Big Data in finance project

Submission 1

Abstract

This document aims at defining our project idea, explain our objective and how we will analyze the problem we will develop in the following submissions. First, we define the problem statement and why it is meaningful. Second we will describe the data, how it was retrieved, and processed, but also how we intend to implement in the following steps. Last, we provide a short exploratory analysis of the data in our possession.

Matteo Rigoni, Mikkel Duif, Laura Caviezel, Davide Callegaro, Lars Horsbøl
Lars Horsbøl, Matteo Rigoni

Problem Definition

Topic and Goal of the Project

The goal of this project is to implement a specific compartmental model, the SIR (Susceptible-Infected-Recovered), on a network to better understand systemic risk and the spread of debt default. We leverage bilateral foreign claims data of 23 countries from the Bank of international Settlements as well as credit ratings as two measures of systemic risk to model default spreading from one national banking system to another. Our aim is to identify factors that influence the speed and spread of a financial crisis by simulating default cascades originating from each of the 23 countries, and spreading through the edges.

We approach the problem in the following steps:

1) build the network,
2) define transmission rate (β, based on foreign claims) and recovery rate (γ, based on credit rating),
3) run simulations with each nation as originator of default to compare the evolution of the resulting financial crises, and
4) draw conclusions about policy initiatives to limit the spread of crises.

Relevance of the Topic

The economy is a highly intercorrelated ecosystem in which many entities and organizations can impact and influence one another across the network. In the recent financial crisis of 2008, we witnessed a perfect example of the inherent danger of systemic risk with the impact of the subprime mortgage crisis on the mortgage-backed securities and commercial papers markets, which spread across the world, affecting numerous markets and industries in different countries (Schwarcz, 2008). Over the last few decades, technological advances, economic developments, and a rise in the complexity of international relations led to increased interdependence in the Banking sector (Brunnermeier, 2010). Interdependence is a great feature to have with regard to risk sharing and economic growth; however, it also brings about a surge in the exposure of a country's economy to domestic and foreign crises, increasing the domestic economy's vulnerability and incrementing its systemic risk. Systemic risk can therefore be viewed as a major social problem because of the potential for spillovers from the financial sector to the real economy (Brunnermeier, 2010). As a result, worldwide interest in systemic risk in the Global Financial system has increased substantially. For the previous reasons, concerns and interest about systemic risk in the financial system is more than justified. In fact, research about this topic can be of great importance to learn more about how shocks (whether positive or negative) spread through the system, but also to help leaders identify what is correct policy-making in regard to systemic risk, and how these policies should be implemented.

Previous literature has shown the transferability of epidemiological models into other research fields, including the financial sector (Kostylenko, 2017). In this project, we want to utilize the SIR model to simulate the spread of a debt crisis. We plan to use a dataset of 23 developed and emerging countries from the Bank for International Settlements and simulate different scenarios in which the crisis emerges from only one of the different countries (Bank for International Settlements, 2019). We hope these simulations will allow us to identify which factors are most impactful with regard to the magnitude of the spread of crises across the global financial system, and gain interesting insights related to their speed and severity.

Data Collection and Processing

As mentioned in the previous section, the data we are going to analyze is the "Summary of foreign claims and other potential exposures" collected from the Bank of International Settlements. This data contains information about the amounts outstanding of the reporting banks. For our analysis the data is aggregated on the country level. With regard to pre-processing, we have performed a few tasks: first we selected 23 countries out of the complete dataset that we are going to analyze throughout the project. Second, we calculated the β and γ values for the countries in our dataset.

We compute β values with the following formula:

$$\beta_i = \frac{\alpha_i}{\sum_{j=1}^{23} \alpha_j}$$

where α_i represents the total claims of country i, while the bottom part of the equation represents the sum of the total claims of the other countries.

Conversely, the probability of recovery γ is defined using Standard & Poor's (S&P) country ratings. These ratings range from D to AAA, so we converted them in a numeric scale between 5 and 100 with increments of 5. Each γ is calculated as follows:

$$\gamma_i = \frac{1}{101 - C_i}$$

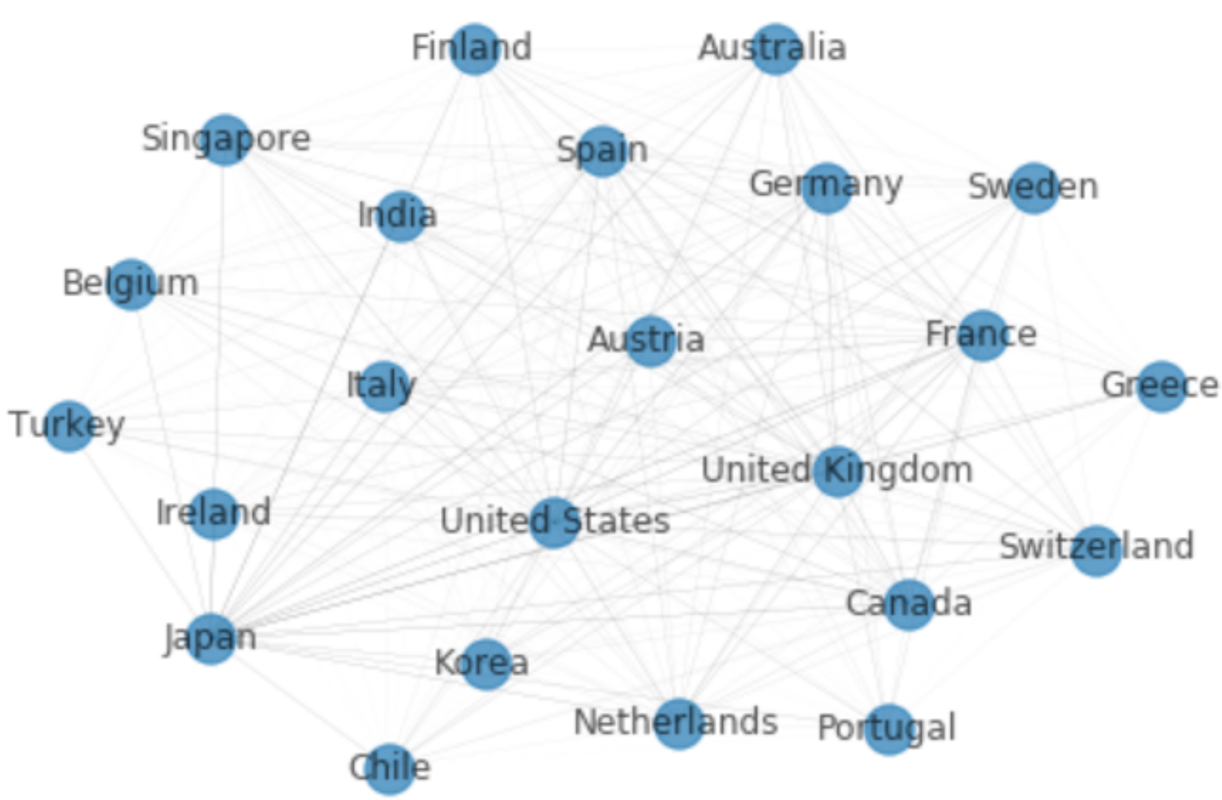
where C_i is the credit score of country I, which is understood as a numeric risk score attributed to that country: the higher the risk score, the longer the recovery time from the crisis.

To represent the countries in an SIR model in a network, we use and undirected graph, hence the probability of default between two countries is symmetric. The connections between all 23 countries are represented as follows:

$$A = \begin{bmatrix} 0 & a_{1,2} & \cdots & a_{1,23} \\ a_{2,1} & 0 & \cdots & a_{2,23} \\ \vdots & \vdots & \ddots & \vdots \\ a_{23,1} & a_{23,2} & \cdots & 0 \end{bmatrix}$$

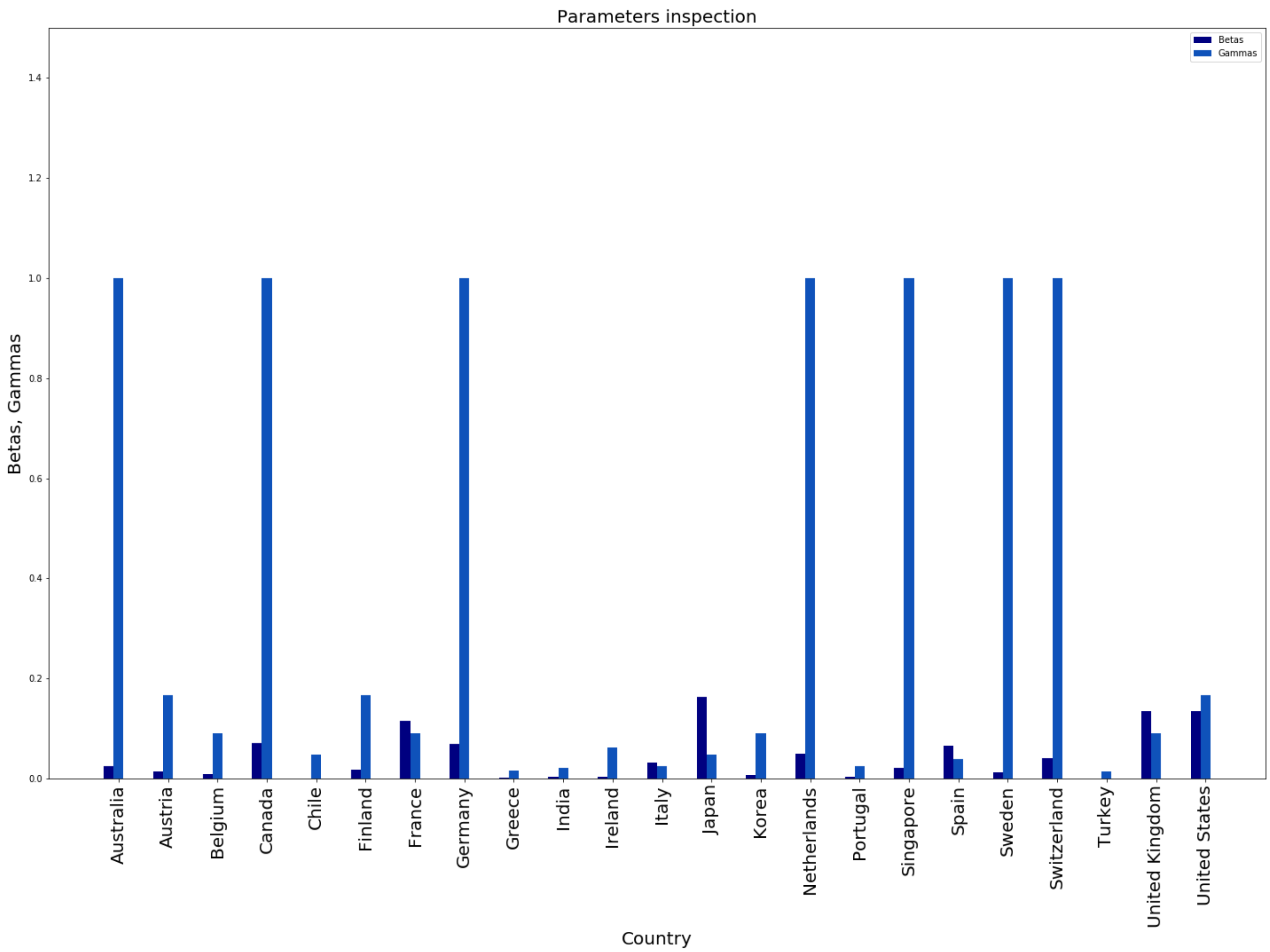
As discussed above, A is a symmetric matrix, resulting in an undirected graph. This approach does not allow full heterogeneity as betas are considered edge rather than node attributes. Consequently, if an edge links country A and B, the corresponding β will be the average of A's and B's β parameters.

The below network graph shows all 23 countries that are in scope, where nodes represent countries and edges represent bilateral claims.



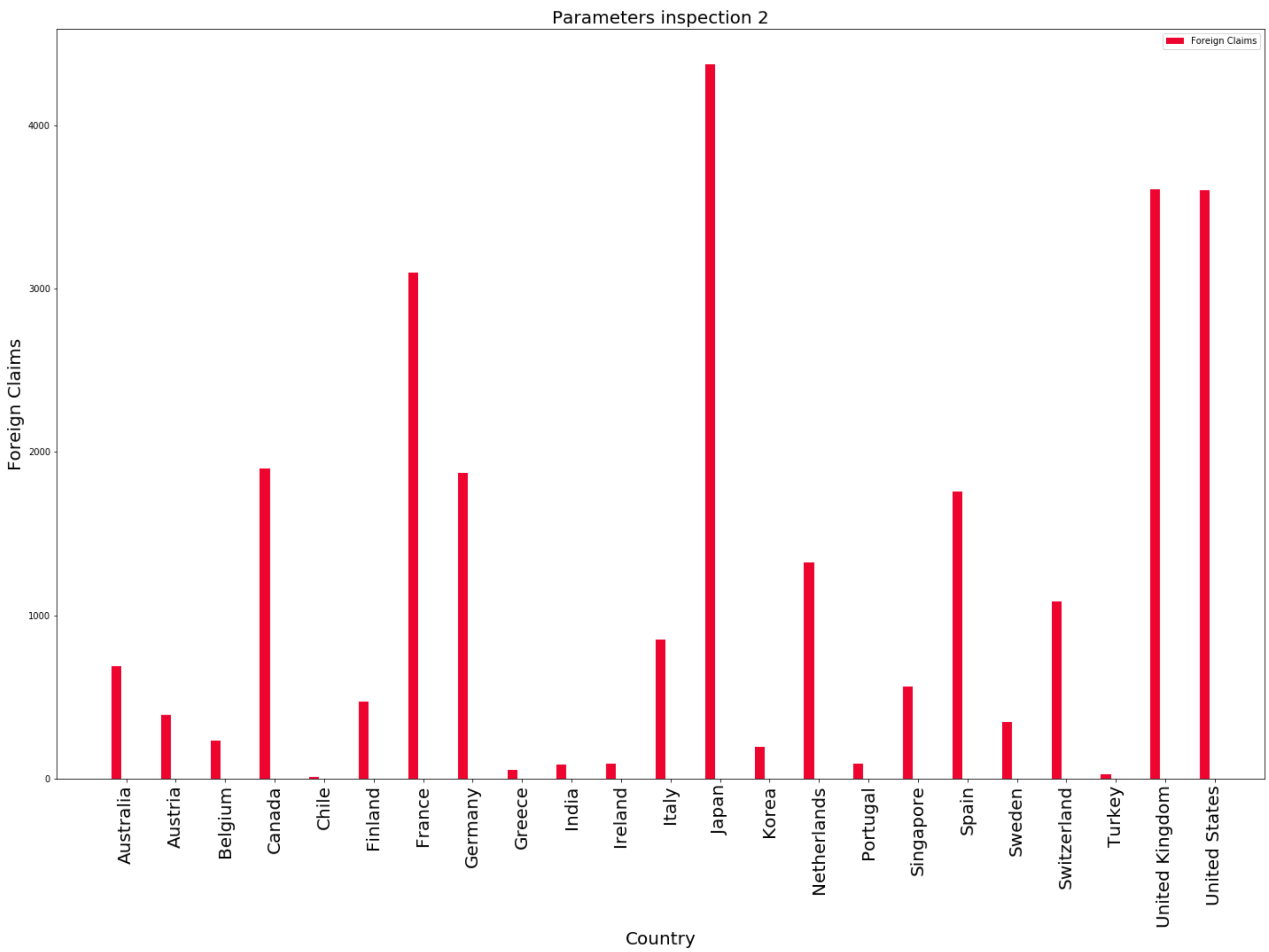
Exploratory Analysis

The first exploratory analysis we performed was the analysis of the β and γ values we calculate for each of the countries. We predicted that β and γ would display an inverse relationship. That is, a country with high susceptibility to default would exhibit slow recovery values and vice versa.



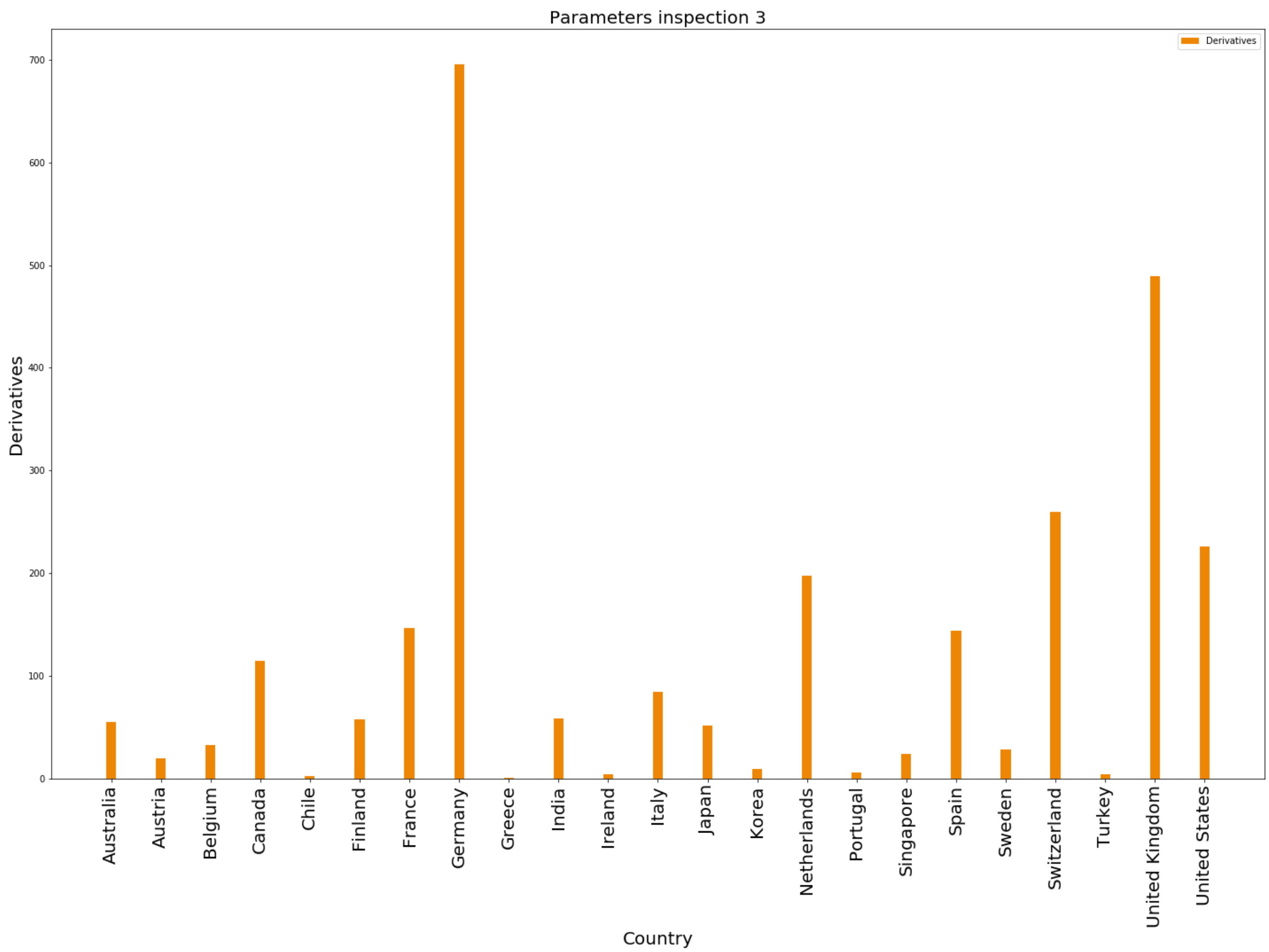
The graph above seems to confirm our hypothesis, with the exception of a few countries, namely Italy, Spain, Greece, India and the United States.

Next, we thought it would be interesting to inspect the level of foreign claims per country (in USD billions). Foreign claims comprise local claims of the bank's offices abroad as well as cross-border claims of the bank's offices worldwide. This graph gives us a good representation of a country's obligations.



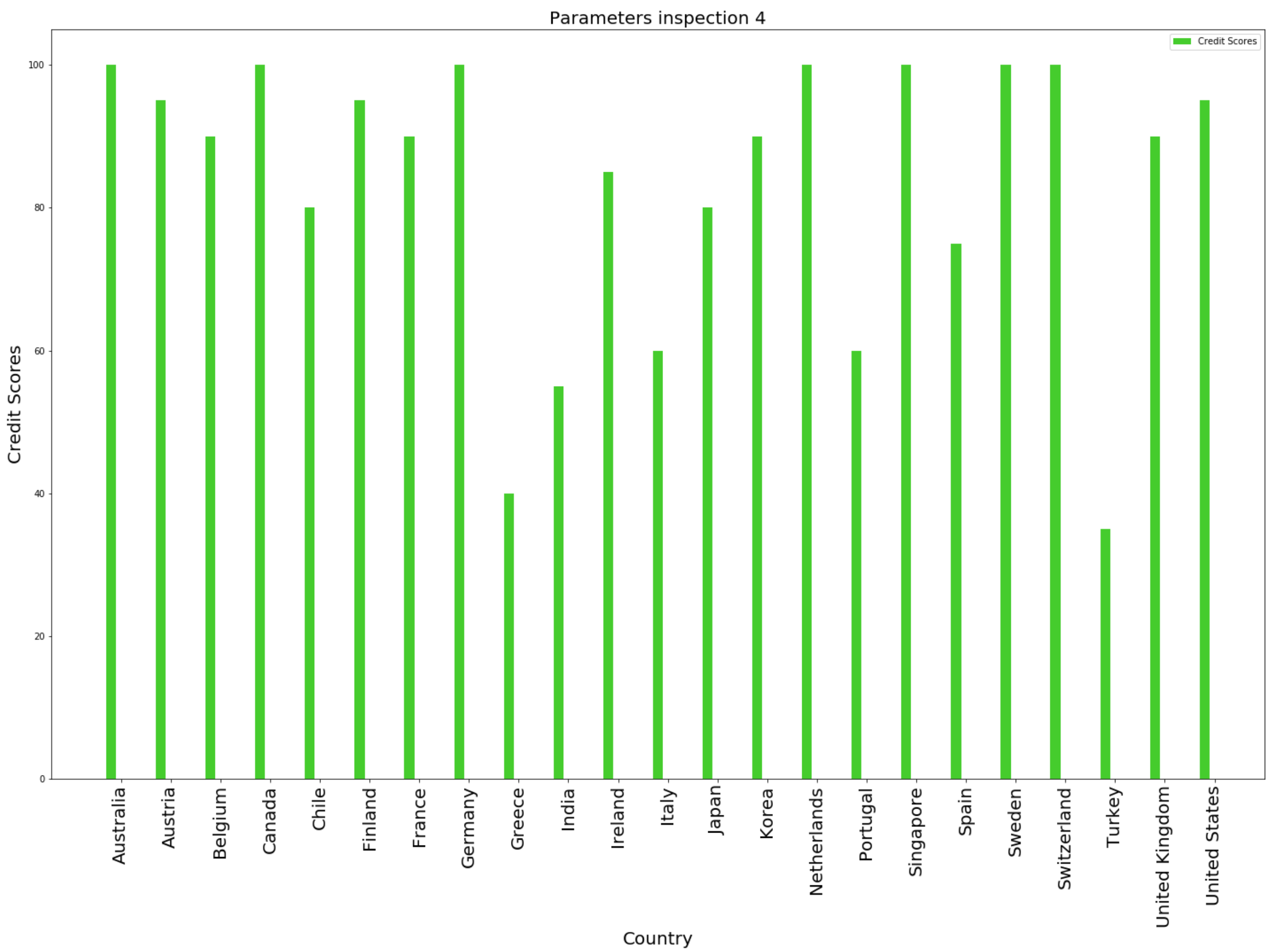
Japan has the highest number of foreign claims (more than USD 4.6 trillion), followed by the United Kingdom and the United States (at roughly USD 3.5 trillion). The countries with the smallest number are Chile, Turkey, and Greece (USD 15, 26, and 60 billion respectively).

Furthermore, we visualized the derivative market by country (in USD billions). After the 2007-2010 financial crisis the derivative market has been blamed for generating an irrational appeal to risk taking. Later on in our project it may be interesting to investigate if and how the derivatives market correlates with a country's susceptibility to default, its recovery or the overall propagation of defaults across the network.



The country with the highest amount of derivatives is Germany (about USD 700 billion), followed by the United Kingdom (USD 500 billion) and the United States (USD 230 billion). The bottom line is occupied again by Greece, Chile, and Turkey with less than USD 10 billion.

Last, we plotted the country's numerical credit score. This score is an indication of the inherent risk of lending to a specific country. Countries with low credit scores may incur in more difficulties when recovering from a default crisis as the costs of borrowing money is higher. A credit score of 100 corresponds to an S&P AAA, the safest rating.



As we can see, only two countries, namely Greece and Turkey, have the lowest credit scores, whereas many countries with high GDP per capita and a stable political situation have a high credit score (Germany, Australia, Canada, The Netherlands, Singapore, Sweden, and Switzerland).

Conclusion

In conclusion, in our project we are going to investigate systemic risk by applying a compartmental model to macroeconomic data about 23 different countries that was retrieved from the Bank for International Settlement. In this paper, we explained the steps that were taking towards preparing the necessary tools that would allows us to complete our objectives. We cleaned and processed the data, and engineered susceptibility and recovery values for the selected countries (β, γ). Moreover, we created the network between the different countries we selected in our dataset. In the network each country corresponds to a node, and the average between two countries's β corresponds to the weight of the edge that connects the two countries.

In the next steps, we will use the SIR model to simulate different scenarios of how a default crisis could propagate across a network of different countries' economies.

BIBLIOGRAPHY

1. STEVEN L. SCHWARCZ, SYSTEMIC RISK, 97 GEORGETOWN LAW JOURNAL 193-249 (2008). RETRIEVED JUNE 4, 2020, FROM DUKE UNIVERSITY LAW
2. BRUNNERMEIER, MARKUS KONRAD AND HANSEN, ET AL. MODELING AND MEASURING SYSTEMIC RISK (OCTOBER 15, 2010). AMERICAN ECONOMIC ASSOCIATION, TEN YEARS AND BEYOND: ECONOMISTS ANSWER NSF'S CALL FOR LONG-TERM RESEARCH AGENDAS. RETRIEVED JUNE 4, 2020, FROM SEMANTICSCHOLAR.ORG
3. BANK FOR INTERNATIONAL SETTLEMENTS (1994). 64TH ANNUAL REPORT, P.177. RETRIEVED APRIL 20, 2020, FROM BIS.ORG
4. DEUTSCHE BORSE GROUP (2008). THE GLOBAL DERIVATIVES MARKET: AN INTRODUCTION [WHITE PAPER], FROM NEW YORK UNIVERSITY
5. KOSTYLENKO, O., RODRIGUES, H.S., & TORRES, D.F. (2017). BANKING RISK AS AN EPIDEMIOLOGICAL MODEL: AN OPTIMAL CONTROL APPROACH. ARXIV: OPTIMIZATION AND CONTROL, 165-176.