Effects of Globalization on Local Stock Markets

SDS3386 Project – Team Les Pingouins
Andrew Pham, Mohamed Ismail Asaklil, Mor Fall Sylla

1. Introduction

Globalization is used to describe global connection and consciousness (Robertson & White, 2007). It represents the transfer of ideas, goods and services across the world and the intermingling of people and culture in a world that is becoming increasingly connected. In the past century, advancements in technology have facilitated communication and trade across the globe and made it easier and more economical. This rapid expansion of globalization has allowed companies to broaden their horizons and expand across borders with less resistance than ever. However, in the past few years, many factors have contributed to reductions in globalization, from the 2008 financial crisis to the COVID-19 pandemic, to geopolitical tensions between the USA and China, and the war in Ukraine. Many scholars and analysts have been warning about the impending slow-down or even "death" of globalization, but what does this mean for the world's economies?

In this study, we search for a link between a country's stock market performance and its level of globalization, be it positive or negative. We aim to achieve this by using a country's stock market index as a proxy for its performance and comparing it to KOF's globalization index data set to determine if there is a relation between globalization indices and stock markets. We also investigate the results of Zaher & Buics (2022), and their findings on the impact of globalization on market volatility in Europe.

2. Methodology

2.1 Datasets and cleaning

The performance of an overall stock market can be captured using an index. An index is a weighted average of the companies that make up the index. The companies vary from index to index and could be selected by arbitrary criteria like historic importance and/or sector, to more objective measures like market cap, and can even represent all stocks in a given stock exchange. The key idea is that a stock index is a way to represent the total performance of a group of stocks, so that one does not have to keep track of each individual stock in that "basket." In this project, we will be using indices representative of a certain number of top companies by market cap (usually denoted by the number in the name, e.g. CAC40 represents the top 40 companies in France), or of the whole stock market (All Shares indices).

Our project used different datasets across the multiple parts of analysis. In the initial stages, we collected data on the BOVESPA (Brazil), CAC40 (France), DAX30 (Germany), Hang Seng (Hong Kong), and Nikkei 225 (Japan) from Macrotrends.net. These datasets contained daily index valuations for every business day, beginning on different dates ranging from 1950 to the 1990s. We also collected data on the S&P500 (USA), Shanghai Composite (China), JSE All Share (South Africa), Moroccan All Shares Index (Morocco) and Nairobi All Shares Index (Kenya) from Investing.com, which contained opening, closing, high/low prices and percentage change on that day, for every business day with dates ranging from 1985 to 2008. However, we only retained data from 1990 to 2020 for use in this project.

The KOF Swiss Economic Institute's KOF globalization index measures globalization across dimensions such as economic, social, political, through various measures such as trade, tourism, migration, patents, embassies, number of NGOs, etc. This globalization index was used as our main dataset for quantifying the level of globalization of a country in a given year.

Here is an overview of the variables used from the dataset:

- **KOFEcGI (KOF Economic Globalization Index):** Concentrates on economic globalization, such as trade, financial flows, investment... It sheds light on nation's economic openness and integration into the global economy.

- **KOFFiGI (KOF Financial Globalization Index):** Measures a country's financial globalization. It provides insight into the global interconnectedness of financial markets.

- **KOFTrGI (KOF Trade Globalization Index):** Focuses specifically on trade-related aspects of globalization. It evaluates a country's integration into the global economy through international trade activities, considering factors like export-import dynamics and trade policies.

- **KOFGI (KOF Globalization Index):** A composite index capturing the overall globalization level of a country. It is calculated by aggregating the indices above as well as others (political, social, etc.).

Note that these globalization indices are split into de facto and de jure measurements; meaning the reality/state of things, and the legal code/environment, respectively.

- **CODE**: Serves as unique identifier for countries adhering to international coding conventions, streamlining country identification in the context of globalization metrics. This is crucial in datasets where the country names might be spelled differently or written in various languages, providing a consistent and machine-readable reference.

- **YoY Change (Year on Year Change):** Measures the percentage change of the index's value over a period of one year. This is calculated as the percentage difference between the closing price on the first trading day of the current year and the closing price on the first trading day of the previous year. Working with percentages allows us to have a consistent measurement across indices, which could be measured in different currencies with vastly different absolute value changes. YoY change is measured as the difference in the index value on the first trading day of the current year minus the value on the first trading day of the last year, over the value on the first trading day of the current year, times 100 to convert to percentage.

$$YoYChange = \left(\frac{\text{Closing Price on First Trading Day of the Year} - \text{Closing Price on First Trading Day of Previous Year}}{\text{Closing Price on First Trading Day of the Year}}\right) \times 100$$

- **Volatility:** Measures the degree of variation in an index over time. High volatility implies larger price swings, while a low volatility suggests more stable and predictable price movements. It is measured as the standard deviation of daily percentage change in an index.

After encountering issues in our initial approach, we brought in a third dataset, the World Bank's S&P Global Equity Indices % Change. This dataset contains the percentage change each year of a country's S&P Global Equity Index. We retained data from 1995 to 2020, giving us a total of 1906 rows.

To prepare our data for analysis, we calculated YoY change and volatility using the formulas above for the datasets from MacroTrends and Investing.com, and later the World Bank data, and merged them with the KOF Globalization Index. These new cleaned and merged datasets were saved into separate .csv files for re-use and consistency. We then created some visualizations to look for visible patterns in the data.

2.2 Regression model and hypothesis testing:
According to Zaher & Buics (2022), the de jure financial globalization index has a significant impact on the volatility of a stock market, specifically in Europe. We sought to replicate the hypothesis:

H_0: Increasing financial globalization reduces stock market volatility.

And see if it applied to countries outside of Europe. To test this hypothesis, we used the following regression model:

$$Volatility_{i,t} = \beta_0 + \beta_1 Volatility_{i,t-1} + \beta_2 KOFFiGI_{i,t} + \epsilon,$$

Where volatility i represents a country and t represents a time period (a certain year), and t-1 is the previous time period (previous year, this is called the lagged volatility).

Zaher & Buics' model was similar, but they added control variables for each country and fixed effects for each year. They found that an increase in a country's financial globalization has a positive effect on its stock market and a negative effect on the volatility of its stock market index. We fit an OLS model predicting volatility with lagged volatility and the financial globalization index, as well as a model predicting YoY Change with lagged YoY change.

After encountering difficulties recreating Zaher & Buics' results with a simpler model, we analyzed the differences in stock market growth before and after the financial crisis of 2008.This analysis is based on the following hypothesis test:

H_0: there is not a significant difference in percentage change before and after the financial crisis of 2008.

To conduct this test, we divided our dataset into post-crisis and pre-crisis observations, with one observation per country, per year (where there was data available), and used a paired, two-sided T-test on the year-on-year (YoY) change, since the measurements compare the YoY changes at different times, and we are looking for any difference, regardless of direction.

2.3 Dimensionality reduction and clustering
Due to the large number of features in the combined KOF Globalization Index and World Bank datasets, we employed several dimensionality and feature reduction techniques to better visualize the data and look for patterns in the data. These techniques include Principal Component Analysis (PCA), t-distributed Stochastic Neighbour Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP). We then proceeded to run various clustering algorithms, including Gaussian Mixture Models using the Expectation-Maximization algorithm, Spectral clustering, and Agglomerative Hierarchical clustering, to cluster the data into two groups: Up and Down, corresponding to the direction in which the stock index moved that year. K-means was not selected due to the results of the dimensionality reduction methods showing a large overlap of points, which reduces the usefulness of k-means. For clustering on the principal components, the GMM method was preferred due to the large overlap of points (See figure 4.1), since mixture models can deal with overlapping clusters, unlike k-means. The classification algorithms were all imported from `sklearn`, while the UMAP implementation was imported from the `umap` package.

T-SNE and UMAP are both manifold learning methods, which map data in higher dimension to lower dimension spaces. For easier visualization, we will be projecting data onto 2 dimensions, so that it is easier to plot on a chart.

2.4 Classification models

With the inconclusive results of the clustering attempts, we decided to train several classifiers to predict whether a given country's stock index would go up or down in a given year. The predictors for this model were the financial globalization index de jure and de facto of the country in question, and the percentage change of the US (S&P500) stock market that year, as well as its de facto and de jure financial index. The response variable was a binary variable representing the direction the stock index will move for that year. The models we built were a logistic regression model, a Support Vector Machine (SVM) model, an artificial neural network (ANN), a Naïve Bayes classifier and a Random Forest (RF) classifier. These models were all built using classes from the `sklearn` package. The logistic regression, naïve Bayes and SVM models were built using package default parameters, RF was built using default number of trees (100), but maximum depth of 1, and the neural network was built using an architecture of (100, 100, 80, 50) nodes, which was determined to be optimal using sklearn's GridSearchCV. The models were evaluated using their 10-fold cross-validation error estimate on the entire training set of 1880 observations.

3. Results

3.1 Initial data exploration

Our initial data exploration began with a scatter plot analysis followed by a heatmap examination. First, we employed a plot to visualize the relationship between the KOF Globalisation Index (KOFGI) and the Year-over-Year (YoY) change (see Figure 3.1). Subsequently, we used a heatmap to examine correlation between all the indices and YoY change.

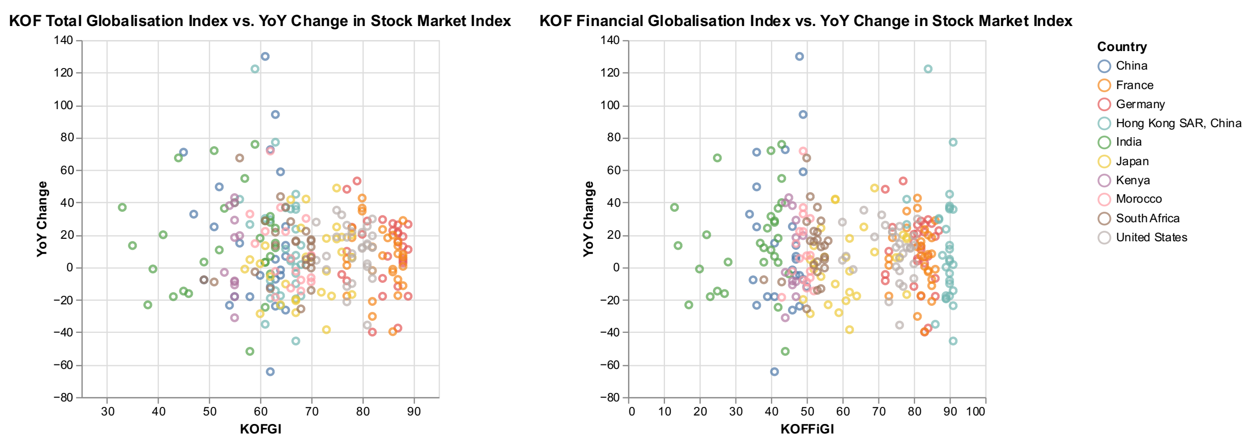

Figure 3.1: Visualizations of KOF globalization indices vs YoY Change from Macrotrends/Investing.com data

As can be seen in Figure 3.1, the scatter plot did not provide a clear indication of a linear relationship between the overall index and the financial index, leading us to the heatmap analysis. However, the heatmap also failed to reveal a noticeable correlation between the YoY change and any of the KOFGI indices, as can be seen in Figure 3.2.

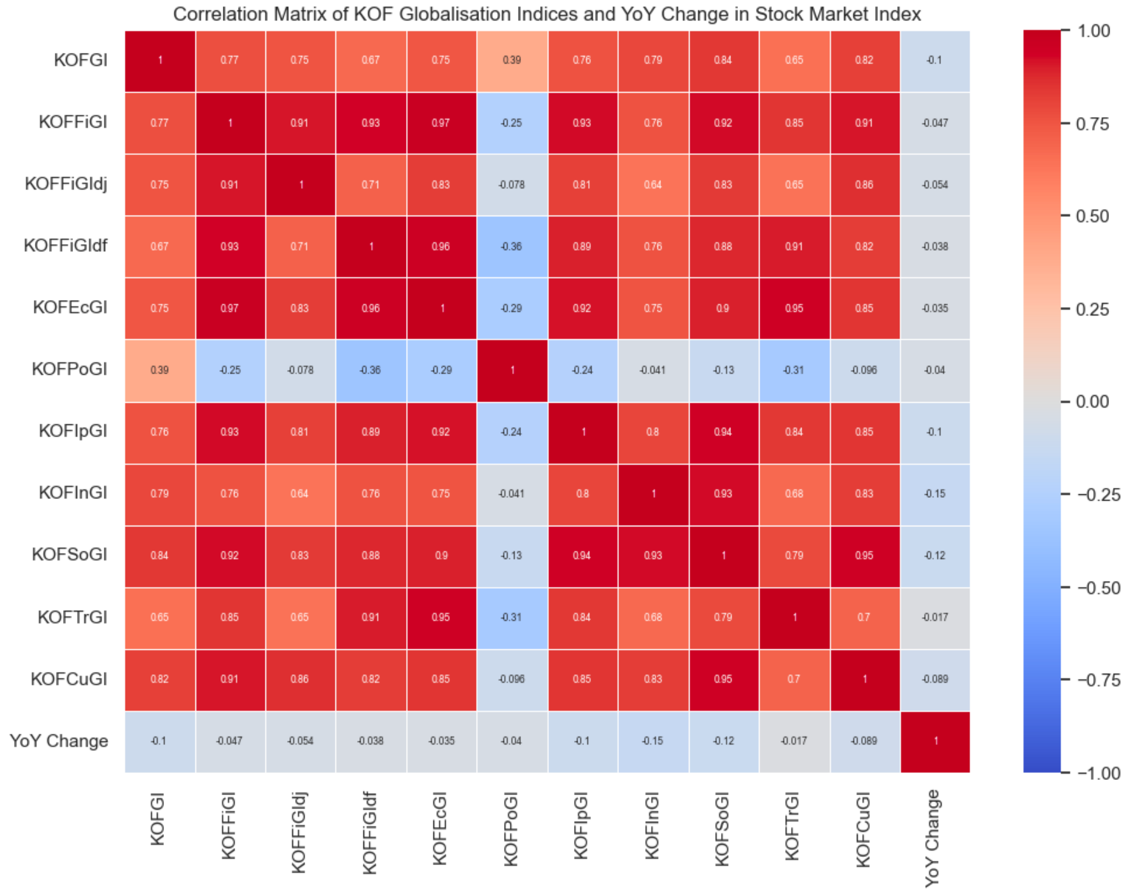

Figure 3.2: Correlation heatmap of combined Macrotrends/Investing.com and KOF Globalization index data

After our initial approach failed to yield results, we brought in the World Bank S&P Equity Index dataset and performed similar analyses on that data. Again, despite the greater number of observations, we were unable to find any readily available correlations between the indices and the stock index yearly change (denoted as Pct Change in the World Bank data), as can be seen in figures 3.3 and 3.4.

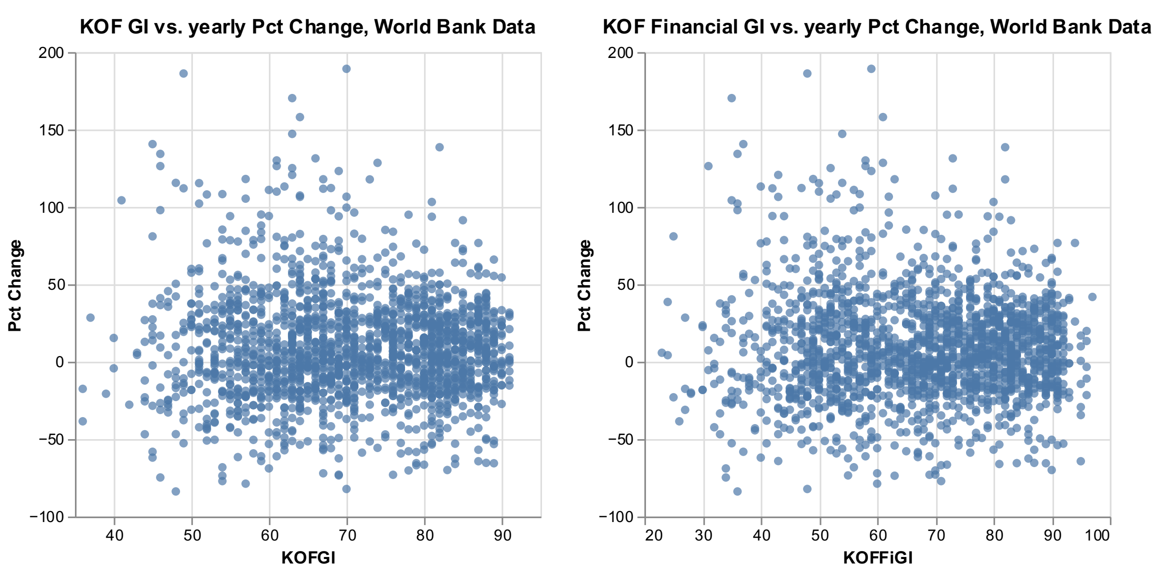

Figure 0-1Visualizations of KOF globalization indices vs YoY Change from World Bank data

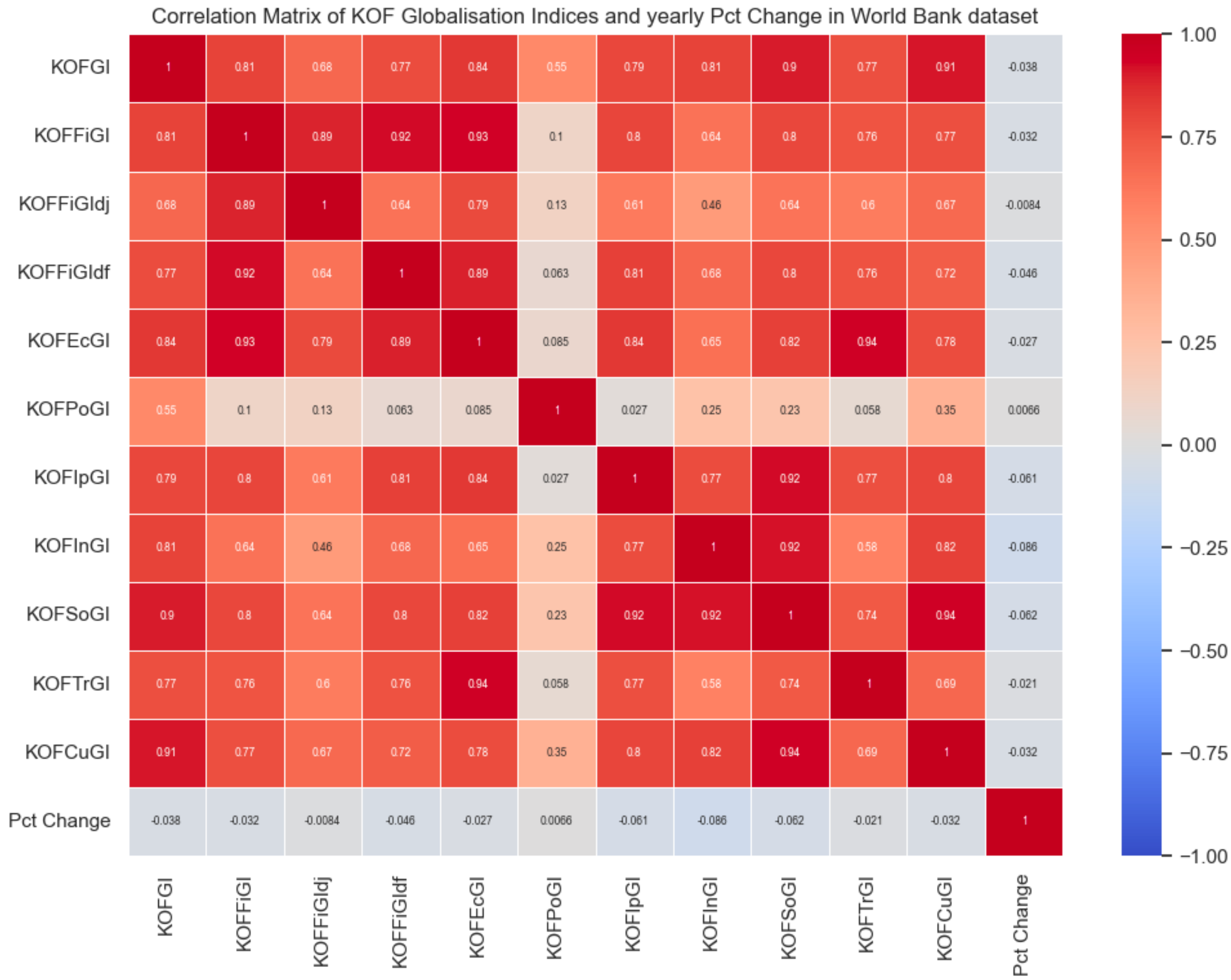

Figure 3.4 Correlation heatmap of combined World Bank and KOF Globalization Index data

3.2 Regression model and hypothesis testing results:

After building the regression model for the Zaher's hypothesis test, we came up with different results in comparison with the article. The KOFGI did not have a significant impact on volatility due to a small F-statistic of 35.17. Despite that, we found positive results in the pre/post-crisis analysis, confirming that there is a significant difference between the pre and post crisis 2008. Given a T-statistic of 8.162 and a small p value (7.604e-16), which was statistically significant at the level $\alpha < 0.01$. The extremely low p-value motivated us to create a visualization of the distributions of Pct Change before and after 2008, which can be seen in Figure 3.5. We can see that on average the post-crisis has a bigger percentage change. We also built a map visualization to see the change of KOFGI over the years. However, we decided not to include it in the report since it would be static, not allowing visualization of index scores over the years. It can be found in the submitted code. This animated map shows us that the number of countries with increasing globalization is less after 2008, and the rate at which it increases is also slower.

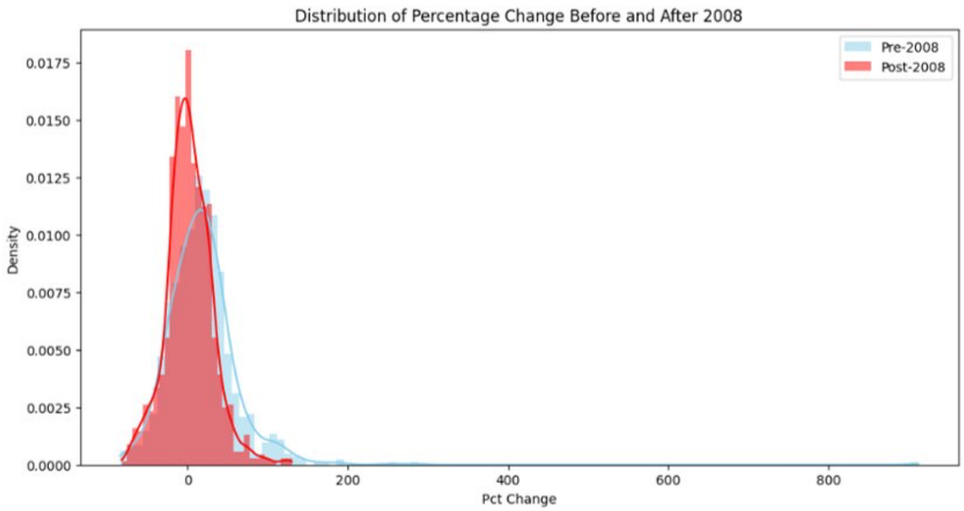

Figure 3.5: Visualization of Distribution of Percentage Change before and after 2008

3.3 Dimensionality reduction and clustering

The results of the dimensionality reduction techniques and clustering were largely inconclusive. As can be seen in figure 3.6, there are no easily visible clusters in the data. Figure 3.10 shows that t-SNE and UMAP also did not reveal any visually distinguishable clusters.

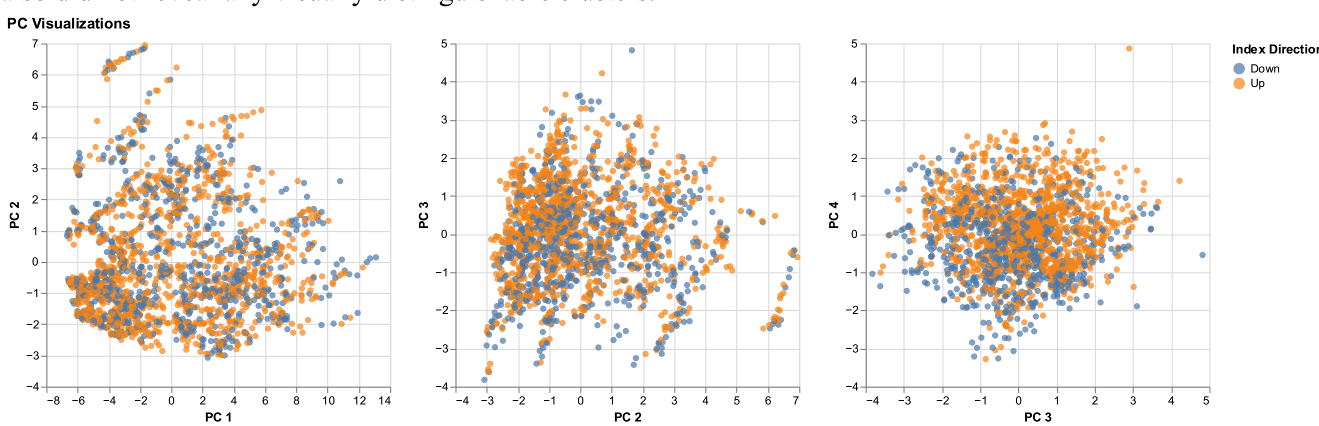

Figure 3.6: Visualizations of PCA Results

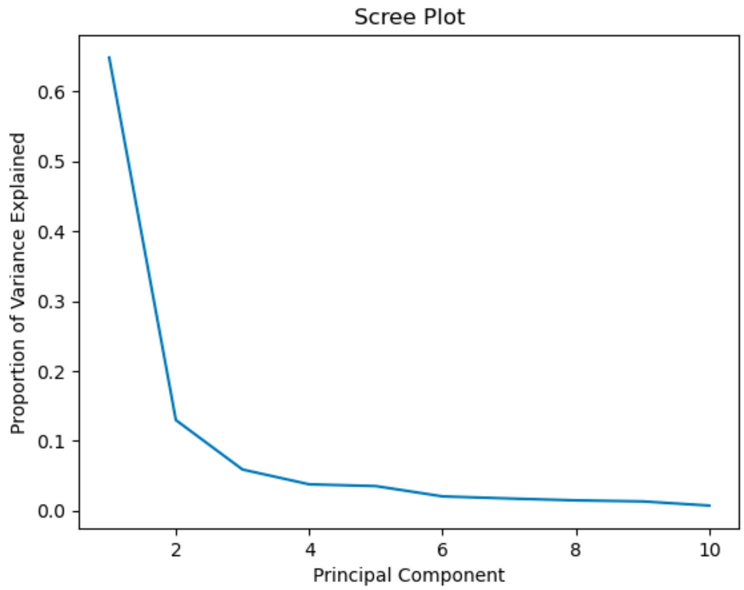

Figure3.7: PCA Scree Plot

While we were not able to discern any visible clusters in the data, we did construct a biplot of the first 2 principal components, as can be seen in Figure 3.8. This biplot shows that the economic index (KOFEcGI) has by far the largest effect on PC1 and therefore explains the most variance, followed by the overall globalization index (KOFGI) and the yearly percentage change (Pct Change) which have more explaining power on PC2.

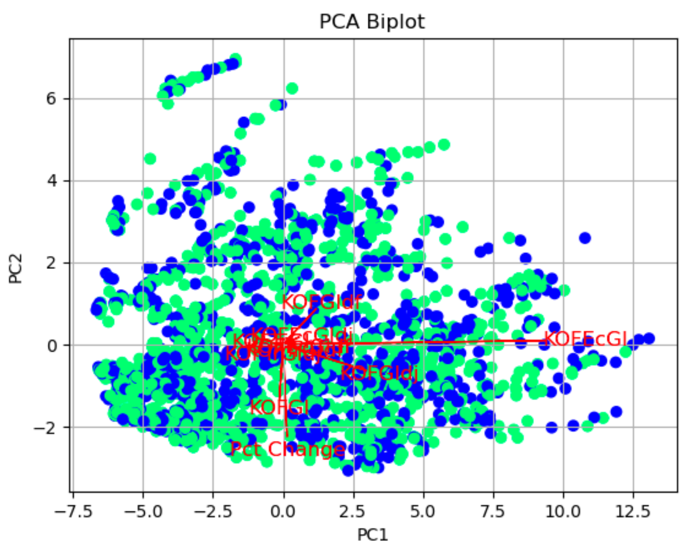

Figure 3.8: PC1 vs PC2 Biplot. Green indicates "Up", Blue indicates "Down."

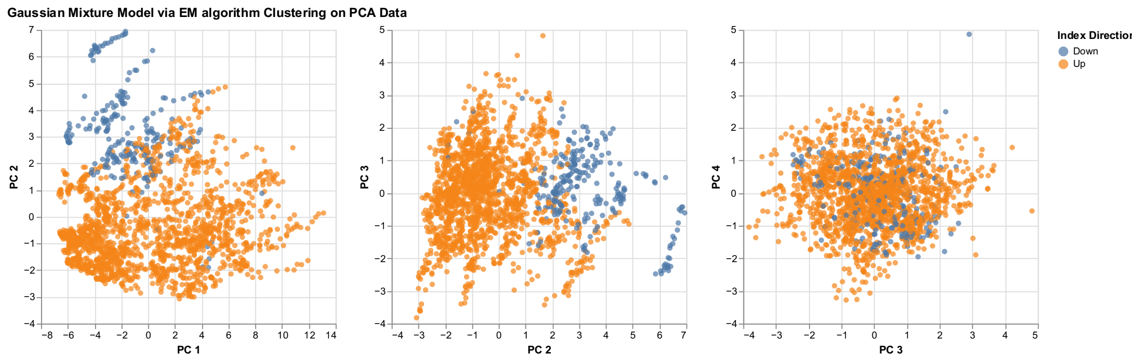

Figure 3.9: GMM Results on PCA data

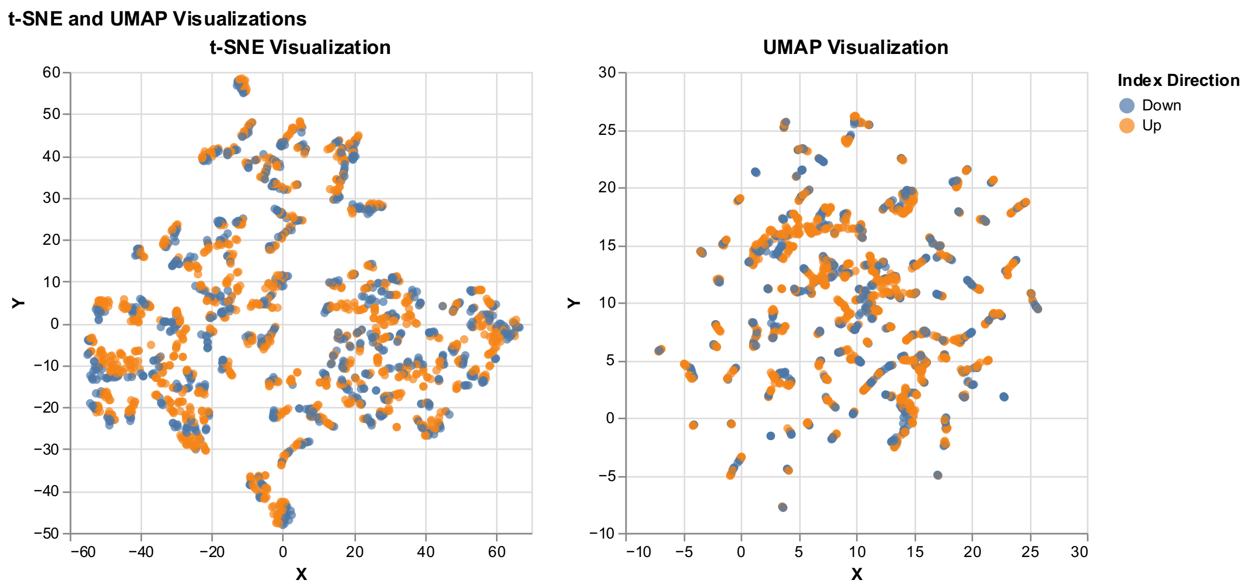

Figure 3.10: t-SNE and UMAP plots

Despite the lack of distinguishable clusters, we proceeded with implementing the clustering algorithms. However, as can be expected, these did not yield accurate results, so the results of these methods were not used in the predictive models. We can see from Figure 3.9, 3.11, and 3.12 that the clustering algorithms failed to provide clustering assignments similar to those in the original visualizations.

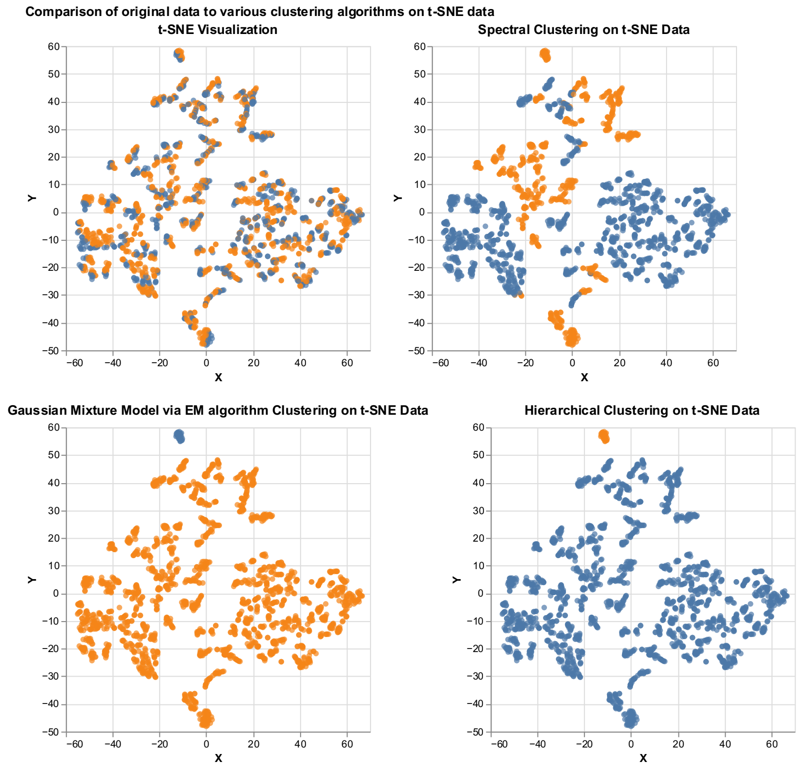

Figure 3.11: t-SNE Clustering Algorithm Comparison. Legend labels left out intentionally, as this graph intends only to show the dissimilarities in cluster assignments.

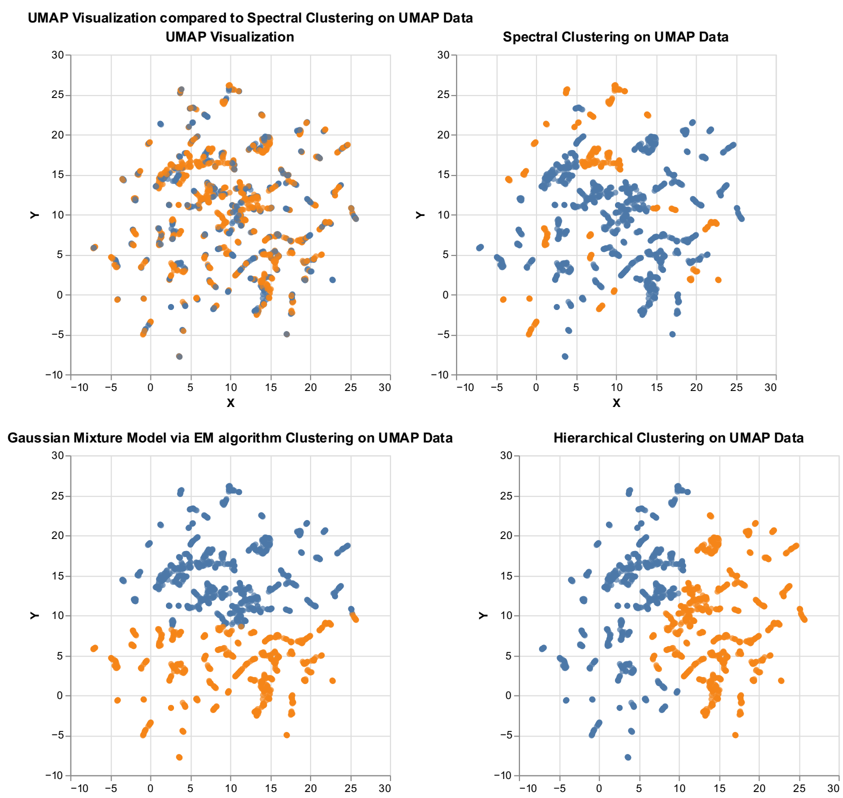

Figure 3.12: UMAP Clustering Algorithm Comparison. Legend labels left out intentionally, as this graph intends only to show the dissimilarities in cluster assignments.

3.4 Classification results

After failing to find distinct clusters and separation between observations where the percentage change was positive and negative, we trained several classifiers to predict whether a given observation represented a country whose stock index went up or down that year. After evaluating the 10-fold CV error estimate for all the models, we found that the random forest model performed best, followed by naïve Bayes, then SVM and ANN, and finally logistic regression, as per table 3.1.

Table 3.1: Accuracy score of various models on World Bank and KOF merged dataset.

Model	Accuracy Score
Logistic Regression	0.676
SVM	0.683
Neural Network	0.683
Naïve Bayes	0.698
Random Forest	0.740

4. Discussion

4.1 Initial data exploration discussion

Our initial data exploration did not yield the results we had initially hypothesized, so we resorted to two measures to continue with our analysis. First, we brought in a new data set, which was the same data set used in Zaher & Buic's 2022 paper, to see if the greater number of observations could improve our results, as perhaps our data collected from MacroTrends and Investing.com were erroneous. However, we found the same result when analyzing those datasets, so we decided to include some strategies for dimension reduction and manifold learning which could possibly bring better results. Overall, we believe that the data provided by the KOF Globalization index was simply insufficient to describe the percentage change of the country stock indices, even somewhat adequately. Even Zaher & Buics' model added control variables for specific countries and years to their model, which we did not account for.

4.2 Regression model and hypothesis testing discussion

We faced many challenges finding insights, especially with the Zaher & Buics' regression model. Therefore, we decided to add a new direction of analysis, focusing on the years before and after the 2008 financial crisis. The results we found were as we expected; there was a significant difference between the percentage change before and after the 2008 financial crisis. This divergence is likely attributable to the pronounced fluctuations characterizing the post-crisis period, as markets underwent a correction and recovery from the 2008 downturn.

4.3 Dimensionality reduction and clustering discussion

Although our techniques of dimensionality reduction and clustering did not yield much information, there are a few things to note. The dimensionality reduction techniques did show that observations of the same country tended to be clustered together. However, for the sake of brevity, we chose not to include this within the report since it is somewhat self-evident that one country's index behaviour will be similar over the years, in comparison to others. It is also difficult to explain other meanings behind this if there are any, due to the difficulty of interpreting manifold learning techniques like t-SNE.

4.4 Clustering model performance

The accuracy of our models turned out to be slightly underwhelming. With most models falling below 70% accuracy, it is clear that these models would not be very useful in the real world, as even if the accuracies were higher, these models simply classify the country's index as having a positive or negative return, not the magnitude. However, this was expected from the start; stock markets have countless factors influencing their valuations, and it would be difficult for students with limited access to data and computing power to build a model that could accurately predict the movements of stock markets, which is a feat yet to be accomplished, at least in the public eye. As for the performance of the individual models, the RF model's improved performance suggests maybe other ensemble learning methods could have performed better than the other models.

5. Conclusion

Our original motivating question was the following: is there a link between a country's level of globalization and the performance of its local stock markets? We found conflicting results between our analysis and the analysis of others, such as Zaher & Buics, and are therefore unable to come to a concrete conclusion that there is or is not a link. Our methods may have been too simple, as we did not account for many external factors unlike Zaher & Buics and other studies. Our search for a simpler explanation, that the KOF globalization index alone could provide enough insight, may have been too broad. Greater depth might have been necessary for a better analysis, such as incorporating more data to account for external conditions that could affect stock markets.

6. Ethical Concerns

While the data for this project was sourced from platforms that provide historical data for analysts and public use under non-commercial terms, it must be acknowledged that we did not obtain explicit permission from the data providers for their use in this study. Recognizing the importance of adhering to ethical standards, we should consider the implications of this oversight. Going forward, it is essential to not only rely on the general permissions granted for non-commercial use but also to seek specific consent for each unique application of such data. This practice would serve to respect the rights and intentions of the data providers, ensure the integrity of our research methods, and reinforce the trustworthiness of our study's conclusions.

7. Contributions

Andrew Pham contributed to section 1: Introduction, section 2.1: Datasets and Cleaning, section 2.3: Dimensionality reduction and clustering, section 2.4: Classification models, section 3.3: Dimensionality reduction and clustering, section 3.4: Classification results, section 4.3: Dimensionality reduction and clustering discussion, section 4.4: Clustering model performance, 5. Conclusion, as well as correction and review of the entire report.

Mohamed Ismail Asaklil contributed to section 2.2: Regression model and hypothesis testing, section 3.2: Regression model and hypothesis testing results, section 4.2: Regression model and hypothesis testing discussion. Section 6: Ethical concerns, as well as a review of the report.

Mor Fall Sylla contributed to section 2.1: Datasets and Cleaning, section 3.1: Initial data exploration, and section: 4.1 Initial data exploration discussion.

References

Robertson, R. and White, K.E. (2007). What is Globalization?. In *The Blackwell Companion to Globalization, G. Ritzer* (Ed.). https://doi-org.proxy.bib.uottawa.ca/10.1002/9780470691939.ch2

Zaher, H.F., Buics, L. (2022) The impact of financial globalisation on stock market volatility in European Union countries. *Hungarian Statistical Review, 5*(1). 109-122. https://doi.org/10.35618/hsr2022.01.en109