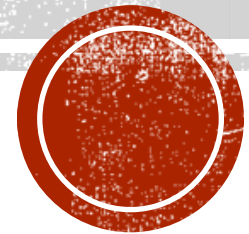


# VIX Predicting Modeling

Jingyi Yang  
Carnegie Mellon University  
The Data Incubator Project



# AGENDA

- Problem Statement
- Process Flow Diagram
  - Data Extraction & Pre-processing
  - Exploratory Data Analysis
- Prediction Result
  - Regression Analysis
  - Stepwise Selection Feature
- Key Insights & Conclusion
- Q & A

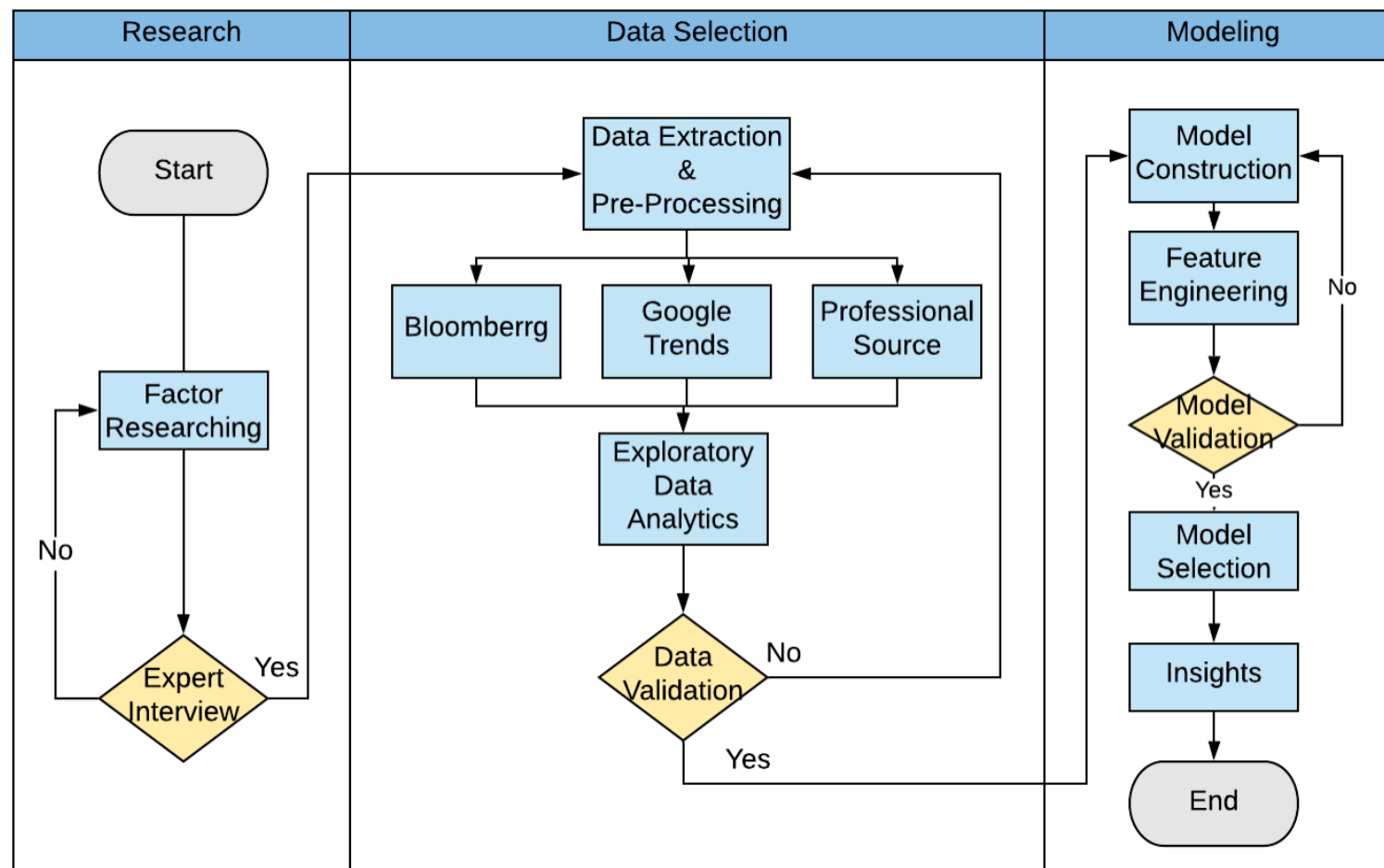


# PROBLEM STATEMENT

- VIX: CBOE Volatility Index, a real-time market index representing the market's expectations for volatility over the coming 30 days
- Investors use the VIX to **measure the level of risk, fear, or stress** in the market when *making investment decisions*
- Project Goal:
  - Use predictive modeling to estimate the change in VIX
  - Research additional relevant factors that will affect VIX changes
- **Dependent Variable:** the VIX “close” difference from day  $t-1$  to day  $t$
- **Predictive Formula:**
$$y = \Delta VIX = VIX_{day1} - VIX_{day0}$$
$$= f(\Delta x_1, \Delta x_2, \dots, \Delta x_n)$$



# PROCESS FLOW DIAGRAM



# DATA EXTRACTION & PRE-PROCESSING

## 1 Data Extraction:

- Collected data from multiple sources: Bloomberg, FRED, CBOE, QUANTIL and Google etc.
- The dataset includes **1546** pieces of training data, **387** pieces of testing data over a **10-year** span
- The model incorporates **20** different factors

## 2 Data Pre-processing:

- Standardize the tables into a single format
- Merge datasets into one data frame using Python (pandas)



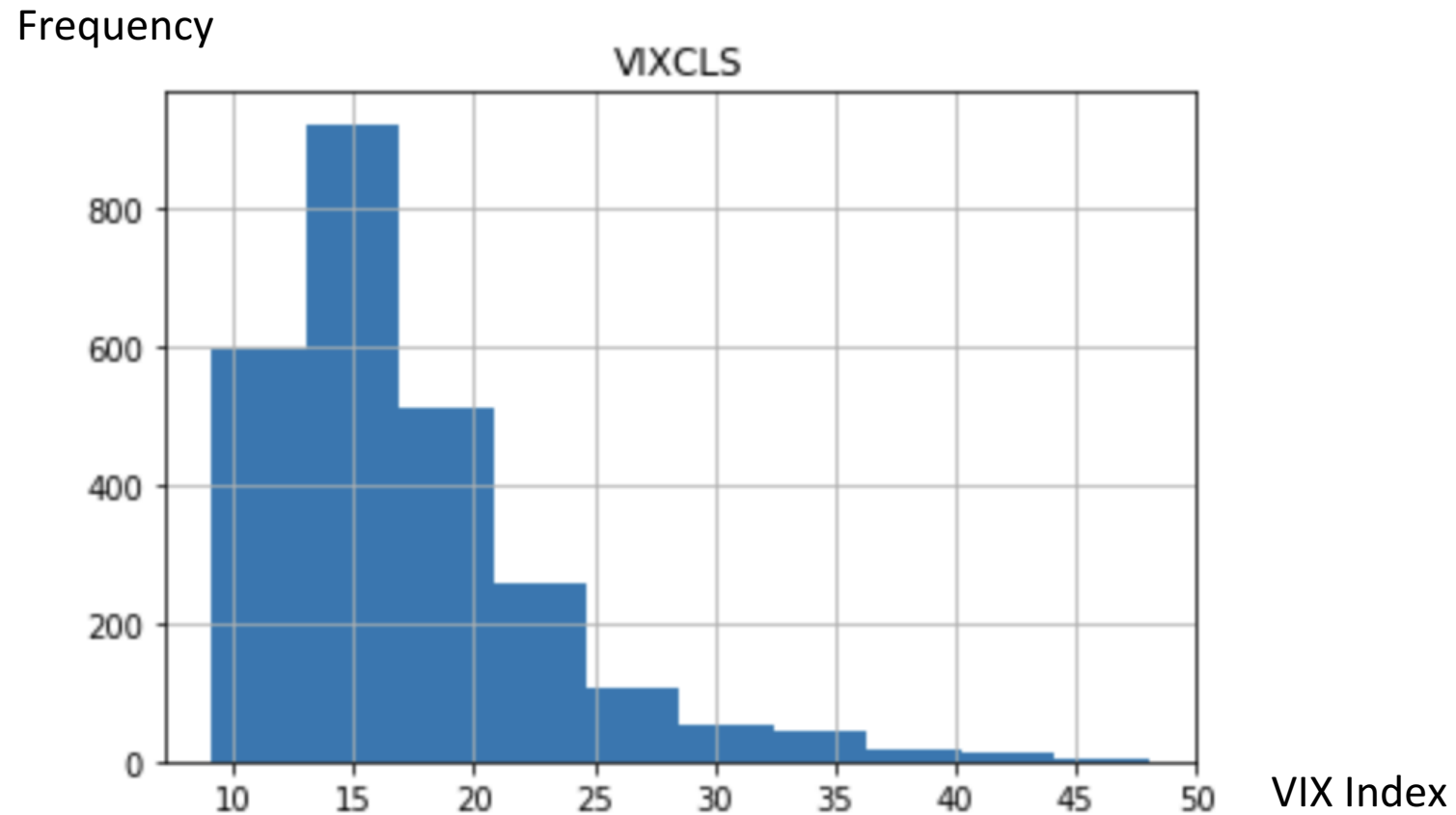
# EXPLORATORY DATA ANALYSIS - TIME SERIES ANALYSIS FOR VIX



Over the past 10 years, VIX mean: 17.08 standard deviation: 5.69  
Using 3 standard deviation, we define the cutoff for outlier at 34.15

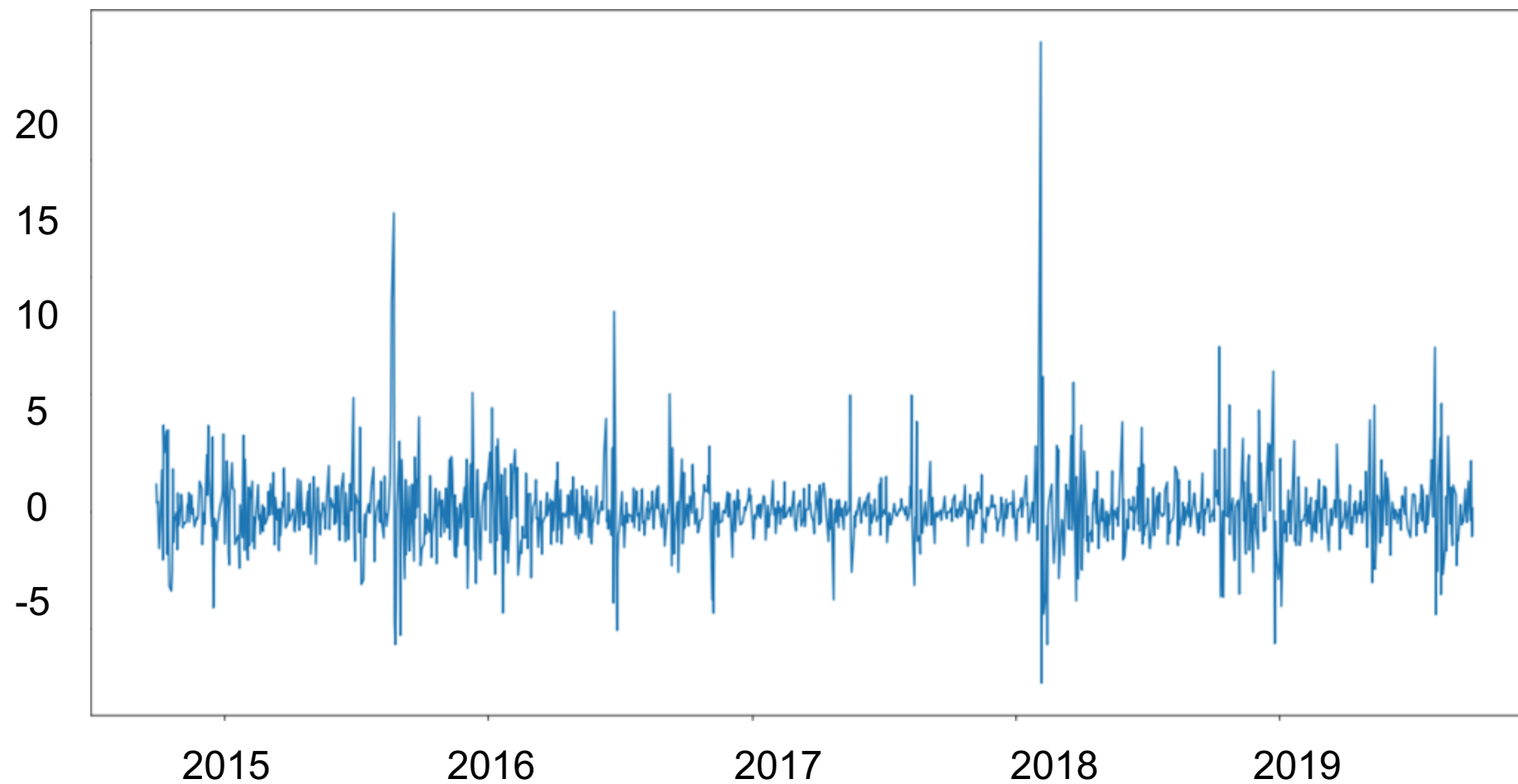


# EXPLORATORY DATA ANALYSIS – VIX DISTRIBUTION HISTOGRAM



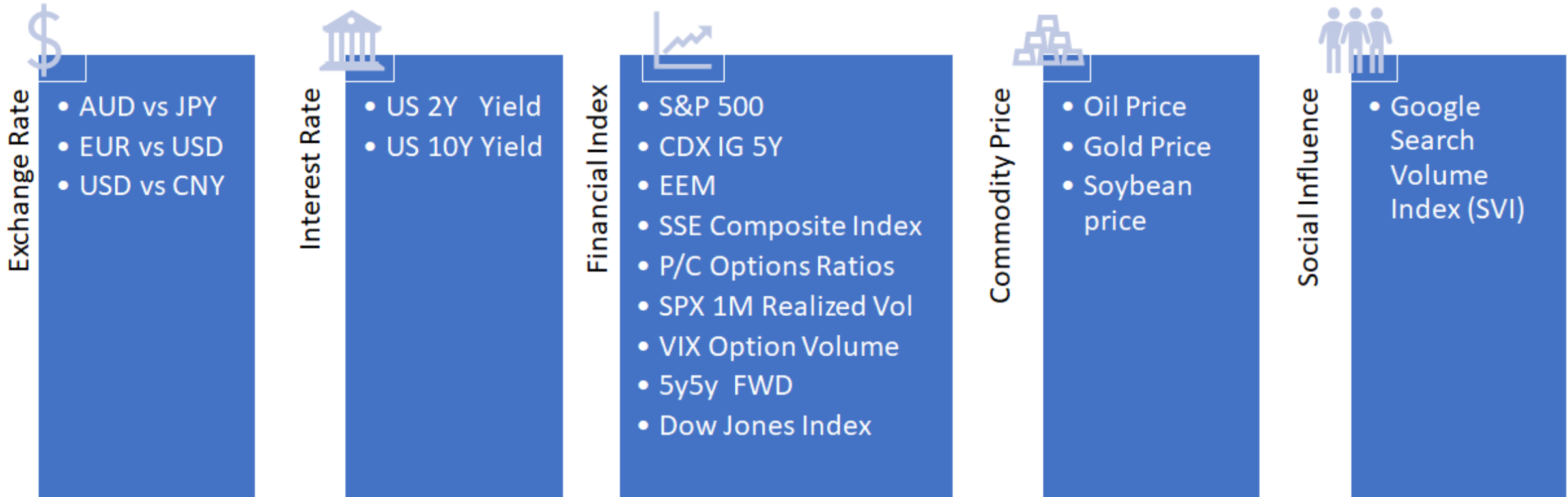
# EXPLORATORY DATA ANALYSIS - DELTA VIX

Delta  
VIX  
Index

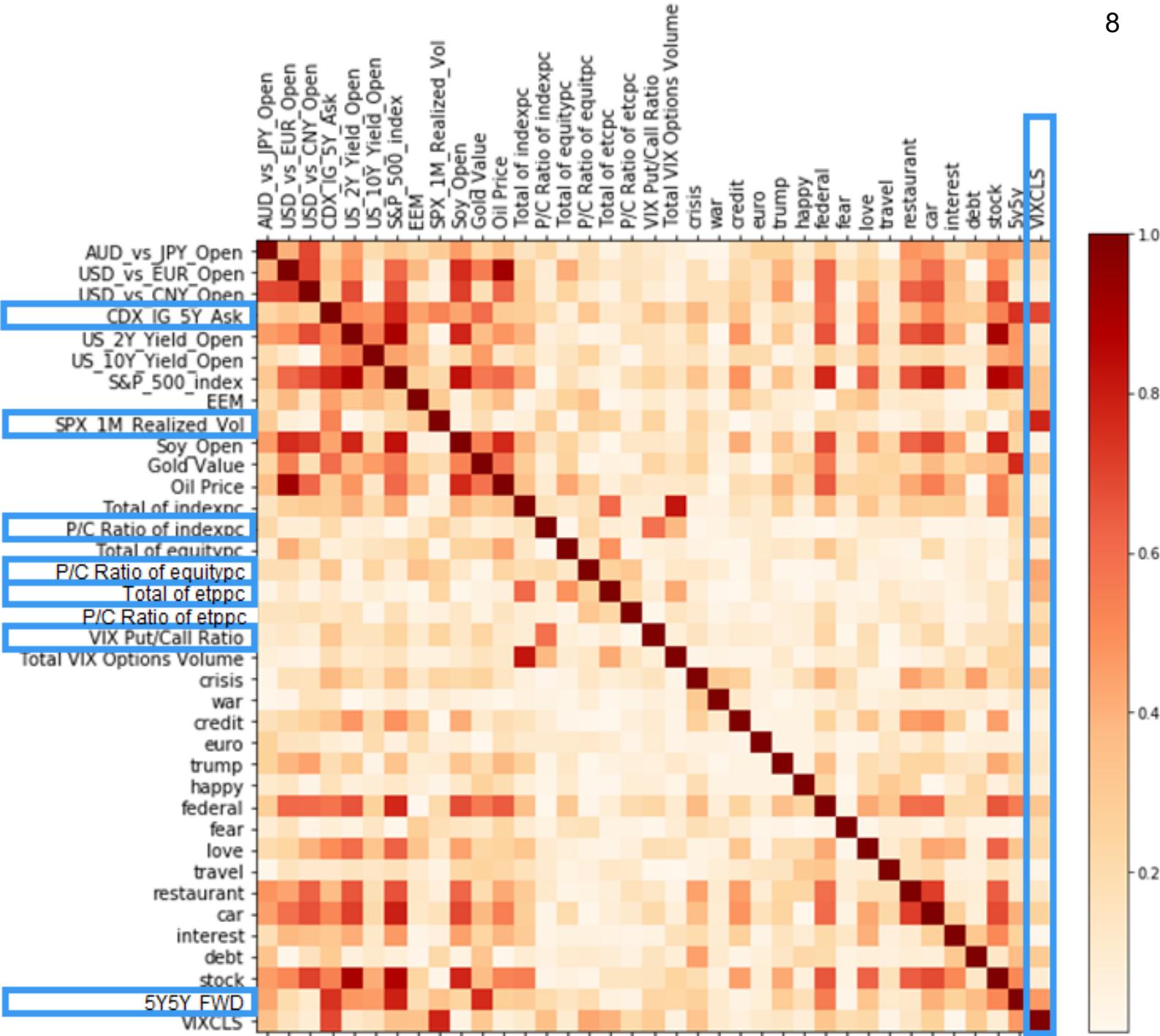




# RELEVANT FACTORS THAT AFFECT VIX CHANGES



# RELEVANT FACTORS' CORRELATIONS



# TEST METHOD

- accuracy of 0.5:  
 $\text{count}(\text{abs}(\text{predict} - \text{real}) \leq 0.5) / \text{total}$
- accuracy of 1:  
 $\text{count}(\text{abs}(\text{predict} - \text{real}) \leq 1) / \text{total}$

## Stepwise

- Feature Selection method



## COMPARISON ACROSS MODELS

	Accuracy of 0.5	Accuracy of 1	Factors
# 22 data factors	47.80%	72.35%	AUD vs JPY, USD vs EUR, USD vs CNY, CDX IG 5Y Ask, US 2Y Yield, US 10Y Yield, S&P 500 index, EEM, SPX 1M Realized Volume, Soy Bean, Gold Value, Oil Price, Total of indexpc, P/C Ratio of indexpc, Total of equitypc, P/C Ratio of equitypc, Total of etcpc, P/C Ratio of etcpc, VIX Put/Call Ratio, Total VIX Options Volume, 5y5y, Dow Jones
# 22 data factors + # 15 google search words with stepwise selection	44.96%	72.61%	Dow Jones, CDX IG 5Y Ask, <i>Fear</i> , SPX 1M Realized Volume, <i>Restaurant</i> , Total etcpc, <i>Trump</i> , <i>Stock</i> , <i>Euro</i> , SPX 1M Realized Volume, USD vs CNY

