DEGREE PROJECT IN ,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2017

The impact of macro-economic indicators on credit spreads

MAROUANE BRAHIMI

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ENGINEERING SCIENCES

The impact of macro-economic indicators on credit spreads

MAROUANE BRAHIMI

Degree Projects in Financial Mathematics (30 ECTS credits)
Degree Programme in Engineering Physics
KTH Royal Institute of Technology year 2017
Supervisor at Amundi Asset Management: Claudia Panseri
Supervisor at KTH: Boualem Djehiche
Examiner at KTH: Boualem Djehiche

TRITA-MAT-E 2017:18
ISRN-KTH/MAT/E--17/18--SE

Royal Institute of Technology
School of Engineering Sciences
KTH SCI
SE-100 44 Stockholm, Sweden
URL: www.kth.se/sci

Contents

Acknowledgements

First, I want to express my sincere gratitude to my tutor Claudia PANSERI for her valuable guidance, sharing her knowledge, her monitoring throughout the period of my internship, and by being always available for providing helpfull advice and trusting me with interesting projects.

I also want to thank the multi-asset engineering team, Jean-Gabriel Morineau and Edouard Van-Yen who provided me with valuable help and comments over my study, both theoretically and practically, but also with a pleasant team environment.

I thank also the Multi-Asset management team that welcomed me and included me as a valuable member of their team, enabling me to work around them, which made our collaboration really pleasant.

Finally I would also like to thank all the Ensimag and KTH professors, especially my tutors Olivier Gaudoin and Boualem Djehiche, not only for their disponibility and advice, but also for the knowledge acquired during these two academic years which revealed to be efficient throughout this internship.

Abstract

A model of credit spreads variations, based on macroeconomic and market variables, has been developed and presented in this paper. Credit spreads of speculative and investment grade bonds have been investigated, leading us to a linear relationship between their quarterly variations. Thanks to their risk contribution we clearly identify government bond rates and a financial conditions index as the most significant variables. Hence, based on macroeconomic views on the market in 2017, we perform some predictions on future variations on spreads based on this model, displaying the flattening of high yield credit spreads and the widening of investment grade spreads in the long run. In addition, a cointegration relationship between spreads, rates and the ISM has been found, meaning that there exists a mean-reverting process representing the spread between credit spreads and a linear combination of these factors. As a consequence, thanks to this process we can conclude about the potential immediate tightening of credit spreads.

Introduction

The purpose of this master thesis is to study the impact of several macroeconomic and market variables on the credit market and more specifically on credit spreads. To do so, we put in place a model of fair-value, to understand what is priced in the spread and figure out if it is overvalued or not. This should enable us to estimate, thanks to views on the market, credit spread changes, for Speculative as well as for Investment grade corporate bonds.

Since studies on fundamental macroeconomic drivers are at the heart of most investment processes, this study on the hypothetical relationship between spreads and economic drivers seems necessary to have ideas on the possible future scenarios or confirm our views. Therefore, we have worked on a model trying to explain credit spreads thanks to mostly key risk factors, growth concerns and some market indicators.

What we call credit spreads or corporate bond spreads, are the difference in yield between any type of bond and what is considered as the risk-free rate (10-Year German government bond in Europe and the US Treasury in the US). The spread indicates the extra premium investors require for the extra credit risk inherent in the corporate bond.

However, there is a common misconception that looking at credit spreads gives you a complete picture of the credit risk of one bond compared to another. Indeed there are other factors that combine with credit risk to make up the spread premium.

$$CreditSpreads = r_i - r_f$$

In this case r_i is is the bond yield and r_f is the risk-free rate. Therefore our objective here is to figure out what is included in these spreads in order to have a more accurate idea about its movement and be able to forecast it correctly.

Generally, credit spreads are studied by considering two different groups of bonds: Investment Grade and High Yield. These are split according to the grades rating agencies (Standard & Poors, Moody's and Fitch) grant them, since both groups often display different characteristics. Investment grade bonds are the ones with the higher grades, that are supposed to be more solvable, whereas the High Yield are bonds with speculative grades.

First a theoretical background of used theory and data is provided, then regression models outlining the impact of significant drivers are explained and presented, afterwards we will analyze the cointegration relation between values to get a mean-reverting process and finally we will discuss the different results obtained for the US Credit market.

1 Theoretical Background and Data

In this chapter, we give a thorough explanation of the theories used to get numerical results, that enable us to conclude about future variations of credit spreads. These theories and methods are often used to handle only stationary time series (for instance the Ordinary Least Squares estimation). As a consequence, we will present the meaning of each one of them and talk about measures of goodness of fit, principal component analysis, to finally consider theories that handle non-stationary processes.

Then, the data used is explained and presented. The meaning of each one of our macro-economic or market data is described as well as its provenance. All the data used is private, in order to have a more complete model, which prevents us from unveiling its content.

1.1 Time Series Analysis

In order to predict future credit spreads, it is necessary to base our model on previous data gathered for all indicators. So in this study, the data used are time series over approximately the last two decades.

Since we are willing to use linear models, and more specifically multiple linear regressions, it seems necessary to begin with studying stationarity for the processes we use. Before explaining the way we tested stationarity among our indicators we will present the concept of stationarity.

1.1.1 Stationary process

A stationary process, is a process which joint probability distribution does not change when shifted in time. Let, for instance, X_t be a stochastic process and $F_X(x_{t_1+r}, ..., x_{t_k+r})$ represent the cumulative distribution function of the joint distribution of X_t at times $t_1 + r, ..., t_k + r$.

Then X_t is said to be strongly stationary if, for all k, for all r, and for all $t_1, ..., t_k$:

$$F_X(x_{t_1+r}, ..., x_{t_k+r}) = F_X(x_{t_1}, ..., x_{t_k})$$

So F_X is not a function of time.

However here we can only consider a weaker form of stationarity, which is characterized by three points :

- $E[X_t] = \mu$ where μ is a constant
- $Var[X_t] = \sigma^2$, where σ is a constant
- $Cov(X_t, X_{t+r}) = h_r$, where h_r depend only on r

1.1.2 Autoregressive processes

What is usually called an *AR* process, is a time-varying process that presents a linear dependence with its own previous values. Mathematically an $AR(p)$ process is defined as :

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t$$

Where the $\phi_1, \ldots, \phi_p$ are the parameters of the model, c a constant and ϵ_t is white noise. Generally in our study we will only consider $AR(0)$ and $AR(1)$ processes.

1.1.3 Augmented Dickey-Fuller Test

The test chosen here to identify a non-stationary process is the *ADF-Test*. Its purpose is to determine if a unit root is present in a time series sample, which represents the null hypothesis of this test. The alternative hypothesis is different depending on which version of the test is used, but here it will be stationarity. The test used is the one computed in **R** through the package *tseries*. The testing process considers the following :

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_{p-1} \Delta y_{t-p+1} + \epsilon_t$$

In our case we consider α as a constant, β equal to 0 and p the lag order of the auto-regressive process. So the hypothesis are as follows :

- $H_0 : \gamma = 0$
- $H_1 : \gamma < 0$

1.2 Linear Regression

To predict credit spreads we have chosen to find a linear relationship between the indicators selected before and credit spreads. Linear regression is a method that assesses whether some variables have a significant impact on the dependent variable.

There exists multiple methods for estimating the unknown parameters in a linear regression model. In this study we chose the most used one which is the Ordinary Least squares method. Its aim is to minimize the sum of squares of the differences between observed responses and the one predicted by the model.

1.2.1 Ordinary least squares estimation

This estimator is known to be consistent when regressors are exogenous and errors are homoscedastic and serially uncorrelated. Under these circumstances the method provides *minimum-variance mean-unbiased* estimation when the errors have finite variances. Moreover, if the errors are normally distributed, then the OLS method is also the maximum likelihood estimator. Therefore we can identify five main assumptions :

- Multivariate normality : It requires that all variables must be multivariate normal.
- No or little multicollinearity : It suggests that predictors are not dependent from each other.
- No auto-correlation : Meaning that the errors are uncorrelated between observations, $E[\epsilon_{t_i}\epsilon_{t_j}|X] = 0$ for $i \neq j$
- Homoscedasticity : Consists in the fact that standard deviations of the error terms are constant and do not depend on indicators.
- Normality : Additionally it assumes that errors have normal distribution conditional on the regressors, $\epsilon|X = N(0, \sigma^2 I_n)$

The dependent variable, y is assumed to be related to the explanatory variables, $x_1, \ldots, x_n$ through the equation:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \epsilon_t$$

The residual error ϵ represents the other factors that aren't captured by explanatory variables and still affect y. It is assumed to be uncorrelated with the explanatory variables and have zero mean:

$$E[\epsilon] = E[\epsilon|x_1, \ldots, x_n] = 0$$

Thus, we choose the coefficients $\beta = (\beta_0, \ldots, \beta_n)$ that minimize the residual sum of squares, knowing that the equation has a unique solution. *OLS* is an unbiased estimation.

1.2.2 Goodness of fit

We present here all the tests used to verify the righteousness of our model and how good they fit real values.

Coefficient of determination R^2 : If we denote by y_t^* the fitted values from regression analysis, the coefficient of determination is defined as $R^2 = Cor^2(y_t^*, y_t)$. We can show that $R^2 = \frac{Var(y_t) - Var(\epsilon_t)}{Var(y_t)}$, where ϵ_t are the residuals and the numerator can be identified as the explained variance.

F-Test : We assume a parametric model for the residuals in order to perform statistical hypothesis testing. The most standard approach assumes the following: $\epsilon \sim N(0, \sigma^2)$. The F-Test tests if a group of variables significantly improve the fit of a regression. Formally, the hypothesis are :

- $H_0 : (\forall j \geq 1, \beta_j = 0)$
- $H_1 : (\exists j \geq 1, \beta_j \neq 0)$

t-tests : This tests if a particular coefficient is null, meaning that it has no impact on the variable we try to explain :

- $H_0 : \beta_j = 0$
- $H_1 : \beta_j \neq 0$

1.2.3 Variable Selection

It consists on selecting a subset of relevant variables among all variables, without deteriorating much the goodness of our model. This is motivated mainly for *prediction accuracy*, since the *OLS* estimates have often low bias but large variance, especially when the number of variables is big. But also for a better *interpretation*.

Therefore when p, the number of predictors, is large, the model can accommodate complex regression functions (small bias). However, since there are many β parameters to estimate, the variance of the estimates may be large. For small values of p, the variance of the estimates gets smaller but the bias increases. As a consequence, an optimal value to find is the one that reaches the so-called *bias-variance trade-off*. There are various methods to find this optimal value (cross-validation, BIC or AIC criterion, ...) and in our case we chose to consider the *AIC* criterion :

Akaike information criterion (AIC) Model selection by AIC picks the model that minimizes

$$AIC = n \log \frac{RSS}{n} + 2p$$

where RSS is the residual sum of squares, n the size of the data and p the number of predictors.

1.3 Model Validation

Once we get our result, we must test it and validate it through multiple tests, to confirm that our result is conform with the OLS theory's hypothesis used in the study. In our case we will focus on the possible collinearity between indicators and then the residual errors to see if they respect the assumptions stated earlier.

1.3.1 Collinearity

Excessive collinearity among explanatory variables can prevent the identification of explanatory variables for a model. Therefore, once we get our model, we must make sure that the cross-correlation among those factors is weak. To test it, we use what is called the *VIF* selection.

VIF : The Variance Inflation Factors, are obtained using the regression's coefficients of determination (r-squared) of that variable against all other explanatory variables through the following formula :

$$VIF_j = \frac{1}{1 - R_j^2}$$

1.3.2 Residuals

The important concern to check when using the OLS method, and a way to verify our results is Homoscedasticity, auto-correlation and normality of residual errors. Indeed, as it was previously indicated, residual errors are considered not to be auto-correlated and follow a normal distribution. To assess this statement we use several tests :

Durbin-Watson test : It is a specific test to detect auto-correlation of order 1 between residual errors, by considering the following :

$$\hat{\epsilon}_t = \rho \hat{\epsilon}_{t-1} + \mu_t$$

Where $\hat{\epsilon}_t$ are the residuals estimated. The null hypothesis of this test is that there is no auto-correlation which is represented by : $H_0 : \rho = 0$.

Quantile-Quantile plot : This probability plot is a graphical method for comparing the distribution of two samples by plotting their quantiles against each other. Therefore in our case we chose to plot the residuals' quantiles against a normally distributed variable to see if they can be considered as normally distributed.

Histogram : This is another graphical representation of the distribution of some numerical data. It lets us have an estimation of the probability distribution, by dividing the entire range of values by series of intervals and plot the number of values in each intervals. Then we can compare it with a normal distribution.

1.4 Cointegration

This being said, once we have identified that our variables aren't stationary, and are moreover integrated of order 1 ($I(1)$) processes, there are other alternatives to study the movement of the variable. We chose here to focus on cointegration theory.

Let $X_t = (x_{1,t}, ..., x_{n,t})$ denote a vector integrated of order 1. X_t is said to be co-integrated if there exists a vector $\beta = (\beta_1, \ldots, \beta_n)$ such that :

$$\beta' X_t = \beta_1 x_{1,t} + \ldots + \beta_n x_{n,t} \sim I(0)$$

So X_t which is not stationary is cointegrated if there exists a linear combination of its components that is stationary. In an economic point of view this relation is often considered as a long-run equilibrium relationship. The fact is we consider generally that $I(1)$ variables cannot drift too far apart from a certain equilibrium.

1.5 Risk Contribution

Generally, risk contributions are considered to explain the different risks caused by several assets composing a certain portfolio. However here we chose this theory to display which variables explain most of the spreads' variation.

Therefore if we consider $\beta_1, \ldots, \beta_n$ to be the coefficients of our regression, $x_1, \ldots, x_n$ the variables associated to these coefficients (which are time series), and finally $R(\beta_1, \ldots, \beta_n)$ a risk measure. We know that if the risk measure is coherent and convex it verifies the following :

$$R(\beta_1, \ldots, \beta_n) = \sum_{i=1}^{n} \beta_i \frac{\partial R(\beta_1, \ldots, \beta_n)}{\partial \beta_i}$$

That is why, risk contribution of a certain factor is defined as :

$$RC_i(\beta_1, \ldots, \beta_n) = \beta_i \frac{\partial R(\beta_1, \ldots, \beta_n)}{\partial \beta_i}$$

In this study we chose the volatility as our risk measure, that satisfies the previous conditions and allows us to give a precise definition of risk contribution as follows :

$$RC_i(\beta_1, \ldots, \beta_n) = \beta_i \frac{(\Sigma\beta)_i}{\sqrt{\beta^T \Sigma \beta}}$$

Where Σ represents the variance/covariance matrix of our different indicators.

1.6 Data

1.6.1 Macro-economic Predictors

As it was said earlier, the main objective here is to be able to have an idea, based on the team's forecasts, of the value of the credit spreads. And to do it we want to focus mainly on fundamental data and limit the presence of market data. So here are the factors that, according to our analysis, drive credit spreads valuation in the US market:

- Gross Domestic Product
- Earnings per share
- Inflation
- Lending Standards
- Financial Conditions Indicator
- Price to Earnings
- Free Cash Flow

Based on their graphs and histograms, we chose to transform the data by taking the logarithm if it was possible (if the data is skewed, and positive).

GDP : It is a monetary measure of the market value of all final goods and services produced in a period (quarterly). It is commonly used to determine the economic performance of a region, which explains why we judged it as a potential good estimator.

Earnings per Share : Represents the portion of a company's profit allocated to each outstanding share of a stock. Here we take the mean of our parent index's ($S\&P500$) earnings per shares. It can be considered as an indicator of companies' health that could influence credit spreads.

Lending Standards : It is a bank lending survey concerning their policies set in place and requirements for potential borrowers. It indicates if companies can leverage more or less easily.

CPI : The Consumer Price Index measures changes in the price level of market basket of consumer goods and services purchased by households. Its variations is a measure of US inflation.

Financial Conditions : To have an estimator of financial conditions we use an index that gauges overall economic activity and related inflationary pressure. It is based on variables that represent leverage, interest coverage and cash flow. These quantities play a key role in firm-value models of credit risk and the ratings methodologies of the major ratings agencies.

PE : Price to Earnings ratio measures the ratio between the current share price to its per-share earnings. Once again we take the mean value over our index. It gathers a certain number of other macro-economic indicators that don't include the ones we selected here.

FCF : The Free Cash Flow is a way of measuring a company's financial performance by calculating its operating cash flow minus capital expenditures. Its importance is due to the fact that it is the money allowing companies to enhance shares value.

1.6.2 Market Variables

However, to capture movements caused by the market, or liquidity, we chose to integrate also some market data to complete the model and have a better explanation of the variance. Therefore we selected a number of them, keeping in mind that we have to minimize the number of market indicators.

- VIX
- 10-Year Treasury Yield
- S&P 500 : Equity index
- Yield Curve
- 2-Year yield (Short term rate)

VIX : It is a volatility index computed as the implied volatility of S&P 500 index options. Commonly used as a gauge of the market's expectation of stock market volatility, and can even be considered as a proxy of liquidity in the market. That's why it seemed necessary to use it in this study.

US Treasury Yield : We use the US government bond of maturity 10 years as a benchmark of the default-free curve in our model.

S&P 500 : To represent variations of equities, we chose to consider the main index as a proxy of their variations. It seems obvious that we should have a strong correlation with credit spreads, since it often concerns similar companies.

Yield Curve : It has been constructed quite simply by taking the 10-Year rate minus the 2-Year government bond rate, in order to capture the slope of the yield curve.

Short Term Rate : In order to capture monetary policy, we integrated also the US 2-Year government bond's yield.

Having selected those market indicators, we have to focus only on the more significant ones and try to limit their number. We will see later that we have performed a PCA to do so.

2 Regression of Credit Spreads on macroeconomic factors

Once we have selected those indicators, the next step is to analyze the data that we gathered, transform it, and then progress on our study by performing our regression. As the European credit market is relatively recent we preferred to focus only on the US market, where we have long historical data.

We adopt a linear regression framework to explore credit spreads dynamics and assess the existence of relationships between markets. Therefore, corporate bonds are considered by groups of investment grade bonds and speculative ones for the US market. From a no-arbitrage standpoint, credit spreads can be justified through two fundamental reasons:

- Risk of default
- The portion of payment received in the case of default

As consequence, we chose to examine how changes in credit spreads react to proxies for these two causes. After screening various possible factors, first we analyze them, to see if they are eligible for this study, but also if we can use them directly or transform the data (stationary and normal or not), then we perform our regression, at an index level, on the data, and try to get the most significant variables and most stable model through various back-tests. Finally, thanks to the co-integration theory, we try to find some *mean-reverting* process as a model of fair-value concerning credit spreads. Therefore the different steps followed in our study are:

- Study the correlation between different variables (macro-economic variables and spreads)
- Test the stationarity of all the variables (ADF test)
- Transform the data, to get conform data for linear regressions
- Select the optimal number of meaningful variables thanks to the AIC criterion

We performed most of this analysis using the software **R** gathering all the data from DataStream Reuters. Then we confirmed each and every result gotten using other software such as EViews or Excel. Finally, we put in place a tool in VBA - Excel allowing us to perform the regression and get the result easily for our estimation.

2.1 Stationary processes

As discussed earlier, before setting any interpretation we must be careful about the data we use, especially when we are considering regression analysis. Indeed using non stationary time series in our regression analysis could easily lead to what we call a spurious regression. When we use non-stationary variables, *OLS* properties don't stand any more. Hence, we could obtain for instance an over estimated R^2 due to the fact that both variables follow a certain trend but doesn't concern the relationship between both variables.

That is why it is fundamental to begin by investigating stationarity among all of our explanatory variables. As a consequence, we applied the Augmented Dickey-Fuller Test to each variable to check if it had a unit root and therefore, wasn't stationary. We summarized the results on the table below:

Level	Log(HY)	Log(IG)	Log(GDP)	EPS	Lendg Stds	CPI	Fin Cond	P/E	FCF
P-Value	0.0103	0.0080	0.492	0.82	0.006	0.0657	0.024	0.14	0.47
T-stat	-3.48	-3.56	-1.57	-0.7745	-3.68	-2.77	-3.18	-2.42	1.62
1%	-3.49	-3.49	-3.49	-3.49	-3.495	-3.49	-3.49	3.49	3.49

Table 1: ADF Test Results

As it can be seen from the table above, most of our variables can't be considered as stationary processes according to this test with a threshold of 1%. With t-statistics bigger than the reference (-3.49), it lets us reject the null hypothesis and prevents us from concluding about the stationarity of our processes.

Therefore, to solve this problem we consider quarterly changes of underlying variables to get stationary data. Since we consider that most of them aren't stationary they all can be represented as follows :

$$\Delta y_t = \alpha + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_{p-1} \Delta y_{t-p+1} + \epsilon_t$$

Then, if we differentiate our time series we could hope to get, this time, stationary processes. Below we have a table outlining results of the ADF test for differenced data.

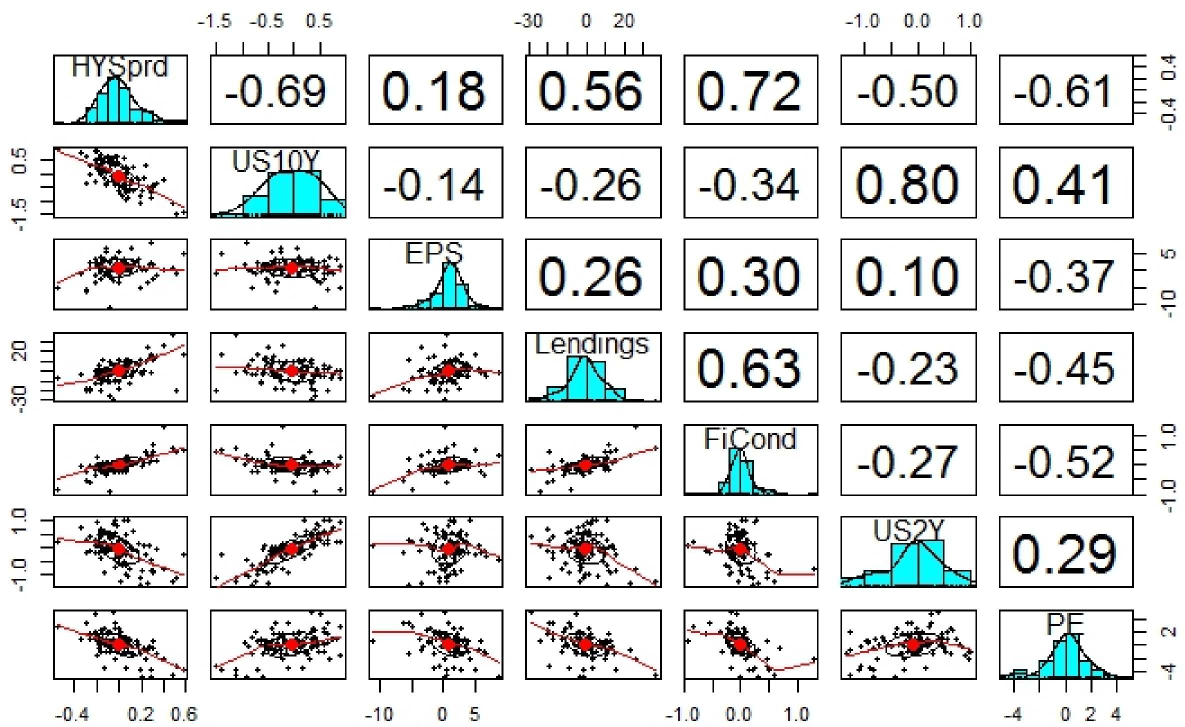
Level	Log(HY)	Log(GDP)	EPS	Lendg Stds	CPI	Fin Cond	P/E	FCF
P-Value	0	0	0	0	0.0657	0.012	0	0
T-stat	-7.77	-4.48	-5	-9.47	-3.42	-7.63	-6.59	-9.32
1%	-3.49	-3.49	-3.49	-3.495	-3.49	-3.49	3.49	3.49

Table 2: ADF Test Results for differenced data

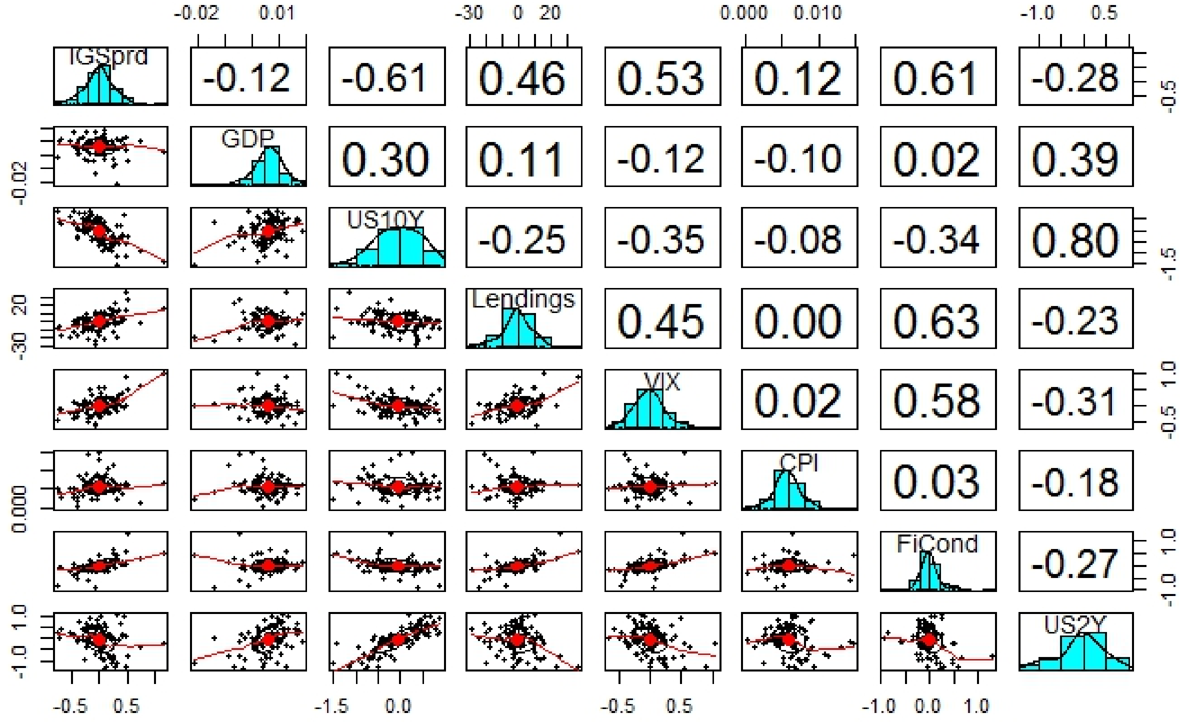
This time, we can see that once the variables have been differentiated, we can consider that we use stationary processes according to the *ADF* test. It encourages us even more to use quarterly changes.

2.2 Correlations

We still have one last step before performing our regression, which is trying to motivate this study by taking a look at correlations between different variables. Here we will have a first look at which variables seem to impact more or less credit spreads. This lets us have an idea on the cross-correlation between explanatory variables and if they are positively or negatively correlated.



High Yield Correlations



Investment Grade Correlations

In the first row of the above matrix, we observe the correlation coefficients between the explanatory variables and credit spreads. This shows us that there are some variables that are strongly correlated with credit spreads, such as : Treasury yield, lending standards, financial conditions or even price to earnings. Therefore it confirms the possible impact of our variables on credit spreads and comforts us in our will to conduct this study. Indeed, except for the VIX and financial conditions, we don't see much cross-correlated data.

2.3 PCA for market variables

In order to choose rigorously which market variables are relevant for our study, we performed a principal component analysis to see which variable explains most of the variance and therefore should be chosen for our model.

Gathering all market variables that have been selected for this study, we perform a *PCA* on them and try to select as few as possible variables that explain most of the variance. Here are the results :

	PC1	**PC2**	**PC3**	**PC4**
Standard Deviation	1.612	0.999	0.612	0.168
Proportion of Variance	0.650	0.249	0.092	0.007
Cumulative Proportion	0.650	0.899	0.993	1.000

Table 3: Variance explanation by axes

According to the table above we judge that enough Variance is explained through the two first axis leading us to only consider both of them since they explain combined 90% of the total variance.

	PC1	PC2	PC3	PC4
S&P 500	0.411	0.113	0.794	0.362
US 10Y	-0.652	-0.027	0.162	0.724
VIX	0.072	-0.993	0.092	0.003
US 2Y	-0.627	0.010	0.578	-0.527

Table 4: Composition of axes

Then on the second table we can easily identify the first component to both rates (US 10Y and US 2Y) and the second one to the VIX since it is clearly its main component here. However we can also notice that the equity index is not negligible in the first component but can be ignored in this study.

2.4 Regression model : High Yield

Once we have our stationary processes, and have selected some market variables, we can try to fit a linear model to our time series to estimate credit spreads. Here we will analyze the results we have for High Yield and Investment Grade bonds.

Regressing on all of our variables we get the following results :

	Coefficients	Standard Error	t-stat	p-Value
Intercept	0,001	0,0261	0,046	0,963
GDP	-2,212	1,693	-1,306	0,195
US10Y	-7,402	10,578	-0,700	0,486
EPS	-0,009	0,004	**-2,141**	0,035
Lending Stds	0,003	0,001	**2,493**	0,014
VIX	0,006	0,045	0,138	0,890
CPI	1,561	3,840	0,406	0,685
Fin Cond	0,286	0,048	**5,999**	0.000
US2Y	7,264	10,578	0,687	0,494
Yield Curve	7,182	10,574	0,679	0,499
PE	-0,017	0,007	**-2,312**	0,023
FCF	0,088	0,087	1,012	0,314

Table 5: HY Regression result against all variables

From the table above we can see that all the variables aren't meaningful in our model, since observing all the t-statistics, most of them are very low, and we agreed to put as a threshold a value of 2. In this case, we only have 4 variables that we can interpret since others have too low t-stats to allow us to get any conclusion.

First concerning Earnings Per Share, and Price to Earnings we have a significantly negative coefficient, which is coherent since a positive variation of earnings is a good sign for the markets pushing credit spreads to tighten. At the opposite lending standards, and financial conditions have a positive positive impact on spreads.

2.5 Regression Model : Investment Grade

Here we regress investment grade credit spreads against all factors and get the following

	Coefficients	Standard Error	t-stat	p-Value
Intercept	-0,071	0,047	-1,498	0,137
GDP	-4,719	3,073	-1,536	0,128
US10Y	-19,385	19,248	-1,007	0,316
EPS	-0,004	0,007	-0,484	0,630
Lending Stds	0,004	0,002	**1,990**	0,049
VIX	0,155	0,081	1,912	0,059
CPI	17,014	6.983	**2,437**	0,017
FiCond	0,322	0,087	**3,719**	0,000
US2Y	19,235	19,248	0.999	0,320
Yield Curve	18,59	19,241	0,980	0,330
PE	-0,002	0,013	-0,137	0,892
FCF	0,002	0,159	0,013	0,990

Table 6: IG Regression result against all variables

If we look at the most significant variables here, we get the same variables that were strongly correlated with High Yield credit spreads. However, once again we have very few significant variables, which not only prevents us from interpreting most of our coefficients, but it also compromises the results concerning coefficients that seem significant. That is why we have to find a way to get rid of other variables. So here we detected mainly two problems :

- A prediction accuracy problem : Since the least squares estimates have low bias but large variance. Prediction accuracy can be improved by variable selection, consisting in choosing a sub-sample of the initial set of predictive variables.
- An interpretation problem : With a large number of predictors, it might be better to sacrifice small details in order to get a more synthetic model with a smaller subset of variables that exhibit the strongest effects

2.6 Variable Selection

In order to obtain a relevant model composed of the most significant indicators, we use the AIC criterion. By trying out multiple models with our variables we select the one that minimizes the AIC criterion on the one hand and that have significant variables on the other hand (which have a *t-stat* bigger than 2, limit fixed with the managers).

This should achieve a trade-off between model fit (represented by the first term of the AIC criterion) and model complexity (represented by the 2^{nd} term). This is a way to avoid over-fitting the model on our training set, which is not what is needed since it must help asset managers judge future moves of spreads. In the table below we summarize the best results concerning the AIC criterion for each value of p (number of variables).

High Yield		Investment Grade	
p	AIC	p	AIC
5		5	-372.55
6	-499.89	6	-371.69
7	-498.01	7	-371.18
8	-495.56	8	-368.47
9	-492.25	9	-364.9
10	-488.59	10	-361.09
11	-484.77	11	-357.24

Table 7: AIC values

In this table we clearly identify both models that we'll be interested in, since they are minimizing the AIC criterion and consider only significant variables. As a result, we consider a 6 variables model for the High Yield, and 5 for the Investment grade.

Once we have our simplified model, we try to regress once again the selected variables only this time against the corresponding credit spreads and get the following.

2.6.1 High Yield

Concerning High Yield credit spreads, we get a relatively good model, with a coefficient of determination equal to 0.790, considered to be significant. Here we present the detailed results :

Regression Statistics	
R Square	0.79
Adjusted R Square	0,78
Standard Error	0,09
Observations	106

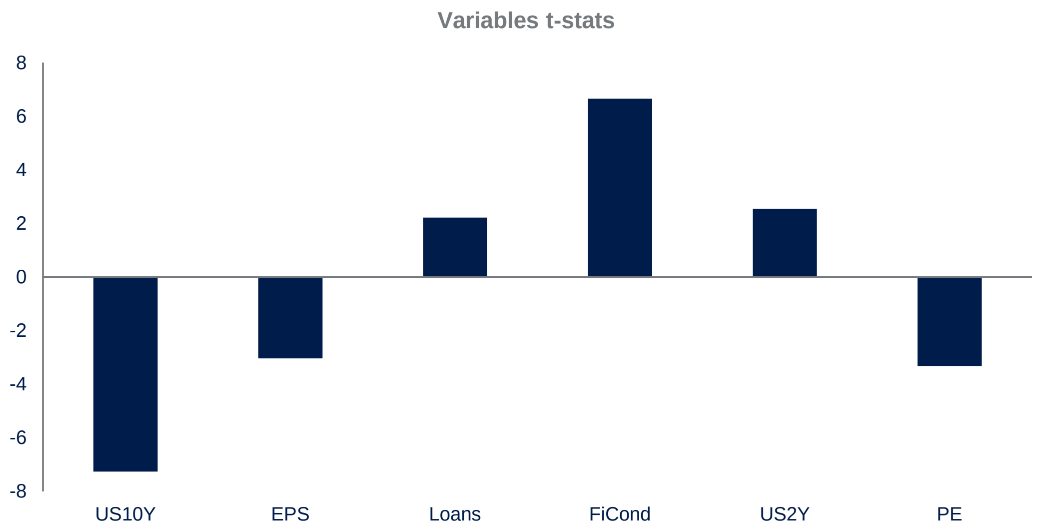
Table 8: Regression summary

	Coefficients	**Standard Error**	**t-stat**	**p-Value**
Intercept	0,001	0,009	0,145	0,885
US10Y	-0,221	0,030	-7,290	0.000
EPS	-0,011	0,004	-3,064	0,003
Lending Stds	0,002	0,001	2,255	0,026
FiCond	0,294	0,044	6,689	0.000
US2Y	0,073	0,028	2,583	0,011
PE	-0,020	0,006	-3,346	0,001

Table 9: Synthetic High Yield Model

This represents the final model that we chose to explain High Yield credit spreads changes, based on fewer indicators as we wanted and giving satisfying results. Let us get a closer look now at the result and coefficients obtained in this study for the high yield credit spreads.

First to get the bigger picture on the impact of the different indicators, the model implies that high yield credit spreads' quarterly changes are mainly functions of long and short term US rates, but also market conditions determining market stress and corporate metrics indicating the global health of companies.



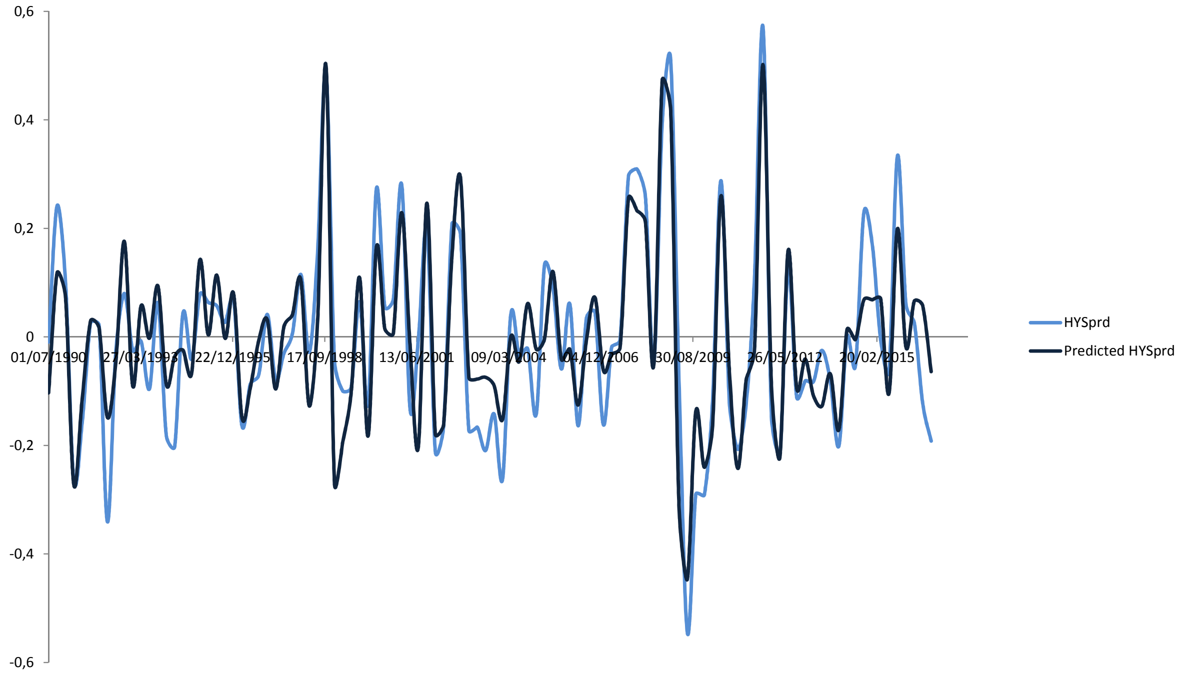
- Treasury Yield : Since credit spreads are defined as credit yields minus the risk-free rate, it is then obvious that credit spreads and the treasury yield should move on the opposite direction since a lower risk-free rate implies a widening of credit spreads

- For any company, earnings growth testifies of the business corporate health. Therefore earnings growth triggers the tightening of credit spreads since investing in them is safer.

- Since lending standards are often used as a strong leading indicator of corporate default rates, it is supposed to be an important explaining factor for High Yield credit. The higher the number, more it is difficult to borrow for companies, therefore credit spreads widen

- The indicator of financial conditions, as it was described before, is positive when financial conditions are bad and negative otherwise. Then we expect spreads to tighten when the indicator decreases, which is exactly what we see, with a positive relation.

- Finally concerning Price over Earnings ratio, it shows that corporates are economically healthy when it raises, which explains why when it increases credit spreads tightens.



High Yield Spread against predicted values

2.6.2 Investment Grade

Now we regress investment grade credit spreads against factors that minimized the AIC criterion previously to get a model with a lower coefficient of determination 0.73. This is still satisfactory according to our team, especially since it matches quite well historical variations. Here we have the detailed results:

Regression Statistics	
R Square	0.726
Adjusted R Square	0,706
Standard Error	0,156
Observations	107

Table 10: Regression summary

	Coefficients	**Standard Error**	**t-stat**	**p-Value**
Intercept	-0,075	0,046	-1,635	0,105
GDP	-4,967	2,776	-1,789	0,077
US10Y	-0,513	0,051	-10.10	0.000
Lending Stds	0,004	0.002	2.039	0,044
VIX	0,172	0,070	2,466	0,015
FiCond	0,313	0,083	3,783	0,000
CPI	17,42	6,686	2,605	0,011
US2Y	0.364	0,050	7,312	0.000

Table 11: Synthetic Investment Grade Model

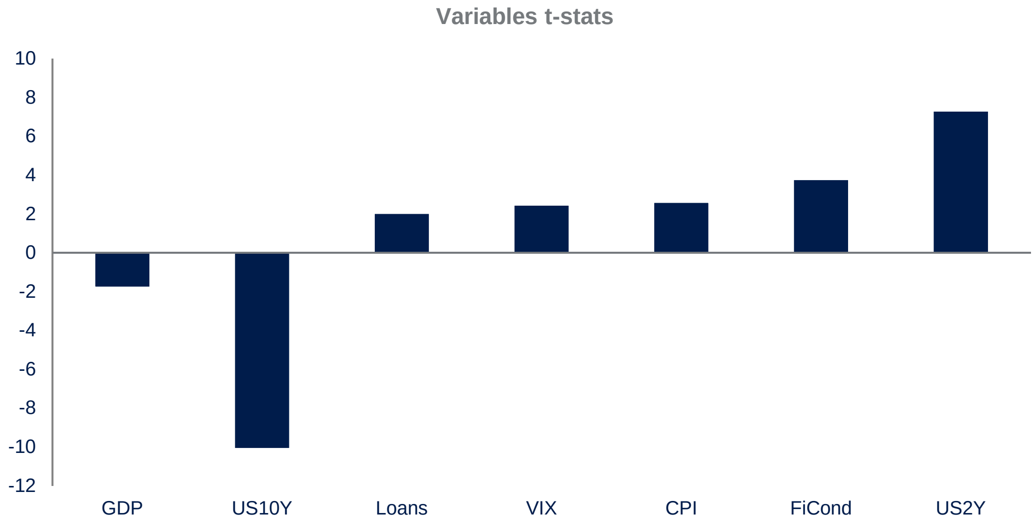
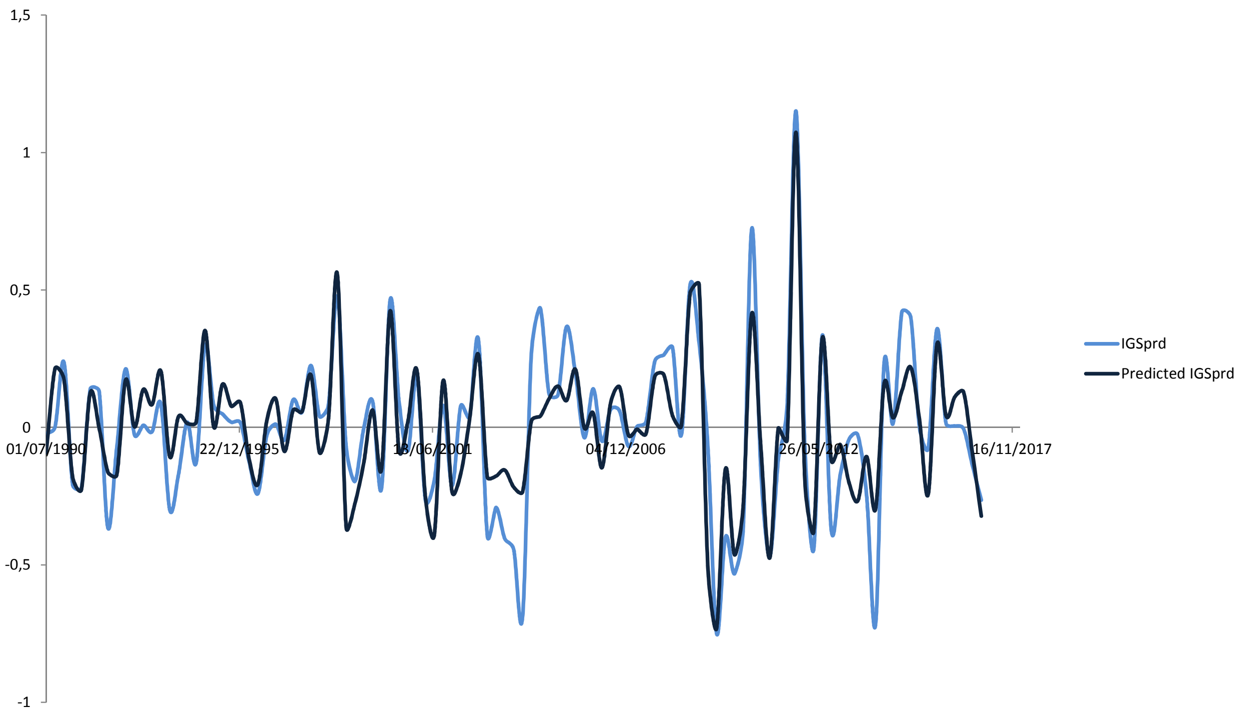


Figure 1: Investment Grade Spread against predicted values

As we can see here both models have common explaining indicators such as, government bond yields, Lending Standards or Financial Conditions, which is completely logic since they are all (except Financial Conditions) directly linked to credit spreads or corporates.

We won't explain once again the factors already significant for the high yield spreads. However, concerning IG spreads, we have more market variables:

- The Gross Domestic Product is significant here, and negatively related to spreads which can be explained by the fact that a higher GDP is a good sign for the economy and and so for corporates.
- A significant portion of investment grade credit spreads should reflect risk premia, which we proxied with the market pricing of future volatility (VIX). Furthermore it should capture large credit spikes that occurred during the crisis.



2.6.3 Confidence Intervals

To confirm our interpretations about factors signs, we display here all the confidence intervals for both models, synthesized in the following table:

High Yield		Investment Grade	
Factor	Interval	Factor	Interval
10-Year Yield	[-0,28 ; -0,16]	GDP	[-10.48 ; -0.001]
EPS	[-0,02 ; 0,00]	10-Year Yield	[-0.61 ; -0.41]
Lending Stds	[0,00 ; 0,01]	Lending Stds	[0.00 ; 0.01]
Fin Conditions	[0,21 ; 0,38]	VIX	[0.03 ; 0.31]
2-Year Yield	[0.02 ; 0.013]	CPI	[4.15 ; 30.68]
PE	[-0,03 ; -0,01]	Fin Conditions	[0.15 ; 048]
		2-Year Yield	[0.27 ; 0.46]

Table 12: Confidence intervals

Those intervals validate our interpretations since all of these coefficients have confidence intervals with extremes having the same sign. Therefore it comforts us in giving conclusions about relationships concerning an indicator with credit spreads.

2.7 Sensitivities to indicators

Having multiple explanatory variables in both of our models, we want to estimate which one of them are necessary to explaining spreads changes and must be carefully studied. Therefore, it will enable us to establish credit spreads' sensitivities to all the variables selected, by ranking them. To do so, based on their time series, we compute their risk contributions, using their volatilities and then add their correlations.

High Yield		Investment Grade	
Factor	Risk Contribution	Factor	Risk Contribution
10-Year Yield	0.087	10-Year Yield	0.189
Fin Conditions	0.063	2-Year Yield	0.065
PE	0.025	Fin Conditions	0.059
2-Year Yield	0.022	VIX	0.029
Lending Stds	0.016	Lending Stds	0.022
EPS	0.006	CPI	0.006
		GDP	0.004

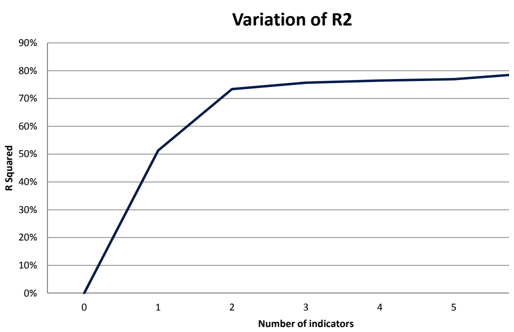
Table 13: Risk Contributions

According to these risk contributions it seems that rates are the principal indicators that influence credit spreads along with the financial conditions index. Credit spreads are more sensitive to risk free rates, short-term rates (especially for Investment Grade spreads which is logic since it can be seen as a proxy of government bonds) and to financial conditions, all of which must be forecasted precisely to get a correct prediction of spreads.

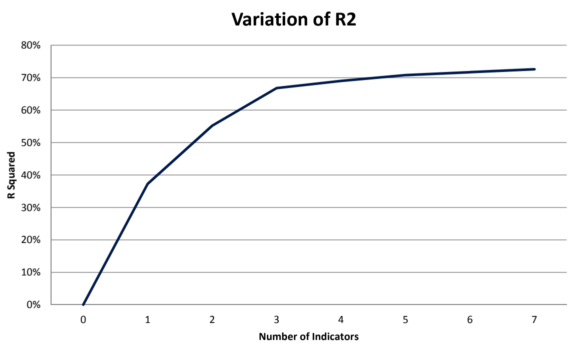
Another way to analyze the importance of indicators is to run a forward selection of variables, by maximizing the coefficient of determination. Here are the results for both models:

High Yield		Investment Grade	
Factor	Cumulated R^2	Factor	Cumulated R^2
Fin Conditions	51.3%	Fin Conditions	37.3%
10-Year Yield	73.4%	10-Year Yield	55.2%
PE	75.7%	2-Year Yield	66.8%
EPS	76.4%	VIX	69.0%
2-Year Yield	77.0%	CPI	70.8%
Lending Stds	79.0%	Lending Stds	71.7%
		GDP	72.6%

Table 14: Forward R Squared



HY R-Squared evolution



IG R-Squared evolution

This confirms our previous conclusions, about the fact that investment grade spreads, are mostly explained by rates and the financial conditions, whereas the high yield is essentially driven by the same factors except short term rate. Moreover both of them already give a sufficient coefficient for our model.

3 Model Validation

Once we have our model with all the significant variables, their coefficients and the equation, we must go further to validate it. Indeed, in this section we will focus on the model validation, by challenging different aspects of our regression to see if it is trustworthy or not. So first we will proceed at an out-of-the-sample cross-validation and then we will try to verify that factors and residuals are conform to hypothesis fixed in the *OLS* theory.

3.1 Cross-Validation

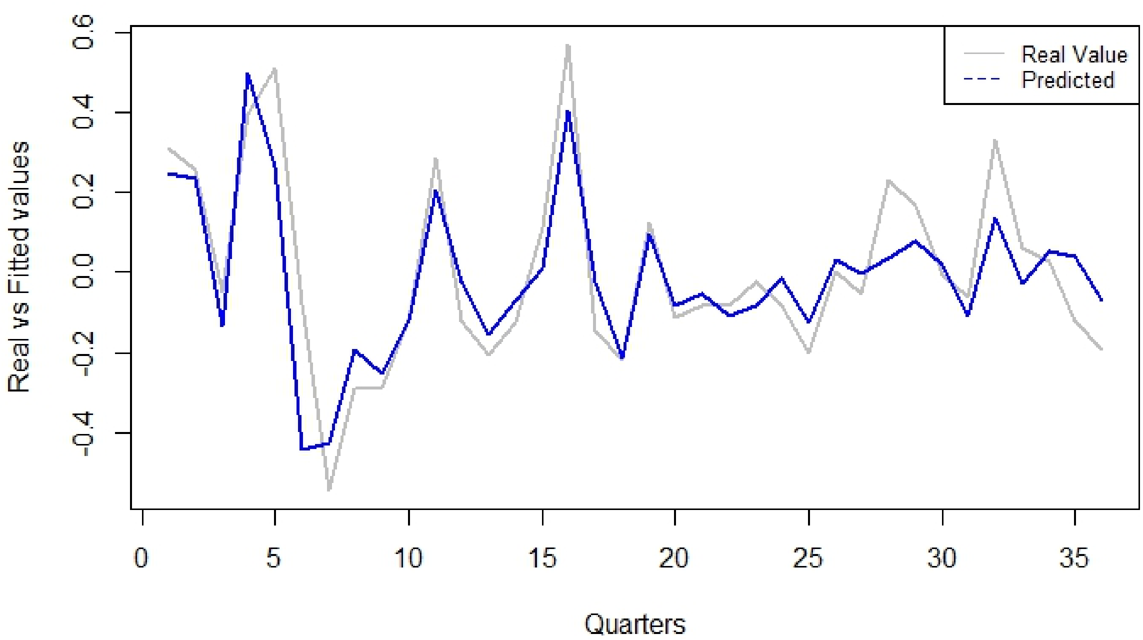
If we are in a data-rich situation, a standard approach to have an idea about the prediction error is to divide the data set into two parts:

- A training set with n_1 points
- A validation set with n_2 points

As a consequence, we consider 70% of the sample as the training set that we will use to calibrate the model and fit it to provide estimates of the regression coefficients. Then we will test its robustness by plotting the result against real values from the validation set.

3.1.1 High Yield

First concerning the High Yield data, as we said it before, we fit the model on 70% of our data set and get a model, which result is tested against the validation set. As a result we get the following function for the regression :



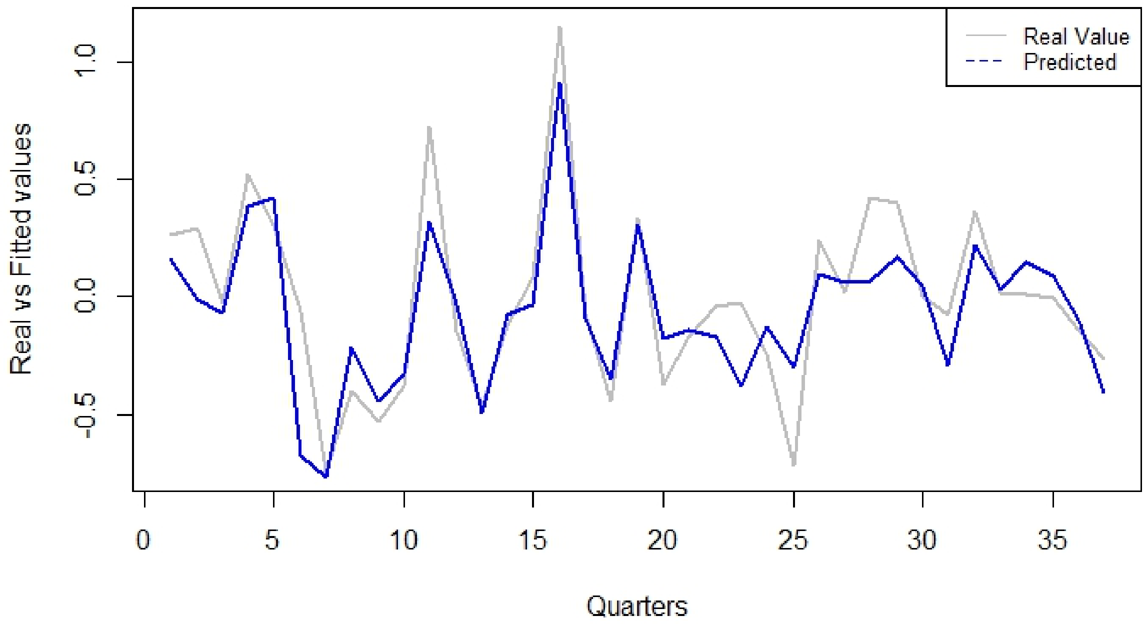
HY

on the validation set

We can see that both curves have the same trend during the whole period. This leads us to assume that the model has good results concerning the trend of credit spread changes. However we can see that even if the the orientation (increase or decrease) is satisfying, the magnitude can be considered as insufficient concerning some periods.

3.1.2 Investment Grade

We apply the same methodology, but this time to the investment grade credit spreads:



IG on the validation set

Once again the result is satisfying with both curves varying the same way for the whole validation set, suggesting that the model is strong and could deliver good previsions concerning future values of credit spreads changes. However one must keep in mind that the model can be very sensitive to the input data (in our case forecasts to get previsions) and therefore gives good results if we use convenient previsions for our indicators.

3.2 Collinearity

The following assessment made here is related to the possible multicollinearity in the data, since linear regression theory assumes that there is no multicollinearity between explanatory variables. Therefore we use the *VIF* selection test in order to determine if we could consider that there are no multicollinearity.

3.2.1 High Yield

Here we consider only the indicators that have been selected thanks to the *AIC* criterion and run the test on them, we get the following :

VIF Selection High Yield		
Variables (in variation)	VIF	$\sqrt{VIF} < 2$
10Y Rate	1.294	False
EPS	1.372	False
Lending Stds	1.738	False
Fin Conditions	1.915	False
Yield Curve	1.239	False
PE	1.64	False

Table 15: VIF HY

We chose to consider as a threshold here the value of 2. If the value given by the square root of the *VIF* test exceeds 2 we consider it to be linearly linked to other variables. In this case, as it is shown in the table above, we can say that variables aren't correlated since all of them have a coefficient lower than 2.

3.2.2 Investment Grade

Now we consider the data explaining investment grade credit spreads, and run our test on these variables.

VIF Selection Investment Grade		
Variables (in variation)	VIF	$\sqrt{VIF} < 2$
GDP	1.259	False
10Y Rate	1.466	False
Lending Stds	1.745	False
VIX	1.627	False
CPI	1.049	False
Fin Conditions	2.090	False
Yield Curve	1.204	False

Table 16: VIF IG

Once again we can conclude that there is no or few cross-correlation between these variables according to our criterion. However we can see that this time we have a coefficient (concerning Financial Conditions) that is slightly bigger than usual, but it still is conform since its square root is significantly lower than 2.

3.3 Residual errors

Once we have set those preliminary checks of the model and got encouraging results, we pursue our model validation, by verifying the *OLS* assumptions. To do so, we begin by studying the normality of residuals, then if the residuals are part of an auto-correlated process.

Residuals here correspond to the difference between the observed value and the estimated value of the quantity of interest:

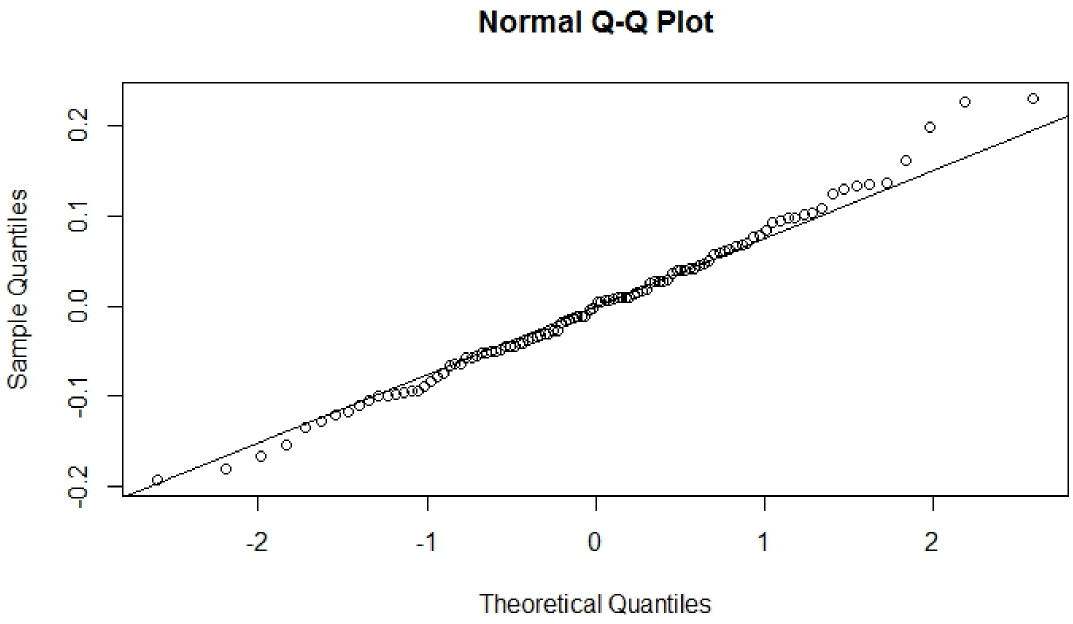
$$\epsilon_t = y_t - (\beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t})$$

3.3.1 Normality of residuals

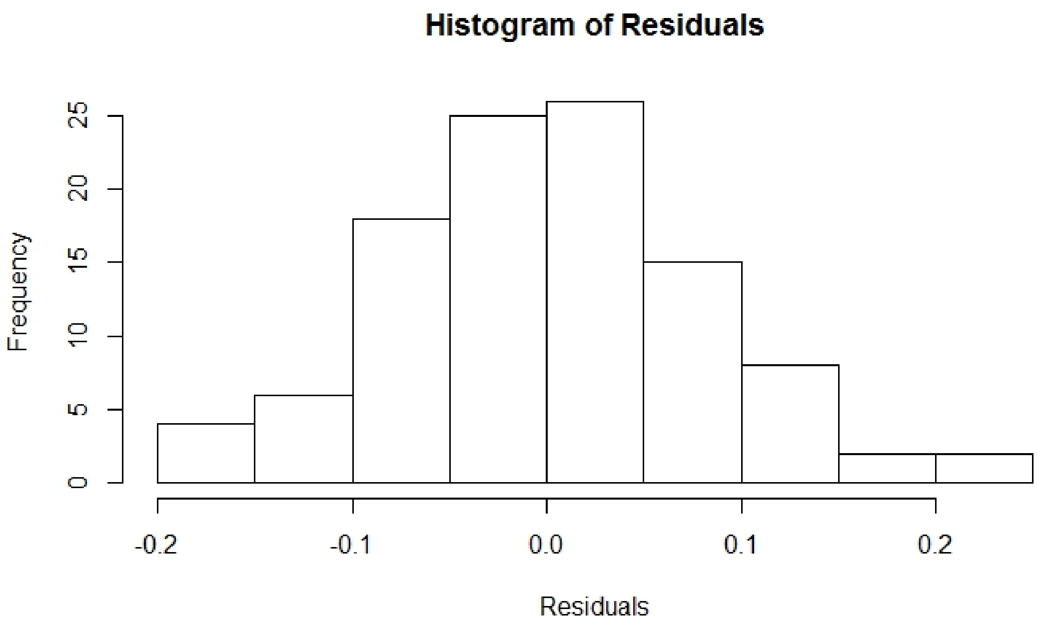
The first thing we do is to test if our residual errors can be considered to be normally distributed. As a consequence for each type of spread we display *Quantile-Quantile* plots and histograms.

High Yield :

Firstly, as it can be seen on the Q-Q plot, theoretical and sample quantiles are relatively in line and the plot of one against the other forms a line passing through the origin. This leads us to think that sample quantiles are normally distributed by being quantiles of a normal distribution.

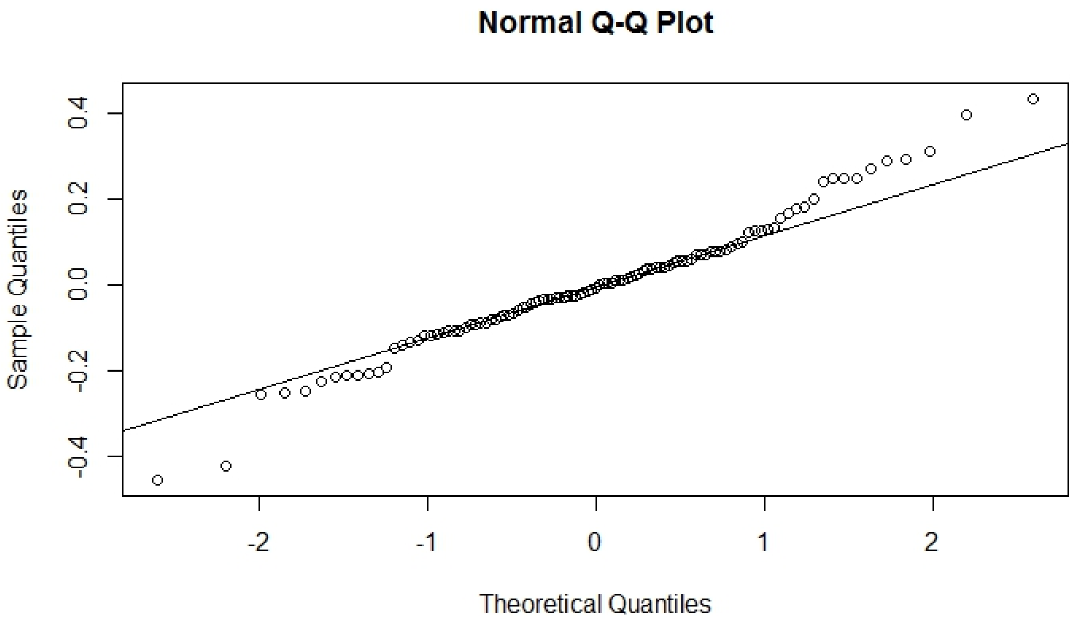


In addition, the histogram confirms our saying, since the shape gotten indicates that the distribution of these residual errors seem to be normal.

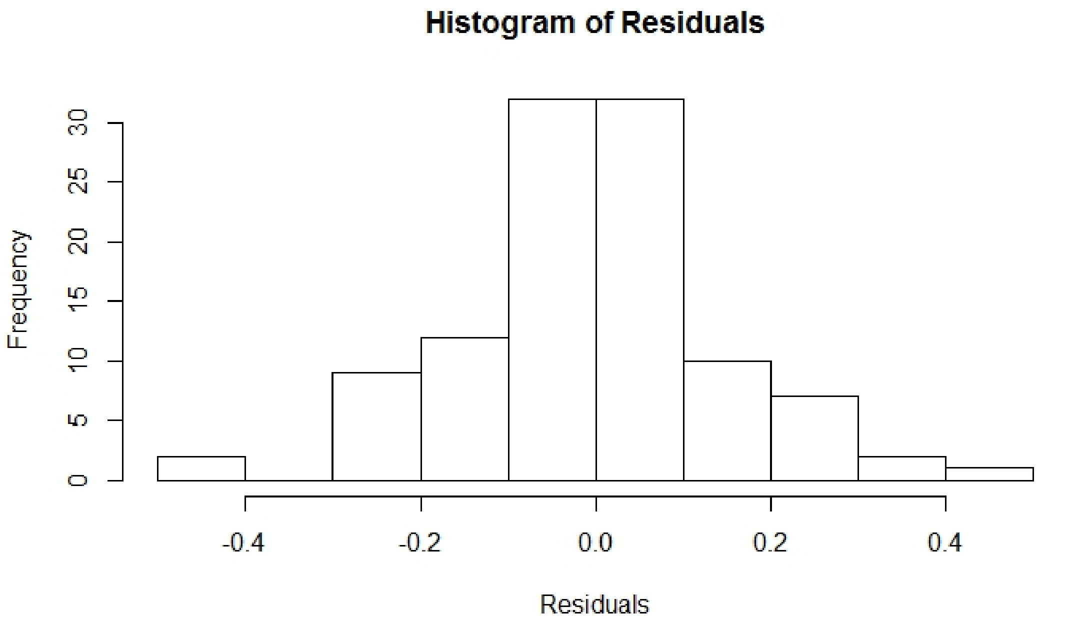


Investment Grade :

This time conclusions are less straightforward. Indeed, even if we can see clearly the same shapes as we did for high yield credit spreads, the graphs contain some errors.



Concerning the Q-Q Plot, except the perfect line in the center outlining the linear relation between sample and theoretical quantiles, we can see that for extreme values we got points that are quite far from the line. However, regarding the histogram, we still have this "normal shape," leading us to us to say that once again residual errors seem to be normally distributed.



3.3.2 Auto-correlation

Now that the normality of residuals has been proven, we focus on the second assumption concerning residual errors, which is to see if they are auto-correlated or not. To detect the presence of autocorrelation in our residual errors we chose to apply the *Durbin-Watson* test statistic.

Since we have just proven that we can consider that residuals are normally distributed and that the other assumptions are verified, we can use it and interpret its results.

High Yield :

Concerning the residual errors obtained in the *High Yield* model we get the following results:

lag	Autocorrelation	D-W Stat	P-Value
1	0.12	1.72	0.14

Table 17: Durbin Watson Test - High Yield

In this case since the *p-value* is equal to 0.14 we can reject the null hypothesis and thus cannot conclude that residual errors are auto-correlated. Which validates even more our model that seems to verify every assumptions.

Investment Grade :

lag	Autocorrelation	D-W Stat	P-Value
1	0.081	1.83	0.368

Table 18: Durbin Watson Test - Investment Grade

Once again the *p-value* is quite high, 0.37, thus we can reject the null hypothesis and cannot conclude that residual errors are auto-correlated.

Thanks to these two tests we can say that the assumptions are verified which validates the results and interpretations that we made about our linear regressions. As a consequence these two models can be used by the managers to predict potential changes of credit spreads. Let us see now if we can get a working model using non-stationary data.

4 Cointegration

Economic theory often implies equilibrium relationship between time series that are integrated of order 1 ($I(1)$). Indeed when we are confronted with data that presents some correlation and a stochastic trend we could hope to find a relationship between those variables that conduct to a stationary process.

That is what we tried to find here with a linear relationship between our credit spreads and some indicators that are $I(1)$. Then we have to find that the usual statistical results hold, and the residual error is $I(0)$ to conclude that our variables are cointegrated and it is not simply a spurious regression.

4.1 Linked Data

In order to find a cointegration between $I(1)$ variables, we need long time series based on monthly rather than quarterly data. Therefore, we replace the GDP by its monthly proxy, the ISM (Manufacturing Index) and keep the VIX and 2-Year and 10-Year rates.

- **ISM** : This manufacturing index is used as a proxy of the GDP and is useful since it is published in a monthly basis.
- **VIX** : Volatility Index
- **Rates** : The 10-Year and 2-Year government bond yield.

These variables were all proven to be integrated of order one and correlated to credit spreads. Therefore, since credit spreads have also unit roots, all the necessary assumptions are respected to verify if they are cointegrated.

4.2 Results

To test for cointegration we chose to use the Engle and Granger test defined earlier. First we perform a linear regression of credit spreads against the selected integrated indicators to estimate the relationship between them. Then once we have significant variables, we test the stationarity of the residual errors thanks to the Augmented Dickey-Fuller test.

4.2.1 High Yield

By regressing the logarithm of credit spreads against levels of previous variables we get the following results:

Regression Statistics	
R Square	0.877
Adjusted R Square	0,875
Standard Error	0,139
Observations	184

Table 19: Cointegration summary

	Coefficients	**Standard Error**	**t-stat**	**p-Value**
Intercept	5.85	0,24	23,99	0,00
ISM	-0.03	0,00	-9,32	0,00
VIX	0,64	0,05	14,10	0,00
US 10Y	0,06	0,02	2,63	0,01
US 2Y	-0,11	0,02	-6,57	0,00

Table 20: Regression Result HY

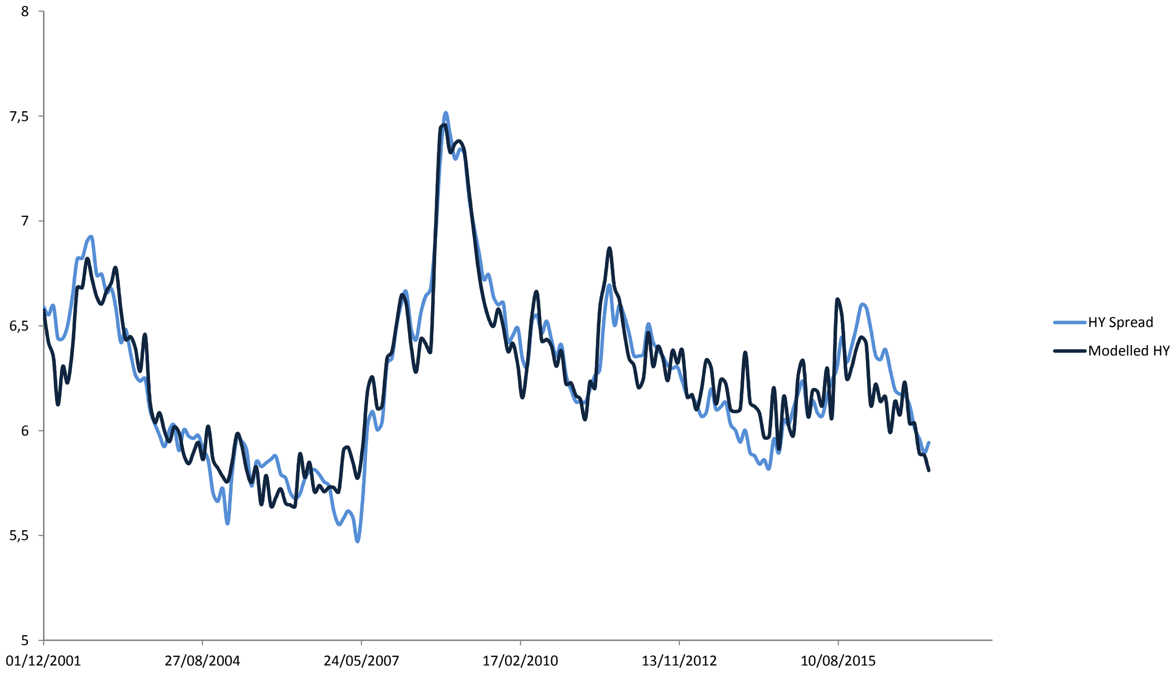
All selected variables have a significant impact according to the results above. Moreover, the R square is very high, bigger than the one found earlier taking stationary processes. However, one has to be careful not to rush in interpreting any result, because as long as we didn't prove the stationarity of the residual errors, we could have a spurious regression.

Therefore we study the obtained residuals and see if it is a stationary process or not:

Level	**Residual Error**
P-Value	0.00
T-stat	-4.44
1%	-3.46

Table 21: ADF test Residuals HY

Since the p-value is approximately equal to 0 we can reject the null hypothesis and consider that the residuals are stationary, and conclude that our variables are cointegrated. As a consequence any divergence in the spread from 0 between observed figures and our forecasts, should be temporary and mean-reverting. This is clear in the graph since the spread between both curves is often equal to 0.



4.2.2 Investment Grade

We try to find the same kind of relationship for investment grade through the same process, regressing against indicators and then test residuals :

Regression Statistics	
R Square	0.842
Adjusted R Square	0,839
Standard Error	0,171
Observations	184

Table 22: Co-integration summary

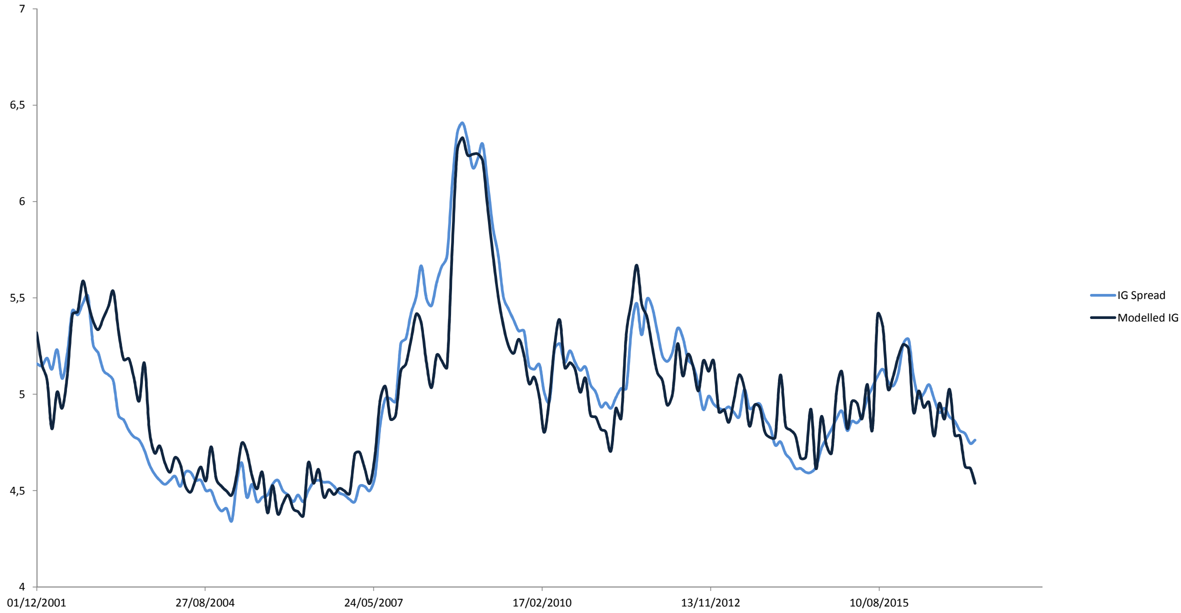
	Coefficients	**Standard Error**	**t-stat**	**p-Value**
Intercept	4.91	0,27	18.15	0,00
ISM	-0.03	0,00	-10.49	0,00
VIX	0,68	0,05	15.01	0,00
US 2Y	-0,07	0,01	-8.03	0,00

Table 23: Regression Result IG

Level	Residual Error
p-value	0.00
t-stat	-4.09
1%	-3.46

Table 24: ADF test Residuals IG

Once again we can consider that these variables are cointegrated, since the residuals are stationary. The spread between both curves (observed data and our forecasts) is a mean-reverting process.



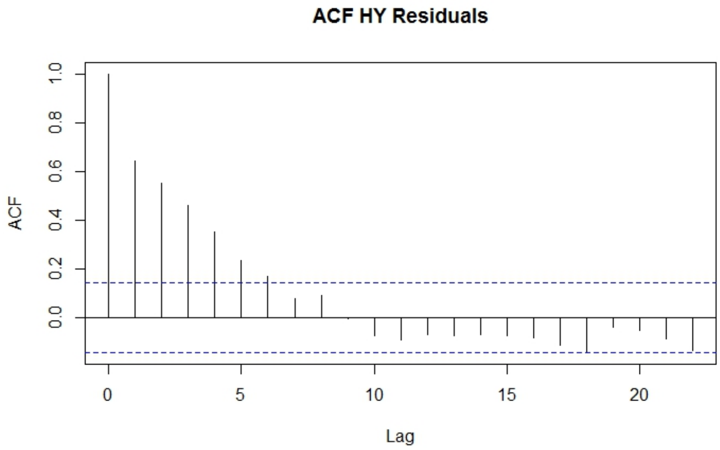
We have demonstrated that in both cases the equilibrium error is $I(0)$ and will rarely drift far from zero. Now that we have shown that it can be considered as a stationary mean-reverting process, let us see if we can identify it to an Auto-regressive process.

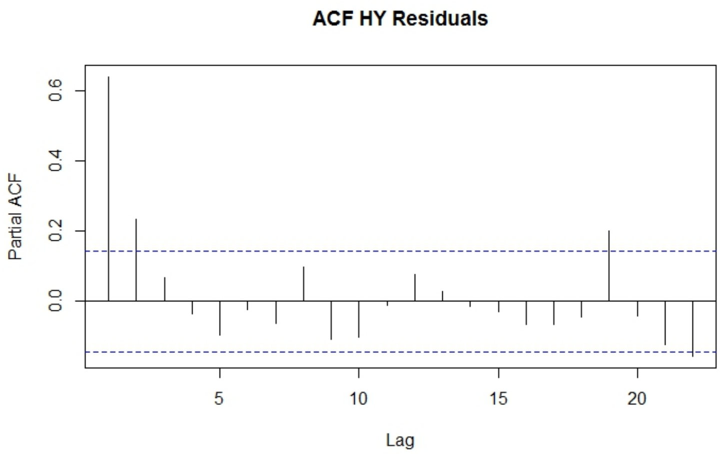
4.3 Residuals continuous model

If we suppose that the spread between the previous cointegrated variables can be modeled by an auto-regressive process of order 1, since its equivalent in continuous time is the Ornstein-Uhlenbeck process, we can model our residuals and identify their parameters. So first we have to find out if our first assumption is reasonable or not. In this section we will only focus on the High Yield case, since similar results were found concerning Investment Grade spreads.

4.3.1 ARMA test

To test the order of our process, we proceed simply by analyzing the *acf* and *pacf* graphs. Below we display both graphs concerning residuals for High Yield and Investment Grade spreads:





In both cases we have the same features, with an ACF diagram that decays slowly and seem to go to 0 at an exponential rate, whereas the PACF becomes negligible after lag 1. Those characteristics let us say that both of them can be considered as $AR(1)$ processes. As a consequence we can continue and try to describe their continuous versions.

4.3.2 Continuous Version

Since the Ornstein-Uhlenbeck process is considered as the continuous counterpart of the $AR(1)$ process, we will try to identify the different parameters. First let us define the dynamics of the process :

$$dX_t = \beta(\alpha - X_t)dt + \sigma dW_t$$

Here W_t is a standard Wiener process, $\sigma > 0$ and α, β are constants, so the process drifts towards α. Then we can derive this equation using Ito's lemma with the function $f(X_t, t) = e^{\beta t} X_t$ that we derive and get :

$$df(X_t, t) = \alpha\beta e^{\beta t} dt + \sigma e^{\beta t} dW_t$$

Which leads to the following solution for the Ornstein-Uhlenbeck process :

$$X_t = X_s e^{-\beta(t-s)} + \alpha(1 - e^{-\beta(t-s)}) + \sigma \int_s^t e^{-\beta(t-u)} dW_u$$

Now that we have the explicit expression of our mean-reverting process, we just have to give its discrete equivalent thanks to Euler method in order to identify each parameter to the ones we have in a AR(1) process.

$$X_{t+\Delta t} = X_t e^{-\beta \Delta t} + \alpha(1 - e^{-\beta \Delta t}) + \epsilon_{t+\Delta t}$$

Where $\epsilon \sim N(0, \sigma^2(\frac{1-e^{-2\beta\Delta t}}{2\beta})$

Then in order to estimate the parameters of mean-reversion, we regress the residual errors at time t on residuals at time (t-1) :

$$x_{n+1} = a x_n + b + \epsilon_{n+1}, \epsilon \sim N(0, \sigma^2(\frac{1 - e^{-2\beta\Delta t}}{2\beta}))$$

And finally identify the parameters through our system of three equations.

Regression Statistics		
	HY	**IG**
R Square	0.414	0.473
Adjusted R Square	0,410	0.470
Standard Error	0,106	0.124
Observations	181	181

Table 25: Residuals Auto-Regression

	Coefficients	**Standard Error**	**t-stat**	**p-Value**
Intercept	0.003	0,01	0.05	0,96
ϵ_n	0.064	0,06	11.30	0,00

Table 26: HY Regression Result Residuals

	Coefficients	**Standard Error**	**t-stat**	**p-Value**
Intercept	0.002	0,01	0.19	0,85
ϵ_n	0.69	0,05	12.75	0,00

Table 27: IG Regression Result Residuals

With the following results from our system of equations :

$$\beta = \frac{1}{\Delta t}\ln(\frac{1}{a}),$$
$$\alpha = \frac{b}{1-a},$$
$$\sigma = \sigma' \sqrt{\frac{\frac{2}{\Delta t}\ln(\frac{1}{a})}{1-a^2}}$$

As a consequence we get the following model as a representation of the continuous version of residuals concerning high yield and investment grades credit spreads:

Parameters		
	HY	**IG**
β	0.44	0.37
α	0.001	0.006
ϵ	0.13	0.15

Table 28: Continuous Version parameters

This representation gives us the choice to select the more adapted model, depending on whether we use daily data or high frequency data, the discrete or continuous model will be preferable over the other.

5 Application

Now that we have established some models and theories to evaluate credit spreads, we put them in use and apply them to predict and analyze credit spreads changes. Based on our macro-economic scenario and our assumptions relative to explanatory variables, we can now predict the different changes for the year to come (2017).

5.1 Visions

First of all, in order to establish any conclusions about where credit spreads will be at the end of this year, we must have strong opinions about the evolution of most indicators that explain credit spreads according to our study. Therefore, the first step is to assume forecasts of some explanatory variables in order to run the linear regression.

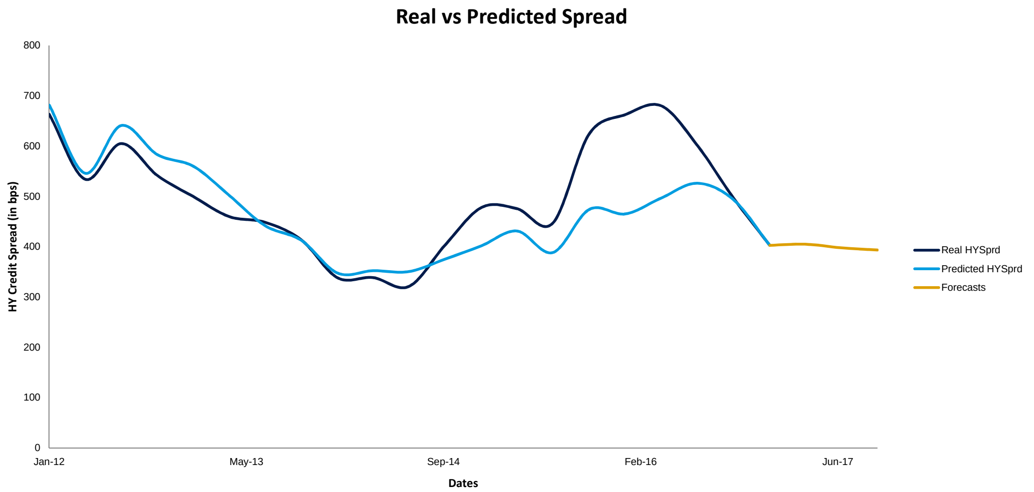
In this case, we established some potential changes for most our variables except 2 : Lending standards and the financial conditions index. Unfortunately, as we saw previously the second variable has a strong impact on the evolution of spreads and the model can be quite sensitive to its variations.

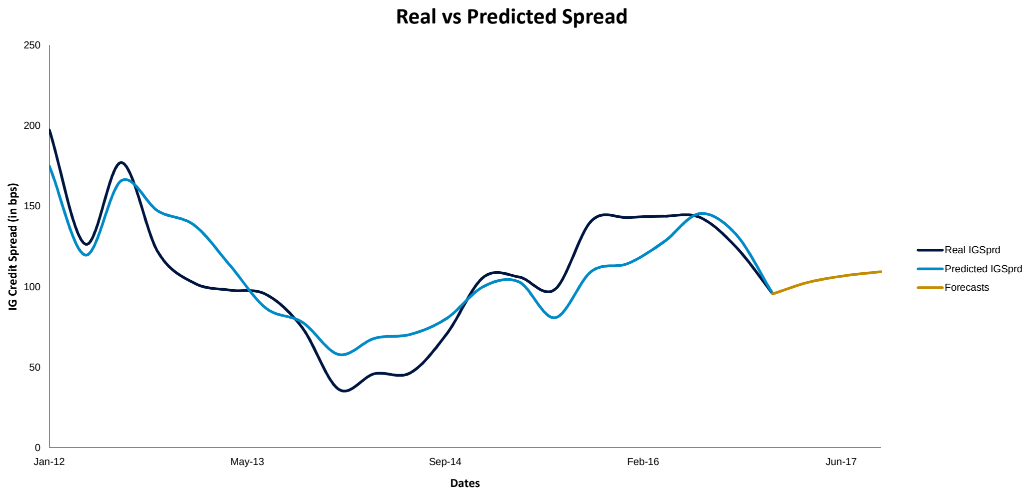
Here are some previsions on variations concerning explaining variables, (we chose not to disclose all of our previsions) :

- GDP : 2% of annual growth
- US 10-Year : +20 basis points
- Lending Stds : No evolution
- US 2-Year : +60 basis points

5.2 Predictions

Once we set up the assumptions, we got :





As a consequence we see that the model results in two different scenarios for investment grade and high yield credit spreads, by getting the following levels of spread in basis points :

Date	**HY Spread**	**IG Spread**
January 2017	402.5	95.6
April 2017	404.75	102.8
July 2017	397.65	106.86
October 2017	393.28	109.35

Table 29: Previsions

On the one hand concerning the high yield, we can see that after a first spike, the spreads tend to tighten and end up below the value at the beginning of the year. However this movement is very weak, so we could consider that it should flatten.

Then, on the other hand, we see a clear widening tendency for investment grade credit spreads that ends at 15 basis points above its starting value. As a consequence, we can clearly conclude that these should widen during this year gradually.

5.3 Verification

After having considered those visions that have been set up in early January, we chose to confront our results given here, with the real data collected in April, and see if it is coherent.

First we display, all different values gathered in April 2017 :

Variables	**Real Variations**	**Predicted Variations**
US 10-Year	-3.37%	$+2.80\%$
EPS	-0.06%	$+2.5\%$
VIX	-11.82%	-7.41%
US 2-Year	$+4\%$	$+10.64\%$
IG Spread	$+3.03\%$	$+7.54\%$
HY Spread	-4.17%	$+0.56\%$

Table 30: Previsions

As expected our previsions, concerning explanatory variables, are not exact. What is worst is that for some variables we didn't get the right variation (increasing instead of decreasing).That is the case, for instance, concerning the 10-Year rate and the earnings. Moreover the variables that we chose to have constant variations, naturally weren't constant (Financial conditions and Lending Standards), and as it was shown before, having a bad vision for the financial conditions index has a big impact on our model and could lead to false interpretations .

This suggests that we won't have correct predictions concerning credit spreads. However we see that, the quarterly variations for investment grade are in the right sens but not at the same extent; concerning High Yield spreads we predicted a flattening whereas it decreased. But as we said it earlier, the model predicts slight tightening of high credit spreads, which is in agreement with the real values.

As a consequence, even if we didn't get exact results, we captured the right tendencies, which outlines the fact that the model is not too sensitive to data and could still deliver relevant results, even if our visions weren't good for every variables. However we should say, that the variables that have been identified to have a big impact weren't necessarily badly predicted and the financial conditions index didn't move too much, which seems crucial for the model.

Conclusion

After having selected several proxies that should closely drive corporate bonds spreads, regression analysis on quarterly variations lets us explain quite well credit spread changes. It is mainly impacted by treasury yields (long and short term) as well as the financial conditions index. We are then able to say if corporate bonds are attractive in the current market conditions.

This doesn't come as a surprise since highly rated bonds are usually used as proxies for risk-free investments and government bonds which explains the narrow relationship between both instruments. Concerning the financial conditions index since it gathers a certain number of economic indicators, it explains naturally the changes of spread.

According to our model we have two different scenarios for both groups of corporate bond spreads:

- US Investment Grade credit spreads are supposed to widen by the end of 2017, which means that the spreads are cheap now. Therefore bond yield's growth, will lead to more expensive prices of bonds. As a consequence it prevents us from buying Investment Grade bonds since their yield will grow their prices will decrease.
- US High Yield credit spreads should flatten or widen a little according to our regression. Once again this is sign indicating that we should not buy corporate bonds (even if it is weaker than for IG spreads). However a way to profit from this situation is to go long Treasury Yields and short high yield. Moreover, spreads between Treasuries and high-yield credit are around lows not seen since 2014, so a normalization of levels of spreads can be expected.

To complete this analysis we found a cointegration relationship between the levels of credit spreads, rates and the ISM. It successfully led us to a mean-reverting process between spreads and a linear combination of the variables we selected. This allowed us to conclude that credit spreads are supposed to tighten in the short term. Therefore, this completes our former conclusions gotten thanks to the regression, outlining the cheapness of bonds in the short term. Moreover we fitted the residuals to an AR(1) process with its continuous version using the Ornstein-Uhlenbeck process to represent the mean-reversion.

Bibliography

[1] Jeffery D Amato and Maurizio Luisi - *Macro Factors in the Term Structure of Credit Spreads*

[2] Frank J. Fabozzi, Sergio M. Focardi - *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*

[3] Daniel Herlemont - *Pairs Trading, Convergence Trading, Cointegration*

[4] Robert F. Engle; C. W. J. Granger - *Co-Integration and Error Correction: Representation, Estimation, and Testing*

[5] Pierre Collin-Dufresne; Robert S. Goldstein; J. Spencer Martin - *The Determinants of Credit Spread Changes*

Appendix

Variables

- GDP : Gross Domestic Product
- EPS : Earnings Per Share
- CPI : Consumer Price Index
- PE : Price / Earnings
- FCF : Free Cash Flow
- US10Y : 10-Year Treasury rate
- US2Y : 2-Year Treasury rate
- VIX : Volatility Index

R Code

Regression Code

```
# ----------------------------------------------------------------------------------------------------
# Multiple Linear Regression
# ----------------------------------------------------------------------------------------------------

# Used to detect the impact of macro-economic indicators on Credit Spreads
# Variables : HY and IG Credit Spreads
# Explaining Variables : GDP; CPI; VIX; EPS; PE; Loans; ...

# ----------------------------------------------------------------------------------------------------

# Library
require(urca)
require(tseries)
require(psych)
require(MASS)
require(adespatial)

#Import Data from csv file

d1 <- data.frame(DataReg)
d <- ts(d1)

#Normalize the vector

#for (i in 1:length(d1)){
#   d1[,i] <- (d1[,i] - mean(d1[,i]))/sd(d1[,i])
#}


#Get cross-correlations
pairs.panels(d1)
```

```
###### If one in nonstationary (at 1%) then differentiate all and test again ######
StatData <- d1
NotStatio = FALSE

for (i in 1:length(StatData[1,])){
  testADF <- summary(ur.df(StatData[,i],"drift",selectlags = "AIC"))
  if (testADF@teststat[1,1] > testADF@cval[1,1]){
    NotStatio = TRUE
    break
  }
}
if (NotStatio){
  d <- StatData[-1,] - StatData[-nrow(StatData),]
  for (j in 1:length(StatData)){
    names(d)[j] = paste("Diff",names(StatData)[j],sep = "")
  }
  StatData <- d
}

#Perform regression with stationary processes
mylm <- lm(StatData[,1]~.,data = StatData[,-1])
summary(mylm)

#Minimize the AIC criterion
steplm <- stepAIC(mylm,direction = c("both"), k = qchisq(0.05,1,lower.tail = F))
coeffs <- steplm$coefficients

pairs.panels(StatData)

# Plots the Real Values against Predicted ones

plot(StatData[,1],type = "n",xlab = "Time",ylab = "Real vs Fitted values")
lines(StatData[,1], col = "grey",lwd = 2.5)
lines(steplm$fitted.values,col = "blue", lwd = 2.5)
legend("topright", legend=c("Real Value", "Predicted"),col=c("grey", "blue"), lty=1:2, cex=0.8)
summary(steplm)


##########  Calculate the volatility of stationary data ##########

VolInd <- c()
for (i in 1:length(StatData)){
  VolInd <- c(VolInd,sd(StatData[,i]))
}


########## Calculate Risk Contributions ############

CovMat <- cov(StatData[,-1])
Betas <- coeffs[-1]
DerivRisk <- CovMat %*% Betas
VarInv <- (1/sqrt(Betas %*% CovMat %*% Betas))
DerivRisk <- VarInv[1] * DerivRisk
```

```
RC <- Betas * DerivRisk

RankedR2 <- forward.sel(StatData[,1],StatData[,-1], alpha = 1)
```

Cross-Validation

```
# ----------------------------------------------------------------------------
# Cross Validation
# ----------------------------------------------------------------------------

# Uses 70% of the data to fit the model as a training set and 30% to plot  against the predicted v
# Variables : HY and IG Credit Spreads
# Explaining Variables : GDP; CPI; VIX; EPS; PE; Loans; ...

# ----------------------------------------------------------------------------


d1 <- data.frame(DataReg)
d <- ts(d1)

# Split our data into 2 sets, one training set and a validation set
UpTraining = floor(0.7*length(d1[,1]))
dtrain <- d1[1:UpTraining,]
dvalid <- d1[(UpTraining +1):length(d1[,1]),]

DataTrain <- dtrain
NotStatio = TRUE

# Use the ADF test to check if it is stationary or not and if not differenciates each serie

for (i in 1:length(DataTrain)){
  testADF <- summary(ur.df(DataTrain[,i],"drift",selectlags = "AIC"))
  if (testADF@teststat[1,1] > testADF@cval[1,1]){
    NotStatio = TRUE
    break
  }
}
if (NotStatio){
  d <- DataTrain[-1,]
  for (j in 1:length(DataTrain)){
    d[,j] <- diff(DataTrain[,j])
    names(d)[j] = paste("Diff",names(DataTrain)[j],sep = "")
  }
  DataTrain <- d
}

mylm <- lm(DataTrain[,1]~.,data = DataTrain[,-1])
summary(mylm)

steplm <- stepAIC(mylm)
summary(steplm)
```

```
#Take the validation set to differentiate it if the training set has been differentiated

DataValid <- dvalid
if (NotStatio){
  d <- DataValid[-1,]
  for (j in 1:length(DataValid)){
    d[,j] <- diff(DataValid[,j])
    names(d)[j] = paste("Diff",names(DataValid)[j],sep = "")
  }
  DataValid <- d
}

ResData <- c()
for (i in 1:length(DataValid)){
  for (j in 1: length(coeffs)){
    if (names(coeffs)[j] == names(DataValid)[i]){
      ResData <- c(ResData,i)

    }
  }
}

DataValidSel <- DataValid[,ResData]

###### Vector predicted by the training data ########

coeffs <- steplm$coefficients
yhat <- c()
yhat = coeffs[1]*rep(1,length(DataValidSel[,1]))

for (i in 1:length(DataValid[,1])){
  k = 2
  for (j in ResData){
    yhat[i] = yhat[i] + coeffs[k]*DataValid[i,j]
    k = k+1
  }
}

plot(DataValid[,1],type = "n",xlab = "Quarters",ylab = "Real vs Fitted values")
lines(DataValid[,1], col = "grey",lwd = 2.5)
lines(yhat,col = "blue", lwd = 2.5)
legend("topright", legend=c("Real Value", "Predicted"),col=c("grey", "blue"), lty=1:2, cex=0.8)
```

Cointegration

```
# Library
require(urca)
require(tseries)
require(psych)
require(MASS)
require(adespatial)

#Import Data from csv file

d1 <- data.frame(DataReg)
d <- ts(d1)

#Perform regression with stationary processes
mylm <- lm(d1[,1]~.,data = d1[,-1])
summary(mylm)

ResCoint <- mylm$residuals


###### ACF / PACF graphs ######

acf(ResCoint,main = "ACF HY Residuals")
pacf(ResCoint,main = "ACF HY Residuals")

###### Auto-Regression of residuals #####

ResCointY <- ResCoint[-1]
ResCointX <- ResCoint[-length(ResCoint)]

OUlm <- lm(ResCointY~ResCointX)
summary(OUlm)

##### Plot Real vs Modeled values #####

plot(ResCointY,type = "n",xlab = "Time",ylab = "Real vs Fitted values")
lines(ResCointY, col = "grey",lwd = 2.5)
lines(OUlm$fitted.values,col = "blue", lwd = 2.5)
legend("topright", legend=c("Real Value", "Predicted"),col=c("grey", "blue"), lty=1:2, cex=0.8)


##### Continuous version parameters estimation #####

beta <- log(1/OUlm$coefficients[2])
Alpha <- OUlm$coefficients[1]/(1 - OUlm$coefficients[2])
epsilon <- sd(OUlm$residuals) * sqrt(2*beta / (1-OUlm$coefficients[2]^2))


##### Johansen Test #####

jotest=ca.jo(d, type="trace", K=2, ecdet="none", spec="longrun")
summary(jotest)
```

TRITA -MAT-E 2017:18
ISRN -KTH/MAT/E--17/18--SE