Progress Report - STAT-615
July 18, 2020

Covid-19: Market Volatility and Liquidity

Abdul Tarawally

Minh-Tuan Nguyen

Natasha Karwal

Progress Report - STAT-615
July 18, 2020

TABLE OF CONTENTS

Progress Report - STAT-615
July 18, 2020

Team

Our team members are:

- Abdul Tarawally at4667a@student.american.edu
- Minh-Tuan Nguyen mnguyen@wcl.american.edu
- Natasha Karwal nk1061a@student.american.edu

Notes:

The team meets at least once a week through zoom and as of now the team is on track to complete the project on time. Materials will be used from course textbook, course handouts and outside related sources which we will later reference on the final project report.

Project Overview

The purpose of this project is to explore the effects of Covid-19 on across the world. Specifically, this data analysis will provide the COVID crisis impact on market volatility and

liquidity. Data has been pulled from John Hopkins's GitHub repository and financial exchanges with APIs using R.

Research Question

What impact will coronavirus indicators have on the stock market of the technology sector?

Brief Description of Dataset

We will be working with live data between January 22, 2020 to present. We will update the dataset using APIs in R. The independent/response variable that we are interested in is the vixclose which is the average closing volatility index of the week. Currently, the dependent variables include: category; stock open, high, low and close; vix open, high, and low; confirmed cases; and, deaths total. As we progress, some of these variables might be taken out and new variables might be added.

COVID-19 data was cloned from John Hopkins's GitHub repository using R. The two datasets utilized were the time_series_19_covid-Confirmed and time_series_19_covid-Deaths. From this data we collected the number of cases that occurred within countries reported by the World Health Organization.

Recent stock market prices for the financial analysis were obtained through the Alpha Advantage API using R. We pulled data market data for two categories: Technology and Hospitality. Technology-related companies include: Alphabet, Amazon, Apple, Facebook, Microsoft, Netflix and Zoom. Hospitality-related companies include: Darden, Hilton, Marriott, McDonald, Royal Caribbean, and Starbucks.

Data Visualization

Our preliminary visualization test includes using histograms, box plots and residual plots to check for normality, linearity and outliers.

Data Transformation and Justification

The data will be converted from daily into weekly and lagged by 1. We will be exploring other transformation methods as we run more tests. Additional transformations will be conducted

when the distribution improves without affecting other distributions with most effective results. Some considerations can be square root transformation. Categorical variables such as category (technology or hospitality) will be converted to factor using as.factor() in R.

Variable of Interest

As of now, confirmed cases and deaths total are the two variables that we are interested in emphasizing since the project goal is to understand the impact of Covid-19 on the financial markets.

Model Fitting and Diagnostics

Our initial model will include total (covid cases) predictors of closing stock prices. We will then perform stepwise model selection in order to select significant predictors for the final best model to ensure that we have included useful variables while omitting ones that do not contribute significantly to predicting closing stock prices.

After the model building, we will look at some diagnostics of our linear model fit. Examining the residuals and finding leverage points that can help find any potential problems with the model. For the Residuals vs fitted plot, we will look for any random scattered and thus verify the independence conditions associated with the plot. We will also construct a Normal Q-Q plot to see how the residuals are approximately normally distributed.There might be some potential points of interest in the plots that may indicate values of increased leverage of outliers.

Results, Time Limit and Material Balance

We will be using relevant graphs, charts and data to present the results. The presentation will be created using Shiny App in R. This will allow us to create an easily digestible dashboard. (subject to change to PowerPoint)

We will use different statistical methods such as:

- Histograms
- Q-Q Plots
- Scatter plots
- Descriptive statistics
- Anova
- Linear regression

Appendix I