Università
Ca'Foscari
Venezia

Double Degree Program

Master's Degree in Economia e Finanza
Università Ca' Foscari di Venezia

Master of Science - Financial Analytics
Stevens Institute of Technology

Final Thesis

The effect of S&P500 additions and deletions on CDS coverage

Supervisor
Ch. Prof. Monica Billio – Università Ca' Foscari di Venezia

Supervisor
Ch. Prof. Stefano Bonini – Stevens Institute of Technology

Graduand
Federico Mecchia
Matricolation number: 875577

Academic Year
2019 / 2020

Acknowledgements

I am profoundly grateful to my Supervisors, Professor Monica Billio and Professor Stefano Bonini (Stevens Institute of Technology), for their precious help and support.

Index

Abstract

The risk associated to financial transactions can be divided into two main groups: market risk and credit risk. To this regard, *Credit Default Swaps* (*CDS*) are a particular type of credit derivatives that are used to mitigate credit risk and that have for this reason played the role of protagonist in the global economic scenario over the most recent years.

Moreover, the *Credit Rating Agencies* have also played a leading role in the economic system for the last few decades. To this regard, *Standard & Poor's* is among the most important and influent *Credit Rating Agencies* worldwide.

This work considers several datasets and data regarding *CDS* and the *S&P500 Index* in order to determine the effect of the *S&P500 Index* additions and deletions on *CDS* and its characteristics.

From the various models and analyses considered it was possible to conclude that the initial hypothesis holds true and therefore that the probability that the companies that are constituents of the *S&P500 Index* are also reference entities of *CDS* contracts is higher than the probability that the companies that are not constituents of the *S&P500 Index* are also reference entities of the *CDS* contracts; in addition to this, several conclusions regarding specific characteristics of *CDS* could also be drawn.

In conclusion, the various models and analyses that were taken into consideration in this work have thus provided interesting results and insights about *CDS* and the *CDS* market.

Keywords: *Credit Derivatives*; *Credit Default Swaps* (*CDS*); *Credit Rating Agencies* (*CRA*); *Over The Counter Markets* (*OTC*); *S&P500 Index.*

---

1. Introduction, Motivation and Objective

Over recent decades technology and globalization have given origin to a very complex global economic scenario. In this globalized economic system, the number and the types of financial transactions have dramatically grown and this has also affected the risk associated to these transactions. This risk can be divided into two main groups: market risk and credit risk.

Different types of credit derivatives have been developed over recent decades in order to mitigate credit risk. Among the different types of credit derivatives, the *Credit Default Swaps*, also known as *CDS*, have become very popular and have thus played the role of protagonist over the most recent years.

It is difficult to determine the exact year during which the *CDS* contracts were born, but according to some Authors (Castellano and Giacometti, 2012; Augustin et al., 2014; Fu et al., 2020) they were probably created by *J.P Morgan* in 1994. Traded on *Over The Counter Markets* (*OTC Markets*), over the years *CDS* contracts have gradually spread worldwide and today they represent about half of the total credit derivatives market.

Besides *CDS*, also *Credit Rating Agencies* have been (and still are) key players of the global economy of the last few years. To this regard, in fact, today *Moody's*, *Fitch Ratings* and *Standard & Poor's*, the three major *Credit Rating Agencies*, represent more than 95% market share (Hill et al., 2010). These agencies all provide ratings regarding the specific entity taken into consideration; more specifically, each agency has its own criteria in order to determine the specific rating that has to be associated to the entity in question. To this regard, for example, *Standard & Poor's* provides alphanumerical ratings regarding the specific entity considered and these ratings therefore define all the aspects of credit risk that are associated to a financial instrument or to a specific entity (Niedziółka, 2019; White, 2018).

Moreover, *Standard & Poor's* also provides the *S&P500 Index*, which is a stock market index that takes into consideration 500 large companies, which are all listed on stock exchanges in the United States, and addresses their stock performance. This work considers the *S&P500 Index*, in particular the companies that are the constituents of the

S&P500 Index, together with other sources of data in order to carry out the different parts of the analysis that will be explained in later sections of this report.

The goal of this work, in fact, is to determine the impact that *S&P500 Index* additions and deletions have on *CDS* and its characteristics.

To this regard, no specific work regarding the way the *S&P500 Index* additions and deletions affect *CDS* and its characteristics could be identified. Over the years several researchers, though, have investigated peculiar aspects of the *Credit Rating Agencies* and of the *CDS* contracts and therefore the studies in question all provided (to different extents) interesting insights regarding some of the most important features of the *CDS* contracts and of the *Credit Rating Agencies*.

More specifically, some authors focused their attention on the *CDS* contracts and on the *CDS* market. To this regard, in fact, Anderson (2010) studied the way the *CDS* contracts can be used in order to mitigate risk and Freeman et al. (2006) analysed the various possible implementations of the *CDS* contracts in the most diverse economic scenarios.

In addition to this, other authors carried out researches regarding the *Credit Rating Agencies*. To this regard, Altman and Rijken (2006) analysed the main criteria that are considered by the agencies in question in order to determine the specific rating of a particular entity. Moreover, Cantor and Mann (2007) studied the various factors that can potentially affect the ratings given to a specific entity and Kiff et al. (2013) determined the way these ratings can change over time.

Furthermore, some authors focused their attention on the *CDS* market and on the stock market. To this regard, in fact, Steiner and Heinke (2001) analysed the impact of the rating announcements on both the *CDS* market and the stock market and determined that these two markets are affected in the same way by the announcements made by the *Credit Rating Agencies*. In addition to this, also Afonso et al. (2011) concluded that the rating announcements actually have an impact on the two markets in question, but also that there are other factors, which are not connected to the *Credit Rating Agencies*, that affect to different extents both the *CDS* market and the stock market. Moreover, Hull et al. (2004) carried out a similar analysis regarding the aforementioned markets and thus understood that both the *CDS* market and the stock market can be taken into consideration in order to determine the way the ratings that are associated to a particular entity will vary in the long term.

As it was stated earlier, the aim of this work is to determine the way the *S&P500 Index* additions and deletions affect *CDS* and its characteristics.

For this reason, the goal of this work can be divided into two main parts.
The first part involves determining the effect of the *S&P500 Index* additions and deletions on *CDS* and therefore, more specifically, the first part of the analysis is aimed at verifying that the probability that a company that is a constituent of the *S&P500 Index* is also a reference entity of *CDS* contracts is higher than the probability that a company that is not constituent of the *S&P500 Index* is also a reference entity of *CDS* contracts. Therefore, the first part of the analysis is aimed at determining whether the following hypothesis holds true or not:

$$p\ (CDS = 1 \mid SP = 1) > p\ (CDS = 1 \mid SP = 0)$$

Moreover, the second part of the analysis is aimed at exploring the characteristics of *CDS*; in this part of the analysis various variables that are included in the dataset will be considered and will therefore be analysed.
For this reason, several models were taken into consideration in order to carry out the various parts of the analysis. These models and parts of the analysis are explained in the following sections of this report.

2. Literature Review

As it was stated in the previous section of this report, the risk involved in financial transactions can be divided into two main groups: market risk and credit risk.

To this regard, Cossin and Pirotte (2001) describe credit derivatives by comparing and contrasting them to the financial derivatives and, in fact, they state that as financial derivatives can be used to manage market risk, analogously credit derivatives can be used as financial assets to transfer credit from one part to another. Over the most recent years the credit derivatives and the credit derivatives market have grown very rapidly and as a consequence also the credit risk and the credit risk market have undergone deep changes (Martire, 2007).

Moreover, the value of credit derivatives is directly connected to the reference entities (which are basically the sovereign issuing companies) and the subject against which the credit exposure can be identified is the underlying of the contract. In addition to this, one of the main peculiar characteristics of credit derivatives is the fact that when the premium is paid only the risk is transferred; for this reason, they are a very effective way of managing credit risk and they today represent one of the most important businesses in banking (Angelini, 2012). Another interesting aspect of credit derivatives is the fact that they are traded or exchanged on the *Over The Counter Markets* (*OTC Markets*) and therefore these contracts do not necessarily adhere to the conditions and rules of the stock market.

Credit Defaults Swaps are one of the most important types of credit derivatives and in the most recent years they have become very popular. *CDS* probably had origin in 1994, when *J.P. Morgan* first invented them (Castellano and Giacometti, 2012; Augustin et al., 2014; Fu et al., 2020).

As it was stated earlier, *CDS* have grown rapidly over the last years; today the *CDS* market is the most liquid credit derivative market and represents about half of the total credit derivative markets (Castellano and Giacometti, 2012; Wagner, 2008). As a consequence, because of the high volumes of *CDS* that are continuously traded worldwide, *CDS* are also considered as important elements that can help identify the perception of risk of bankruptcy on the international markets.

CDS have thus become a fundamental part of today's global economic scenario and for this reason several articles and papers about specific aspects of *CDS* have been published.

To this regard, Das and Hanouna (2006) focused their attention on the pricing of *CDS* contracts, Stulz (2010) then determined the impact of *CDS* on the financial crisis and Jarrow (2011) studied *CDS* and their potential application to the actuarial insurance markets. Moreover, Bolton and Oehmke (2013) determined the way *CDS* influence individual market participants and other Authors assessed the impact of the *CDS* market on the financing choices and the cost of capital of firms (Ashcraft and Santos, 2009; Saretto and Tookes, 2013; Subrahmanyam et al., 2014).

Furthermore, Reyngold et al. (2007) analysed the way the most peculiar characteristics of the *CDS* market change over time and Alexander and Kaeck (2008) studied the way the stock volatility affects the *CDS* market. Moreover, Blanco et al. (2005) determined that in the short term the credit risk of a specific entity can be derived by the *CDS* quotes. In addition to this, Acharya and Johnson (2007) concluded that the *CDS* contracts can be taken into consideration in order to determine the financial quality of the specific entity considered and Longstaff et al. (2005) also determined that the economic quality of a particular entity can be deduced by taking into account particular aspects of the *CDS* contracts.

As it can be easily inferred, *CDS* have played the role of protagonists in many different aspects of the global economic scenario over the most recent years, but they have not been the sole protagonists. To this regard, in fact, also the *Credit Rating Agencies* have also played a leading role in the global economic system over recent decades.

The *Credit Rating Agencies*, in fact, provide specific ratings regarding the particular entity considered and therefore they are of fundamental importance in many different situations and contexts (Jacobs et al., 2016).

Today *Moody's*, *Fitch* and *Standard & Poor's* are the three most important *Credit Rating Agencies* and in fact they together represent more than 95% market share (Hill et al., 2010).

In this work *Standard & Poor's* is taken into consideration. To this regard, in fact, as it was stated earlier, *Standard & Poor's* is among the most important *Credit Rating Agencies* worldwide and is therefore a key player of today's economy; for this reason, *Standard & Poor's* and the other two *Credit Rating Agencies* therefore have the capability of deeply influencing the opinions and consequently also the behaviours of the investors (Schoreder, 2015). More specifically, *Standard & Poor's* provides alphanumerical codes that describe the credit risk that is associated to a particular entity (Niedziółka, 2019). These alphanumerical codes are ranked from best to worst and therefore investors or, more generally, any user can easily understand the credit risk associated to the particular entity taken into consideration.

In addition to this, *Standard & Poor's* is also known for the *S&P500 Index*, which is an index that considers 500 large companies in the United States that are listed on the stock exchanges; this index is often taken into consideration to describe and summarize the performance of the U.S stock market.

In this work the 500 companies that are constituents of the *S&P500 Index* are taken into consideration to assess the impact that the additions and the deletions from this index have on *CDS* and its characteristics.

Therefore, because the *Credit Rating Agencies* have always been of fundamental importance over recent decades, several authors have thus studied various characteristics of the *Credit Rating Agencies* and the impact they can have on the global economic scenario.

To this regard, for example, Weinstein (1977), Pinches and Singleton (1978) studied the impact of the changes in specific credit ratings on the corporate asset prices and Cantor and Packer (1996) determined the effect of the *Credit Rating Agencies* announcements on the daily sovereign bond price and they thus concluded that, to a certain extent, the market spreads are influenced by the ratings of the *Credit Rating Agencies*. Moreover, Hite and Warga (1997) investigated the way the announcements of the *Credit Rating Agencies* influenced the corporate and sovereign bond prices, Dichev and Piotroski (2001) and Vassalou and Xing (2004) studied the effect of the announcements made by *Credit Rating Agencies* on equity prices and Hamilton and Cantor (2007) assessed, in general terms, the impact of ratings on *CDS* spreads.

In addition to this, also other authors carried out various researches regarding different aspects of the *CDS* spreads. To this regard, Ammer and Clinton (2004) concluded that the rating announcements have an impact on the *CDS* spreads and Ismailescu and Kazemi (2010), who took into consideration only the most important emerging sovereign markets, also determined that the changes in *CDS* spreads are mainly affected by the *Credit Rating Agencies* announcements.

Furthermore, Zhang et al. (2009) found that in the short run the changes in *CDS* spreads are the result of the fluctuations of volatility and Boss and Scheicher (2002), who considered only the European countries in order to carry out their research, concluded that the risk-free rate and the volatility are the two main causes of the changes in *CDS* spreads in the short term.

Therefore, over the years numerous studies have investigated the credit derivatives, the *CDS* and also the *Credit Rating Agencies* and also the impact of these financial instruments and entities on several aspects of the economic system. No specific study regarding the impact and the effect of the *S&P500 Index* additions and deletions on *CDS*, though, could be identified, but the aforementioned articles and studies each proved to be useful in understanding single, specific and particular aspects that are taken into consideration in this work.

3. Materials and Methods

Several datasets and databases as well as models and methods were considered in order to assess, study and investigate the various questions that were explained in the introduction. This part of the report explains the dataset that was used as well as the models that were considered, created, tested and implemented.

The following paragraphs describe the dataset that was created in order to carry out the various parts of the analysis; to this regard the software STATA was used to clean the data and also to create the dataset in question.

The dataset was implemented by taking into consideration the data regarding the companies that are reference entities of *CDS* contracts and the data regarding the companies that are the constituents of the *S&P500 Index*; the former was taken from the database named *Markit* and the latter was taken from the database named *Compustat*.
The data regarding the companies that are reference entities of *CDS* contracts that was taken from *Markit* consisted in quarterly data from year 2001 until year 2018; therefore, 72 quarters in total were considered. The data of each quarter was then exported into a single file (and so 72 files were thus created).
Moreover, as it was stated earlier, the data regarding the companies that are the constituents of the *S&P500 Index* was taken from *Compustat*. Also in this case 72 files containing the quarterly data of the 18 years considered were created.
In order to build the dataset that had to be considered in order to carry out the various parts of the analysis, the data taken from *Markit* was merged with the data taken from *Compustat*. As it can be inferred, because the data was taken from different databases, no criteria could be used to accurately and automatically merge the 72 files taken from *Marki*t with the 72 files taken from *Compustat*; this is due to the fact that the identifiers that are used in the *Markit* datasets are different from the identifiers that are considered in the *Compustat* datasets. For this reason, a manual merge and selection procedure were enhanced in order to merge the data taken from *Markit* with the data taken from *Compustat*. Therefore, each company of each of the 72 files taken from *Markit* was taken into consideration and was eventually associated (if possible) to the corresponding

company that could be found in the 72 files taken from *Compustat*. In the end 72 new files were thus created.

The aim of this manual merge was basically to identify which companies that are reference entities of the *CDS* contracts are (or are not) companies that are constituents of the *S&P500 Index* and also which companies that are constituents of the *S&P500 Index* are (or are not) reference entities of *CDS* contracts.

All the remaining steps involved in the creation of the dataset consisted in using the software STATA. In other words, the aforementioned part is the only part that required manual merging and selection procedures. For this reason, specific commands in STATA were then used to add new data to this dataset; this could be obtained automatically since the 72 datasets in question all contained particular identifiers and other variables that had been created and that could therefore be used in order to merge these datasets with the new data.

The new data consisted of 18 years (from year 2001 until year 2018) of quarterly data (therefore 72 quarters in total) and contained the companies of NASDAQ and NYSE.

After merging the new data with the 72 datasets in question, 72 new files were thus created; these 72 files were then all combined together into one single file.

The final dataset therefore contains 18 years (from 2001 until 2018) of quarterly data and so contains 72 quarters in total.

As it was stated earlier, the final dataset also includes specific variables that will be used in the analysis. To this regard, for example, the final dataset contains data regarding the *spread5y*, the *recovery*, the *currency* and the *sector* of the various companies included in the final dataset (Table 1).

Table 1: Description of the variables contained in the dataset.

VARIABLE	**DESCRIPTION**
spread5y	The par spread (no coupon spread) for the tenor in 5 years
recovery	The composite recovery rate associated with the *CDS* curve
ccy	The currency of which the *CDS* is priced on
sector	The sector of the company considered

In addition to this, the final dataset also contains two columns, named *CDS* and *SP*.

More specifically, the column named *CDS* takes into consideration whether the company considered is a reference entity of *CDS* contracts and the column *SP* considers whether

the company in question is a company that is constituent of the *S&P500 Index*. The values contained in these two columns can either be 0 or 1; the value 1 in the column *CDS* indicates that the company taken into consideration is a reference entity of the *CDS* contracts and analogously the value 1 in the column *SP* indicates that the company considered is a company that is constituent of the *S&P500 Index*. Moreover, the value 0 in the column *CDS* indicates that the company considered is not a reference entity of *CDS* contracts and the value 0 in the column *SP* indicates that the company taken into consideration is not a constituent of the *S&P500 Index*.
In other words, the following four cases can be found in the final dataset:

- *CDS* = 0 and *SP* = 1: this means that the company in question is not a reference entity of *CDS* contracts but also that the specific company considered is a constituent of the *S&P500 Index*;
- *CDS* = 1 and *SP* = 0: this means that the company in question is a reference entity of *CDS* contracts, but also that the company in question is not a constituent of the *S&P500 Index*;
- *CDS* = 1 and *SP* = 1: this means that the company in question is a reference entity of *CDS* contracts and that it is also a constituent of the *S&P500 Index*;
- *CDS* = 0 and *SP* = 0: this means that the company in question is not a reference entity of *CDS* contracts and also that it is not a constituent of the *S&P500 Index*.

As it was stated earlier, the final dataset was taken into consideration and used in order to carry out all the various parts of the analysis.

Moreover, the software STATA was used to create and implement the different models that were considered in the various parts of the analysis; to this regard, the following paragraphs explain the main models that were taken into consideration.

The first model that was considered in the analysis is the *Logit model*, which is a non-linear regression model that is used when the dependent variable is dichotomic and which is therefore implemented in order to model the probability that a particular event, expressed as a binary variable, occurs.
More specifically, from a mathematical perspective, consider a row vector x_i; the first column of this vector is 1, in other words $x_i = (1, x_1, x_2, \ldots, x_{ip})$ with p explanatory

variables. Moreover, the vector β can be defined as $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$, where β_0 is the intercept.

Therefore, for $i = 1, \ldots, n$ (where n indicates the number of observations), the model can thus be defined as:

$$logit(P(Y_i = 1|x_i)) = \log\left(\frac{P(Y_i = 1|x_i)}{P(Y_i = 0|x_i)}\right) = \log\left(\frac{P(Y_i = 1|x_i)}{1 - P(Y_i = 1|x_i)}\right) = x_i\beta$$

Or equally as:

$$P(Y_i = 1|x_i) = \mathrm{F}(x_i\beta) = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}}$$

Where $F(z) = \frac{e^z}{1+e^z}$ is the distribution function of the standard logistic distribution.

Moreover, in the various parts of the analysis different variables were taken into consideration. To this regard, for example, some models were created by considering the variables *CDS* as y (which can be equal to either 0 or 1), *SP* as x_1 and *currency* as z. Since the variable *currency* is a categorical variable with several levels, the variable in question is split into two variables 0 and 1, one identified by x_2 and the other one identified by x_3. The model can thus be defined as:

$$logit(P(Y_i = 1|x_i)) = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}$$

Furthermore, also the *Random Effects Logit model* was taken into consideration in the analysis.

From a mathematical perspective, consider the row vector x_{it}; in this case the first column is 1, in other words $x_{it} = (1, x_{it1}, x_{it2}, \ldots, x_{itp})$ with p explanatory variables. Moreover, consider $i = 1, \ldots, n$ (where n is the number of observations) and $t = 1, \ldots, T_i$ (where t indicates the time). Therefore, the *Random Effects Logit model* can be defined as:

$$logit(P(Y_{it} = 1|x_{it}, v_i)) = \log\left(\frac{P(Y_{it} = 1|x_{it}, v_i)}{P(Y_{it} = 0|x_{it}, v_i)}\right) =$$
$$= \log\left(\frac{P(Y_{it} = 1|x_{it}, v_i)}{1 - P(Y_{it} = 1|x_{it}, v_i)}\right) = x_i\beta + v_i$$

Or equally as:

$$P(Y_{it} = 1|x_{it}) = F(x_{it}\beta + v_i) = \frac{e^{x_{it}\beta+v_i}}{1+e^{x_{it}\beta+v_i}}$$

Where $F(z) = \frac{e^z}{1+e^z}$ is the distribution function of the standard logistic distribution.

In this type of model, the particular characteristics of a specific variable that cannot be observed and that are shared by all the observations are described as a random variable, named *random effect*. In addition to this, the *random effects* that are related to distinct variables are considered to be independent.

Moreover, v_i represents the effect of a specific subject considered and, since the subjects are randomly extracted, the various effects identified by v_i can be considered to be independent and identically distributed random variables with average equal to zero; in other words $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$.

To this regard, in the various parts of the analysis that were carried out, the subject was always represented by the specific company taken into consideration.

Furthermore, also the *Fixed Effects Logit model* was taken into consideration in the various parts of the analysis.

In this case, the same assumptions that were made for the *Random Effects Logit model* are still valid for the *Fixed Effects Logit model*. The only difference is the fact that v_i, which represents the effect of the specific subject considered (a specific company in this case), is not a random variable but it is considered to be a parameter that has to be estimated by the model in question. Therefore, the *Fixed Effects Logit model* can simply be defined as:

$$logit\big(P(Y_{it} = 1|x_{it}, v_i)\big) = \log\left(\frac{P(Y_{it} = 1|x_{it}, v_i)}{P(Y_{it} = 0|x_{it}, v_i)}\right) =$$
$$= \log\left(\frac{P(Y_{it} = 1|x_{it}, v_i)}{1 - P(Y_{it} = 1|x_{it}, v_i)}\right) = x_i\beta + v_i$$

Or equally as:

$$P(Y_{it} = 1|x_{it}) = F(x_{it}\beta + v_i) = \frac{e^{x_{it}\beta+v_i}}{1 + e^{x_{it}\beta+v_i}}$$

Where $F(z) = \frac{e^z}{1+e^z}$ is the distribution function of the standard logistic distribution. Therefore, in this case the parameters v_i and β are not known and so have to be estimated by the model in question. To this regard, it is interesting to observe the fact that as $n \to \infty$, also the number of parameters v_i increases. For this reason, in some cases the model does not produce significant results (in some cases it does not converge); this is known as the *incidental parameters problem*.

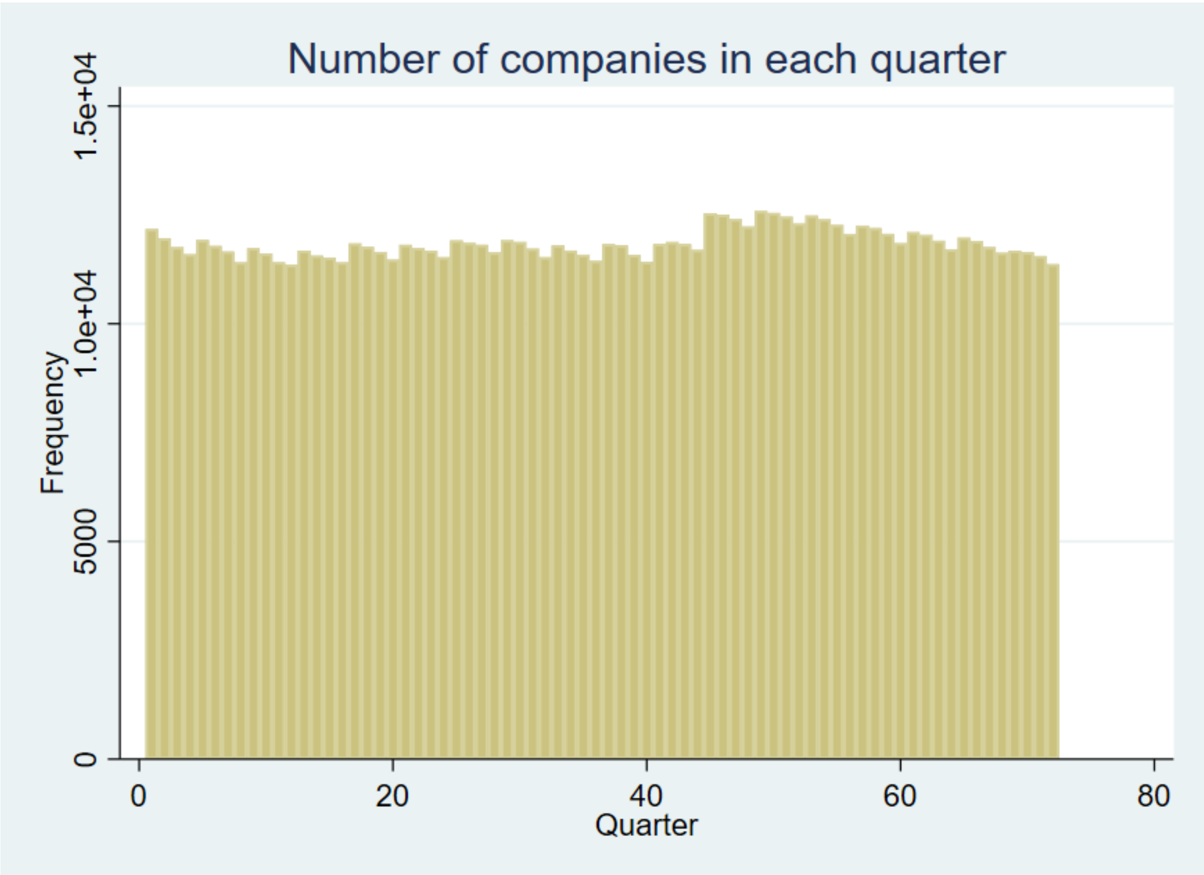
The following sections of this report explain the various parts of the analysis that were carried out by taking into consideration the aforementioned dataset and these models.

4. Analysis and Discussions

Several analyses were carried out in order to answer and assess the different questions explained in the previous sections of this report. To this regard, the following paragraphs explain the various parts of the analysis and also include several visualizations of the dataset that was explained in the previous section of this report.

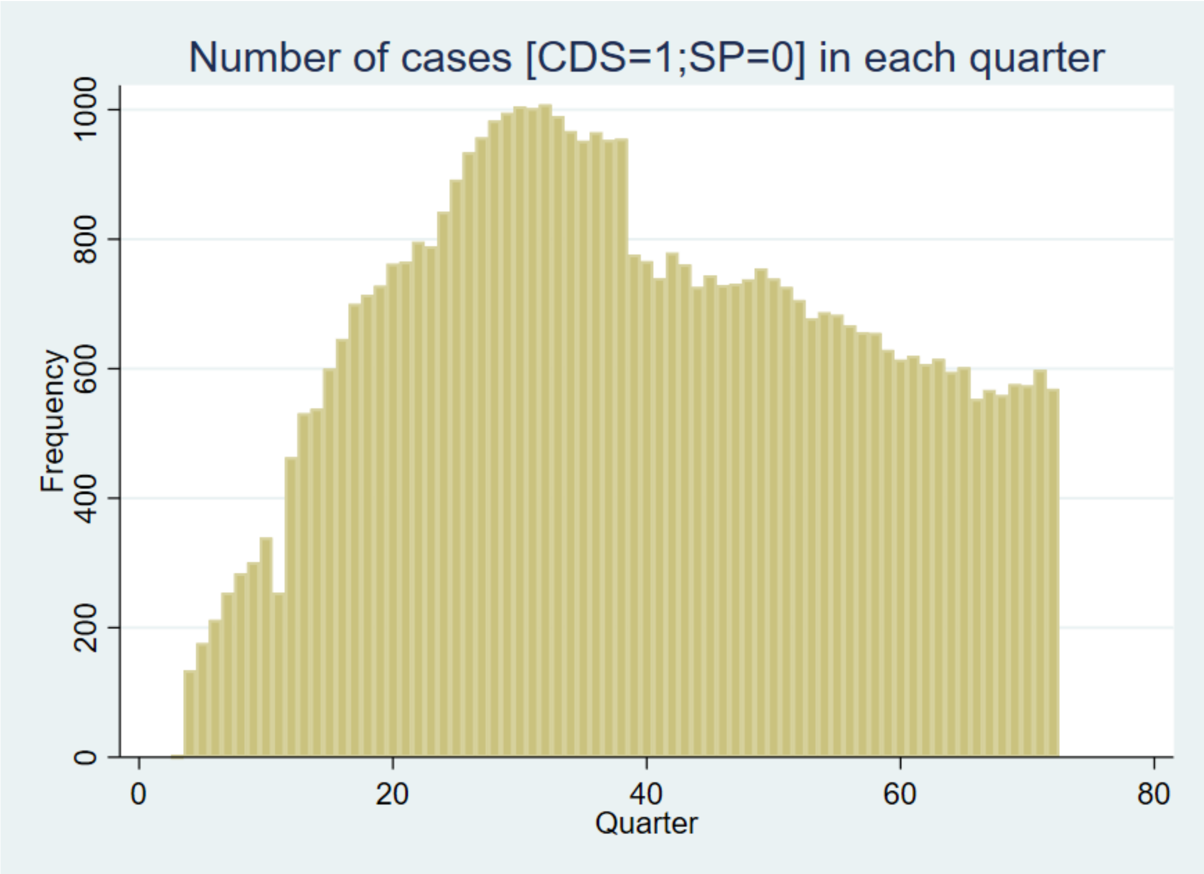
First, some visualizations regarding the dataset in question were created. To this regard, the following visualization display the number of companies in each of the 72 quarters considered (Graph 1).

Graph 1: Number of companies in the period 2001-2018.

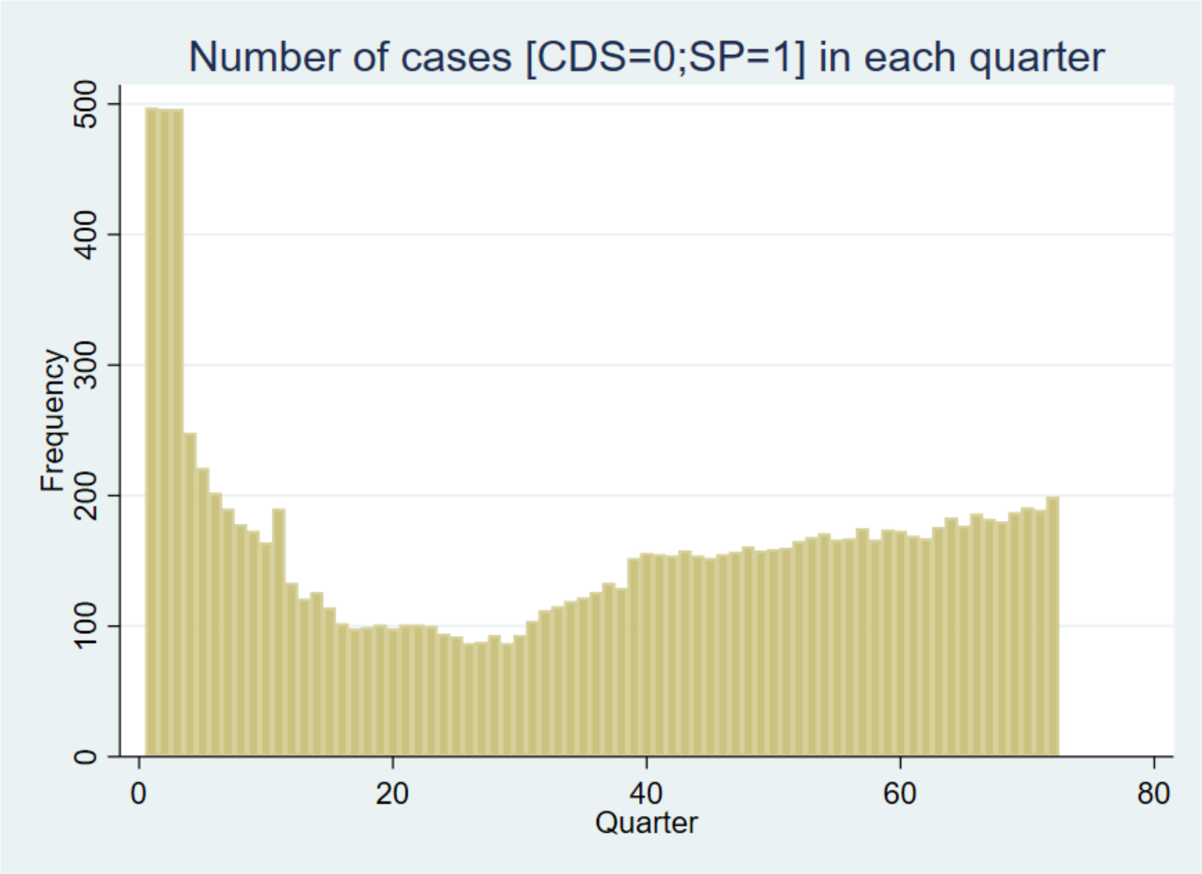


Moreover, the following visualizations display the number of cases [*CDS*=1; *SP*=0] (Graph 2) and [*CDS*=0; *SP*=1] (Graph 3) in each quarter.

Graph 2: Number of cases [*CDS*=1; *SP*=0] in each quarter.



Graph 3: Number of cases [*CDS*=0; *SP*=1] in each quarter.



As it can be observed from these visualizations, the number of companies in each quarter can be considered to be almost constant over the 18 years considered, while the numbers of cases regarding the different possible combinations of the variables *CDS* and *SP* vary over the time span considered.

Furthermore, as it was stated in the previous sections of this report, the first part of the analysis is aimed at determining whether the companies that are constituents of the *S&P500 Index* have a higher probability of being reference entities of *CDS* contracts. Therefore, the first part of the analysis is aimed at determining whether the following hypothesis holds true or not:

$$p\ (CDS = 1 \mid SP = 1) > p\ (CDS = 1 \mid SP = 0)$$

Three models were implemented in order to determine whether the aforementioned hypothesis holds true or not. To this regard, in fact, the following models were considered:

- *Logit model*;
- *Random Effects Logit model*;
- *Fixed Effects Logit model*;

First the *Logit model* was considered with regard to the columns named *CDS* and *SP* and the following command was therefore used in STATA:

logit CDS SP

These results were thus obtained (Table 2):

Table 2: *Logit model* applied to the whole dataset (only *CDS* and *SP* are considered).

Iteration 5: log likelihood = -202783.67			
Logistic regression	Number of obs	=	853,116
	LR chi2(1)	=	84819.10
	Prob > chi2	=	0.0000
Log likelihood = -202783.67	Pseudo R2	=	0.1730

CDS	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
SP	3.521776	0.0122248	288.08	0.000	3.497816	3.545736
_cons	-2.794143	0.0047462	-588.71	0.000	-2.803445	-2.78484

From these results it is possible to observe the fact that the coefficient of *SP* is positive and so this means that the probability that a company that is a constituent of the *S&P500 Index* is also a reference entity of *CDS* contracts is higher than the probability that a company that is not a constituent of the *S&P500 Index* is a reference entity of *CDS* contracts. In other words, the results of this model confirm the fact that the aforementioned hypothesis holds true.

In addition to this, since the coefficient of *SP* is positive, it is also possible to state that the probability that a company that is a reference entity of *CDS* contracts is also a company that is constituent of the *S&P500 Index* is higher than the probability that a company that is not a reference entity of *CDS* contracts is a company that is constituent of the *S&P500 Index*. In other words, from these results it is possible to state that also the following hypothesis holds true:

$$p\ (SP = 1 \mid CDS = 1) > p\ (SP = 1 \mid CDS = 0)$$

Moreover, as it was stated earlier, also the *Random Effects Logit model* was considered with regard to the variables named *CDS* and *SP* and therefore the following command was used in STATA:

xtlogit CDS SP,re

In this way the following results were obtained (Table 3):

Table 3: *Random Effects Logit model* applied to the whole dataset (only *CDS* and *SP* are considered).

```
Iteration 18: log likelihood = -30440.596

Random-effects logistic regression        Number of obs      =    853,116
Group variable: newid                     Number of groups   =     27,141

Random effects u_i ~ Gaussian             Obs per group:
                                          Min                =          1
                                          Avg                =       31.4
                                          Max                =         72

Integration method: mvaghermite           Integration pts.   =         12

                                          Wald chi2(1)       =    3734.39
Log likelihood = -30440.596               Prob > chi2        =     0.0000
```

CDS	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
SP	4.015183	0.0657046	61.11	0.000	3.886405	4.143962
_cons	-16.60779	0.0816273	-203.46	0.000	-16.76778	-16.44781
/lnsig2u	4.970724	0.0165022			4.93838	5.003067
sigma_u	12.00546	0.0990583			11.81287	12.20119
rho	0.9776839	0.00036			0.9769672	0.9783787
LR test of rho=0: chibar2(01) = 3.4e+05					Prob >= chibar2 = 0.000	

Also in this case it is possible to observe that the coefficient of *SP* is positive and so the conclusions that could be drawn in the case of the *Logit model* are valid also in this case.

Furthermore, the following command was used in STATA in order to implement the *Fixed Effects Logit model* with regard to the columns named *CDS* and *SP*:

xtlogit CDS SP, fe

These are the results obtained by considering the *Fixed Effects Logit model* (Table 4):

Table 4: *Fixed Effects Logit model* applied to the whole dataset (only *CDS* and *SP* are considered).

Iteration 4: log likelihood = -17031.517			
Conditional fixed-effects logistic regression	Number of obs	=	52,262
Group variable: newid	Number of groups	=	886
	Obs per group:		
	Min	=	4
	Avg	=	59.0
	Max	=	72
	LR chi2(1)	=	5547.99
Log likelihood = -17031.517	Prob > chi2	=	0.0000

CDS	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
SP	3.61043	0.0606142	59.56	0.000	3.491629	3.729232

These results display the fact that the coefficient of *SP* is positive and therefore the conclusions of the previous two models are true also in this case.

Therefore, all the three models considered display a positive coefficient for *SP*. For this reason, it is possible to state that:

- The probability that a company that is a constituent of the *S&P500 Index* is also a reference entity of *CDS* contracts is higher than the probability that a company that is not a constituent of the *S&P500 Index* is a reference entity of *CDS* contracts;
- The probability that a company that is a reference entity of *CDS* contracts is also a constituent of the *S&P500 Index* is higher than the probability that a company that is not a reference entity of *CDS* contracts is a constituent of the *S&P500 Index*.

Even if all the three models display a positive coefficient of *SP* and therefore lead to the same conclusions, the *Random Effects Logit model* was chosen as the best model, because:

- The aim of this work (and more specifically of this part of the analysis) is to determine an overall effect or pattern between the columns (or variables) named *CDS* and *SP* and not the single effects or patterns regarding each single company considered in the dataset;
- The time span considered is long and in fact the dataset includes 18 years of quarterly data (therefore 72 quarters in total);
- The dataset is very large and, in fact, it contains more than 800,000 rows;
- As it will be shown in the next section of the analysis, in some cases the other models (for instance the *Logit model* and the *Fixed Effects Logit model*) do not always produce significant results.

As it was stated in the previous sections of this report, the second part of the analysis is aimed at studying and exploring other variables included in the dataset.

Therefore, in order to carry out the second part of the analysis, some preliminary data analyses were also considered.

To this regard, different currencies could be identified in the dataset; for this reason, the variable that was originally named *ccy* was thus renamed *Currency2* and the various currencies that were part of the dataset were grouped into the following three groups:

- *Other*;
- *CAD*;
- *USD*.

Different models were then considered with regard to the variables named *CDS*, *SP* and *Currency2*.

The first model that was taken into consideration was the *Logit model*. To this regard, as it was stated earlier, the variables *CDS*, *SP* and *Currency2* were included in the model in question and the group named *Other* of the variable named *Currency2* was taken as reference for the other two groups of the variable *Currency2* (for instance *CAD* and *USD*). This command was used in STATA to create this model:

logit CDS SP ib2.Currency2

The following results were thus obtained (Table 5):

Table 5: *Logit model* applied to the whole dataset (*CDS*, *SP* and *Currency2* are considered).

Iteration 6: log likelihood = -120631.06			
Logistic regression	Number of obs	=	853,116
	LR chi2(3)	=	249124.31
	Prob > chi2	=	0.0000
Log likelihood = -120631.06	Pseudo R2	=	0.5080

CDS	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
SP	3.581928	0.0144286	248.25	0.000	3.553649	3.610208
Currency2						
CAD	-11.65096	0.2782351	-41.87	0.000	-12.19629	-11.10563
USD	-10.97088	0.2775097	-39.53	0.000	-11.51479	-10.42697
_cons	7.541806	0.2774228	27.19	0.000	6.998068	8.085545

These results display a positive coefficient for the variable *SP* and negative coefficients for *CAD* and *USD*. This means that also the *Logit model* that considers the variables named *CDS*, *SP* and *Currency2* confirms the fact that the probability (adjusted for the variable *Currency2*) that a company that is constituent of the *S&P500 Index* is also a reference entity of *CDS* contracts is higher than the probability that a company that is not a constituent of the *S&P500 Index* is also a reference entity of *CDS* contracts.

Moreover, since the coefficients of *CAD* and *USD* are negative, it is possible to state that, given a specific value of *SP* (therefore either 0 or 1), the probability that a company with either *CAD* or *USD* as currency is a reference entity of *CDS* contracts (therefore the value associated to *CDS* would be 1 in this case) is lower than the probability that a company with *Other* as currency is a reference entity of *CDS* contracts.

Furthermore, also the *Random Effects Logit model* was considered with regard to the variables named *CDS*, *SP* and *Currency2* and with the group *Other* of the variable *Currency2* taken as reference in the model.

The following command was used in STATA:

xtlogit CDS SP ib2.Currency2, re

These are the results that were obtained (Table 6):

Table 6: *Random Effects Logit model* applied to the whole dataset (*CDS*, *SP* and *Currency2* are considered).

Iteration 12: log likelihood = -25623.738			
Random-effects logistic regression	Number of obs	=	853,116
Group variable: newid	Number of groups	=	27,141
Random effects u_i ~ Gaussian	Obs per group:		
	min	=	1
	avg	=	31.4
	max	=	72
Integration method: mvaghermite	Integration pts.	=	12
	Wald chi2(3)	=	3075.12
Log likelihood = -25623.738	Prob > chi2	=	0.0000

CDS	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
SP	2.885575	0.0660816	43.67	0.000	2.756057	3.015092
Currency2						
CAD	-40.80244	1.250108	-32.64	0.000	-43.25261	-38.35228
USD	-41.63995	1.24753	-33.38	0.000	-44.08506	-39.19484
_cons	26.91321	1.245264	21.61	0.000	24.47254	29.35389
/lnsig2u	4.44697	0.0161625			4.415292	4.478648
sigma_u	9.239475	0.0746666			9.094284	9.386984
rho	0.9628925	0.0005775			0.9617439	0.9640079
LR test of rho=0: chibar2(01) = 1.9e+05				Prob >= chibar2 = 0.000		

Also in this case, the coefficient of *SP* is positive and the coefficients of *CAD* and *USD* are negative and so the same conclusions can be drawn also in this case.

In addition to these two models, also the *Fixed Effects Logit model* was considered with regard to the variables named *CDS, SP* and *Currency2* (and also with the group *Other* of the variable named *Currency2* as reference).

Therefore, the following command was used in STATA:

xtlogit CDS SP ib2.Currency2, fe

These results were thus obtained (Table 7):

Table 7: *Fixed Effects Logit model* applied to the whole dataset (*CDS, SP* and *Currency2* are considered).

convergence not achieved					
Conditional fixed-effects logistic regression			Number of obs	=	52,262
Group variable: newid			Number of groups	=	886
			Obs per group:		
			min	=	4
			avg	=	59.0
			max	=	72
			LR chi2(3)	=	10970.83
Log likelihood	=	-14320.095	Prob > chi2	=	0.0000

CDS	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
SP	2.532968	0.0632978	40.02	0.000	2.408906	2.657029
Currency2						
CAD	-9.118329	1.593203	-5.72	0.000	-12.24095	-5.995709
USD	-11.26838	1.587016	-7.10	0.000	-14.37887	-8.157881

Therefore, in this case the result states *"converge not achieved"* and so the model does not provide significant results. As it was stated earlier, this is also one of the reasons why the *Random Effects Logit model* was preferred to the other models in the first part of the analysis (when just the variables *CDS* and *SP* were taken into consideration).

The fact that the *Fixed Effects Logit model* does not produce significant results (and therefore displays *"convergence not achieved"*) is known as the *Problem of monotone likelihood* and this is basically caused by the fact that, because the dataset contains some particular cases (for example particular combinations of the values of *CDS*, *SP* and *Currency2*), the value of the *log likelihood* does not improve as the number of iterations increases and in the same way it does improve as the various coefficients and intercepts are estimated.

Finally, the *Logit model* of the first part of the analysis (that thus considered just the variables *CDS* and *SP*) was compared to the *Logit model* that considered the variables *CDS*, *SP* and *Currency2*. To this regard, these are the commands that were used in STATA to save and store the *Logit model* of the first part of the analysis (named *relog_tot*) and the *Logit model* of the second part of the analysis (named *relog_tot1*):

estimates store relog_tot

estimates store relog_tot1

Moreover, this is the command that was used in STATA to compare the two models in question:

lrtest relog_tot relog_tot1, stats

These are the results of this comparison (Table 8):

Table 8: Comparison between the two aforementioned *Logit Models*.

Model	N	ll(null)	ll(model)	df	AIC	BIC
relog_tot	853,116	-245193.2	-202783.7	2	405571.3	405594.6
relog_tot1	853,116	-245193.2	-120631.1	4	241270.1	241316.8

From these results, it is possible to observe that the *Logit model* that took into consideration the variables named *CDS*, *SP* and *Currency2* must be preferred to the *Logit model* explained in the first part of the analysis, because the values of *AIC* and *BIC* are both lower when the *Logit model* with the variables *CDS*, *SP* and *Currency2* (named *relog_tot1*) is taken into consideration.

In a similar way a comparison between the *Random Effects Logit model* of the first part of the analysis (that took into consideration just the variables *CDS* and *SP*) and the *Random Effects Logit model* of the second part of the analysis (that took into consideration the variables *CDS, SP* and *Currency2*) was considered. The following commands were used in STATA to name the *Random Effects Logit model* of the first part of the analysis *rand_tot* and the *Random Effects Logit model* of the second part of the analysis *rand_tot1*:

estimates store rand_tot

estimates store rand_tot1

This command was then used to compare the two models:

lrtest rand_tot rand_tot1, stats

These results were thus obtained (Table 9):

Table 9: Comparison between the two aforementioned *Random Effects Logit Models*.

Model	N	ll(null)	ll(model)	df	AIC	BIC
relog_tot	853,116	.	-30440.6	3	60887.19	60922.16
relog_tot1	853,116	.	-25623.74	5	51257.48	51315.76

Also in this case it is possible to observe that the *Random Effects Logit model* of the second part of the analysis (named *rand_tot1)* must be preferred to the *Random Effects Logit model* of the first part of the analysis (named *rand_tot*) since the *AIC* and *BIC* values are both lower when the model of the second part of the analysis is considered.

Finally, for the same reasons that were explained in the first part of the analysis, also in this second part of the analysis the *Random Effects Logit model* was preferred to *Logit model*.

Therefore, this second part of the analysis confirmed the fact that the coefficient of the variable *SP* is positive (and therefore the aforementioned conclusions can be drawn) and also that the coefficients of the groups *CAD* and *USD* of the variable named *Currency2* are negative (and so also in this case the aforementioned conclusions are valid).

Furthermore, a subset of the whole dataset was created in order to carry out the other parts of the analysis. More specifically, the subset in question contains all the companies that have the value of *CDS* equal to 1 and so both the cases [*CDS*=1; *SP*=0] and [*CDS*=1; *SP*=1] are taken into consideration in this subset.
In addition to this, in this specific analysis all the companies whose values of *spread5y* could not be identified were removed from the subset in question.
Moreover, the variable regarding the various sectors was named *sect2* and the various sectors were thus grouped into these groups:

- *service*;
- *utilities*;
- *industry*.

Therefore, the subset in question has the following characteristics:

- It contains only the cases [*CDS*=1; *SP*=0] and [*CDS*=1; *SP*=1];
- The companies whose values of *spread5y* could not be identified were removed from the subset in question;
- The various sectors that could be identified were grouped into three main groups, named *service*, *utilities* and *industry*, and the variables regarding the sectors was named *sect2*.

Also in this case, for the reasons explained in the previous sections of this report, the *Random Effects Logit model* was considered with regard to the variables *spread5y*, *Currency2*, *sect2* and *recovery* and with the group named *industry* of the variable named *sect2* and the group named *Other* of the variable named *Currency2* taken as references. The following command was therefore used in STATA:

xtlogit resp spread5y i.sect2 recovery ib2.Currency2,re

These results were thus obtained (Table 10):

Table 10: *Random Effects Logit model* applied to the subset in question (*spread5y*, *sect2*, *recovery* and *Currency2* are considered).

Iteration 14: log likelihood = -3267.3381				
Random-effects logistic regression		Number of obs	=	49,198
Group variable: newid		Number of groups	=	2,239
Random effects u_i ~ Gaussian		Obs per group:		
		Min	=	1
		Avg	=	22.0
		Max	=	69
Integration method: mvaghermite		Integration pts.	=	12
		Wald chi2(6)	=	714.83
Log likelihood = -3267.3381		Prob > chi2	=	0.0000

resp	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
spread5y	-9.25254	1.537725	-6.02	0.000	-12.26643	-6.238653
sect2						
service	0.5821071	0.2978063	1.95	0.051	-0.0015825	1.165797
utilities	-0.4503189	0.2709301	-1.66	0.096	-0.9813322	0.0806943
recovery	-4.57001	1.239066	-3.69	0.000	-6.998536	-2.141485
Currency2						
CAD	-1.401503	0.8443613	-1.66	0.097	-3.056421	0.2534149
USD	-6.694774	0.2700507	-24.79	0.000	-7.224063	-6.165484
_cons	-4.406706	0.5727235	-7.69	0.000	-5.529223	-3.284188
/lnsig2u	6.922974	0.1311539			6.665917	7.180031
sigma_u	31.86433	2.089565			28.02113	36.23464
rho	0.9967703	0.0004222			0.9958275	0.9975006
LR test of rho=0: chibar2(01) = 5.4e+04					Prob >= chibar2 = 0.000	

These results display the fact all the coefficients of all the variables considered are negative, except for the coefficient of *service*, which is positive.

Therefore, assuming all the other variables are kept constant, if the value of s*pread5y* of the company considered (which is a reference entity of *CDS* contracts) is higher than the value of *spread5y* of another company (which is also a reference entity of *CDS* contracts), then the probability that the company considered is a constituent of the *S&P500 Index* is lower than the probability that the other company is a constituent of the *S&P500 Index*. The same idea applies also to the variable named *recovery*.

Moreover, assuming all the other variables are kept constant, the probability that a company (that is a reference entity of *CDS* contracts) with either *CAD* or *USD* as currency is a constituent of the *S&P500 Index* is lower than the probability that a company (that is also a reference entity of *CDS* contracts) with *Other* as currency is a constituent of the *S&P500 Index*.

In the same way, it is also possible to state that, assuming all the other variables are kept constant, the probability that a company (that is a reference entity of *CDS* contracts) with *utilities* as sector is also a constituent of the *S&P500 Index* is lower than the probability that a company (that is also a reference entity of *CDS* contracts) with *industry* as sector is a constituent of the *S&P500 Index*.

Lastly, it is also possible to state that, assuming all the other variables are kept constant, the probability that a company (which is a reference entity of *CDS* contracts) with *service* as sector is also a constituent of the *S&P500 Index* is higher than the probability that a company (that is also a reference entity of *CDS* contracts) with *industry* as sector is a constituent of the *S&P500 Index*.

Furthermore, also the *Logit model* and *Fixed Effects Logit model* were considered with regard to the same subset and the same variables. Therefore, the following commands were used in STATA:

Logit model

logit resp recovery spread5y i.sect2 ib2.Currency2

Fixed Effects Logit model

xtlogit resp spread5y recovery i.sect2 ib2.Currency2, fe

These results were thus obtained (Table 11 and Table 12):

Table 11: *Logit model* applied to the subset in question (*spread5y*, *sect2*, *recovery* and *Currency2* are considered).

convergence not achieved			
Logistic regression	Number of obs	=	49,198
	LR chi2(6)	=	5598.82
	Prob > chi2	=	0.0000
Log likelihood = -30261.654	Pseudo R2	=	0.0847

resp	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
recovery	-4.990334	0.2061586	-24.21	0.000	-5.394397	-4.586271
spread5y	-31.57281	0.6756234	-46.73	0.000	-32.89701	-30.24862
sect2						
service	0.6160945	0.0255414	24.12	0.000	0.5660343	0.6661547
utilities	0.411129	0.0255872	16.07	0.000	0.360979	0.4612789
Currency2						
CAD	-0.5296582	0.057229	-9.26	0.000	-0.6418251	-0.4174914
USD	0.7336633	0.0202229	36.28	0.000	0.6940272	0.7732994
_cons	1.390929	0.0882705	15.76	0.000	1.217922	1.563936
Note: 102 failures and 0 successes completely determined. convergence not achieved						

Table 12: *Fixed Effects Logit model* applied to the subset in question (*spread5y*, *sect2*, *recovery* and *Currency2* are considered).

convergence not achieved			
Conditional fixed-effects logistic regression	Number of obs	=	3,321
Group variable: newid	Number of groups	=	89
	Obs per group:		
	Min	=	2
	Avg	=	37.3
	Max	=	69
	LR chi2(5)	=	364.96
Log likelihood = -1245.3844	Prob > chi2	=	0.0000

resp	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
spread5y	-6.555104	1.358333	-4.83	0.000	-9.217387	-3.892821
recovery	-3.719972	1.010299	-3.68	0.000	-5.700122	-1.739822
sect2						
service	9.768057	3883.971	0.00	0.998	-7602.676	7622.212
utilities	-10.40637	.	.	.	.	.
Currency2						
CAD	5.53845	18620.55	0.00	1.000	-36490.07	36501.14
USD	-16.92792	127.1073	-0.13	0.894	-266.0537	232.1979

Both the two models do not provide significant results and, in fact, the results of the *Logit model* and of the *Fixed Effects Logit model* display *"convergence not achieved"*. This is caused by the *Problem of monotone likelihood* that was explained in the previous paragraphs.

Furthermore, a new subset of the whole dataset was created. This new subset contains all the companies that have the value of *CDS* equal to 1 and therefore contains the cases [*CDS*=1; *SP*=0] and [*CDS*=1; *SP*=1].

In addition to this, the values of *spread5y* that could not be identified in the subset in question were replaced with the average value of *spread5y* of the specific company considered. If a particular company did not have any value of *spread5y*, the company in question was removed from the subset.

Moreover, also in this case the variable regarding the sectors was named *sect2* and the various sectors were grouped into three main groups, named *service, utilities* and *industry* and also the variable regarding the currencies was named *Currency2* and the various currencies were grouped into three groups, named *Other*, *CAD* and *USD*.

Therefore this new subset has these characteristics:

- Only the cases [*CDS*=1; *SP*=0] and [*CDS*=1; *SP*=1] are contained in the subset in question;
- The values of *spread5y* that could not be identified were replaced with the average value of *spread5y* of the specific company considered and if a particular company did not have any value of *spread5y* the company in question was removed from the subset;
- Also in this case the variable regarding the sector was named *sect2* and the various sectors that could be identified were grouped into three main groups, named *service*, *utilities* and *industry* and the variable regarding the currencies was named *Currency2* and the various currencies were grouped into three main groups, named *Other*, *CAD* and *USD*.

The *Random Effects Logit model* was then considered with regard to the variables *spread5y*, *sect2*, *recovery* and *Currency2* and the group named *Other* of the variable named *Currency2* and the group named *industry* of the variable named *sect2* were taken as references.

The following command was thus used in STATA:

xtlogit resp spread5y i.sect2 recovery ib2.Currency2,re

These results were obtained (Table 13):

Table 13: *Random Effects Logit model* applied to the subset in question (*spread5y*, *sect2*, *recovery* and *Currency2* are considered).

Iteration 16: log likelihood = -4246.1704			
Random-effects logistic regression	Number of obs	=	68,694
Group variable: newid	Number of groups	=	2,247
Random effects u_i ~ Gaussian	Obs per group:		
	Min	=	1
	Avg	=	30.6
	Max	=	70
Integration method: mvaghermite	Integration pts.	=	12
	Wald chi2(6)	=	648.01
Log likelihood = -4246.1704	Prob > chi2	=	0.0000

resp	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
spread5y	-5.290359	1.105797	-4.78	0.000	-7.457681	-3.123038
sect2						
service	1.833369	0.2432565	7.54	0.000	1.356595	2.310143
utilities	-0.0901133	0.207149	-0.44	0.664	-0.4961178	0.3158913
recovery	-0.9811413	0.5367541	-1.83	0.068	-2.03316	0.0708773
Currency2						
CAD	-1.387245	0.7486255	-1.85	0.064	-2.854524	0.0800342
USD	-4.300359	0.2033647	-21.15	0.000	-4.698946	-3.901771
_cons	1.852963	0.3223601	5.75	0.000	1.221149	2.484777
/lnsig2u	7.133578	0.1003782			6.93684	7.330316
sigma_u	35.40273	1.776832			32.08601	39.0623
rho	0.997382	0.0002621			0.9968146	0.9978486
LR test of rho=0: chibar2(01) = 7.3e+04				Prob >= chibar2 = 0.000		

From these results it is possible to observe that all the coefficients are negative, except for the coefficient of *service*, which is positive; therefore, the signs of the coefficients are the same as the signs of the coefficients of the analysis that was previously explained. For this reason, the conclusions that could be drawn previously are valid also in this case.

Overall several analyses were carried out. These analyses confirmed that the initial hypothesis holds true and therefore confirmed the fact that the probability that the companies that are constituents of the *S&P500 Index* are also reference entities of *CDS* contracts is higher than the probability that the companies that are not constituents of the *S&P500 Index* are also reference entities of the *CDS* contracts. In addition to this, it is also possible to state that the probability that a company that is reference entity of *CDS* contracts is also a company that is constituent of the *S&P500 Index* is higher than the probability that a company that is not a reference entity of *CDS* contracts is a company that is constituent of the *S&P500 Index* (Table 14).

Table 14: Summary of the *Logit model*, *Random Effects Logit model* and *Fixed Effects Logit model* applied to the whole dataset (with *CDS* and *SP).*

Model	**CDS**	**Coef.**	**Sign of Coef.**
Logit model	SP	3.521776	Positive
	_cons	-2.794143	Negative
Random Effects Logit model	SP	4.015183	Positive
	_cons	-16.60779	Negative
Fixed Effects Logit model	SP	3.61043	Positive

Moreover, the other parts of the analysis provided interesting results and insights (Table 15).

Table 15: Summary of the *Random Effects Logit model* (*spread5y*, *sect2*, *recovery* and *Currency2* are considered).

resp	**Coef.**	**Sign of Coef.**
spread5y	-5.290359	Negative
sect2		
service	1.833369	Positive
utilities	-0.0901133	Negative
recovery	-0.9811413	Negative
Currency2		
CAD	-1.387245	Negative
USD	-4.300359	Negative
_cons	1.852963	Positive

To this regard, in fact, from the results it is possible to conclude that, assuming all the other variables are kept constant, if the value of s*pread5y* of the company considered (that is a reference entity of *CDS* contracts) is higher than the value of s*pread5y* of another company (that is also a reference entity of *CDS* contracts), then the probability that the company considered is a constituent of the *S&P500 Index* is lower than the probability that the other company is a constituent of the *S&P500 Index*. In addition to this, the same idea applies also to the variable named *recovery*, since also the sign of the coefficient of *recovery* is negative. In a similar way, it is possible to state that, assuming all the other variables are kept constant, the probability that a company (which is a reference entity of *CDS* contracts) with either *CAD* or *USD* as currency is a constituent of the *S&P500 Index* is lower than the probability that a company (that is also a reference entity of *CDS* contracts) with *Other* as currency is a constituent of the *S&P500 Index*.
Moreover, if the sector of the company is taken into consideration, the probability that, assuming all the other variables are kept constant, a company (which is a reference entity of *CDS* contracts) with *utilities* as sector is also a constituent of the *S&P500 Index* is

lower than the probability that a company (that is also a reference entity of *CDS* contracts) with *industry* as sector is a constituent of the *S&P500 Index*.

Finally, it is also possible to state that, assuming all the other variables are kept constant, the probability that a company (that is a reference entity of *CDS* contracts) with *service* as sector is also a constituent of the *S&P500 Index* is higher than the probability that a company (that is also a reference entity of *CDS* contracts) with *industry* as sector is a constituent of the *S&P500 Index*.

5. Conclusion

CDS and *Standard & Poor's* have played the role of protagonists in the global economic scenario over recent years. The objective of this work was to determine the effect of *S&P500* additions and deletions on *CDS* and its characteristics.

In particular, the results of the analysis confirmed that the initial hypothesis holds true and therefore the probability that the companies that are constituents of the *S&P500 Index* are also reference entities of *CDS* contracts is higher than the probability that the companies that are not constituents of the *S&P500 Index* are also reference entities of *CDS* contracts. Moreover, it is also possible to state that the probability that the companies that are reference entities of *CDS* contracts are also companies that are constituents of the *S&P500 Index* is higher than the probability that the companies that are not reference entities of *CDS* contracts are companies that are constituents of the *S&P500 Index*. From an economic perspective, large companies obviously attract the attention of many market participants, including those interested in *CDS* contracts. To this regard, in fact, in the current economic system the companies that are reference entities of *CDS* contracts are usually large companies and they also often belong to the *S&P500 Index*.

In addition to this, from the other parts of the analysis it was possible to make some conclusions concerning the other variables considered. To this regard, if two companies (that are reference entities of *CDS* contracts), named *Alpha* and *Beta*, are considered, it is possible to state that if the value of *spread5y* of *Alpha* is higher than the value of *spread5y* of *Beta*, then the probability that *Alpha* is a constituent of the *S&P500 Index* is lower than the probability that company *Beta* is a constituent of the *S&P500 Index*. The same conclusion can be drawn with regard to the variable named *recovery*. To this regard, in fact, the variables *spread5y* and *recovery* are integral parts of *CDS* and, therefore, they both influence the *CDS* contracts and consequently also the *CDS* market.

In a similar way, it is possible to state that, assuming all the other variables are kept constant, the probability that a company (which is a reference entity of *CDS* contracts) with either *CAD* or *USD* as currency is a constituent of the *S&P500 Index* is lower than the probability that a company (that is also a reference entity of *CDS* contracts) with *Other* as currency is a constituent of the *S&P500 Index*. Also in this case, the currency of which *CDS* is priced on has an impact on the *CDS* contracts and the *CDS* market.

Moreover, if the sector of the company, which is a reference entity of *CDS* contracts, is taken into consideration, the probability that, assuming all the other variables are kept constant, a company with *utilities* as sector is also a constituent of the *S&P500 Index* is lower than the probability that a company (that is also a reference entity of *CDS* contracts) with *industry* as sector is a constituent of the *S&P500 Index*.

Finally, it is also possible to state that, assuming all the other variables are kept constant, the probability that a company (that is a reference entity of *CDS* contracts) with *service* as sector is also a constituent of the *S&P500 Index* is higher than the probability that a company (that is also a reference entity of *CDS* contracts) with *industry* as sector is a constituent of the *S&P500 Index*. Therefore, also the sector to which the company belongs has an impact on both the *CDS* contracts and the *CDS* market.

Lastly, all the various analyses that were taken into consideration in this work confirm that the initial hypothesis holds true and also offer interesting results and insights regarding the main aspects of *CDS* contracts and the *CDS* market.

As it was stated earlier, the *CDS* and the *CDS* market have grown rapidly over recent decades and they will for sure continue to change in the coming years.

In conclusion, future work could involve studying these future changes and also understanding the impact of these changes on the global economic scenario.

6. References

Acharya, V.V. and Johnson, T.C. (2007), "Insider trading in credit derivatives", *Journal of Financial Economics,* 84(1), pp. 110–141.

Afonso, A., Gomes, P. and Rother, P. (2011), "Short and long-run determinants of sovereign debt credit ratings", *International Journal of Finance and Economics,* 16(1), pp. 1-15.

Alexander, C. and Kaeck, A. (2008), "Regime dependent determinants of credit default swap spreads", *Journal of Banking and Finance,* 32(6), pp. 1008-1021.

Altman, E.I. and Rijken, H.A. (2006), "A point-in-time perspective on through-the-cycle ratings", *Financial Analysts Journal*, 62(1), pp. 54-70.

Ammer, J. and Clinton, N. (2004), "Good news is no news? The impact of credit rating changes on the pricing of asset-backed securities", *FRB International Finance Discussion Paper.* No. 809.

Anderson, R.W. (2010), "Credit default swaps: what are the social benefits and costs?", *Banque de France Financial Stability Review - Derivatives, Financial innovation and stability*, 14, pp. 1–14.

Angelini, E. (2012), "Credit default swaps (CDS) and their role in the credit risk market", *International Journal of Academic Research in Business and Social Sciences,* 2(1), pp. 584-593.

Ashcraft, A.B. and Santos, J.A.C. (2009), "Has the CDS market lowered the cost of corporate debt?", *Journal of Monetary Economics,* 56(4), pp. 514-523.

Augustin, P., Subrahmanyam, M.G., Tang, D.Y. and Wang, S.Q. (2014), "Credit default swaps – A survey", *Foundations and Trends in Finance,* 9(2), pp. 1–196.

Blanco, R., Brennan, S. and Marsh, I.W. (2005), “An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps”, *The Journal of Finance,* 60(5), pp. 2255–2281.

Bolton, P. and Oehmke, M. (2013), “Strategic conduct in credit derivative markets”, *International Journal of Industrial Organization,* 31(5), pp. 652-658.

Boss, M. and Scheicher, M. (2002), “The determinants of credit spread changes in the Euro Area”, *BIS Papers chapters in: Bank for International Settlements (ed.), Market functioning and central bank policy*, 12, pp. 181-199.

Cantor, R. and Mann, C. (2007), “Analyzing the tradeoff between ratings accuracy and stability”, *The Journal of Fixed Income,* 16(4), pp. 60-68.

Cantor, R. and Packer, F. (1996), “Determinants and impact of sovereign credit ratings”, *Economic Policy Review,* 2, pp. 37-53.

Castellano, R. and Giacometti, R. (2012), “Credit default swaps: implied ratings *versus* official ones”, *4OR - A Quarterly Journal of Operations*, 10, pp. 163-180.

Cossin, D. and Pirotte, H. (2001), *Advanced credit risk analysis: financial approaches and mathematical models to assess, price and manage credit risk*, J. Wiley, Chichester.

Das, S.R. and Hanouna, P. (2006), “Survey of the literature: credit default swap spreads”, *Journal of Investment Management,* 4(3), pp. 93-105.

Dichev, L.D. and Piotroski, J.D. (2001), “The long-run stock returns following bond ratings changes”, *The Journal of Finance*, 56(1), pp. 173-203.

Freeman, M.C., Cox, P.R. and Wright, B. (2006), "Credit risk management: the use of credit derivatives by non-financial corporations", *Managerial Finance,* 32(9), pp. 761-773.

Fu, X., Li, M.C. and Molyneux, P. (2020), “Credit default swap spreads: market conditions, firm performance, and the impact of the 2007–2009 financial crisis”, *Empirical Economics*, pp. 1-23.

Hamilton, D.T. and Cantor, R. (2007), “Adjusting corporate default rates for rating withdrawals”, *Journal of Credit Risk,* 3(2), pp. 1-28.

Hill, P., Brooks, R. and Faff, R. (2010), “Variations in sovereign credit quality assessments across rating agencies”, *Journal of Banking & Finance,* 34(6), pp. 1327-1343.

Hite, G. and Warga, A. (1997), “The effect of bond-rating changes on bond price performance”, *Financial Analysts Journal*, 53(3), pp. 35-51.

Hull, J.C., Predescu, M. and White, A. (2004), “The relationship between credit default swap spreads, bond yields, and credit rating announcements”, *Journal of Banking and Finance*, 28, pp. 2789–2811.

Ismailescu, J. and Kazemi, H. (2010), “The reaction of emerging market credit default swap spreads to sovereign credit rating changes”, *Journal of Banking & Finance*, 34 (12), pp. 2861–2873.

Jacobs, M. Jr., Karagozoglu, A.K. and Layish, D.N. (2016), “Credit risk signals in CDS market *vs* agency ratings”, *The Journal of Risk Finance,* 17(2), pp. 194-217.

Jarrow, R.A. (2011), “The economics of credit default swaps”, *Annual Review of Financial Economics*, 3, pp. 235-257.

Kiff, J., Kisser, M. and Schumacher, L. (2013), “Rating through-the cycle: what does the concept imply for rating stability and accuracy?”, *IMS Working Paper*, 13(64), pp. 1-28.

Longstaff, F.A., Mithal, S. and Neis, E. (2005), "Corporate yield spreads: default risk or liquidity? New evidence from the credit default swap market", *The Journal of Finance*, 60(5), pp. 2213–2253.

Martire, R. (2007), *Credit derivatives. Fondamenti teorici e modalità di pricing*, Editrice UNI Service, Trento.

Niedziółka, P. (2019), "The Visegrád Group Countries' Ratings – Do the credit rating agencies make the decisions independently or piggyback themselves?", *Journal of Economics and Business*, 2(2), pp. 306-333.

Pinches, G.E. and Singleton, J.C. (1978), "The adjustment of stock prices to bond rating changes", *Journal of Finance*, *American Finance Association*, 33(1), pp. 29-44.

Reyngold, A., Kocagil, A.E. and Gupton, G.M. (2007), "Fitch CDS implied ratings (CDS-IR) model", *Quantitave Financial Research Special Reports*, pp. 1-14.

Saretto, A. and Tookes, H.E. (2013), "Corporate leverage, debt maturity and credit supply: the role of credit default swaps", *Review of Financial Studies*, 26(5), pp. 1190-1247.

Schoreder, S.K. (2015), *Public credit rating agencies. Increasing capital investment and lending stability in volatile markets*, Palgrave Macmillan, New York.

Steiner, M. and Heinke, V.G. (2001), "Event study concerning international bond price effects of credit rating actions", *International Journal of Finance and Economics*, 6(2), pp. 139-157.

Stulz, R.M. (2010), "Credit default swaps and the credit crisis", *Journal of Economic Perspectives*, 24(1), pp. 73-92.

Subrahmanyam, M.G., Tang, D.Y. and Wang, S.Q. (2014), "Does the tail wag the dog?: The effect of credit default swaps on credit risk", *The Review of Financial Studies*, 27(10), pp. 2927–2960.

Vassalou, M. and Xing, Y. (2004), "Default risk in equity returns", *The Journal of Finance,* 59(2), pp. 831-868.

Wagner, N. (2008)*, Credit risk: models, derivatives, and management,* Chapman & Hall/CRC Taylor & Francis Group, New York.

Weinstein, M.I. (1977), "The effect of a rating change announcement on bond price", *Journal of Financial Economics*, 5(3), pp. 329-350.

White, L.J. (2018), *The credit rating agencies and their role in the financial system*, eds. Oxford University Press, New York.

Zhang, B.Y., Zhou, H. and Zhu, H. (2009), "Explaining credit default swap spreads with the equity volatility and jump risks of individual firms", *The Review of Financial Studies*, 22(12), pp. 5099–5131.

Websites

Wharton WRDS:
https://wrds-www.wharton.upenn.edu/

Wharton WRDS – Compustat:
https://wrds-web.wharton.upenn.edu/wrds/ds/compustat/index.cfm

Wharton WRDS – Markit:
https://wrds-web.wharton.upenn.edu/wrds/ds/markit/index.cfm

7. Appendix

Since the entire code created in STATA is very long, just the most significant parts of the code in question are included in this Appendix.

```
* The file paths and folder paths that can be found in my PC are taken into
* consideration in this code



**********************************************************
*****************Creation of the dataset******************
**********************************************************



******************
****march 2001****
******************
use "C:\Users\Federico\Desktop\2001march_mkmerge.dta", clear

drop date tier gvkey gvkeyx from thru tic Year qdate _merge Quarter

sort CompanyName
quietly by CompanyName: gen dup1 = cond(_N==1,0,_n)
list CompanyName if dup1!=0
drop if dup1>1
drop dup1

gen Year=2001
gen Quarter = 1
save "C:\Users\Federico\Desktop\2001march_ab.dta", replace
use "C:\Users\Federico\Desktop\data2001_1.dta", clear

drop datadate fiscalyear fiscalquarter industryformat levelofconsolidationcompanyinter
populationsource dataformat calendardatayearandquarter fiscaldatayearandquarter
rename globalcompanykey glo_comp_key
rename cusip co_cusip
rename companyname CompanyName
rename isocurrency ccy
rename activeinactivestatusmarker A_I_status
gen Year =2001
gen Quarter=1
gen region="N.Amer"
gen country ="United States"
```

```
replace country = "Canada" if ccy=="CAD"
sort CompanyName
quietly by CompanyName: gen dup1 = cond(_N==1,0,_n)

drop if dup1>1
drop dup1
merge m:1  CompanyName using "C:\Users\Federico\Desktop\2001march_ab.dta"

save "C:\Users\Federico\Desktop\2001march_merge.dta", replace
sort co_cusip
drop if (co_cusip=="" | co_cusip=="#DIV/0!")



quietly by co_cusip: gen dup1 = cond(_N==1,0,_n)

replace  glo_comp_key=glo_comp_key[_n-1] if glo_comp_key==.
replace  tickersymbol=tickersymbol[_n-1] if tickersymbol==""
replace  ccy=ccy[_n-1] if ccy==""
replace  region=region[_n-1] if region==""
replace  A_I_status=A_I_status[_n-1] if A_I_status==""
replace  country=country[_n-1] if country==""


drop if (dup1>0 & CDS==. & SP==.)

save "C:\Users\Federico\Desktop\2001march_merge1.dta", replace
use "C:\Users\Federico\Desktop\2001march_merge.dta", clear
sort co_cusip
drop if (co_cusip!="" & co_cusip!="#DIV/0!")
sort tickersymbol
quietly by tickersymbol: gen dup1 = cond(_N==1,0,_n)
save "C:\Users\Federico\Desktop\2001march_merge2.dta", replace
use "C:\Users\Federico\Desktop\2001march_merge1.dta", clear
append using "C:\Users\Federico\Desktop\2001march_merge2.dta"

replace CDS=0 if CDS==.
replace SP=0 if SP==.

save "C:\Users\Federico\Desktop\2001march_op.dta", replace
```

```
use "C:\Users\Federico\Desktop\2018dec_op.dta", clear

save dec2018.dta,replace
use "C:\Users\Federico\Desktop\2018sept_op.dta", clear

append using dec2018.dta
save dec2018.dta,replace
use "C:\Users\Federico\Desktop\2018june_op.dta", clear

append using dec2018.dta
save dec2018.dta,replace
use "C:\Users\Federico\Desktop\2018march_op.dta", clear

append using dec2018.dta
save dec2018.dta,replace

use "C:\Users\Federico\Desktop\2017dec_op.dta", clear

append using dec2018.dta
save dec2018.dta,replace

use "C:\Users\Federico\Desktop\2017sept_op.dta", clear

append using dec2018.dta,force
save dec2018.dta,replace

use "C:\Users\Federico\Desktop\2017june_op.dta", clear

append using dec2018.dta,force
save dec2018.dta,replace

use "C:\Users\Federico\Desktop\2017march_op.dta", clear

```

```
use "C:\Users\Federico\Desktop\2001sept_op.dta", clear

append using dec2018.dta
save dec2018.dta,replace

use "C:\Users\Federico\Desktop\2001june_op.dta", clear

append using dec2018.dta
save dec2018.dta,replace

use "C:\Users\Federico\Desktop\2001march_op.dta", clear

append using dec2018.dta
save dec2018.dta,replace

save "C:\Users\Federico\Desktop\final_dataset.dta", replace


use"C:\Users\Federico\Desktop\final_dataset.dta", clear

sort co_cusip
replace co_cusip=redcode if co_cusip==""
replace co_cusip=string(glo_comp_key) if co_cusip==""
replace co_cusip=string(glo_comp_key) if co_cusip=="#DIV/0!"
list if co_cusip==""


save "C:\Users\Federico\Desktop\final_dataset.dta", replace

```

```
* The variable currency is grouped into three groups


gen Currency ="Other"
replace Currency=ccy if ccy=="USD"
replace Currency=ccy if ccy=="CAD"
encode Currency, generate(Currency2)

xtset newid Quarter

```

```
* Random Effects Logit model

xtlogit CDS SP ib2.Currency2, re

*Iteration 12:  log likelihood = -25623.738

*Random-effects logistic regression              Number of obs     =    853,116
*Group variable: newid                           Number of groups  =     27,141
*
*Random effects u_i ~ Gaussian                   Obs per group:
*                                                              min =          1
*                                                              avg =       31.4
*                                                              max =         72
*
*Integration method: mvaghermite                 Integration pts.  =         12
*
*                                                Wald chi2(3)      =    3075.12
*Log likelihood  = -25623.738                    Prob > chi2       =     0.0000
*
*------------------------------------------------------------------------------
*         CDS |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
*-------------+----------------------------------------------------------------
*          SP |   2.885575   .0660816    43.67   0.000     2.756057    3.015092
*             |
*   Currency2 |
*         CAD |  -40.80244   1.250108   -32.64   0.000    -43.25261   -38.35228
*         USD |  -41.63995    1.24753   -33.38   0.000    -44.08506   -39.19484
*             |
*       _cons |   26.91321   1.245264    21.61   0.000     24.47254    29.35389
*-------------+----------------------------------------------------------------
*    /lnsig2u |    4.44697   .0161625                      4.415292    4.478648
*-------------+----------------------------------------------------------------
*     sigma_u |   9.239475   .0746666                      9.094284    9.386984
*         rho |   .9628925   .0005775                      .9617439    .9640079
*------------------------------------------------------------------------------
*LR test of rho=0: chibar2(01) = 1.9e+05               Prob >= chibar2 = 0.000

estimates store rand_tot1
```

```
* Random Effect Logit model

xtlogit resp spread5y  i.sect2 recovery ib2.Currency2,re

*Iteration 16:  log likelihood = -4246.1704
*
*Random-effects logistic regression              Number of obs     =     68,694
*Group variable: newid                            Number of groups  =      2,247*
*Random effects u_i ~ Gaussian                    Obs per group:
*                                                              min =          1
*                                                              avg =       30.6
*                                                              max =         70
*
*Integration method: mvaghermite                  Integration pts.  =         12
*
*                                                 Wald chi2(6)      =     648.01
*Log likelihood  = -4246.1704                     Prob > chi2       =     0.0000
*
*------------------------------------------------------------------------------
*        resp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
*-------------+----------------------------------------------------------------
*    spread5y |  -5.290359   1.105797    -4.78   0.000    -7.457681   -3.123038
*             |
*       sect2 |
*     service |   1.833369   .2432565     7.54   0.000     1.356595    2.310143
*   utilities |  -.0901133    .207149    -0.44   0.664    -.4961178    .3158913
*             |
*    recovery |  -.9811413   .5367541    -1.83   0.068     -2.03316    .0708773
*             |
*   Currency2 |
*        CAD  |  -1.387245   .7486255    -1.85   0.064    -2.854524    .0800342
*        USD  |  -4.300359   .2033647   -21.15   0.000    -4.698946   -3.901771
*             |
*       _cons |   1.852963   .3223601     5.75   0.000     1.221149    2.484777
*-------------+----------------------------------------------------------------
*    /lnsig2u |   7.133578   .1003782                       6.93684    7.330316
*-------------+----------------------------------------------------------------
*     sigma_u |   35.40273   1.776832                      32.08601     39.0623
*         rho |    .997382   .0002621                      .9968146    .9978486
*------------------------------------------------------------------------------
*LR test of rho=0: chibar2(01) = 7.3e+04                Prob >= chibar2 = 0.000
```