plots of all companies* (aggregated by mean)



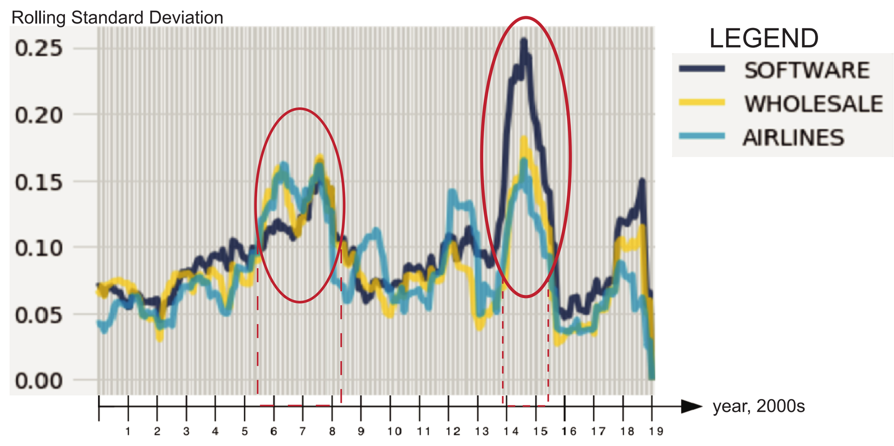
An autocorrelation function (ACF) plot is **highly useful** for a time series analysis. ACF plot gives the values of auto-correlation, measuring the internal association between observations in a time series.

No correaltion
-1 0 1
Strong and Postive correaltion *Weak and Negative correaltion*

DATA EXPLORATION

Exploring data of Industries and Correlation of Features

We want to compare the rolling standard deviation over different industries from the year 2000 to 2019 to find specific industries that are more sensitive to volatility changes.

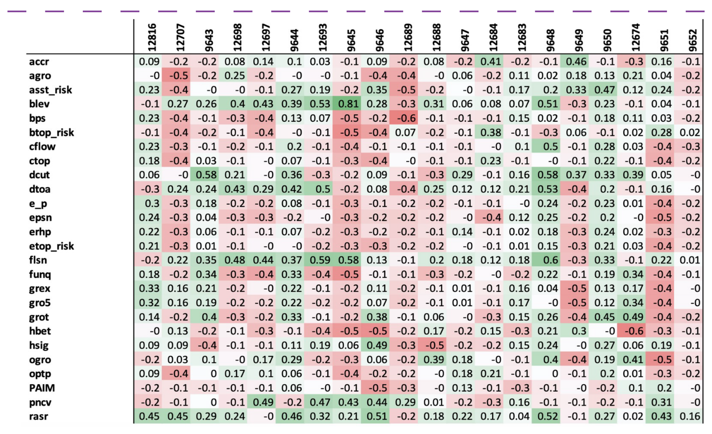


Graph 1
Comparsion of Rolling Standard Deviation between Software, Wholesale and Airlines industry over the year 2000 to 2019.

These graphs shows **sudden spikes and plummets** in the rolling standard deviation for the year 2006 to 2008 and for the year 2014 to 2015. These extreme changes can be accounted by the 2008 financial crisis and the 2015 chinese stock market boom.

IMPACT

By identifying the industries that are more sensitive, clients can use this information to justify investment decisions while relating to our ARIMA model for a more informed investment experience.



Correlation matrix heat map

Measure the degree of the relationship between features by checking if the features have similiar R values across companies with respect to standard deviation

Green **means positive**
Red **means negative**

The **stronger** the color, the **larger** the correlation magnitude.

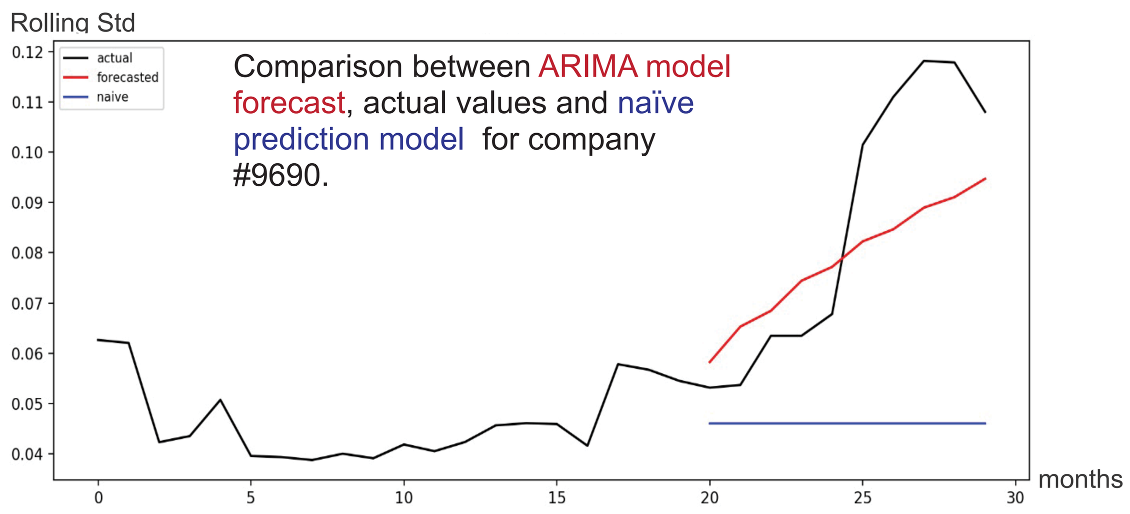
IMPACT

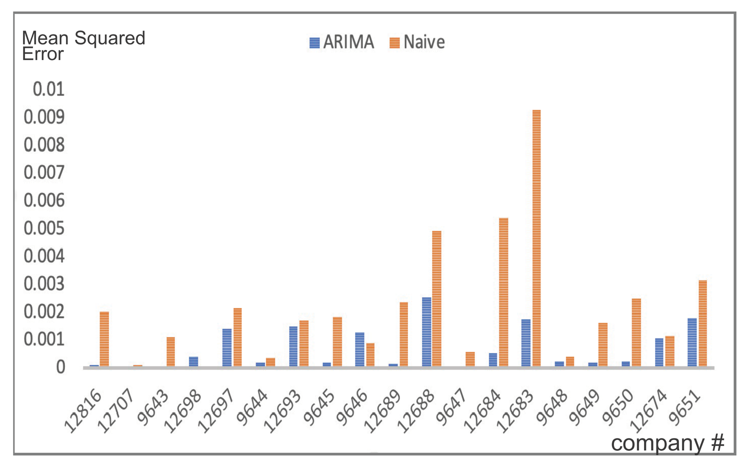
The correlation matrix heat map of the features of the given dataset shows that the features have **no visible correlation** to the rolling standard deviation. Thus, we decided **not** to continue looking at regression models and focused more on **time series models.**

Analysis of ARIMA MODEL

Main Result

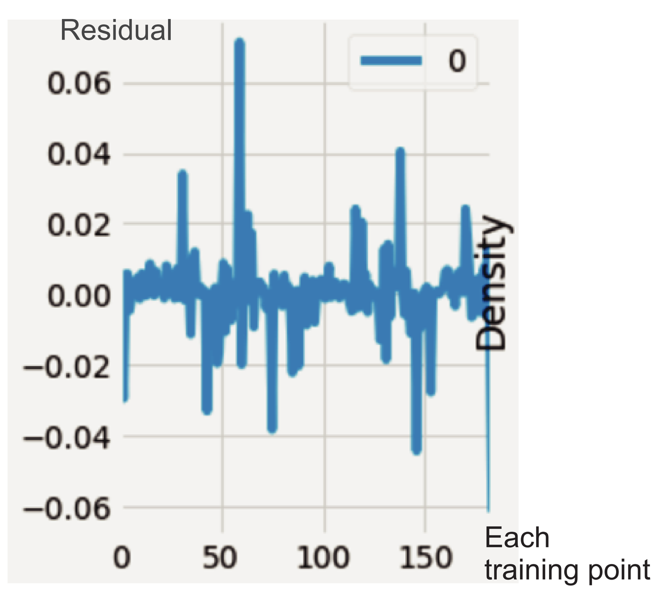
We used the ARIMA model for forecasting #company9690's future volatility for the next 12 months. We want to **compare the accuracy** of our ARIMA predictive model with a **naïve prediction model**. Thus, we created an **unbiased naïve forecasts** by setting all forecasts to be the **mean value** of the last 12 months observations.

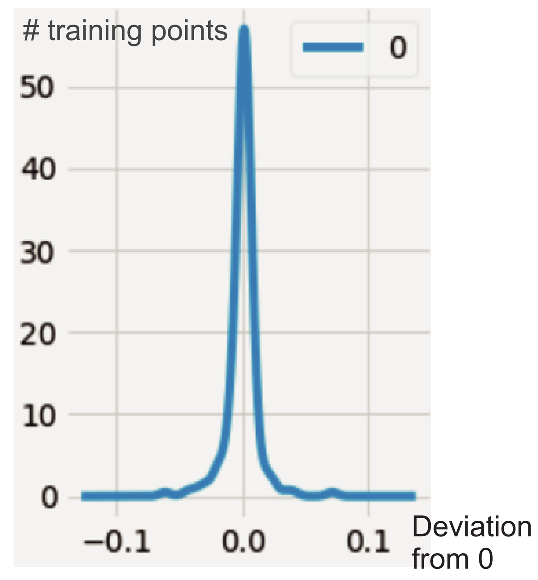




IMPACT

Since the Mean Squared Error of the Arima predictive model VS Actual values is **lower** than the Mean Squared Error of the naïve model VS Actual values, this shows that the Arima predictive model is accurate.



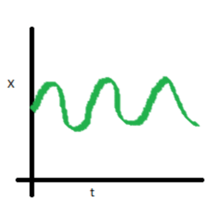


Residual plot for the ARIMA model shows **random pattern** with **density centered at 0**, showing that the residuals are **uncorrelated** and have **zero mean**.

IMPACT

This means that our model is a good estimation **without bias** and that our model has adequately captured the information in the data.

LIMITATIONS



ARIMA : Data throughout the years are generally stationary, with the exception of the year 2008 and 2015.

ARIMA: Our model trains on companies with at least 15 years of data. Companies with fewer data points may suffer inaccuracy

CONCLUSION

We created a **simple but fairly accurate predictive model.** ARIMA model has its advantages in time-series analyses. The trend, autocorrelation is easily controlled by autoregressive and moving average without performing complicated transformations. Rolling Standard Deviation is a good statistical measurement of market volatility. All in all, the **entire methodology** is **well described** and is made to be **easily replicable.**