Aug 2017

Anshuman Sinha, Benedict Tan, Meng Huat Pee

WFVA CVA Proxy

[Document subtitle]

Table of Contents

[Introduction 2](#_Toc491875722)

[Statistical and Descriptive Results 3](#_Toc491875723)

[The Cross-Section Methodology 10](#_Toc491875724)

[Description 10](#_Toc491875725)

[Types of Datasets 12](#_Toc491875726)

[Dataset 1 12](#_Toc491875727)

[Dataset 2 12](#_Toc491875728)

[Dataset 3 12](#_Toc491875729)

[Methods 13](#_Toc491875730)

[Method 1 - Ordinary Least Squares 13](#_Toc491875731)

[Method 2 - Least Squares with regularisation 13](#_Toc491875732)

[Method 3 - Least Squares with -regularisation 14](#_Toc491875733)

[Results 15](#_Toc491875734)

[Robustness Check 21](#_Toc491875735)

[Approach 1: Regression 23](#_Toc491875736)

[Approach 1(A) 23](#_Toc491875737)

[Approach 1(B) 23](#_Toc491875738)

[Approach 1(C) 24](#_Toc491875739)

[Results 25](#_Toc491875740)

[Approach 2: Nelson Siegel Fitting 27](#_Toc491875741)

[Motivation 27](#_Toc491875742)

[Nelson Siegel Curve Fitting 27](#_Toc491875743)

[Methodology 27](#_Toc491875744)

[Results 28](#_Toc491875745)

[Approach 3: Classification 31](#_Toc491875746)

[Motivation 31](#_Toc491875747)

[Model Comparison 32](#_Toc491875748)

[-Fold Cross Validation 32](#_Toc491875749)

[Loss function 32](#_Toc491875750)

[Cross Validation Error 33](#_Toc491875751)

[Results 33](#_Toc491875752)

[Conclusion 40](#_Toc491875753)

[References 41](#_Toc491875754)

Introduction

CDS spreads have become the market practice to calculate default intensities. These default intensities are then used in advanced modelling approaches to get the valuation of several exotic products.

For a large number of entities, CDS quotes cannot be retrieved from the market. When the credit spread is not liquid and hence, not observable in the market, institutions are required to proxy the credit spread having regard a few common factors of the counterparty.

The European Banking Authority (EBA) proposed an intersection model to calculate a proxy CDS or Bond spread. The method proposed is a simple linear model, where the proxy spread for an illiquid entity belonging to a particular Sector, Region and Implied Rating are defined as the mean of the available liquid CDS spreads of entities belonging to the same Sector, Region and Implied Rating. This intersection method could face potential drawbacks in practice when there are certain entities that cannot be mapped to any set. Furthermore, the model is not stable over time.

In addition, Chourdakis(2013) proposed a cross-sectional regression model, which has an advantage of providing more robust and stable results without the loss of transparency. The method is a cross sectional regression approach, aims to solve the problem by an idea of extrapolation. This method meant that, given a Sector, Region and Implied Rating one can estimate what the *average* CDS spread should be.

Lastly, we have also thought an approach by means of an unsupervised machine learning approach. It takes the form of classification and is further explained in a later chapter below.

In our document, we explore the regression approach with a few additions, and along with a few ideas that we have came up with. We have found that for some approaches the fit is good with low out-of-sample error and for others, the fit is not good with higher out-of-sample errors.

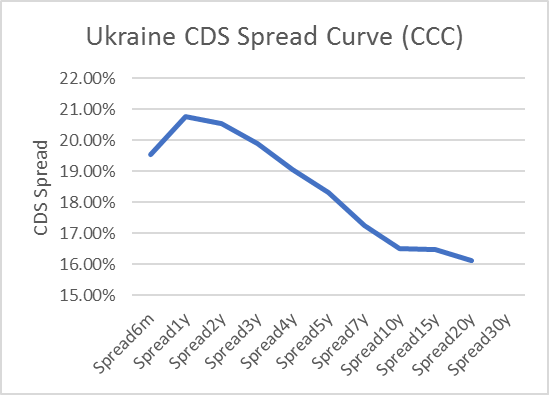
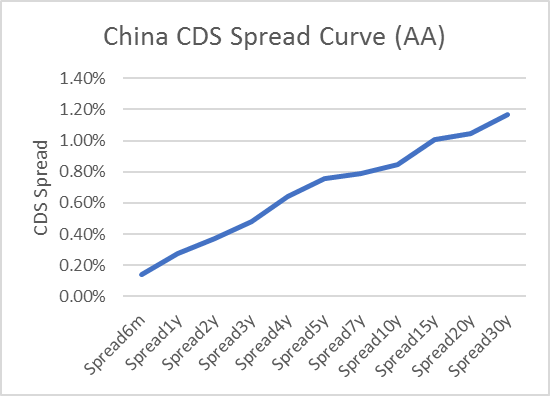
Our target is to get the CDS spreads for illiquid names whereby the data is sparse. What we have is data that is plentiful in the “region” whereby the entities are considered of better credit rating.

# Statistical and Descriptive Results

Markit produces a daily file of liquid CDS spreads and recovery rates, together with a number of contributors. The daily file contains details of the CDS spreads, for example, sectors, regions, Average Rating and Implied Ratings. This gives us a high quality and independent data source for the calibration of the proxy spread factors.

Below we display some characteristics in the dataset that we have obtained. These data characteristics are model independent.

* The average spread level increases as the crediting rating of obligor decreases
* The variance of the spread increases as obligor’s credit rating decreases. (See Figure 1)
* There are mostly rated CDS contracts in the daily file. As the credit rating decreases, the number of CDS contracts decreases.
* CDS obligors of good credit Rating usually have an increasing curve, while CDS obligors of poor credit rating have humped shape or decreasing curve.
* Most of the data comes from the region of N. America, Europe and Asia.
* Most of the data comes from the sector of Others, which does not include Government and Financials.



Following the idea of Chourdakis and the factors proposed by EBA, we have identified the 3 explanatory factors for the CDS Proxy spread to be of Sector, Region and ImpliedRating. These factors are readily available in the dataset provided by Markit.

In addition, since we are dealing with only extracting the CDS Spreads for sovereigns only, we have grouped the rest of the sectors in the “Sector” category to be Others.

Below we also list some descriptive statistics:

|  |  |  |
| --- | --- | --- |
| Sector | Region | ImpliedRating |
| Financials  Government  Others | **Asia**  **N.America**  **Europe**  **Oceania**  **OffShore**  **E.Europe**  **Lat.America**  **MiddleEast**  **Caribbean**  **Africa**  **India**  **Supra** | **AA**  **A**  **BBB**  **BB**  **B**  **CCC** |
| Total Count: 2621 | | |

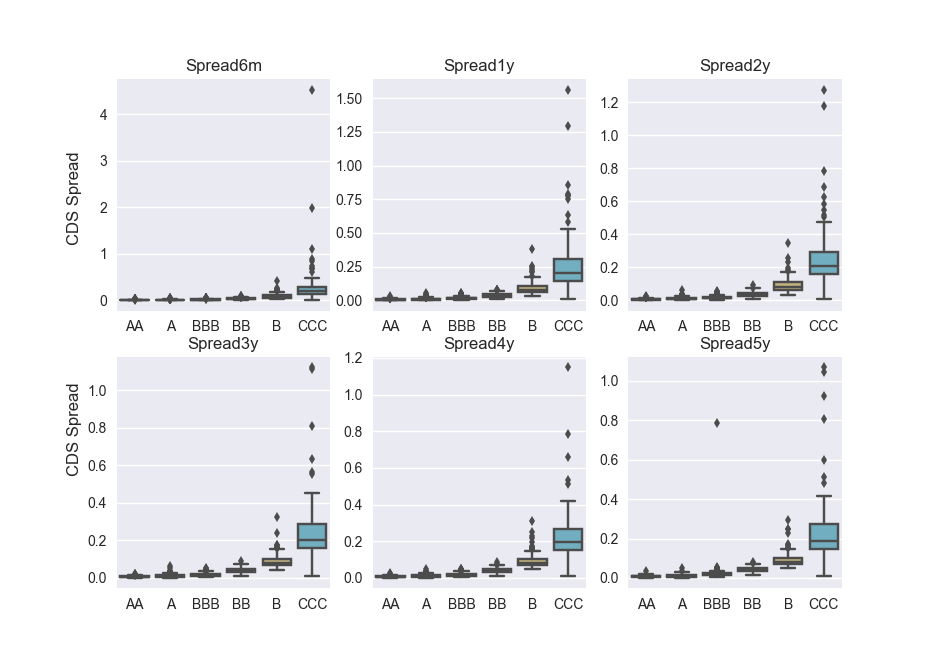
Table

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sector | Region | ImpliedRating | Num Obilgors | Sector | Region | ImpliedRating | Num Obilgors |
| Financials | **Africa** | **A** | 1 | **Government** | **Africa** | **AA** | 1 |
| **Asia** | **AA** | 67 | **A** | 3 |
| **A** | 20 | **BBB** | 3 |
| **BBB** | 4 | **BB** | 1 |
| **BB** | 6 | **B** | 1 |
| **B** | 7 | **Asia** | **AA** | 31 |
| **Caribbean** | **BBB** | 2 | **A** | 16 |
| **E.Eur** | **BBB** | 1 | **BBB** | 1 |
| **BB** | 2 | **BB** | 4 |
| **B** | 3 | **B** | 1 |
| **CCC** | 13 | **Caribbean** | **BB** | 3 |
| **Europe** | **AA** | 110 | **B** | 2 |
| **A** | 41 | **E.Eur** | **AA** | 5 |
| **BBB** | 10 | **A** | 1 |
| **BB** | 13 | **BBB** | 2 |
| **B** | 15 | **BB** | 13 |
| **CCC** | 2 | **B** | 3 |
| **India** | **AA** | 2 | **CCC** | 3 |
| **A** | 2 | **Europe** | **AA** | 27 |
| **CCC** | 1 | **A** | 5 |
| **Lat.Amer** | **AA** | 1 | **BBB** | 2 |
| **A** | 5 | **B** | 2 |
| **B** | 1 | **India** | **A** | 6 |
| **MiddleEast** | **AA** | 1 | **BB** | 1 |
| **A** | 8 | **CCC** | 1 |
| **BBB** | 2 | **Lat.Amer** | **A** | 10 |
| **BB** | 4 | **BBB** | 4 |
| **B** | 1 | **BB** | 3 |
| **N.Amer** | **AA** | 76 | **B** | 1 |
| **A** | 50 | **CCC** | 4 |
| **BBB** | 20 | **MiddleEast** | **AA** | 1 |
| **BB** | 24 | **A** | 6 |
| **B** | 28 | **BBB** | 5 |
| **CCC** | 26 | **BB** | 8 |
| **Oceania** | **AA** | 8 | **B** | 2 |
| **A** | 7 | **N.Amer** | **AA** | 19 |
| **BB** | 2 | **A** | 22 |
| **CCC** | 1 | **BBB** | 12 |
| **OffShore** | **AA** | 3 | **BB** | 6 |
| **BB** | 1 | **Oceania** | **AA** | 3 |
| **B** | 5 | **Pacific** | **B** | 1 |
| **CCC** | 2 | **Supra** | **AA** | 7 |
|  |  |  |  | **BBB** | 1 |

Table : Distribution of CDS Tickers **across** Sector, Region, ImpliedRatings

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sector | Region | ImpliedRating | Num Obilgors | Sector | Region | ImpliedRating | Num Obilgors |
| Others | **Africa** | **BB** | 4 | Others | **N.Amer** | **AA** | 80 |
| **B** | 7 | **A** | 277 |
| **Asia** | **AA** | 55 | **BBB** | 212 |
| **A** | 108 | **BB** | 148 |
| **BBB** | 56 | **B** | 100 |
| **BB** | 50 | **CCC** | 71 |
| **B** | 29 | **D** | 1 |
| **Caribbean** | **BBB** | 1 | **Oceania** | **AA** | 6 |
| **BB** | 2 | **A** | 13 |
| **E.Eur** | **AA** | 1 | **BBB** | 15 |
| **A** | 1 | **BB** | 5 |
| **BBB** | 2 | **B** | 2 |
| **BB** | 6 | **OffShore** | **AA** | 3 |
| **B** | 5 | **A** | 2 |
| **CCC** | 1 | **BBB** | 1 |
| **Europe** | **AA** | 66 | **BB** | 3 |
| **A** | 109 | **B** | 4 |
| **BBB** | 110 | **CCC** | 3 |
| **BB** | 48 | **Supra** | **B** | 1 |
| **B** | 31 |  |  |  |  |
| **CCC** | 13 |  |  |  |  |
| **India** | **BBB** | 7 |  |  |  |  |
| **BB** | 6 |  |  |  |  |
| **B** | 13 |  |  |  |  |
| **CCC** | 10 |  |  |  |  |
| **Lat.Amer** | **BBB** | 19 |  |  |  |  |
| **BB** | 15 |  |  |  |  |
| **B** | 8 |  |  |  |  |
| **CCC** | 1 |  |  |  |  |
| **MiddleEast** | **BBB** | 4 |  |  |  |  |
| **BB** | 2 |  |  |  |  |
| **B** | 1 |  |  |  |  |

Table 2: Distribution of CDS Tickers **across** Sector, Region, ImpliedRatings



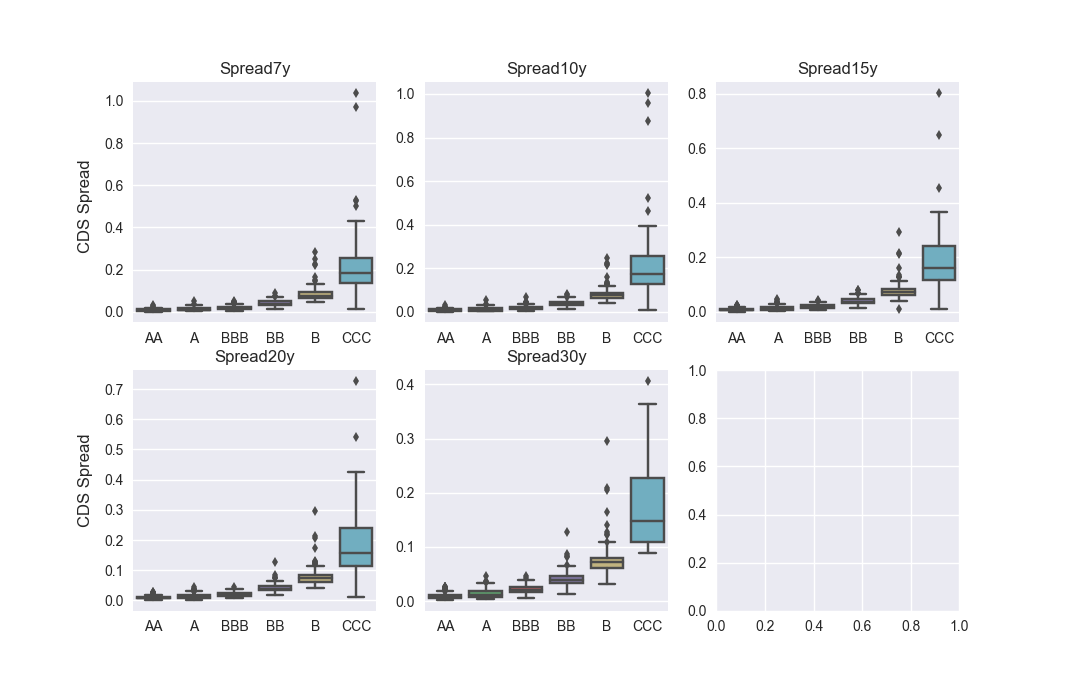


Figure : Boxplot of CDS Spreads across ImpliedRatings

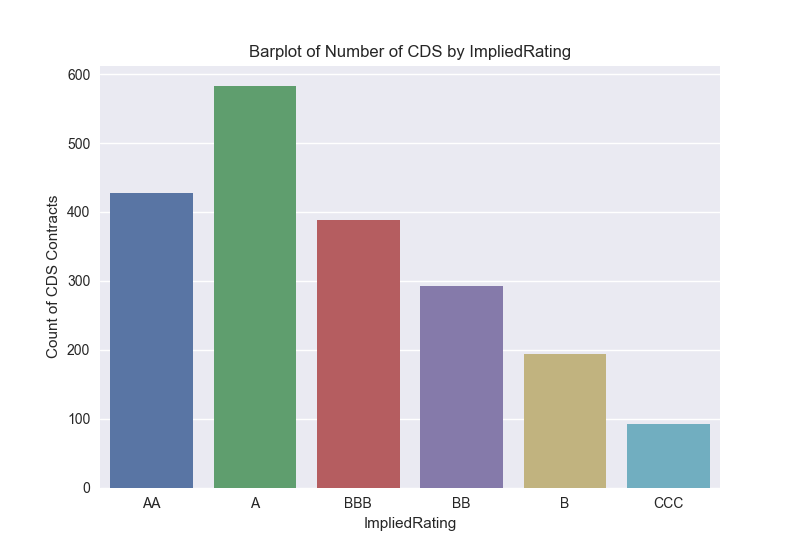


Figure : Barplot of #CDS Contracts across Ratings

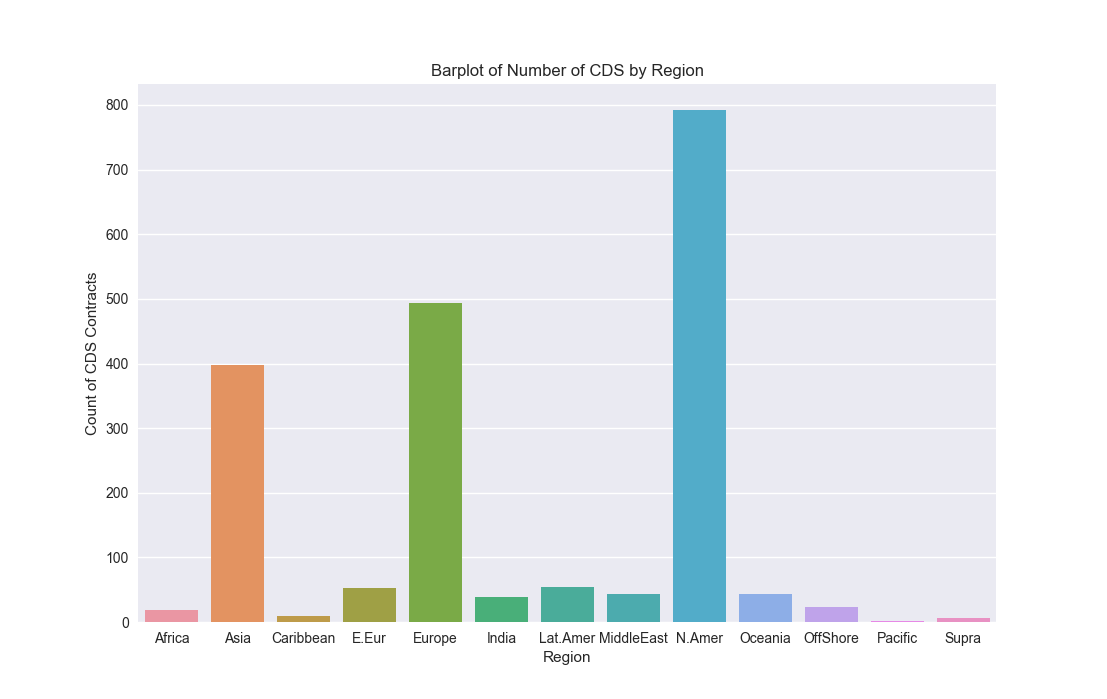


Figure : Carplot of #CDS Contracts across Region

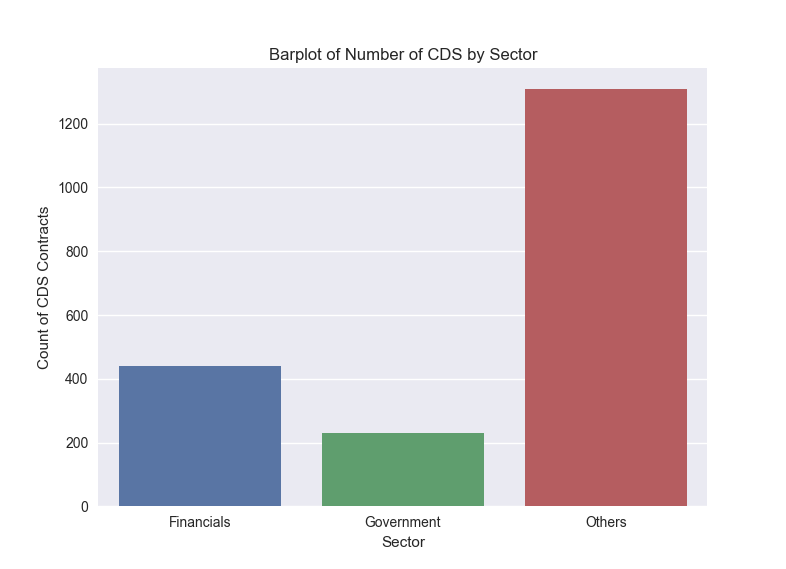


Figure : Barplot of #CDS Contracts across Sectors

# The Cross-Section Methodology

## Description

Our objective is to get a term structure of CDS Spreads for a given country that has no CDS Spreads.

As our concern is mainly about the risk profile of sovereigns only, our methodology differs from Nomura’s paper by which we can only choose factors that affect our proxy spread.

We have chosen our regression factors to be:

* Global factor
* Factor for sector of the obligor
* Factor for Region of the obligor
* Factor for ImpliedRating of the obligor

In the symbols we can write the proxy spread of obligor as:

where

* sector of obligor
* region of obligor
* ImpliedRating of obligor
* global factor

For example, for a European Financial Company rated BBB, we would have

Because our data is not of a binary form, we cannot apply classification techniques to get the value of . We resort to different regression models to estimate our value of from the common factors cited above.

We propose to work on 3 different datasets, and also a few models of regression to look for our proxy .

Our data, Sector, Region and Implied Rating are not numeric. Sector and Region are categorical data, while Implied Rating is ordinal data. They are encoded into binary form in the matrix below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **T icker** | Sector\_Financials | Sector\_Government | Sector\_Others | Region\_Africa | Region\_Asia | Region\_Caribbean | Region\_E.Eur | Region\_Europe | Region\_India | Region\_Lat.Amer | Region\_MiddleEast | Region\_N.Amer | Region\_Oceania | Region\_OffShore | Region\_Supra | ImpliedRating A | ImpliedRating AA | ImpliedRating B | ImpliedRating BB | ImpliedRating BBB | ImpliedRating CCC |
| AMR | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| ATFJ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| BNDES | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| CAG | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

In this example, we see that the ticker AMR, corresponding to AMR Corp belongs to Sector Others and it belongs to N.America, and the ImpliedRating is BB.

We want to find the optimal that makes the proxy spreads as close as possible to the market spreads

## **Types of Datasets**

In our following dataset, we filter only take **CDS Obligors of Seniority Rating: SNRFOR, Currency: USD and DocClause: CR**.

### Dataset 1

Since our objective is to get a termstructure of CDS Spreads for our proxy.

Dataset 1 consists of removing **any** rows of CDS Tickers that have missing CDS Spreads at any tenor.  
This gives us a smaller subset left, which is around 1000 CDS Tickers.

From the example below:

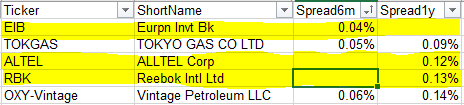


Figure : Dataset 1 Example

When regressing on every of the tenors, we exclude the CDS Tickers EIB, ALTEL and RBK.

### Dataset 2

Since our objective is to get a termstructure of CDS Spreads for our proxy, and we are regressing on each CDS Tenor one by one, Dataset 2 consists of only removing rows of CDS Tickers that have missing spreads **at each tenor**. On average, we have around 1343 CDS Tickers per Tenor.

From the example below:

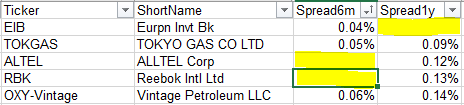


Figure : Dataset 2 Example

While regressing on the Spread6m Tenor, we only leave out CDS Tickers ALTEL and RBK.  
When regressing on the Spread1y Tenor, we only leave out the CDS Tickers EIB.

### Dataset 3

Dataset 3 is a more refined version of how Dataset 2 is defined. We are still regressing on each CDS Tenor one by one. Dataset 3 aims to remove the bias and outliers that might be present in the data.

For each tenor, Dataset 3 takes only CDS Tickers within standard deviations in each ImpliedRating Category.

## Methods

There are 5 methods that we have tried for regression.  
For each of the methods we calibrate our model on the different datasets that we have. In addition, we calculate the proxy spread tenor by tenor to get our whole curve.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Dataset 1 | Dataset 2 | Dataset 3 |
| Method 1 | Ordinary Least Squares Regression | Ordinary Least Squares Regression | Ordinary Least Squares Regression |
| Method 2 | Regression with -regularisation | Regression with -regularisation | Regression with -regularisation |
| Method 3 | Regression with -regularisation | Regression with -regularisation | Regression with -regularisation |

The following description is a brief methodology of the models that are used in the regression approach.

### Method 1 - Ordinary Least Squares

This model aims to solve:

where

* our data matrix of and ,

After solving for , our proxy spreads can be explained by:

where

Solving the least squares problem can require different methods. Because the matrix is singular, inverting the matrix will be problem if we solve for it directly. We have chosen to factorize the matrix using Singular Value Decomposition, and then solve for .

### Method 2 - Least Squares with regularisation

This model aims to solve:

where

* our data matrix of and ,
* regularization parameter

After solving for , our proxy spreads can be explained by:

where

The addition of the -regulariastion gives solution uniqueness of and also prevents the problem of overfitting when there are too many parameters.

### Method 3 - Least Squares with -regularisation

This model aims to solve:

where

* our data matrix of and ,
* regularization parameter

After solving for , our proxy spreads can be explained by:

where

The addition of the penalty function helps the model to estimate sparse coefficients. It will be useful in siuations due to its tendency to prefer solutions with fewer parameter values, effectively reducing the number of variables upon which the given solution is dependent.

For further details, one can refer to the link below to gain a better understanding.

<http://statweb.stanford.edu/~tibs/lasso.html>

## Results

Below we have run our regressions on each of the dataset, with the corresponding models. The spread factors are displayed in the tables below.

Below we show the weights that are obtained from solving the different models.

Some observations to expect:

* One can expect to see that the Spread Factors for Implied Ratings to be monotonic. For Sector and Region, there should be no patterns expected.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset 1 | Model 1  Spread Factor | Model 2  ()  Spread Factor | Model 3  ()  Spread Factor | Model 4  Spread Factor | Model 5  ()  Spread Factor |
| Sector\_Financials | -7.64E+10 | -0.41193 | 0 | -0.41997 | -0.41497 |
| Sector\_Government | -7.64E+10 | -0.65852 | -0.11053 | -0.66062 | -0.65754 |
| Sector\_Others | -7.64E+10 | -1.0077 | -0.53294 | -1.01823 | -1.03555 |
| Region\_Africa | -2.11E+09 | -0.09663 | 0 | -0.08892 | -0.21556 |
| Region\_Asia | -2.11E+09 | -0.36483 | -0.17537 | -0.31818 | -0.28587 |
| Region\_Caribbean | -2.11E+09 | -0.14544 | 0 | -0.36041 | -0.40343 |
| Region\_E.Eur | -2.11E+09 | -0.31953 | 0 | -0.31692 | -0.43053 |
| Region\_Europe | -2.11E+09 | -0.09179 | 0.031219 | -0.04926 | -0.02735 |
| Region\_India | -2.11E+09 | 0.039017 | 0 | 0.077975 | 0.155971 |
| Region\_Lat.Amer | -2.11E+09 | -0.06475 | 0 | -0.03305 | -0.00528 |
| Region\_MiddleEast | -2.11E+09 | -0.19488 | 0 | -0.18444 | -0.22234 |
| Region\_N.Amer | -2.11E+09 | -0.18027 | 0 | -0.14203 | -0.1125 |
| Region\_Oceania | -2.11E+09 | -0.19308 | 0 | -0.15314 | -0.1336 |
| Region\_OffShore | -2.11E+09 | -0.41389 | 0 | -0.46313 | -0.4276 |
| Region\_Supra | -2.11E+09 | -0.05208 | 0 | -0.06731 | 0 |
| ImpliedRating\_AA | 5.49E+11 | -1.89581 | -1.76629 | -1.92857 | -1.91517 |
| ImpliedRating\_A | 5.49E+11 | -1.41398 | -1.2935 | -1.43405 | -1.4117 |
| ImpliedRating\_BBB | 5.49E+11 | -0.75042 | -0.61441 | -0.76507 | -0.70237 |
| ImpliedRating\_BB | 5.49E+11 | -0.08047 | 0 | -0.08789 | -0.07655 |
| ImpliedRating\_B | 5.49E+11 | 0.611831 | 0.637719 | 0.61778 | 0.607568 |
| ImpliedRating\_CCC | 5.49E+11 | 1.450697 | 1.400438 | 1.498981 | 1.390155 |
| Global | -4.70E+11 | -2.07815 | -2.80645 | -2.09881 | -2.10807 |

Table : Results Dataset 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset 2 | Model 1  Spread Factor | Model 2  ()  Spread Factor | Model 3  ()  Spread Factor | Model 4  Spread Factor | Model 5  ()  Spread Factor |
| Sector\_Financials | 3.04E+12 | -0.37255 | 0 | -0.29025 | -0.31137 |
| Sector\_Government | 3.04E+12 | -0.67292 | -0.11125 | -0.59423 | -0.69075 |
| Sector\_Others | 3.04E+12 | -0.98113 | -0.55331 | -0.9003 | -0.97386 |
| Region\_Africa | 1.55E+12 | -0.04228 | 0 | -0.00661 | -0.13903 |
| Region\_Asia | 1.55E+12 | -0.3476 | -0.12065 | -0.29985 | -0.30686 |
| Region\_Caribbean | 1.55E+12 | -0.16978 | 0 | -0.23177 | -0.28265 |
| Region\_E.Eur | 1.55E+12 | -0.27097 | 0 | -0.25018 | -0.27933 |
| Region\_Europe | 1.55E+12 | -0.17006 | 0 | -0.12332 | -0.14098 |
| Region\_India | 1.55E+12 | 0.047353 | 0 | 0.063323 | 0.079248 |
| Region\_Lat.Amer | 1.55E+12 | -0.05527 | 0 | -0.0149 | -0.02466 |
| Region\_MiddleEast | 1.55E+12 | -0.14681 | 0 | -0.11312 | -0.12648 |
| Region\_N.Amer | 1.55E+12 | -0.22405 | 0 | -0.18133 | -0.18024 |
| Region\_Oceania | 1.55E+12 | -0.21901 | 0 | -0.18104 | -0.19447 |
| Region\_OffShore | 1.55E+12 | -0.37755 | 0 | -0.39051 | -0.37135 |
| Region\_Supra | 1.55E+12 | -0.05058 | 0 | -0.05549 | -0.00918 |
| ImpliedRating\_AA | -2.87E+00 | -1.96901 | -1.82297 | -2.34917 | -2.0729 |
| ImpliedRating\_A | -2.41E+00 | -1.51658 | -1.38162 | -1.88976 | -1.58438 |
| ImpliedRating\_BBB | -1.73E+00 | -0.83535 | -0.677 | -1.20634 | -0.87976 |
| ImpliedRating\_BB | -9.84E-01 | -0.09831 | 0 | -0.46243 | -0.18546 |
| ImpliedRating\_B | -2.98E-01 | 0.575508 | 0.60117 | 0.22152 | 0.481794 |
| ImpliedRating\_CCC | 5.69E-01 | 1.407341 | 1.348927 | 1.086092 | 1.264716 |
| Global | -4.59E+12 | -2.0266 | -2.76476 | -1.78478 | -1.97598 |

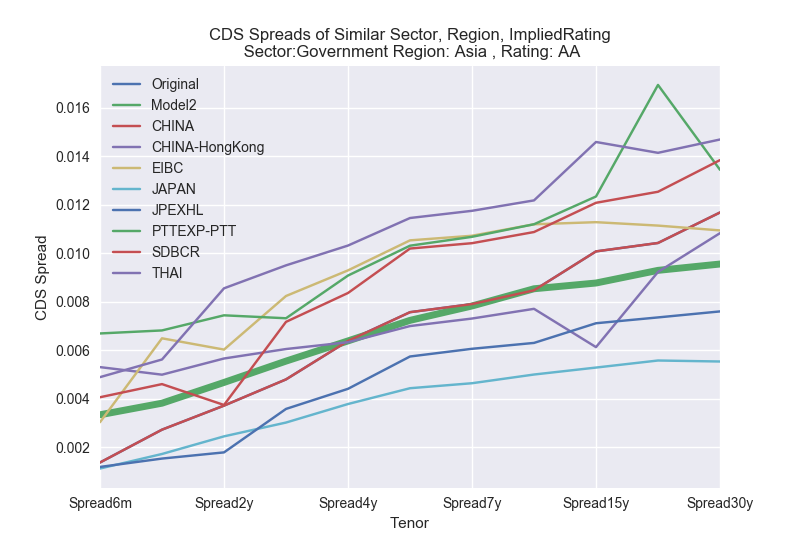
Table : Results Dataset 2

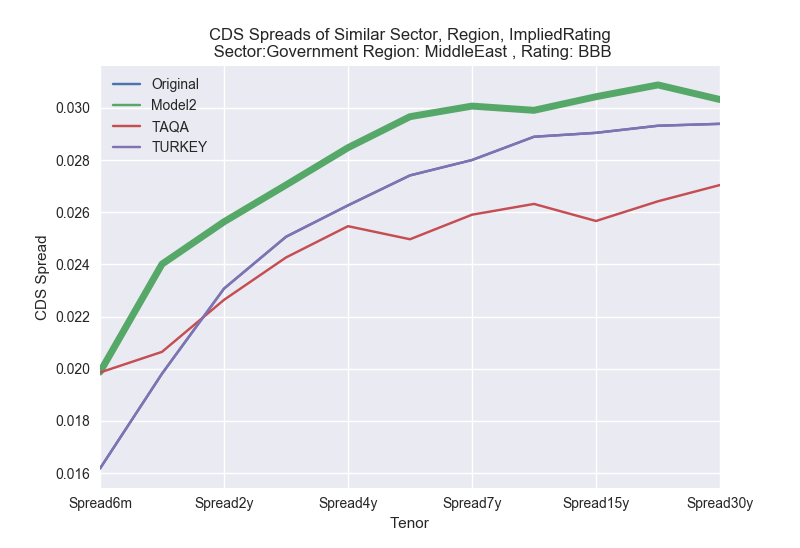
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset 3 | Model 1  Spread Factor | Model 2  ()  Spread Factor | Model 3  ()  Spread Factor | Model 4  Spread Factor | Model 5  ()  Spread Factor |
| Sector\_Financials | 6.52E+12 | -0.3902 | 0 | -0.3926 | -0.3502 |
| Sector\_Government | 6.52E+12 | -0.6918 | -0.1117 | -0.6952 | -0.7297 |
| Sector\_Others | 6.52E+12 | -0.9932 | -0.5481 | -0.9972 | -1.0090 |
| Region\_Africa | -1.12E+12 | -0.0398 | 0 | -0.0277 | -0.1382 |
| Region\_Asia | -1.12E+12 | -0.3618 | -0.1307 | -0.3369 | -0.3241 |
| Region\_Caribbean | -1.12E+12 | -0.1639 | 0 | -0.2452 | -0.2741 |
| Region\_E.Eur | -1.12E+12 | -0.2637 | 0 | -0.2664 | -0.2794 |
| Region\_Europe | -1.12E+12 | -0.1757 | 0 | -0.1518 | -0.1480 |
| Region\_India | -1.12E+12 | 0.0019 | 0 | 0.0127 | 0.0016 |
| Region\_Lat.Amer | -1.12E+12 | -0.0548 | 0 | -0.0381 | -0.0289 |
| Region\_MiddleEast | -1.12E+12 | -0.1456 | 0 | -0.1358 | -0.1264 |
| Region\_N.Amer | -1.12E+12 | -0.2291 | 0 | -0.2089 | -0.1915 |
| Region\_Oceania | -1.12E+12 | -0.2231 | 0 | -0.2079 | -0.1963 |
| Region\_OffShore | -1.12E+12 | -0.3713 | 0 | -0.4057 | -0.3760 |
| Region\_Supra | -1.12E+12 | -0.0484 | 0 | -0.0733 | -0.0076 |
| ImpliedRating\_AA | -6.87E+10 | -1.8998 | -1.8174 | -1.9210 | -1.9198 |
| ImpliedRating\_A | -6.87E+10 | -1.4500 | -1.3791 | -1.4643 | -1.4376 |
| ImpliedRating\_BBB | -6.87E+10 | -0.7684 | -0.6732 | -0.7802 | -0.7289 |
| ImpliedRating\_BB | -6.87E+10 | -0.0314 | 0 | -0.0365 | -0.0359 |
| ImpliedRating\_B | -6.87E+10 | 0.6279 | 0.5899 | 0.6328 | 0.6213 |
| ImpliedRating\_CCC | -6.87E+10 | 1.4465 | 1.3247 | 1.4841 | 1.4120 |
| Global | -5.33E+12 | -2.0752 | -2.7718 | -2.0850 | -2.0888 |

Table : Results Dataset 3

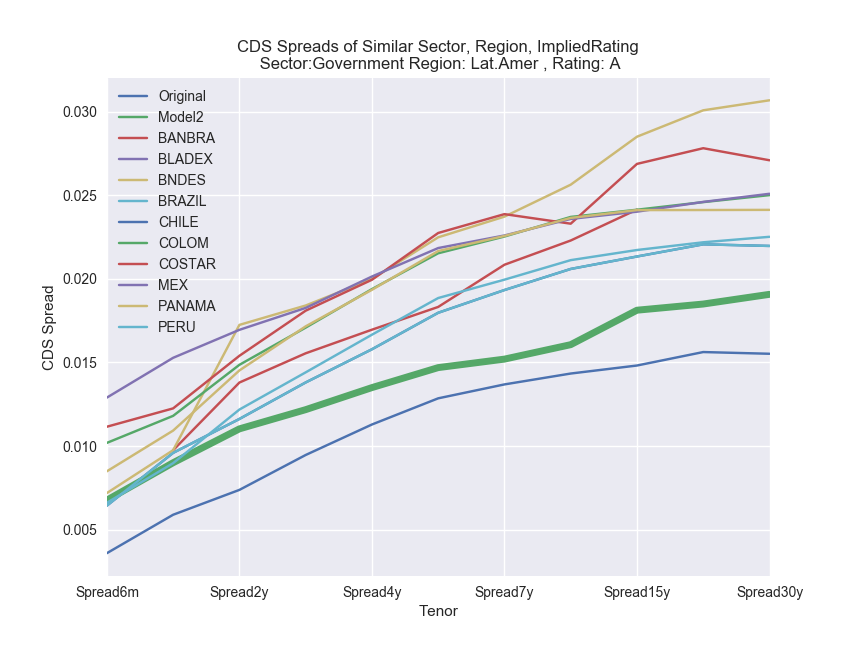
As seen in the above tables, the Spread Factors of Model 1 are clustered into the 3 different factors. For each factor they only differ in amounts of decimal places.

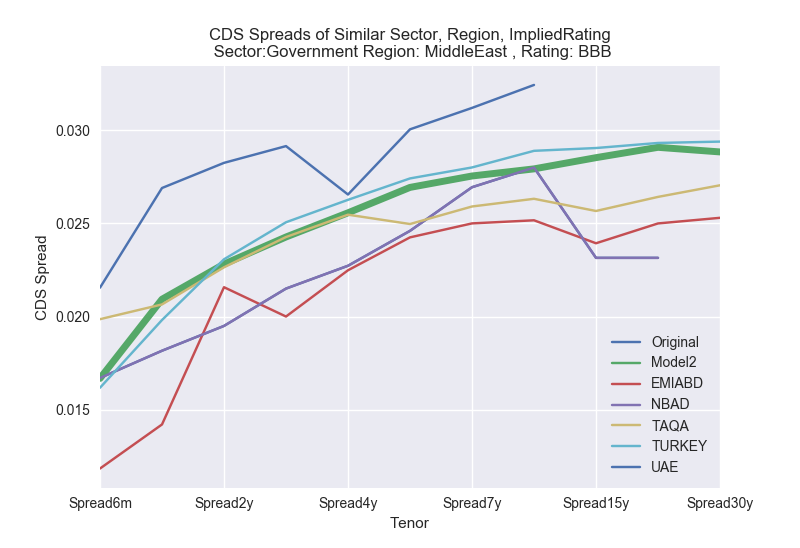
Dataset 1



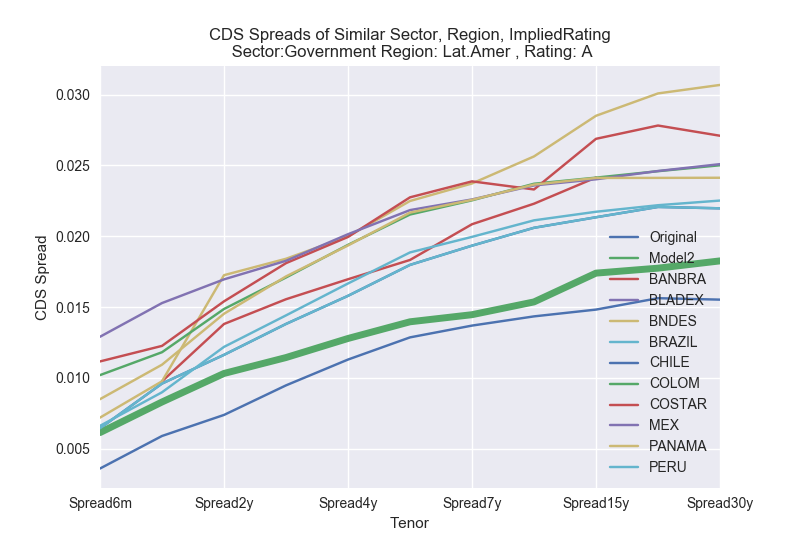


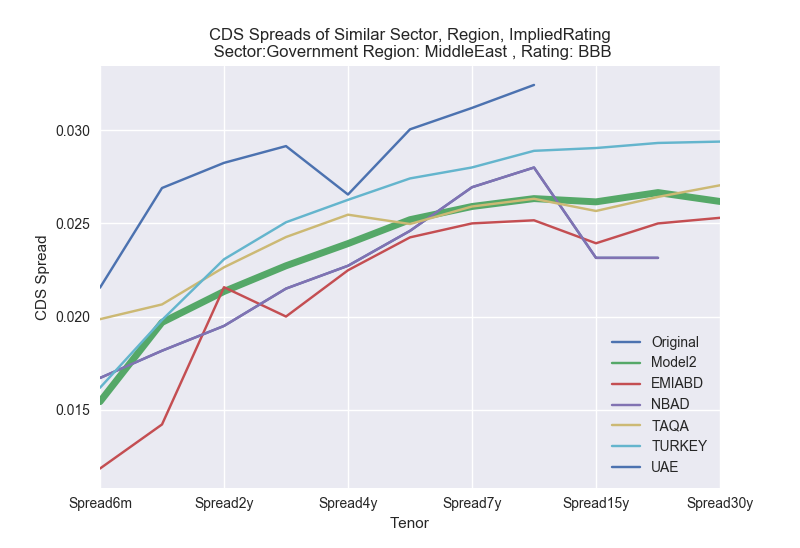
Dataset 2





Dataset 3





## Robustness Check

We conducted a robustness check to see if our models correspond. We fix the sector and rating, and vary Implied Ratings across, to see if the CDS spreads increase when ratings decrease.

Below are the results for the robustness check for **Model 2, adopted to Dataset 2.**  
We displayed the results for 1 Liquid Area and 1 Illiquid Area, where a Liquid Area is classified by having many CDS obligors in Sector and Region.

Results below show that our robustness checks have succeeded.

**Liquid Area**

* **Sector:** Financials
* **Region:** N.America

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ImpliedRating:  Tickers: | AA  JPM | A  USB | BBB  C | BB  JANUS | B  AVBDGT | CCC  SREGL |
| Spread6m | 0.0071 | 0.0122 | 0.0270 | 0.0603 | 0.1311 | 0.2860 |
| Spread1y | 0.0079 | 0.0136 | 0.0292 | 0.0636 | 0.1336 | 0.2942 |
| Spread2y | 0.0089 | 0.0147 | 0.0307 | 0.0660 | 0.1344 | 0.2926 |
| Spread3y | 0.0095 | 0.0153 | 0.0312 | 0.0665 | 0.1297 | 0.2768 |
| Spread4y | 0.0102 | 0.0158 | 0.0315 | 0.0658 | 0.1266 | 0.2624 |
| Spread5y | 0.0107 | 0.0163 | 0.0320 | 0.0650 | 0.1232 | 0.2497 |
| Spread7y | 0.0106 | 0.0158 | 0.0303 | 0.0616 | 0.1128 | 0.2271 |
| Spread10y | 0.0106 | 0.0152 | 0.0285 | 0.0575 | 0.1039 | 0.2040 |
| Spread15y | 0.0110 | 0.0161 | 0.0286 | 0.0552 | 0.1006 | 0.1858 |
| Spread20y | 0.0116 | 0.0167 | 0.0292 | 0.0549 | 0.0982 | 0.1787 |
| Spread30y | 0.0120 | 0.0168 | 0.0291 | 0.0526 | 0.0929 | 0.1671 |

**Illiquid Area**

* **Sector:** **Govt**
* **Region: Middle East**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ImpliedRating:  Tickers: | AA | A  ISRAEL | BBB  TAQA | BB  LEBAN | B  IRAQ | CCC |
| Spread6m |  | 0.0076 | 0.0167 | 0.0372 | 0.0808 |  |
| Spread1y |  | 0.0097 | 0.0209 | 0.0455 | 0.0956 |  |
| Spread2y |  | 0.0109 | 0.0228 | 0.0491 | 0.1000 |  |
| Spread3y |  | 0.0119 | 0.0243 | 0.0517 | 0.1008 |  |
| Spread4y |  | 0.0128 | 0.0256 | 0.0533 | 0.1026 |  |
| Spread5y |  | 0.0138 | 0.0269 | 0.0547 | 0.1037 |  |
| Spread7y |  | 0.0143 | 0.0275 | 0.0561 | 0.1026 |  |
| Spread10y |  | 0.0149 | 0.0279 | 0.0563 | 0.1017 |  |
| Spread15y |  | 0.0160 | 0.0285 | 0.0550 | 0.1002 |  |
| Spread20y |  | 0.0166 | 0.0291 | 0.0546 | 0.0977 |  |
| Spread30y |  | 0.0167 | 0.0288 | 0.0522 | 0.0922 |  |

From the 2 graphs shown above, we can conclude that our model is robust to changes in the Implied Ratings.

## **Approach 1: Regression**

In the following, we note down 3 different approaches that we have tried. For any of the approach below, we can choose

* Any Dataset choice in the options above
* Any Method choice in the options above

### Approach 1(A)

In Approach 1(A) we just pick a dataset and method, and regress upon Sector, Region and Implied Ratings to get the CDS Proxy curve. Upon getting the Proxy Curve, we bootstrap directly to get our hazard rates and survival probabilities. We do not add any further complexity to the model.

#### Methodology to get Hazard Rates

1. Pick a Dataset and Method.
2. Regress the CDS Spreads Tenor by Tenor against Sector, Rating, ImpliedRating choosing a preferred method.
3. Bootstrap the CDS Proxy curve to get desired Hazard Rates

### Approach 1(B)

We follow the approach 1(A), and upon getting the bootstrapped hazard rates, we smooth the hazard rates points with a Nelson Siegel smoothing function.

#### Methodology to get Hazard Rates

1. Pick a Dataset and Method.
2. Regress the CDS Spreads Tenor by Tenor against Sector, Rating, ImpliedRating choosing a preferred method.
3. Bootstrap the CDS Proxy curve
4. Smooth the Hazard Rates using smoothing technique Nelson Siegel

More details of the Nelson Siegel Method is explained in the chapter below.

### Approach 1(C)

In Approach 1(C), we follow the steps of Approach 1(A), but upon getting the CDS Proxy Curve, we smooth the proxy curve before bootstrapping them to get the desired hazard rates.

Our technique of smoothing follows the same way as above, using the Nelson Siegel method.\

#### Methodology to get Hazard Rates

1. Pick a Dataset and Method.
2. Regress the CDS Spreads Tenor by Tenor against Sector, Rating, ImpliedRating choosing a preferred method.
3. Smooth the CDS Proxy Curve using smoothing technique Nelson Siegel
4. Bootstrap the smoothed CDS Proxy Curve

# Results

Below we display the Hazard Rates that are being bootstrapped from the CDS Proxies obtained from their respective models.

# **Approach 2: Nelson Siegel Fitting**

## Motivation

Our motivation is driven by regressing our explanatory factors based on the level, slope and curvature of given fitted curves.

We use the Nelson Siegel Fitting approach to fit each CDS Ticker, with more than or equal to 5 data points for sufficient fitting, to obtain a list of parameters for each the level, slope and curvature. Based on the fitted curves, we can hopefully get our “hopeful” parameters from list of parameters above by regressing on the respective sector, region and ImpliedRating.

With such parameters, we can thus get our desired CDS Proxy spread curve at any time point.

## Nelson Siegel Curve Fitting

Our fitted curve for each

Where

* CDS Spread with maturity of at time
* decay factor. We fix
* time to maturity

The parameter governs the exponential decay rate. Small values of produce slow decay and can fit the curve at long maturities, while large values of produce fast decay and can fit the curve at shorter maturities. also governs where the loading on achieves its maximum.

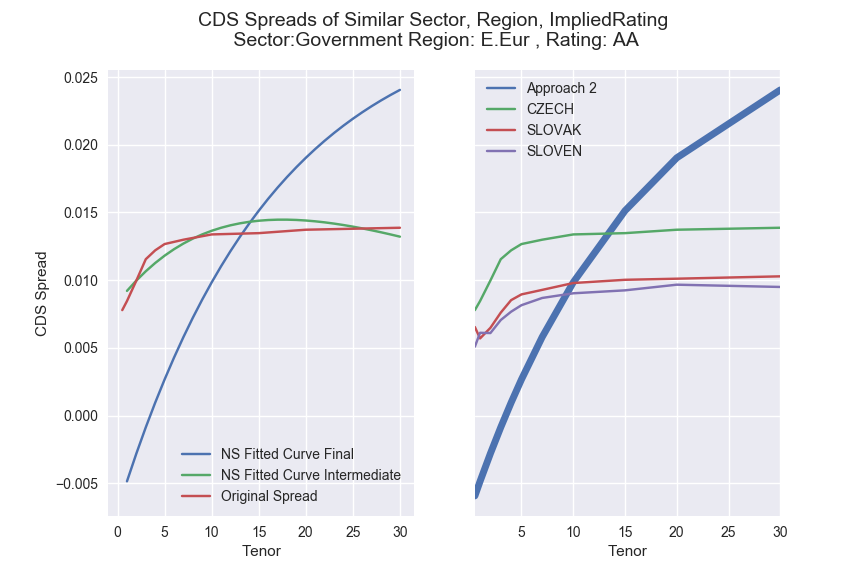
## Methodology

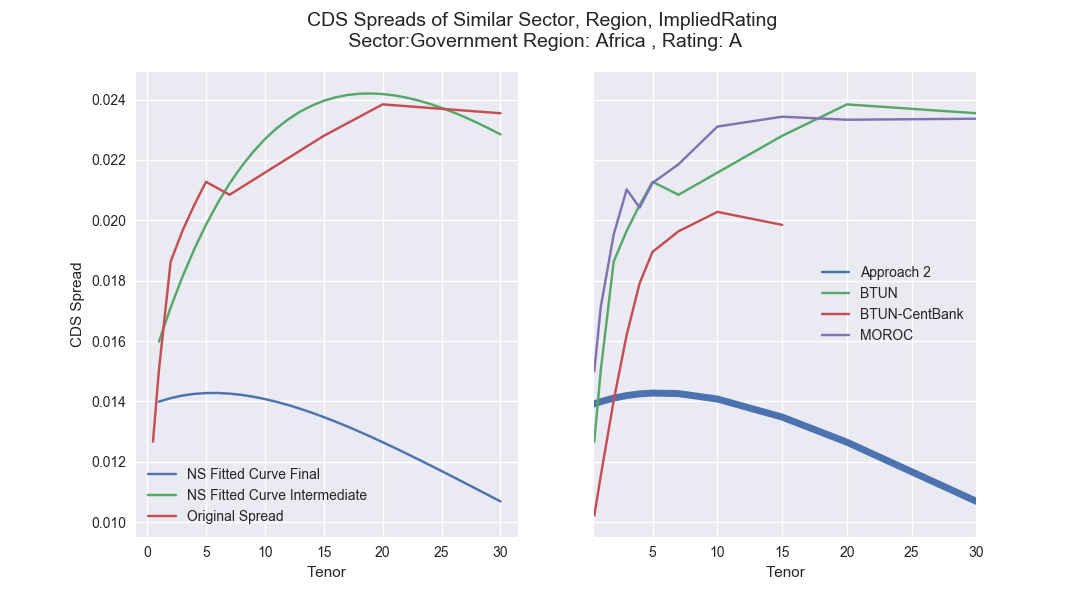
To get an estimated CDS Proxy curve for a given selected ,

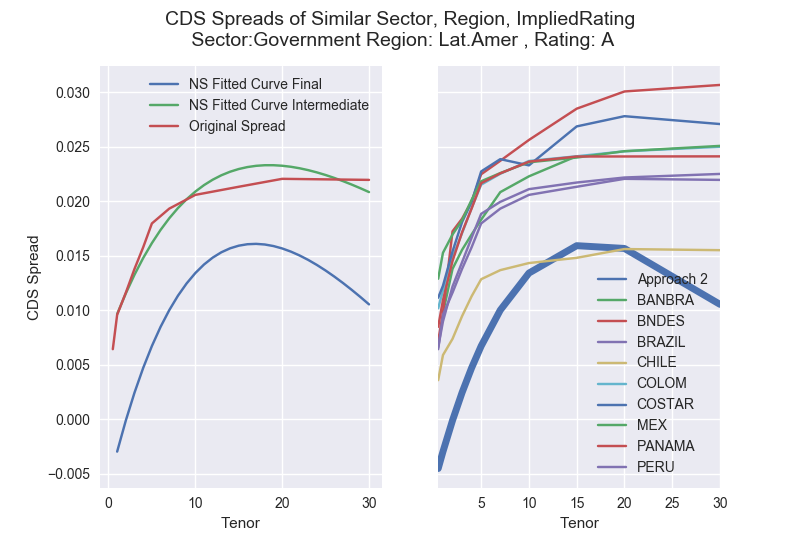
1. At time , For each of the CDS Ticker with more than available CDS Spreads, we fit a Nelson Siegel curve to get a set of parameters
2. We regress on each of the set of paramters , and with our regressor factors to be Sector, Region and ImpliedRatings.

Results  
  
Below is a picture of some of the fits after the first step of calibrating to each CDS Curve.









As seen in the above graphs, we can see that:

1. The initial fitting to the original CDS spread curve seems to fit well with the proposed value of .
2. The final fitted CDS Proxy curve to the factors Sector, Region and Implied Rating separately does not seem to be close enough to the other curves in the same buckets.

One reason could be due to the regression of on the level, slope and curvatures factors alone. For a given value of , we see that the slope, level and curvature obey some form of correlation.

One possible alternative is that in the first step, we optimize for such that makes to have zero correlation. Then, we can proceed to regress on each factors separately.

# **Approach 3: Classification**

## Motivation

Approach 3 takes on a different route altogether from Approach 1 and 2. We propose to estimate the CDS Proxy curve by means of classification approaches.

Our approach is to find given a set of parameters e.g , we can find a closest CDS Ticker to proxy it.

This approach will somewhat be similar to an unsupervised machine learning approach.

Since our data is of categorical, ordinal, and of numeric nature, we need to find means to calculate the distance within each factor and across factors. One metric that could potentially answer this problem is the gower distance.

We have yet to document our result and this will remain to be left for future work.

# Model Comparison

Here we compare the models that we have proposed for above.

### -Fold Cross Validation

Cross validation is a model validation technique to access how the results of a statistical analysis will generalize to an independent dataset.

In our case, we apply a -Fold cross validation to each of our models.

The original sample is randomly partitioned into equal sized groups. Of the equal sized groups, only single group is retained as the validation data for testing the model, and the equal sized groups are used as training data. The validation process is then repeated times, of which the results can then be averaged to produce a single estimation.

For example, setting  results in -fold cross-validation. In -fold cross-validation, we randomly shuffle the dataset into five sets ,, so that all sets are equal size (this is usually implemented by shuffling the data array and then splitting it in five sets). We then test on  and test on , followed by testing on  and training on ,. Then we do likewise for the other sets.



### Loss function

Since the output variable is not a binary variable, I use the Loss function as:

where

In our specific example, our elements will be the CDS Spread at a specific tenor.

### Cross Validation Error

For each tenor, the Cross Validation error is defined as

IDEA of Cross Validation:

Where

* Actual spread at for element in test group
* Model predicted spread at for element in test group
* corresponds to the th subfold of the Cross validation split.
* corresponds to the number of subfolds
* corresponds to the number of elements in each subfold

We have used random shuffle and according to an initial seed, so the shuffle every time is constant.

To get a single number for the Cross Validation result, we have averaged the Cross Validation Error across all Tenors to get a single number.

### Results

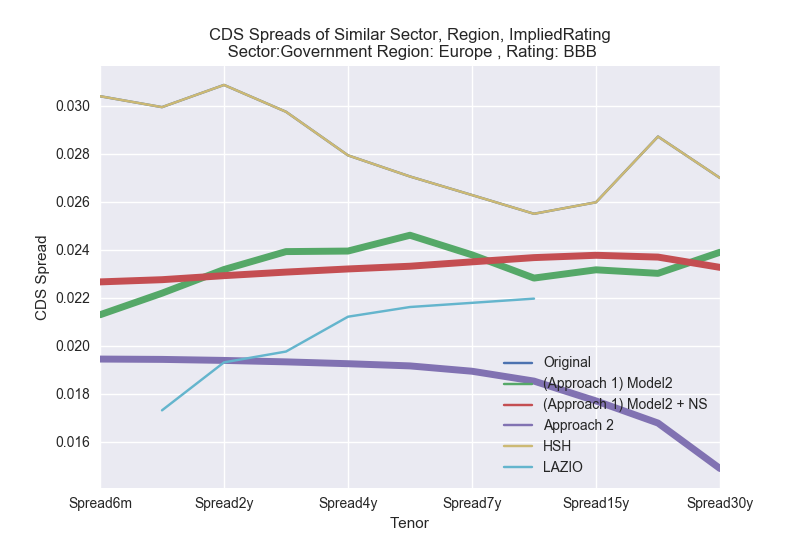
For the country of Brazil

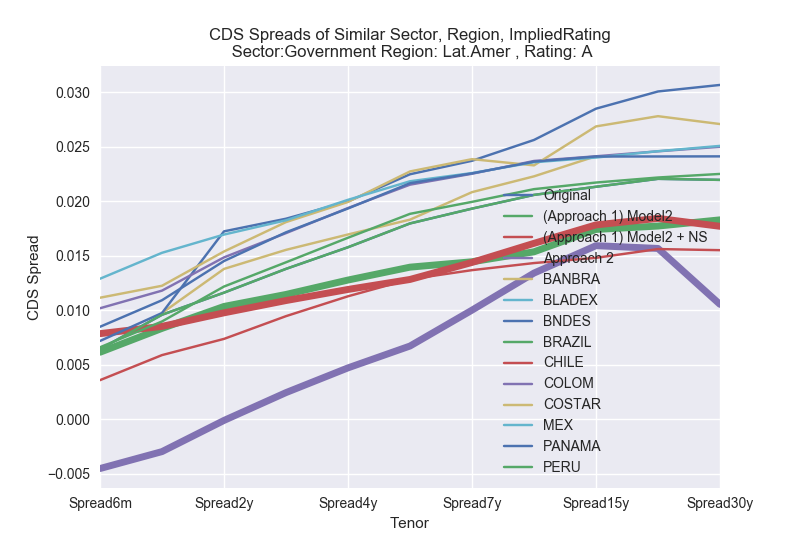
Dataset 2

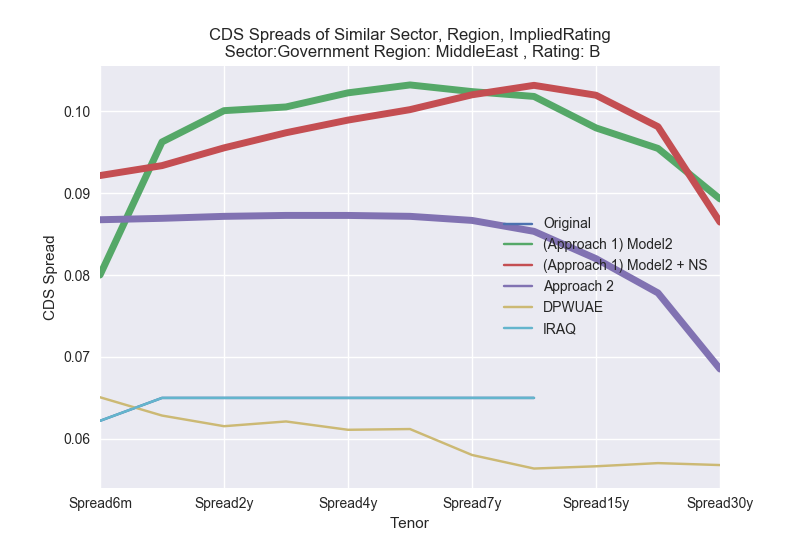
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cross Validation Results | Method | CV Error | CV Error | CV Error |
| Approach 1  (A) | Standard Least Squares |  |  |  |
| Approach 1  (A) | -regularisation | 0.003328923 | 0.003312877 | 0.003298 |
| Approach 1  (A) | -regularisation | 0.00267435 | 0.0026630868 | 0.0026553048496 |
| Approach 1  (C) | Standard Least Squares +  Nelson Siegel Fitting |  |  |  |
| Approach 1  (C) | -regularisation +  Nelson Siegel Fitting | 0.002791541 | 0.002792467 | 0.002791 |
| Approach 1  (C) | -regularisation+  Nelson Siegel Fitting | 0.00285044 | 0.00285139819 | 0.002850351 |
| Approach 2 | Nelson Siegel Fitted | 0.6860 | 0.3411 | 0.170 |

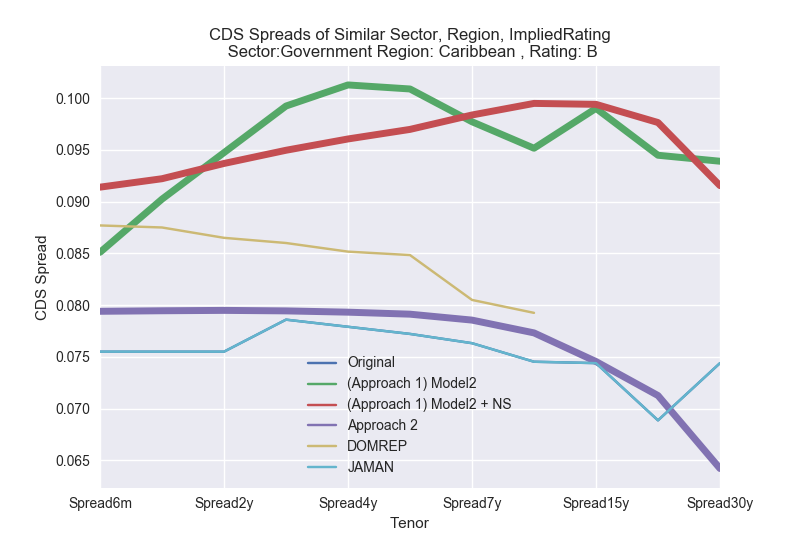
The error should be the lower the better.

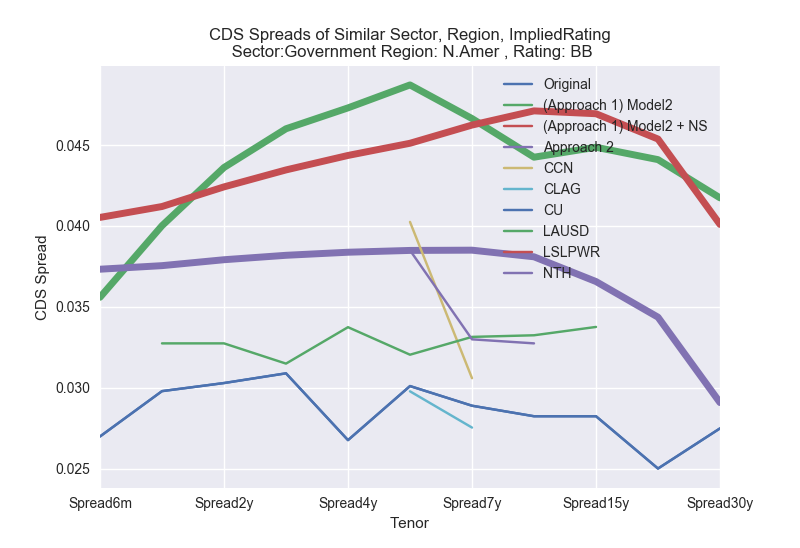
Some overall pictures below:

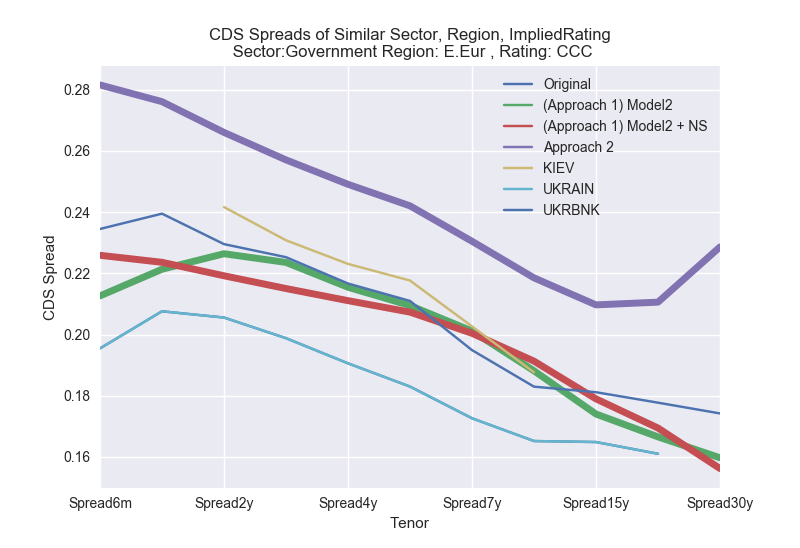
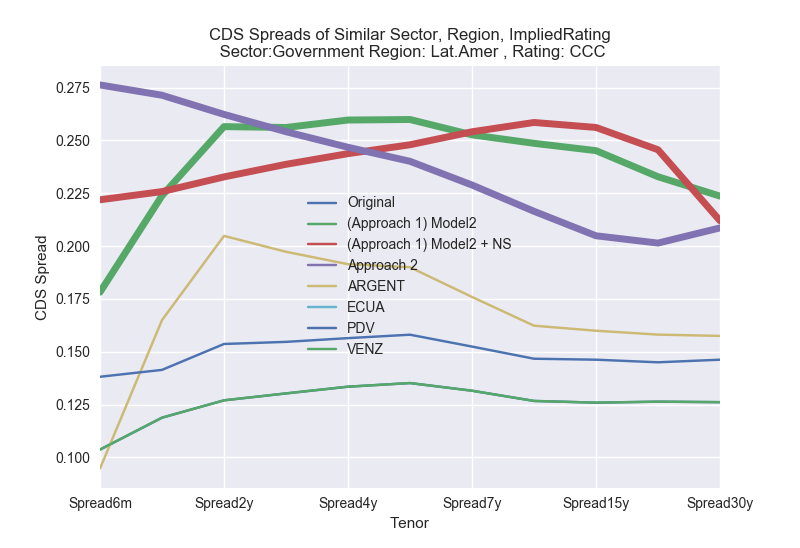
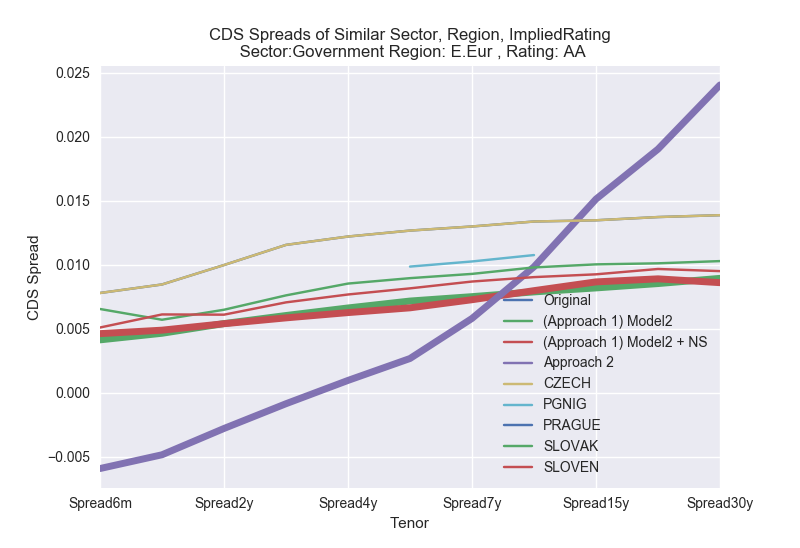
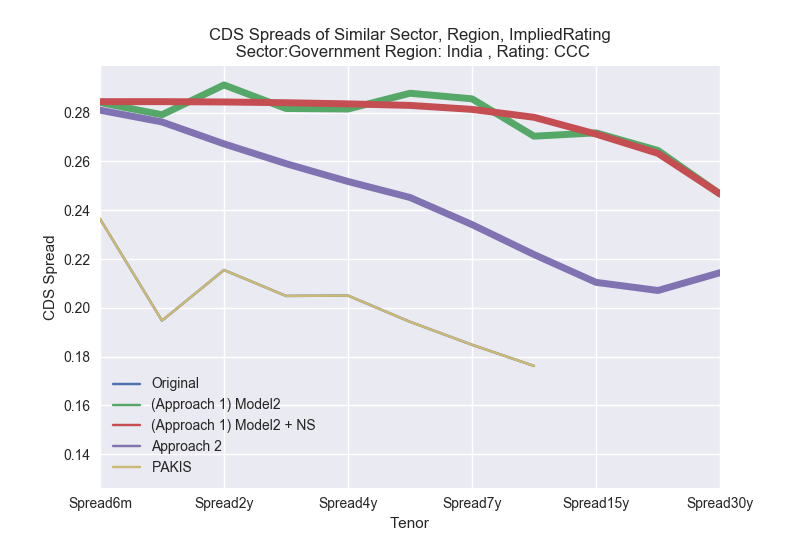
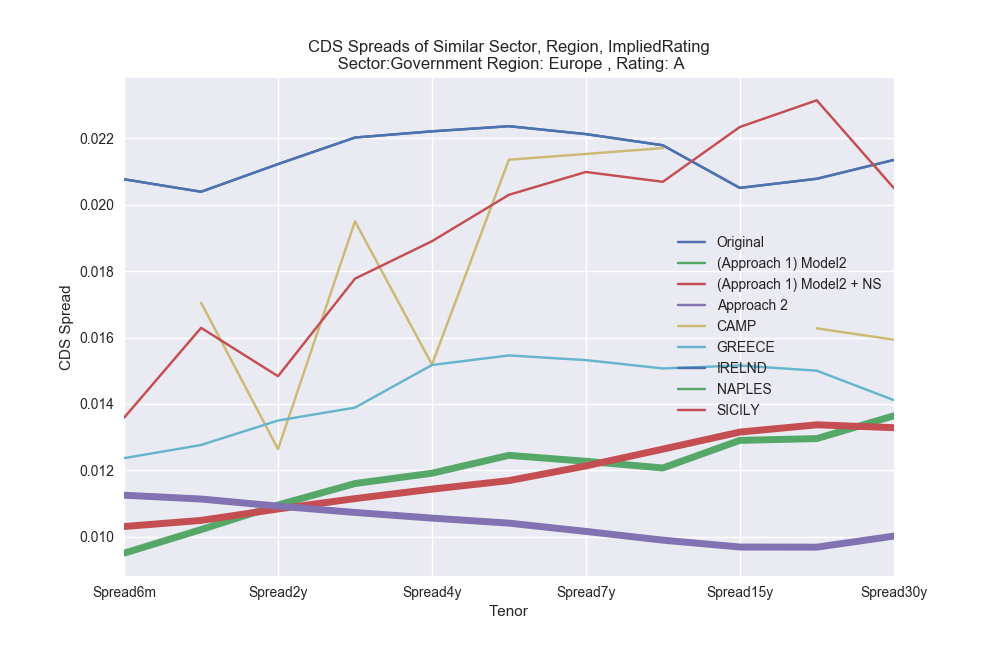












# Conclusion

Using the cross-sectional approach for regression only does not guarantee a CDS proxy spread that is arbitrage free, as seen from some of the cases above. We have added a smoothing parameter to fit the curve smoothly, where there be might some market inefficiencies.

In the case of fitting the parameters smoothly, we have tested and seen that

* Smoothing the Curve after bootstrapping the CDS Proxy Curve

Since our main aim is to get us a reasonable Hazard Curve, we can just smooth the Hazard Curve right away after bootstrapping the CDS Proxy curve.

* Smoothing the CDS Proxy Curve before getting Hazard Rates

The method is intuitive. As our method goes to produce CDS Spreads tenor by tenor sequentially, the CDS Proxy curve obtained might be rough or jagged. Smoothing the curve is a straightforward approach, but the results obtained are not stable for some of the CDS curve. Some of the Hazard Rates blow up.

Under the 2nd approach, we see that CDS Proxy spreads can become negative. This is a major drawback for our purpose as negative spreads is never seen in the market.

Below is a summary of results have decided to go with **Dataset 2, Method 2.**

|  |  |  |
| --- | --- | --- |
| Dataset 2, Method 2 | Pros | Cons |
| Approach 1(A)  -regression | Simple Model with no additional Complexity | Bootstrapped Hazard Rate can be rough or jagged. |
| Approach 1(B)  -regression + Smoothing Hazard Rates | Intuitive |  |
| Approach 1(C)  -regression + Smoothing CDS Proxy Curve | Intuitive | Some of the bootstrapped hazard rates can blow up and do not remain stable. |
| Approach 2 | Able to explain the CDS Proxy Curve by the 3 parameters described by Nelson Siegel function. | Does not fit well to approximate liquid CDS Spreads available in the market. |

Under the Regression Approach, we have decided to go with **Dataset 2, Method 2, Approach 1(B).**

# References

Bank of International Settlements. (2015). *Review of the Credit Valuation Adjustment (CVA) risk framework*. Retrieved from http://www.bis.org/bcbs/publ/d325.htm.

Chourdakis, K., Epperlein, E., Jeannin, M., & Mcewen, J. (2013). *A cross-section across CVA - Nomura.*

Diebold, F., & Li, C. (2005). Forecasting the term structure of government bond yields. *Journal of Econometrics*.

European Banking Authority. (2017). *Amending RTS on CVA proxy spread*. Retrieved from https://www.eba.europa.eu/regulation-and-policy/market-risk/amending-rts-on-cva-proxy-spread.

Sourabh, S., Hofer, M., & Kandhai, D. (2017). Liquidity risk in derivatives valuation: an improved credit proxy method. *Journal of Quantitative Finance*.

# Appendix

