CREDIT DEFAULT SWAP SPREAD MODEL: DESCRIPTIVE AND PREDICTIVE

INSIGHT IN THE DETERMINANTS OF CDS SPREADS

[Public version]

AUTHOR
Danny Wemmenhove

DATE
12 March 2009

SUPERVISORS KEMPEN&CO
Richard Klijnstra
Rik den Hartog

1ST SUPERVISOR UNIVERSITEIT TWENTE
Emad Imreizeeq

2ND SUPERVISOR UNIVERSITEIT TWENTE
Berend Roorda

Universiteit Twente
de ondernemende universiteit

KEMPEN CAPITAL MANAGEMENT

Contents page

Universiteit Twente
de ondernemende universiteit

KEMPEN CAPITAL MANAGEMENT

ACKNOWLEDGEMENTS

I started my Master graduation project at the 1st of August 2008. In the beginning of 2008 the financial world was already going through some rough times, but during my internship a crisis of a gigantic size has shocked the financial system. Although a crisis is difficult to connect with positive aspects, for me, it has been a very interesting and valuable experience. I think I have learned a lot from this period from which we have not seen the end yet.

I would like to thank Kempen Capital Management for giving me this opportunity to do my Master graduation project. Doing an assignment for them has given my project extra value for which I am very grateful. The experience of working at the fixed income desk has given me extra insights into the fascinating world of finance.

Without good supervisors, an internship is not complete. First of all, I would like to thank Rik den Hartog for all the help he gave me during my internship as my supervisor at Kempen Capital Management. I really enjoyed our discussions and the fact that he always found time for my questions. Furthermore, I would like to thank Richard Klijnstra as head of fixed income and as my additional supervisor at Kempen Capital Management for his guidance and help. Especially the insight he gave me in setting up the entire regression model was very helpful. I would also like to thank the entire fixed income team for not only a very interesting time, but also for the enjoyable aspects of working with them.

From the University Twente, I would like to thank my first supervisor Emad Imreizeeq. His quick responsiveness to my questions, our interesting meetings and his guidance in setting up this Master thesis have been of indescribable value. In addition, I would like to thank my second supervisor Berend Roorda, for his added value to my Master graduation assignment. His theoretical insights were very helpful.

Last but not least, I would like to give special thanks to my parents. Their help and support during my entire university period and through this last project have made me the proud man I am today.

Author:

Danny Wemmenhove

MANAGEMENT SUMMARY

Credit default swaps have started trading at the beginning of this century. There is still a lot to be discovered on the subject of credit default swaps and especially their spreads. The spreads are considered to be a price indication of the risk associated to the credit default swap. Or so to say; the risk premium one is willing to pay. How this risk premium is determined and what makes it go up and down are questions which are mostly still unanswered.

This study investigates the determinants of credit default swap spreads. The aim of this project is to find out which factors describe a credit default swap spread, and which factors can give an indication in terms of forecast.
The factors taken into account are divided into fundamental, market and macro variables. Each variable is weekly based over the period from 2004 Q1 to 2008 Q2. In this research 146 non-financial European companies have been used for firm-specific data. To find the determinants, the ordinary least-squares methodology has been applied. Through the usage of correlation matrices, univariate regressions and multivariate regressions the variables which are used in the final model have been determined. Together with the requirements set by Kempen Capital Management, the variables that show its worth in describing the spread are: Net Debt divided by EBITDA, Return On Assets, ln(total Assets), Implied volatility over 3 months, the yield difference between AAA corporates and BBB corporates, and as a correlation factor with the market; Beta. This model seems to perform quite well. Applied to several companies, the model gives an indication of where the spread should be. Testing the model the out-of-sample period of 2008 Q3, resulted in increasing confidence in the performance of the describing model.

For the forecasting model, the same variables turned out to be the most suitable to use, with Beta left out of the determinants. Although the R^2 of the forecasting model is lower than the describing model, it still seems to perform well. Applying the forecasting model to the period of 2007 Q1 to 2008 Q3 with a one month forecast, companies proved to be profitable. This profit is achieved by setting a short/long trading rule determined by the forecasting model. Using the forecast model only on investment grade companies results in higher profits.

Although the profits look very promising, one has to be careful in putting all his faith in the forecasting model. Credit default swap spreads are difficult to interpret. However, the describing model and the forecasting model are a step in the right direction.

1. INTRODUCTION

In the last decade the financial world witnessed a very turbulent period in which the sub prime and credit crisis have left their mark in the newspapers. For bankers/investors it is important to retrieve an indication of how the markets will perform. One would perhaps prefer to have a possible look into the future. Yet, this last period showed once again that the events of the financial world are hard to predict.

A factor where the market made an unexpected turn was in the credit default swap (CDS) spreads. The confidence and trust in the financial world completely vaporized after a sequence of disturbing events. When the credit worthiness of financial institutions was questioned with the forced sale of Bear Stearns and the fall of Lehman Brothers, this had its impact on the spreads of CDSs. To give an example of how one can look at CDS spreads; Corporation ABC may have its credit default swaps currently trading at 265 basis points (bp). In other words, the cost to insure 10 million euros of its debt would be 265,000 euros per annum. If the same CDS had been trading at 170 basis points a year before, it would indicate that markets now view ABC as facing a greater risk of default on its obligations.

In a CDS, there are three parties involved. One of these parties is a reference firm who is issuing bonds. Another party is called the protection buyer who acquirers the bond and wants to buy protection for the bond purchased. The third party is the protection seller, who receives from the protection buyer a fixed premium each period until either default occurs or the swap contract matures. In return, if the reference firm defaults on its debt, the protection seller is obligated to buy back the defaulted bond at its par value from the buyer (Longstaff et al 2003) or to settle it in cash. The spread (or periodic payment), taken as a percentage of the notional value of the CDS contract, is a metric of the credit risk of the reference firm (Das et al 2007). In chapter 2, there will be a further explanation on the concept of the spread.

To give an insight on the movements of the market in CDS spreads in 2007, the graph on the next page shows how several financial institutions went from a stable low spread to an exploding spread at the end of 2007. Spreads between the 100 and 200 bp were already considered quite exceptional, one can imagine what an impact a spread of 500 bp and above had on investors.

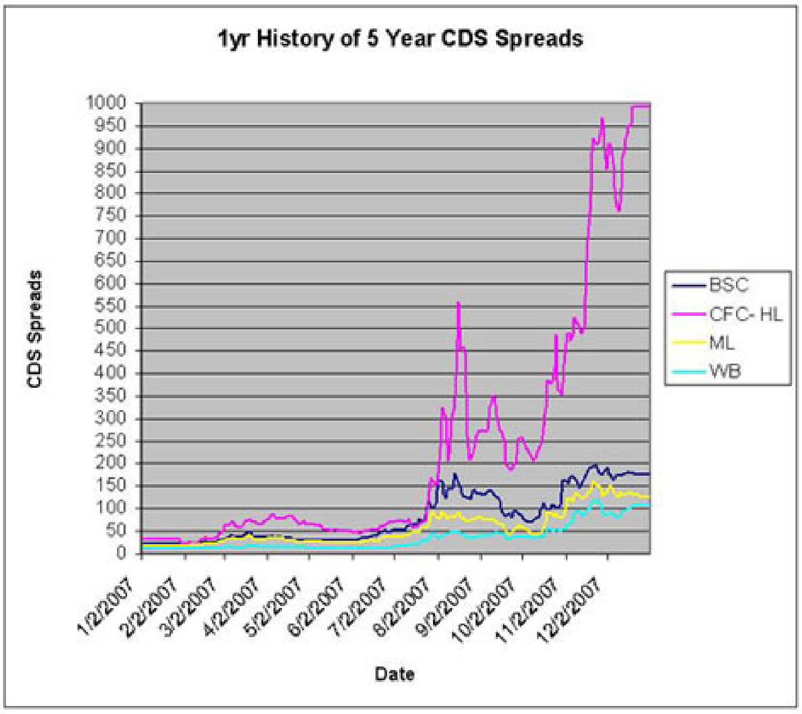


Figure 1 – This figure shows several large (former) US financial institutions such as Merrill Lynch (ML) and Bear Stearns (BSC) with their 5 Year CDS Spreads of 2007. As one can see, the spreads during the year are increasing where Countrywide Financial Corporation (CFC, an American Home Loan Lender) even is exploding due to the sub prime crisis.

The global market for credit derivatives has expanded tremendously in the last decade, with CDSs being a large part of derivatives traded. This increase can mainly be described by the growing desire to improve the management of credit risk. This desire does not only come from bankers or investors, it also comes from the government and regulators. For instance, the Basel II accord has forced all financial institutions to take a close look at implementing and improving the modelling and management of credit risk. Credit risk can be defined as the potential that a borrower will fail to meet its obligations in accordance with agreed terms (Sobehart and Stein 2000).

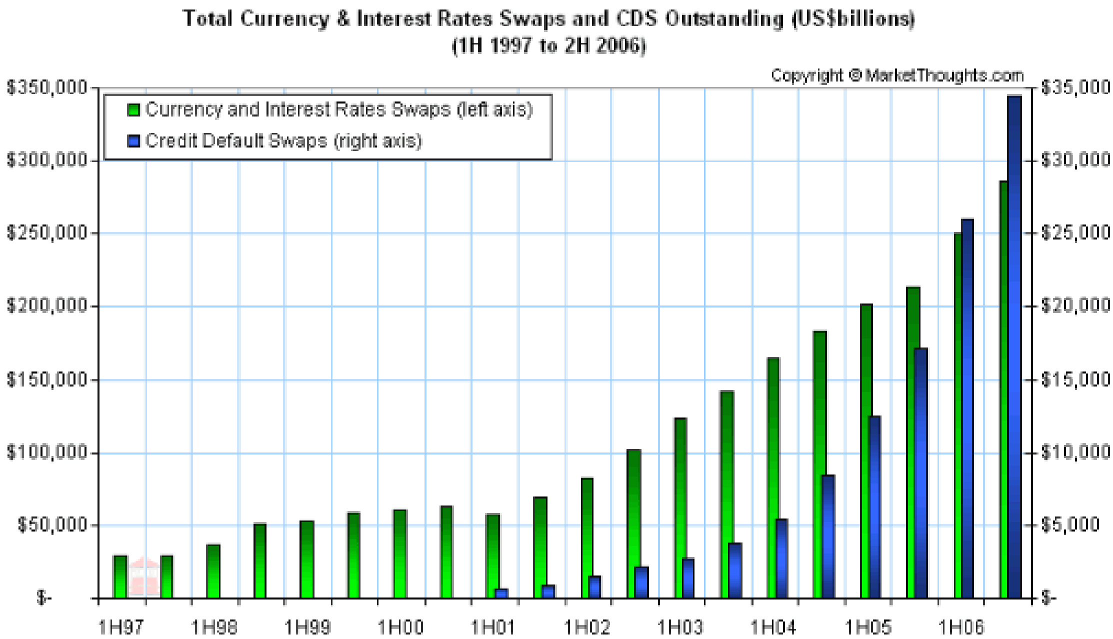


Figure 2 – This figure shows the increase in the last decade in the amount of CDS outstanding compared to total currency and interest rate swaps outstanding (in US $ billions). Where the increase of the latter is approaching linearity, the rate of CDS seems to approach an exponential increase.

Several research papers have been dedicated to set out the content of credit risk and trying to create models which should give an indication on the credit spread of an entity. When companies try to make an estimation of the credit risk of an entity, they usually use their own methods or models to sketch a view on the credit risk. Next to this approach, there are also external ways to get a credit risk view. Rating agencies try to assign credit ratings for issuers for certain types of debt obligations to represent their opinion on the issuers' credit risk situation. The three largest and popular rating agencies are Standard & Poor's (S&P), Moody's, and Fitch. In the past, a lot of companies relied on the ratings published by these agencies for making their decision which seemed to work. However, since the crisis of 2007, it was proven that the rating agencies could be wrong. A lot of their published ratings did not comply with the enormous credit crisis which the financial market went through.

Companies are turning increasingly to their own modelling principles. A lot of different credit score models have been set up especially by banks, but also by the recovering rating agencies. A credit score can be defined as a numeric expression which is statistically derived to represent the credit worthiness of an entity to meet its obligations. These kinds of models look especially at factors like the probability of default. Currently there is a trend to try to expand the view on the credit risks not only using a credit score model, but also to try to get an indication on the spread. It is difficult to define the content of a spread. Some part of it will consist of credit risk. That is the probability to default (PD), loss given default (LGD) and exposure at default (EAD), but that definitely does not cover the whole lot. What determines the rest of a spread is a hot issue in current discussions. Since this trend is a recent development, not many articles have been written on the subject.

This thesis will be on the development of a model that can make an assessment of the spread of a credit default swap of an enterprise. This model will be based on the analysis of factors from the Kempen Credit Score Model and of additional factors. To set up this research, the starting point is the determination of the factors that will be included in the entire research. The second step is to select the industries with the accompanying data that will be taken in the research. Then, we shall test the model according to the demands of Kempen Capital Management (KCM). The main research questions that we address here are:

- Do the factors following from the Kempen Credit Score Model have additional value in predicting the spread?
- Does the spread model contain additional factors that play a role in the determination of the risk premium?
- Is it possible with the spread model to recognize potential profit opportunities?

Since there are already spread models available (but only very few), one could wonder why KCM does not use an existing spread model. Reasons for this are that the current spread models do not work according to their demands, the costs of purchasing such a spread model are very high and they would like to have a model that contains their view on spreads. Since the spread models have been developed only recently, it is the general opinion that there are a lot of questions unanswered and there is still a lot to gain in the field of spread modelling.

In the next chapter, several literature articles are being reviewed on the subject. Furthermore the different available methods for modelling will be described. Chapter 3 will describe the industries and data used in this research. In chapter 4 the methodology used for the testing and modelling will be explained. Chapter 5 will show the results of the describing model and several tests. Chapter 6 will then go in on a forecasting model. Finally chapter 7 will give the conclusion and final thoughts on the results and recommendations for further research.

Because this is the public version of the thesis, some of the results are not available due to confidentiality.

2. LITERATURE REVIEW

2.1. Introduction

As mentioned in the first chapter, not a lot has been written specifically on spread models. There are however some research articles on spreads, and modelling as such. Spreads are dealt with in different categories. Most of the spreads studied are corporate bond spreads. One can define credit spreads in this case as the difference between the yield of bond *i* and the associated yield of the Treasury curve at the same maturity (Collin-Dufresne et al 2001).

There are some differences between a bond spread and a credit default swap spread. For instance the CDS spread data provided by a broker, consists of a firm's bid and often quotes from dealers. Once a quote has been made, the dealer is committed to trading a minimum principal at the quoted price. On the other hand, the bond yield data available are usually indications from dealers; there is no commitment from the dealer to trade at the specified price. Another difference is that CDS spreads do not require an adjustment, whereas bond yields require an assumption about the appropriate benchmark risk-free rate before they can be converted into credit spreads. CDS spreads are already credit spreads (Hull et al 2004). Before determining the spread of a bond, the risk-free rate has to be determined. This rate is mostly chosen as the rate of government bonds, but there is no unified rule determining which government rate has to be used as the risk-free rate.

Despite these differences, several studies, like Elton et al (2001) and Avramov et al (2007), have taken the corporate bond spreads as data to test their model instead of CDS spreads. Arguments for this approach are among others, that corporate bonds retrieve a far more representative calibration of market-wide parameters (like a Market Sharpe Ratio and the size premiums) by covering a wider range of names. As shown in figure 2 CDS spreads started trading in 2001. Corporate bonds however have been traded for a longer time, which results in a longer history of data available for testing. Furthermore, the characteristics of these corporate bonds are fairly well understood in the industry, which could help in creating insight. The use of CDS data is not mentioned often in the current research literature. The advantages of using CDS data are however discussed in Das et al (2007), Jakovlev (2007). As stated earlier, the fact that the CDS quotes already represent credit spreads makes it more clear cut. In addition, there is no need for CDS data to remove instruments of bonds like coupons and call provision effects.

This research focuses only on the CDS spreads. This means that there won't be the same amount of historical data available for research as there would be if this research was on corporate bond spreads. As the data for CDS spreads is mostly available from 2004, the data for corporate bond spreads have more data available before 2004.
A model is mostly determined by the factors included. The amount and type of factors specify how many variables a model has to cope with and whether the model is strictly quantitative or perhaps also contains some qualitative factors. There are several research articles (Das et al 2007, Jakovlev 2007, Greatrex 2008) on which factors should be included in a model to get a good indication on the CDS spread. As one can imagine, not all the research articles agree on which factors should be considered important, and therefore should be included in the model. Putting extra variables in the model depends on the opinion of the researcher, whether the variables appear to be significant and which kind of statistical methodology for modelling is used. The next section will go more in depth on this last part.

2.2. Methodologies

2.2.1. Statistical Model

Ordinary Least Square (OLS) model
OLS (Larsen and Marx 2001, Fraas and Newman 2003) is one of the easiest to use and oldest statistical model to find a 'fit' to the data, or so to say to find the linear model that minimizes the sum of the squared errors. In the regression, the independent variables x_j are used to calculate the expected dependent variable *y*. In addition there is a random error term ε_i and the parameters β_j, which has to be estimated.

$$y = \sum x_j\beta_j + \varepsilon_i = x\beta + \varepsilon_i \qquad (1)$$

The OLS model attempts to retrieve solutions by choosing those β_j's that minimize the sum of squares of the vertical distances from the data points to the function *y*.

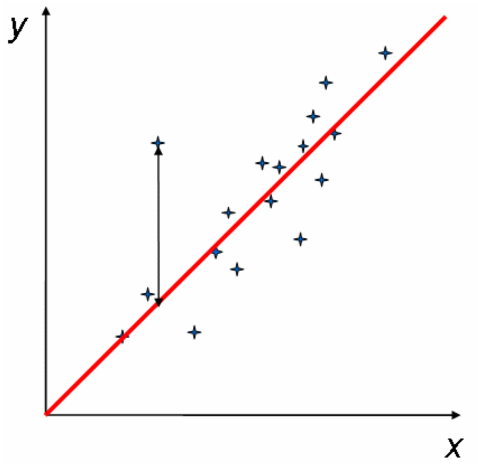


Figure 3 – The graph displays an example of the vertical distance to the presumed distribution function. Important feature is whether certain data points should perhaps be considered as outliers and should be taken out of the analysis.

As mentioned, this type of regression has certain advantages such as the easiness of applying this method. Compared to other methods, it also has more intuitive appeal. It is easier to determine whether a particular positive value is high or a negative value low. There are however some drawbacks that have initiated researchers to invent new methods. One of the drawbacks is that the OLS model assumes multivariate normal distribution with equal variance-covariance matrices of the independent variables. This has empirically been proven to be a false assumption. Applying OLS with the assumption of normality could therefore lead to inconsistent and inefficient estimates. Not often recognized in literature, but a second drawback (Morgan and Teachman 1988), which is linked to the first drawback, is the assumption that standard errors of the coefficients in most cases will be incorrect due to serial correlation. This assumption leads to wrongfully-stated conclusions regarding statistical significance. This all is triggered by two aspects. The first is that estimators do not have the smallest standard error, which can be recognised as inefficiency. The second aspect is that estimators do not converge to their values when the sample size increases, which can be stated as inconsistency. The last drawback in this discussion is the fact that the predicted values of **Y** may fall outside the range of 0 to 1, which is especially inconvenient for predicting a probability like the default. However in the case of estimating the spread of CDSs this last drawback is not of importance.

Logit (probit) model
Further possibilities as statistical methods to use are the logit or probit model. The most commonly used of the two is the logit method. There are not a lot of differences between the logit and the probit model. One main distinction is that probit models use the normal cumulative function for weighing the independent variables and assign scores in the form of

a default probability, whereas logit models use logistic distribution. An advantage for a default probability model of using the logistic distribution is that it automatically bounds the dependent variable between 0 and 1.
The logit model is used for prediction of, for example, a default event by fitting data to a logistic curve. The independent variables in the model can be either of numerical value or as categorical value.
Regressions of the type probit/logit are performed by means of a non-linear maximum likelihood procedure. To give an impression on the usage of these kinds of models, one should start with the logistic function f(z):

$f(z) = \frac{1}{1+e^{-z}}$, where $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k$. (2)

The z-value is calculated for the f(z) by the formula above. The f(z) represents the probability of an outcome that is specified by the modeller. The z-value represents a measure of the total contribution of all the risk factors included in the model. These risk factors are represented by the variable *x* and the weight that the risk factors have in the model is represented by β_j 's. If the β_j is positive, this will result in the risk factor increasing the probability expressed by f(z). Vice versa, a negative β_j results in a decrease in f(z). As mentioned before, as one of the advantages of the logit model, it does not matter how large or small the z-value turns out to be. Even if the z-value approaches infinity or negative infinity, the probability of a specified outcome represented by f(z) will always be in the range of 0 to 1. Other advantages of the logit model are its possible application to panel data, the appealing non-linear shape and the allowance for qualitative data with categories by means of dummies.

Studies by Altman (1968) and by Ohlson (1980) bringing forward respectively the popular Z-Score and O-Score, have given statistical models a boost in usage. Both models have been set up by choosing as independent variables risk factors that have been noted often in research studies. In addition, some independent variables were added by personal view on relevant factors. These statistical models are all fundamental (accounting) based. All the independent variables are factors that are derived from a firm's balance sheet and income statement. This implies that the models take account for firm specific risk factors. Market or macro factors are not included in the statistical models. Although empirical evidence has shown that the statistical models proved to be useful and with good performance, the exclusion of market and/or macro factors has lead to quite some discussions and further research for improving the models.

2.2.2. Structural Model

Structural models were introduced by Merton (1974). The idea behind this approach is based on option theory. Merton-model published in 1974 uses the assumption that a firm has a single class of debt and equity. As aspects of a firm, the equity is viewed as a call on firm's

assets, whereas debt is seen as a risk-free bond and a put on the firm's assets. The debt is derived by taking the total market value of assets and by subtracting from it the value of the option (an option-type structure that represents equity).
The reason that the debt is seen as a risk-free debt and a put on the firm's assets in this model is that it is believed that a regular risky debt with an asset guarantee is the same as risk-free debt. This asset guarantee can be seen as a put on the firm's assets.
An entity is considered to be in default when the value of the assets falls below a default threshold. This approach of defining default indicates that the structural model relies on firm's financial specifications as timely market value-dependent variables. Important factors for the structural model are a firm's equity, assets and the accompanying volatilities. With the option theory, the option pricing model of Merton can be used:

$$E = A * N(d_1) - D * e^{-rT} * N(d_2), \quad (3)$$

where $d_1 = \frac{\ln(A/D) + \left(r + \frac{\sigma_a^2}{2}\right) * T}{\sigma_a * \sqrt{T}}$, and $d_2 = d_1 - \sigma_a * \sqrt{T}$

A is the firm's value, E is the equity, D is the debt, r is the risk-free interest rate, T is the time to maturity and σ_a is the volatility of the assets.
To calculate the volatility of the equity (σ_e), the following formula can be used:

$$\sigma_e * E = \left(\frac{\partial E}{\partial A}\right) * \sigma_a * A \quad (4)$$

Table 1 describes the two different scenarios of a default and no default of the reference firm. When the firm's value is above the threshold (the debt of the firm) then the payoff for the debt holders is simply the amount of debt. Consequently, the payoff for the equity holders is the remaining part of asset minus the debt. In the case of a default, according to the structural model, the value of the assets has dropped below the threshold. This would mean that the only payoff the debt holders can retrieve is the value of the assets. Since in this case there is more debt than assets there is nothing left as payoff for the equity holders.

Table 1 – Payoff Structure

Nature	Asset value	Equity holders	Debt holders
No default	A > D	A - D	D
Default	A ≤ D	0	A

It is often believed that the power of the structural model lies in its ability to explain how the capital structure of a firm and the market environment in which it operates influence the prices and risks of its equity and debt. However the structural model as described in the paper of Merton (1974) had some restrictions like the assumption that a firm has only a single class of debt and equity. These restrictions were not highlighted that much, because the approach as such was a breakthrough in the financial world at that time. In the time till now, several researchers have offered extensions or reformulations of Merton model like

Jarrow and Turnbull (1995), and Longstaff et al (2003). One of the extensions was that no longer there was a single class of debt, but a distinction was made between short-term debt, long-term debt (2-5 year) and other long-term liabilities (i.e. convertible bonds or perpetual capitals). This distinction in terms of debt can be quite effective in the prediction of CDS spreads. It is assumed for example that entities with a higher amount of short-term debt compared to entities with a higher amount of long-term debt have a higher spread.

Another adjustment to the Merton model has been the removal of the assumption that a default only can occur when the debt reaches maturity. This assumption was set up by the usage of European options in the Merton model. A European option can only be exercised at time of maturity. Since empirical evidence has shown that defaults often occur before maturity is reached, researchers tried to overcome this by changing from European options to American options. American options do have the possibility to exercise before maturity is reached, but these options require an adjusted formula in equation (3).
A final important adjustment has been the implementation of a variable risk-free rate. Where the Merton model from 1974 assumes a constant risk-free rate, this does not comply with the real world. A variable risk-free rate should improve the representation of the real risk-free interest rate. Normally, such a risk-free interest rate is set equal to a Treasury bill rate or a swap rate.

Looking at the general effects that the structural model is assumed to have on a CDS curve, three cases are described in table 2. In the first case, a higher firm value results in a higher equity price and also an increase in debt. This results in the CDS curve to tighten, because the leverage of the firm decreases. If the firm's value decreases, it is the other way around. In the case the asset's volatility increases, the spread across the curve widens and the equity price will increase. This is caused by the fact that the debt holders are short volatility and the equity holders are long.

Table 2 – Parameter Effects

Parameter move	Equity price	Debt present value	CDS Curve
Higher firm value	Up	Up	Tightens
Lower firm value	Down	Down	Widens & inverts
Higher volatility	Up	Down	Widens

Although these adjustments have improved the overall performance of the structural model, there remain some drawbacks to the model. These drawbacks will be discussed in chapter 4.

2.2.3. Reduced-form Model

The reduced-form approach is another model which is developed by Litterman and Iben (1991), Jarrow et al (1997) and Duffie and Singleton (1999). The reduced-form approach differs from structural models by abandoning the direct reference to the firm's asset value process. For this model, the occurrence of default and recovery rate at default determines the credit risk. Market data is used to retrieve the parameters of these two components. In essence, the structural model and the reduced-form model are really the same model. The main difference does not depend on the determination of the default time, but what kind of information is known by the modeler. Reduced-form models assume that the modeler has an incomplete knowledge of the firm's condition, but having the same information set by the market. In most cases, with a low level of firm's specific information, it is hardly possible to determine the default time. In contrast, structural models assume that modeler does have the same information set as a firm's manager would be likely to have. This firm specific information leads in most situations to a predictable default time.

Which models are used for the purpose of pricing and hedging? It seems that the reduced-form model is the most appropriate model. This is based on the fact that prices are set by the market and the balance of the market is determined by information that it is available to make decisions. For other aspects, like marking-to-market or to assign market risk, reduced-form model is the preferred modeling method. In the case that one is part of, or represents, the management of a firm and wants to judge the possibility of default for capital reasons, then a structural model may be preferred.
Since the reduced-form side is dependent on the observable data from the market, it relies heavily on the quality and quantity of the data. In this way, traded issuers will not be well modeled unless they issue more traded debt. A prediction for the CDS spreads is that in cases where an issuer has many traded bonds in the market (dependent on market data), the reduced-form model tends to work as the best of the two.

2.2.4. Hybrid Model

While the statistical models are fundamental based, the Merton model is market based. Both models have their advantages and their shortcomings. The hybrid model tries to combine these advantages and leave out some of the disadvantages. Hybrid models have only recently been introduced by Sobehart and Stein (2000), Tudela and Young (2003), and Benos and Papnastasopoulos (2005). Hybrid models try to combine the timely market value-dependent variables from structural models with the fundamental variables of companies as input for the model. In addition, statistical methods are used to retrieve and estimate a model that best fits the historical data. Also the addition of macroeconomic variables makes hybrid models a tool which tries to capture as much of relevant information as possible.

Since there are not a lot of research results available, it is difficult to state at this moment whether hybrid models already have proven their worth. The inclusion of accounting, market and macro variables covers quite a lot of factors that could be of importance to the model and eventually to the determination of the CDS spread. The only problem with the hybrid model known today is that the large amount of variables taken into the model makes it a very data-heavy tool. This could have its effect on the time-consumption of the model.

2.3. Variables

As stated in the previous sections, in the literature and the models, there is a distinction made between fundamental, market and macroeconomic variables. Fundamental variables are in general accounting data and financial ratios on a firm-level. The reason that this kind of variable is included is because they represent the financial condition and performance of a specific company. The general idea behind fundamentals is that the figures from the financial statement of a firm should indicate if and when a company is likely to default. This also contributes partly to the determination of the spread. Fundamentals can be divided in five different categories. These categories represent the most important areas which can determine the probability that a distress event arises. Empirical evidence of different researches has proven that it is common to use at least one variable from every category. These five fundamental categories are leverage, solvency, liquidity, profitability and efficiency.

The idea behind market variables is that a company's value is not only determined by the firm-level factors, but also by the opinion of the market on the firm's financial health. Most commonly used to get a perspective of the market on a firm's health is its stock price. Different aspects of the stock price can be retrieved that can be useful for modelling. Aspects that can be considered are for instance the daily stock return and its standard deviation. These can all be derived from the markets history, so one could arrange a desired period set of historical data on the returns and the volatility.

Macroeconomic variables represent the economic environment where the company finds itself in. For instance in case of an economic recession, it is likely that there will be more defaults and that the spreads will widen. Some of the more commonly used macroeconomic factors in models are the interest rate, the inflation rate and the slope yield curve. These factors are also the first used by researchers when thinking about macroeconomic variables. However one can also think of including the Gross National Product (GNP) or the monthly S&P 500 return in the model. These last two factors can give an indication on what the status is of the economic situation in the current financial environment. Useful indices specifically set up to show the macro view on volatilities are for example the VDAX and the VIX. These less individual variables can show the overall macro trend where the company at hand is situated in.

3. DATA

3.1. Universe

The universe, from which the data will be retrieved in this research, has been set by my supervisors at KCM. Their intention was to investigate the sectors which they were dealing with in the CDS market. For this research the universe was restricted to European companies. This is their main target area. The addition of American companies would create an enormously large data set. The sectors taken into account are the Automotives, Basic Resources, Chemicals, Construction Materials, Consumer Products, Healthcare, Industrials, Traveling & Leisure, Media, Retail, Telecom and Utilities. This resulted in a total of 146 European firms that will be included in the research. The total list of all the companies is not displayed due to confidentiality. For this research, Insurance and Banks will not be grouped as Corporates. The reason for this distinction is because banks and insurance companies require a different approach in determining the CDS spreads. From interviews at KCM, it appeared that for an indication on the CDS spreads of corporates mostly the same factors were taken into account for different sectors, whereas for banks and insurance companies different factors were considered important in determining the CDS spread. The general opinion is that it is difficult to determine the credit worthiness of financial institutions by just looking at their fundamentals or their stock return for example. As especially appeared in the credit crisis of August and September 2008, the substance of the assets and liabilities of these particular companies is very complex. Even banks themselves do not exactly know what kind of loans, bonds, structured products, etc. they have in their books. The general thought in the last decades was that financial institutions were very trustworthy, large banks like the U.S. bank Lehman Brothers appeared to be in great debts and filed for bankruptcy during the crisis. Investors and other financial institutions were in quite a shock, and the so solidly-seeming trust was suddenly gone. This led to tremendous increases in the CDS spreads of financial institutions. In the past however, the spreads of banks were quite tight and not very volatile.

3.2. Variables

As variables to be taken into the model, the determination has mainly been based on previous studies like Hartog (2007), Collin-Dufresne et al (2001), Das et al (2007), Jakovlev (2007) and Greatrex (2008). Their findings of significant variables in the determination of CDS spreads show a lot of similarities. The fact that these articles are recent, adds to the level of confidence in adding these variables to this research. Other variables added to the research model have been derived from interviews at KCM. Their experience and insight in the CDS spreads are believed to be of much added value to this research. There were also some personal factors they thought to be influential. Some factors they take into account are sector specific, which for corporates would be excluded, but which are crucial for sectors like banking and insurance. The list of factors used for the corporates in this research is shown in table 3. The emphasis for the fundamental variables lies especially on the free cash flow, capital expenditure and ratios. The interpretation of these variables might differ from their usual definitions. The used definitions were set according to definitions set by KCM. In general, Net Debt/EBITDA and Free Cash Flow margin are their key indicators. For market variables, the main items are stock return and the implied volatility of the companies. The macro variables are quite common used variables with the focus on the European market. The main variables for this segment are the yield difference, the slope of the rate curve and the ISM Purchasing Managers index.

To give a more detailed insight in the variables of corporates, they are set out below. First, the fundamental variables for the tests will be described. The abbreviations behind the factors are the terms used in the regression formula (6).

- Net Debt / EBITDA (NDE) is considered to be an important variable for the CDS spread. In the case this ratio increases, it shows that the debt of a company is growing larger in comparison to the Earnings Before Interest and Depreciation and/or Amortization (EBITDA). This ratio gives an indication on what parts of a company's earning can the debt be fulfilled with. One can imagine that the higher this ratio is, the less confidence there is that a company can pay off its debts, which is expected to result in the spread to widen.
- Free Cash Flow (FCF) margin is set up according to the specifications given by KCM. This variable is considered as an indicator for the liquidity of the company. With an increase in the FCF to settle debt, the FCF margin increases and it is expected this is accompanied by a decrease in the spread.
- The interest coverage ratio (IC) is one of the indicators for the solvency of a company. The better the interest expense is covered by the EBIT, the better things look for a company. It is assumed that an increase in the interest coverage ratio is accompanied by a tightening of the spread.
- Income growth (IG) is a variable that speaks its name. When the growth is increasing, one interprets a company to perform well, which results in the expectation that this will decrease the spread.
- Current ratio (CR) is a ratio in the class of determining the liquidity of a company, where the main target is to describe the relatively short term in assets and liabilities on the balance sheet of a company. The higher the current ratio, the more liquid a company is considered to be. With this statement, the expectation is that an increase in the current ratio will give a lower spread.
- Return on Assets (ROA) and Return on Equity (ROE) are profitability indicators for a company. If a company has a high return, for both indicators, the company is considered to perform well in its profits. Such an increase in these ratios creates the expectation that the spread will decrease.

- EBITDA margin (EM) is also a profitability variable and has a similar indication as for the ROA and ROE. With an increase in the EBITDA margin of a company, the spread is considered likely to decrease.
- Debt ratio (DR) is a variable in the factor class of leverage of a company. This variable should give an indication in how the asset side of the balance sheet can cope with the liabilities side. When the debt ratio is increasing, this means that the liabilities side is increasing in comparison to the total assets. This increase in leverage is expected to increase the spread.
- Asset size (Size) is there to try to display the size factor of a company. The asset size should indicate how large a company is in comparison to other companies. An increase in the total assets indicates an increase in the size of the firm, which sets the expectation to a decrease in the spread of a firm.
- EBIT volatility (Evol) is considered as a measure of stability for a company. The more an Earnings Before Interest (EBIT) is in line with its EBITs in previous years, the more stable a company is considered to be. Stability brings trust in the world of investors, which sets the expectation that the spread will decrease when the stability increases.
- The dummy variable for the EBIT (EBITdummy) negative or positive is to give value to the event when a company is non-performing. When the EBIT is negative, the value is 1 and when EBIT is positive the value is 0. In the case a company is non-performing it is expected that the market will cause an increase in the spread.

The market variables are far outnumbered by the fundamental variables and macro variables in this research. The equity return (Ret3M) of a company is considered to be a leverage factor. Companies with high leverage are considered risky which can be displayed in the equity return. In contrast to the debt ratio, an increase in the return on company's stock will decrease the spread. For this research the implied and historical volatility were considered. After checking the correlation between the two, it was decided to only consider the implied volatility (Ivol3M) because of its statistical significance. As was the case with the EBIT volatility and as will be the case with every component showing its volatility, it is expected that a low volatility suggests a stable performance of the specific component and should be rewarded with a decrease in the spread.
The market component, described by Beta, is an interesting variable which is not often used in research articles. The variable sets out the spread of company *i* against the spread of the iTraxx main. The latter is an index on the average of the CDS spreads in certain universes. The expectation is that if the variable increases, the spread will also increase. The argument for this suggestion is that in case the spread of company *i* increases not as fast as the iTraxx main, the company is considered more secure than the market overall, which should boost the spread to decrease. In the case Beta increases, the spread of company *i* is increasing more than the iTraxx main, which indicates that investors require a higher risk premium than they need for the iTraxx main. Therefore the spread of the CDS is expected to widen even more.

Turning to the macro variables, the first variable, the spread difference (Sdiff) is not often mentioned in research articles. It was mentioned in the research of Jakovlev (2007) and it was found significant. When mentioning this variable, the supervisors of KCM were not familiar with this factor, but they were quite interested in the results of this variable and therefore it was included in this research. The spread difference (called spread difference in articles, but it is perhaps more a yield difference) is derived by subtracting the yield on AAA-rated bonds from the yield on BBB-rated bonds. The idea is that this difference should represent the risk premium which is required by investors for bearing the additional credit risk. With an increase in this difference, the risk premium required by investors has apparently increased which should result also in an increase in the spread.
Information on the business condition is one of the aspects that the slope of the rate curve (slope) is considered to be able to give an indication on. When the curve increases, the business conditions are considered more positive, giving power to the expectation that this will decrease the spread. For the risk-free interest rate (Rf), the same idea holds, with an expectation that an increase in the interest rate will decrease the spread. The volatility index taken into account in this research is the VIX. As mentioned earlier the general expectation is that an increase in the volatility will result in a widening of the spread.
ISM Purchasing Managers index is considered as a measure of economic condition. This is one of the leading and most followed indicators for the world economy. Since the majority of the companies used in this research are affected by the events in the world's economy, this indicator is applicable for this research. It is considered that if this index decreases it is a sign of decrease in the world's economy. So with an increase in this variable, it is expected that the spreads will decrease.
The last variable of the list is the credit score resulting from the Kempen Credit Score model (KCSM), set up by Rik den Hartog. The scores resulting from this model should represent the default probability of the implemented CDSs. As one can imagine, an increase in the default probability increases the credit score and most likely should increase the spread of a CDS.

To sum these expectations up of the expected signs in the regression, the variables with their signs have been set in table 3 on the next page. The variables are described as they are set in the regression formula (6). As an additional note, the predicted signs comply mostly with other studies like Das et al (2007) and Jakovlev (2007).

Table 3 – Variables and their predicted signs in the regression with the CDS spread as the dependent variable. i is for the company indication, t is the indication of time

Variable	Description	Data Source	Predicted Sign
NDE_{it}	Net Debt / EBITDA	Reuters	+
FCF_{it}	Free Cash Flow margin	Reuters	-
IC_{it}	EBIT / Interest Expense	Reuters	-
CR_{it}	Current Assets / Current Liabilities	Reuters	-
ROA_{it}	Net income / Total Assets	Reuters	-
ROE_{it}	Net income / Total Equity	Reuters	-
DR_{it}	Total Debt / Total Assets	Reuters	+
$Size_{it}$	Ln(Total Assets)	Reuters	-
IG_{it}	(Net income – previous quarter's net income) / Total Assets	Reuters	-
EM_{it}	EBITDA / Total revenue	Reuters	-
$Evol_{it}$	Volatility of EBIT over the last 4 periods	Reuters	+
$EBITdummy_{it}$	Positive or negative EBIT, 1 if negative or 0 if positive	Reuters	+
$Ivol_{it}$	Implied volatility, 1M/3M/6M/1Y call	Bloomberg	+
$Beta_{it}$	Spread i / spread iTraxx main	Datastream	+
VIX_t	Volatility index	Datastream	+
$Sdiff_t$	Yield Difference AAA corporates vs BBB corporates	Datastream	+
$Slope_t$	10yr German government bond – 2yr German government bond	Datastream	-
R_{f_t}	10yr German government bond	Datastream	-
Ret_{it}	Stock (equity) Return 1M/3M/6M	Datastream	-
ISM_t	ISM Purchasing Managers Index	Datastream	-
$KCSM_{it}$	Kempen credit score model	Reuters / Datastream	+

Some variables were left out and therefore not mentioned in the table above for three reasons. The first is that some of the variables derived from research articles were proven to be insignificant in several external tests. The second reason is that the added value of some of the variables for the tests was not clear and the third reason is that some of the variables had a level of correlation that was considered too high. It would be more valuable to apply only one of them as was the case with implied and historical volatility.

As mentioned in the previous section, banking and insurance companies require different factors to determine the CDS spread. Moody's uses for banks and insurance companies a so-called financial strength rating to determine the CDS spread. This rating is considered to be a good indicator and fits to be used as an indication on CDS spreads. Many of the variables however used to get this rating are qualitative and difficult to determine. Further interviews indicated that diversification is an important factor. The diversification of the funding and the business lines is seen as an important indication of the stability of a bank. Leverage is an important distinction between banks and corporates. Banks are in comparison to corporates highly leveraged (25%-40%). It is therefore of importance not to take leverage into a research for banks in the same way as is done for corporates. Because one could spend an entire research just on banks and insurance companies, this research only focuses on corporates.

3.3. Data

The data required for the variables in the corporates model are retrieved from several reliable resources. For most of the fundamental variables Reuters Knowledge was used to retrieve the data of the selected 146 European companies. Where sometimes data was not available Bloomberg was used to fill the missing gaps. Due to the fact that fundamental variables rely on the frequency of companies reporting, the variables are only available quarterly or semi-annually. For structural variables, Datastream and Bloomberg were used to compare and create a certain level of confidence on the data. Also for the macro variables and for the historical CDS spreads, Datastream and Bloomberg were used. This data can be retrieved on a daily basis. The data retrieved is for the period of January 1st, 2004 till July 31st, 2008. It would be better if it was possible to get data before 2004, but as mentioned in the beginning of the thesis most historical data of CDS spreads can not be found before 2004. At the time of gathering data, there was already data available for the months August and September. These data points were however left out of the regressions, because several variables depend on financial reports of the third quartile which was not available at the time. The fact that the month July is taken into account (being part of the third quartile) is because the effects of the third quartile are not considered to have an impact on the figures.

For the CDS spreads, as the dependent variable, the spreads of the CDS 5 year maturity were used. The reason for this is that the 5 year CDS is considered the most liquid of all the maturities available for CDSs.

For each variable, there has been a screening to find outliers that could harmfully impact the results. Per variable the data was checked and a lower and upper bound were set if necessary. Data exceeding the bounds would get respectively the value of the lower bound or the value of the upper bound. The bounds per variable can't be displayed due to confidentiality. An interesting issue was the variable Net Debt / EBITDA. As mentioned earlier, if the value of this variable is low, it should indicate that a firm is performing well and should thus have a positive relation with the spread. Some firms have reported a negative Net Debt in their financial reports. In combination with a positive EBITDA, this variable was a negative. A negative Net Debt indicates a positive situation for a firm, because it shows the firm has no Net Debt to pay, but to receive. If the variable goes negative, this indicates a low Net Debt / EBITDA and should therefore lead to a lower spread.

There are however also some cases were a firm reports a positive Net Debt, but in addition a negative EBITDA. In this case a firm is considered to be nonperforming and should result in an increasing spread. However, with the negative EBITDA, the variable will also turn to negative and would therefore contribute to a lower spread. In these cases, the variables were set to the upper bound as to indicate a high variable which should lead to a higher spread.

In the scarce events that both Net Debt and EBITDA were negative, these variable values were set equal to the upper bound value. It was considered more important that a firm was nonperforming in comparison to a negative Net Debt.

4. METHODOLOGY

4.1. Introduction

As described in chapter 2, there are several models available for the purpose of this research. Since it is the intention to use variables of the classes' fundamental, market and macro, some of the models are not useful in single-use like the structural model. A hybrid model would perhaps be a good suggestion to implement all the variables. Having the advantage of using a statistical model in combination with the Merton model creates the possibility to profit from the advantages of using these two models in one single model. Looking however at what the set of variables are for the analysis, the amount of fundamental and macro variables is more compared to the three structural variables. So one can wonder what the added value would be of using a hybrid model compared to a simpler statistical model. It is not surprising that the majority of research papers on CDS spreads use a statistical model, an OLS regression to be specific. The use of hybrid models would perhaps be more useful if there was extra data available for the structural variables or in the case that there were more structural variables included in the research. In this research the advantages of the statistical models are considered of more added value than introducing a hybrid model. In this research, we will therefore commit to the statistical model to test the variables to get an indication on CDS spreads.

Looking at the research questions, the main target is to retrieve an indication on CDS spreads and not on a probability. It is therefore not a plain decision to choose for instance a logit model instead of an OLS model. The simplicity, the descriptiveness and the number of times used in other research articles make the OLS model an applicable model for this research. The fact that it is a model that functions properly is for instance shown in the research article of Das et al (2007). In this particular article a similar model is applied, with however different variables for U.S. companies' dataset. The results of the OLS model appear to create a proper working model in determining CDS spreads. There is however one important issue in applying the OLS regression and that is that the model assumes the data to have a normal distribution. However when we look at the distribution of the CDS spreads, as shown in figure 4, the distribution does not seem to fit a normal distribution. Looking at one of the statistics of the CDS spread data; the skewness is 4.482085. The skew for normal distributions is zero. A positive skew gives an indication that a lognormal, gamma or Weibull distribution might have a better fit. After applying Probability-Probability (P-P) plots to several possible distributions, the lognormal distribution resulted in the best fit. In Appendix A1 the P-P plots of the normal and lognormal distribution are displayed. Applying the lognormal formula distribution:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\frac{\ln(x)-\mu}{2\sigma})^2}, \qquad (5)$$

where μ represents the mean and σ is the standard deviation, resulted in the distribution as displayed in figure 4. With the Chi-square test as a goodness-of-fit test, failing to reject the H_0 hypothesis with α=0.01, it was decided to apply the lognormal factorization on the CDS spreads in the OLS regression.

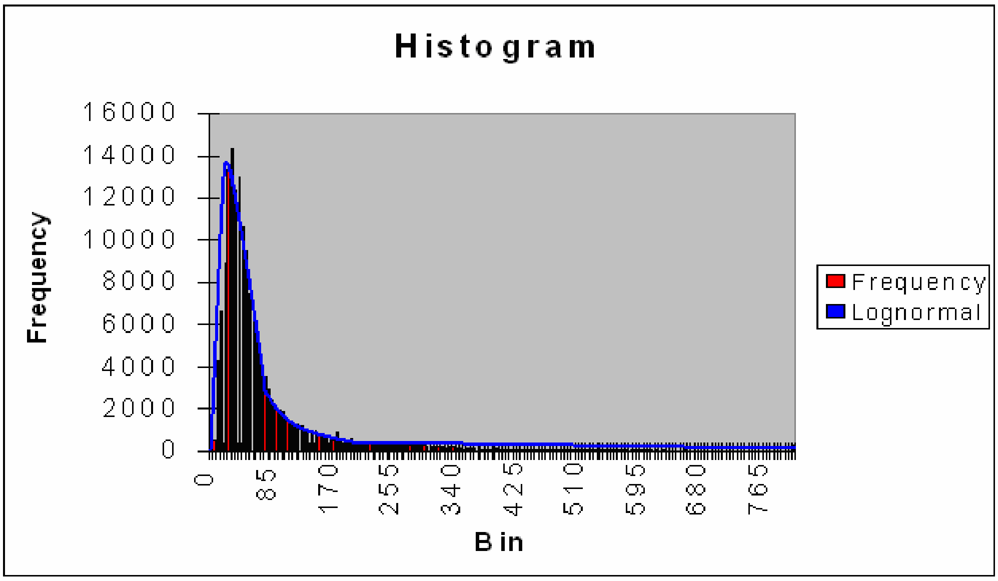


Figure 4 – This figure displays a histogram of all the gathered CDS spreads. The lognormal distribution has implemented in the same graph to display the fit. (μ=46.68, σ =0.8638)

One can not simply implement the values of the variables in one single regression if the variables are expressed in different meanings. For instance one variable could be expressed as a ratio, where another could perhaps be expressed as a value in a currency amount. To format all the variables in such a manner that they all can be implemented in one single regression model every variable had to be either a ratio, a lognormal value or expressed in points (of an index). Fortunately most of the variables are expressed or defined as a ratio. For the variables on size and on the volatility of EBIT, the lognormal value was taken as was also the case for the CDS spread. An index like the ISM is an index already expressed in points. With these characteristics the variables could all be implemented into one regression model.

Combined with all the other variables results into the following regression formula:

$$\begin{aligned} \ln CDS_{it} = {} & \beta_0 + \beta_1 NDE_{it} + \beta_2 FCF_{it} + \beta_3 IC_{it} + \beta_4 IG_{it} + \beta_5 CR_{it} + \beta_6 ROA_{it} + \beta_7 ROE_{it} + \beta_8 EM_{it} \\ & + \beta_9 DR_{it} + \beta_{10} Size_{it} + \beta_{11} Evol_{it} + \beta_{12} EBITdummy_{it} + \beta_{13} Ret_{it} + \beta_{14} Ivol_{it} \\ & + \beta_{15} Sdiff_t + \beta_{16} Slope_t + \beta_{17} r_{f_t} + \beta_{18} VIX_t + \beta_{19} Beta_{it} + \beta_{20} ISM_t \\ & + \beta_{21} KCSM_{it} + \varepsilon_{it} \end{aligned} \tag{6}$$

,where the β_0 represents the intercept and the ε_{it} is the error component. *i* is an indication for company *i* and *t* marks the time. CDS_{it} indicates the CDS spread for company *i* at time *t*. The other variables have been described in table 3. This formula is the starting point of the variables testing procedure.

4.2. Test for stationarity

For financial data, such as the time series of the variables in this research, it is common to not have a constant mean or variance. This indicates that the time series are likely to be non-stationary. Stationary series are time series where all of its statistical properties do not vary with time. If the data of the variables are non-stationary series, this could lead to drawing the wrong conclusions from the results of a regression. Non-stationary series could lead to wrong interpretation of the significance of several variables.

To test the variables for stationarity, the unit root test has been applied. This is a statistical test which accepts or rejects the null hypothesis of a time series having a unit root. If the null hypothesis is rejected, there appears to be no unit root in the series and so the series is considered to be stationary. There are several unit root tests available, but the most commonly used tests are the augmented Dickey-Fuller test (ADF) and the Philips-Perron test (PP). There is not much difference between the two except that the PP incorporates an automatic correction to allow for autocorrelated residuals. The tests however usually will give the same conclusion on the aspect of stationarity. The more negative the value is for a time series, the stronger the rejection will be of the null hypothesis that there is a unit root. To be complete, both two tests have been used to test the data for stationarity. In table 4 a sample of the variables are shown. As can be derived from the ADF and PP test, none of the shown variables are non-stationary at a 1% significance level. The test-statistic values are relatively quite negative and the statistical p-values are all close to zero. In line with these results are the unit root test results of the other variables which also appear to be stationary.

Table 4 – This table displays a sample of all the variables which were tested for stationarity. The upper test displays the ADF test and the lower results are from the PP test.

	NDE	FCF	ROA	DR	SIZE	BETA	RET3M	VIX	ISM
ADF Test Results									
Levels									
test-statistic value	-10,6251	-17,9763	-21,9881	-10,6713	-16,2270	-12,6692	-41,9052	-28,7173	-22,5692
p-value	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)
PP Test Results									
Levels									
test-statistic value	-11,5463	-27,5928	-22,9216	-11,0134	-15,5928	-11,1308	-38,2407	-30,9426	-14,2088
p-value	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)
Conclusion	**I(0)**	**I(0)**	**I(0)**	**I(0)**	**I(0)**	**I(0)**	**I(0)**	**I(0)**	**I(0)**

Significance is tested at the 1% level. Critical value for 1% level is both for ADF and PP: -3.4311

5. DESCRIBING MODEL

5.1. Introduction

In this chapter, the variables will be tested for their significance as being a determined of CDS spreads. In chapter 5 only the power to describe CDS spreads will be tested. The purpose of the results retrieved from these tests is to examine whether the model can create an indication on whether a spread is perhaps too high or too low. This mispricing between the actual spread and the fitted spread could lead to opportunities from which advantage can be taken. In the following chapter the predictive power of the variables will be tested.

5.2. Univariate analysis

Before trying to create a single model from a multivariate regression, it is interesting to examine how the variables relate to CDS spreads and to each other. All the variables have been individually tested in a univariate regression to get a first insight to the relation to the dependent variable CDS spread. Due to confidentiality, these results are not allowed to be displayed in this public thesis. Scatter plots have been created to get an indication on the relation of variables with CDS spreads. Three of these plots are displayed below.

Figure A: CDS spread and Size

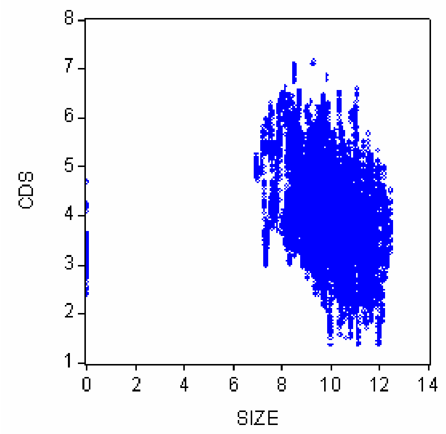


Figure B: CDS spread and Ivol3M

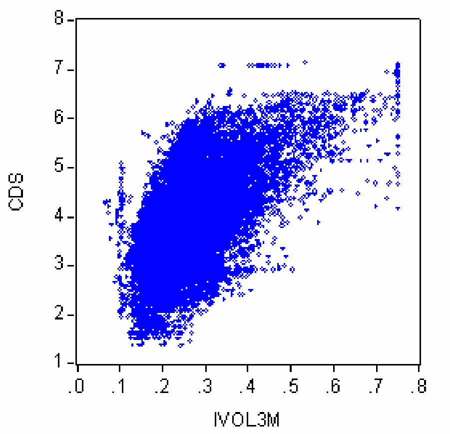


Figure C: CDS spread and Beta

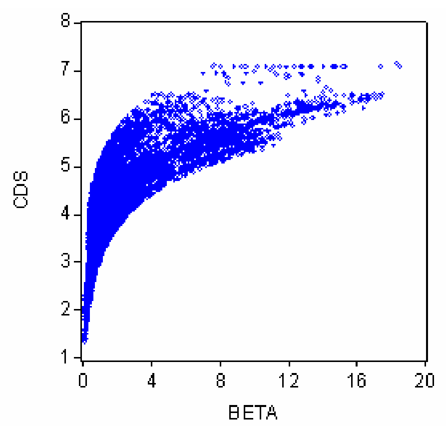


Figure 5 A, B and C – These scatter plots display the relation of respectively size, implied volatility and beta with spreads. As confirmation of the expected signs, size displays a negative relation and implied volatility and beta display a positive relationship with spreads.

Figure A and B of the scatter plots show a large area of dots close together. The large area tends to display a direction indicating the relationship between the independent variables and the dependent variable. In figure C however, there appears to be something distinctive.
The shape of the dotted area suggests that the CDS spreads are bounded by the independent variable Beta. This scatter plot however has this shape due to the fact that in the definition of Beta, the CDS spread is divided by the iTraxx main. The reason that the scatter plot displays a curve and not straight lines is due to the dependent variable CDS spread being expressed in lognormal values.

From the univariate regressions one can already get an indication on the level of significance of a variable as being a determinant for CDS spreads. According to the univariate regression the variables: Current ratio, Income growth, Slope and surprisingly Free cash flow margin are not significant at a 1% significance level. The insignificance of Free cash flow margin is a surprise because in general this variable is used as a strong indicator for the spreads of CDSs. The insignificance of the yield curve slope is a bit of a surprise as well, since some research articles found that this variable ought to be significant. The statistical insignificance of these variables is difficult to explain. Especially for the Free cash flow margin and the Slope it may be due to the set up of the variables. These two variables are often used as determinants for CDS spreads, however not always with the same definition. In the case of Free cash flow margin, the specific definition of KCM has been used. This can cause differences in regression results. With a different set up of these variables in their definition the statistical significance outcome might be different from the results retrieved in this research. It has to be kept in mind that although these univariate give a good first indication, variables might display different results in significance when combined with other variables.

Some other surprising results are the signs of the coefficients of the following variables: Free cash flow margin (FCF), Current ratio (CR), EBIT volatility (Evol) and the slope yield curve (Slope). For the Slope, strong arguments can be found that the sign of their coefficients should be positive instead of negative. When the economy is going through difficult times, people tend to invest more in Treasury bills than in corporate obligations. The effect of this switch is that the prices of these bills increase, which consequently decreases the short-term interest rate. The long-term interest rate will also be affected, but will decrease not as strongly. The slope is defined as the difference between the long-term rate and the short-term rate. As a consequence of the situation described, the slope will therefore increase.
That the sign of the coefficient of Evol is negative is quite surprising. As mentioned earlier in chapter 3, when a company's performance or the market index is volatile, the situation is considered unstable which could result in spreads to widen. A possible reason for a negative sign is that this variable, functions more as a size indicator than as a stability indicator. In case a firm has a relatively large Evol, it is perhaps assumed that it has this large fluctuation in the EBIT since it is a large firm that has a relatively high EBIT compared to other firms.

In other words; a firm has to be large of size to achieve relatively a large EBIT volatility. As the variable Size has a negative sign, so should, according to this argument, also the variable Evol have a negative sign.
The reason for the FCF to have a positive sign is less complicated. In the last period of 2007 and in 2008 a large amount of firms have reported a negative free cash flow. This in turn results in negative FCF values, which consequently has changed the sign to positive. The reason that the sign of Current ratio is positive is most likely due to a coefficient very close to zero.
From the univariate regression results, no direct consequences have been drawn as to which variables should be taken in the multivariate regressions. An important factor for implementing variables in a multivariate regression is the correlation between variables. If for instance the correlation between two variables is high, it would be inefficient to implement both variables in the multivariate regression, because their explanatory power would most likely be quite similar. In the correlation-matrix shown in table 5 (next page) there are some correlations marked either by having a black line around them or by being completely red. The correlations with a black lining are considered high. These are correlation values below -0.6 and above 0.6. The correlations marked red are logically high correlated because these variables contain the same information but set up with a different rounding period. These variables are the implied volatilities and the equity returns, which were not intended to be implemented together with other implied volatilities or equity returns in a multivariate regression. As explained as results from the univariate regressions, the variable Size and Evol have a high correlation. These two variables will therefore not simultaneously be included in a regression. Also the ROA and ROE have a high correlation. This is most likely due to the strong relation between assets and equity. From the univariate regression and from discussions at KCM, it has been decided to drop the variable ROE and only apply ROA in a regression.
The slope and risk-free rate also appear to have a high correlation, which is mainly due to the fact that the risk-free rate is used in calculating the slope. Although the slope was found insignificant from the univariate regression and it has this high correlation, the variable is not dropped because there is interest to see how the slope works in combination with other variables.

The high correlation between Beta and CDS spreads can be explained by the definition of Beta. For Beta the CDS spread of company *i* is divided by the spread of the iTraxx. Checking the correlation between Beta and CDS spreads is therefore not surprisingly high. Due to the high correlation and the set up of Beta, it has been decided to insert the Beta in every regression as an important indicator. The high correlation of Sdiff with the VIX and the ISM is perhaps less straight forward. The Sdiff has probably a high positive correlation with the VIX because in case the economy is unstable, the VIX will be high. In unstable periods investors require a higher risk premium for the risk they are taking and therefore the

Sdiff will also increase. The ISM is an indicator of how the macro environment is performing. If the macro environment is performing well, then the ISM will most likely increase. In a well performing environment, investors tend to require less risk premium and so the Sdiff decreases. Since the VIX is also correlated with the implied volatilities, the decision has been made to only implement VIX or implied volatility to the regression as a volatility measure without concerning whether Sdiff or ISM is already in the regression. Since the relationship between Sdiff and ISM is highly correlated, the two variables will not be together in a regression model.

Table 5 – Correlation matrix – This table displays the correlations of the independent variables with each other and with the dependent variable. The red marked figures are by definition high. The squared figures have a correlation lower than -0.6 or higher than 0.6.

	CDS	NDE	FCF	IC	CR	ROA	ROE	DR	SIZE	IG	EM	EVOL	EBITDU	IVOL1M	IVOL3M	IVOL6M	IVOL1Y	BETA	VIX	SDIFF	SLOPE	RF	RET1M	RET3M	RET6M	ISM	KCSM
CDS	1,00																										
NDE	0,10	1,00																									
FCF	0,06	-0,08	1,00																								
IC	-0,28	-0,25	-0,02	1,00																							
CR	-0,07	-0,28	0,13	0,08	1,00																						
ROA	-0,32	-0,27	0,00	0,27	0,05	1,00																					
ROE	-0,23	-0,13	-0,02	0,13	-0,08	0,78	1,00																				
DR	0,16	0,38	-0,09	-0,25	-0,37	-0,13	0,22	1,00																			
SIZE	-0,32	0,19	-0,08	0,10	-0,21	-0,01	-0,03	-0,04	1,00																		
IG	-0,06	-0,15	0,04	0,05	-0,01	0,53	0,37	-0,06	0,04	1,00																	
EM	-0,24	0,04	0,07	0,12	0,02	0,33	0,29	-0,16	0,21	0,21	1,00																
EVOL	-0,29	0,04	-0,03	0,03	-0,06	0,07	-0,03	-0,15	0,64	0,02	0,18	1,00															
EBITDUMMY	0,12	-0,07	0,02	-0,18	0,04	-0,22	-0,21	-0,04	0,02	-0,09	-0,17	0,09	1,00														
IVOL1M	0,57	-0,15	0,05	-0,06	0,05	-0,15	-0,18	-0,03	-0,13	-0,04	-0,24	-0,15	0,07	1,00													
IVOL3M	0,59	-0,17	0,05	-0,06	0,08	-0,16	-0,19	-0,04	-0,13	-0,03	-0,25	-0,15	0,07	0,96	1,00												
IVOL6M	0,60	-0,17	0,05	-0,05	0,09	-0,15	-0,19	-0,04	-0,12	-0,03	-0,25	-0,14	0,08	0,94	0,98	1,00											
IVOL1Y	0,58	-0,20	0,06	-0,04	0,10	-0,14	-0,18	-0,04	-0,10	-0,02	-0,27	-0,13	0,09	0,88	0,93	0,95	1,00										
BETA	0,66	0,01	0,04	-0,25	0,03	-0,29	-0,25	0,08	-0,37	-0,01	-0,20	-0,30	0,21	0,22	0,24	0,24	0,24	1,00									
VIX	0,41	0,01	0,05	0,01	-0,05	-0,01	-0,02	-0,02	0,07	-0,05	0,00	-0,01	-0,01	0,59	0,58	0,59	0,55	-0,10	1,00								
SDIFF	0,51	0,03	0,07	0,03	-0,06	-0,04	-0,05	0,01	0,07	-0,06	0,00	0,02	-0,03	0,56	0,58	0,60	0,57	-0,08	0,72	1,00							
SLOPE	0,08	0,00	0,01	-0,02	-0,02	-0,02	0,02	0,01	-0,04	-0,04	0,01	0,04	0,01	-0,10	-0,13	-0,15	-0,16	-0,03	-0,08	-0,01	1,00						
RF	0,04	0,00	0,01	0,02	-0,01	0,03	0,00	-0,02	0,07	0,02	0,00	-0,04	-0,01	0,31	0,33	0,35	0,34	-0,02	0,43	0,21	-0,61	1,00					
RET1M	-0,22	0,01	-0,05	0,01	0,02	0,07	0,07	-0,01	0,03	0,02	0,01	0,04	-0,01	-0,35	-0,33	-0,32	-0,28	0,00	-0,45	-0,26	0,00	-0,15	1,00				
RET3M	-0,33	0,00	-0,08	0,02	0,03	0,13	0,13	-0,01	0,05	0,06	0,01	0,07	-0,02	-0,45	-0,45	-0,45	-0,39	-0,04	-0,52	-0,46	-0,10	-0,13	0,62	1,00			
RET6M	-0,42	0,01	-0,08	0,03	0,03	0,19	0,19	0,00	0,06	0,09	0,04	0,09	-0,05	-0,49	-0,51	-0,51	-0,47	-0,07	-0,51	-0,58	-0,02	-0,14	0,50	0,78	1,00		
ISM	-0,27	-0,03	0,00	-0,03	0,02	-0,01	0,03	0,01	-0,07	0,00	0,00	0,01	0,02	-0,40	-0,42	-0,44	-0,44	0,03	-0,54	-0,64	0,38	-0,44	0,14	0,26	0,35	1,00	
KCSM	0,47	0,12	0,04	-0,17	-0,12	-0,29	-0,19	0,23	-0,03	-0,07	-0,09	-0,11	0,12	0,55	0,52	0,51	0,44	0,12	0,56	0,53	0,02	0,15	-0,41	-0,57	-0,59	-0,33	1,00

5.3. Multivariate analysis

5.3.1. Serial Correlation

We now start experimenting with the remaining variables with the constraints set in chapter 5.2, in several multivariate regressions. The Durbin-Watson test shows that there could be serial correlation in the residuals of the variables. The lower bound value of 0 indicates according to Durbin-Watson a high negative serial correlation, a value of 4 indicates a high positive serial correlation, and the value 2 indicates that there does not appear to be any serial correlation. A possible reason for this serial correlation indication might be the fact that the values of the fundamentals are kept fixed over similar periods.

Serial correlations can affect the results of a regression (Pindyck and Rubinfeld 1997). However, the serial correlation does not affect the unbiasedness of the coefficients of the variables. It will however, affect their effectiveness in fitting a model. In case of a positive serial correlation, the error terms will be estimated smaller than they actually are. These small residuals will tend to give the idea that the coefficients are performing quite precise. This situation is displayed in figure 6.

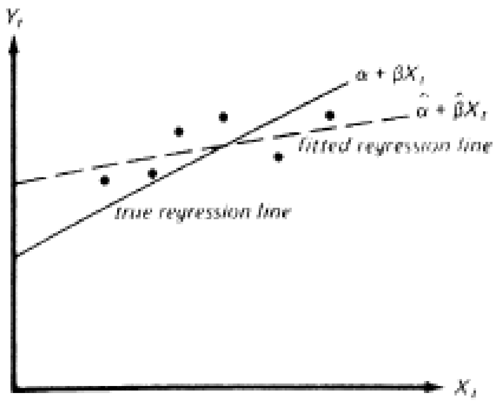


Figure 6 – This graph displays the case where there appears to be a positive serial correlation. The dots are the actual error terms. This is merely an example.

Figure 6 displays how a fitted regression can differ from the true regression line due to serial correlation. What appears to happen is that the fitted regression line, fits the points more precise which will result in a R^2 that is larger than it should be, and the error variance will be smaller than in reality. Due to the serial correlation, the fitted regression coefficients will not comply with how the actual regression coefficients should be.
To decrease the effect of serial correlation in the regressions, the Newey-West technique has been applied to every regression. This technique is available in the used econometric

software Eviews[1] for running the regressions. The Newey-West technique will directly adjust the error terms of the regression estimates to take account of serial correlation. Due to serial correlation, it is possible that in a regression certain variables are found insignificant which in reality are significant and vice versa. This issue is resolved by the Newey-West technique.
Furthermore, the data sets have been tested in randomized order, meaning that all the dates and companies have been shuffled. Of course all the variables per date have been kept the same, only the arrangement of the dates and companies have been shuffled. This resulted in a Durbin-Watson statistic value of 1.8068, which is in a fair range of 2 so there appears to be little serial correlation.

5.3.2. Fundamental, market and macro regressions

To start with the multivariate regressions, first the distinction has been made between the fundamental, market and macro variables. The supervisors at KCM added as specification requirements for the final model that it should contain about 6 variables, from which only one macro variable and preferably not more than two market variables. With the distinction made for the regressions, an indication could be found on how certain variables work together in a regression and how powerful they are in describing CDS spreads. As stated earlier, although this distinction between the three categories of variables has been made, the variable Beta has been implemented in every regression. The variable VIX was used in both the macro segment and in the market segment, to check it performance compared to implied volatility. Examples of some regression models found are not displayed in this public thesis due to confidentiality. An important conclusion that can be drawn from regressions in the macro segment is that the variable Sdiff tends to outperform the other variables only in combination with Beta and in combination with more variables. The KCSM and the risk-free rate appear to perform the worst as macro variables in the regressions.

From the fundamental variable regressions, it appeared that ROA and Size are very powerful factors in describing CDS spreads. ROA clearly seems to outperform Evol. The variables FCF, CR, EM and EBITdummy are found to have small effects. IG works reasonably as a factor, but is outperformed by the variables NDE and DR. The NDE and DR seem to add value in the regressions however, only when they are not in the same model.

As results from the market variable regressions, the VIX and implied volatilities achieve high R^2s, but not with each other. It appeared that the implied volatility of 3 months (Ivol3M) achieves the best results compared to the other implied volatility variables. The equity return of 3 months (Ret3M) also showed better results than the equity return of 1

[1] Eviews 5.1, © 1994-2005 Quantitative Micro Software, Enterprise Edition Feb 2007, www.eviews.com

month and 6 months. However, the Ret3M showed little significance in the regression models.

5.3.3. Regression models

As a consequence of these regression results, it has been decided to use the following variables for further multivariate analysis without the distinction between fundamental, market and macro variables: NDE, ROA, DR, Size, Ivol3M, VIX, Sdiff, ISM and Beta. The constraints in the use of these variables are that the variable pairs NDE and DR, Ivol3M and VIX, Sdiff and ISM may not be used simultaneously in one regression. The regression results of all the possible models are not displayed due to confidentiality.

M1: ROA, DR, Size, Ivol3M, Sdiff, Beta
M2: ROA, DR, Size, Ivol3M, ISM, Beta
M3: ROA, DR, Size, VIX, Sdiff, Beta
M4: ROA, DR, Size, VIX, ISM, Beta
M5: NDE, ROA, Size, Ivol3M, Sdiff, Beta
M6: NDE, ROA, Size, Ivol3M, ISM, Beta
M7: NDE, ROA, Size, VIX, Sdiff, Beta
M8: NDE, ROA, Size, VIX, ISM, Beta

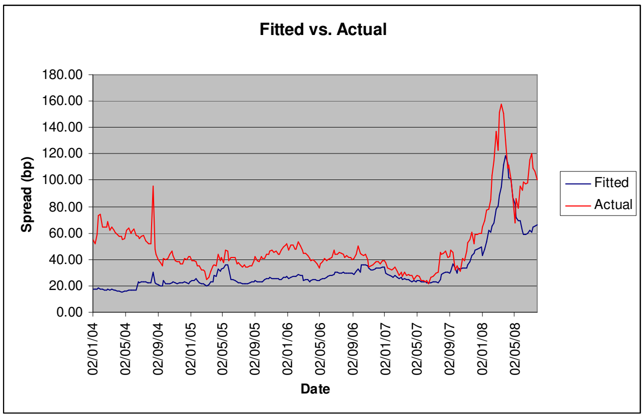
The models with the variable ISM perform worse than when Sdiff is included according to the R^2, so ISM will be dropped and Sdiff will be used as a variable in the final model. The performance difference between NDE and DR or Ivol3M and VIX is difficult to determine. The NDE seems to perform best in combination with Ivol3M. DR achieves the highest results in combination with VIX. The decision for the final model will therefore go between model M3 and M5.

To try to check the robustness of both models, they both have been applied in an out-of-sample check. In this check, the data of 2008 is considered to be unknown and the regressions are applied to the data from 2004 to 2007. The coefficients and adjusted R^2 however do not differ much from the previous results. They both tend to equally perform on the out-of-sample check.

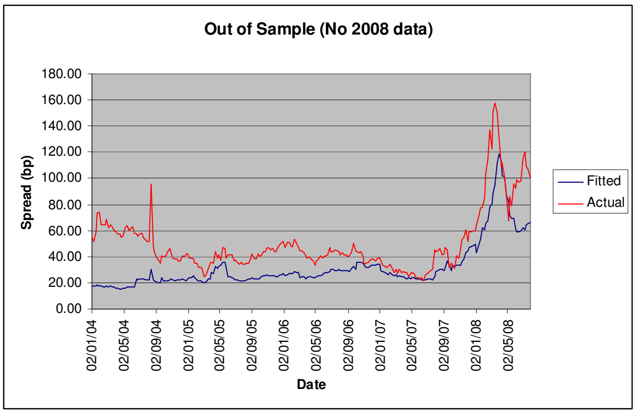
When applying both M3 and M5 to data on 2008 Q3 as an out-of-sample test, it appears that M5 outperforms M3. Especially in cases where companies show a large increase in spreads, M5 tends to remain in a small range of the spread, whereas M3 mostly overshoots the spreads, leading sometimes to spreads of 1000+ bp. In the out-of-sample tests the adjusted R^2 of M3 and M5 are respectively 0.71 and 0.78. The most logical explanation for this issue is found in the difference in applying VIX or implied volatility. While VIX is an indication of the volatility of the market, implied volatility is a company specific estimate. As in 2008 Q3 the market was very volatile, this is represented in the VIX by a high score. However, not all the companies in this research experienced the same volatility. In this case implied

volatility gives a better representation of how a company performed in 2008 Q3. Since the expectations of KCM are that the coming periods will still contain turbulence and that the spreads will remain very volatile, the decision has fallen to apply M5 as the model for describing a CDS spread. Derived from the test results, this model could give an indication on whether a spread is considered by the model to be too tight or perhaps too wide.
To give an indication of how a CDS spread is described by the model on a company, there are three graphs displayed below. The first gives the result of applying M5 to the normal data of a company. The second graph displays the out-of-sample case where the 2008 data is considered unknown. In the third graph M5 has been applied to the normal data and to the extra data of 2008 Q3.

Model example:



Out of Sample example:



Out of Sample example:

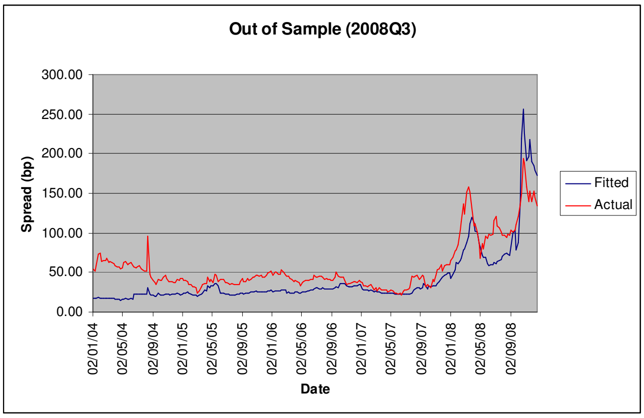


Figure 7 – These graphs display the difference between the actual spreads and the fitted spreads during a certain period.

As can been seen in the graphs, there appears in all three cases to be large difference at the beginning of the data till the beginning of 2005. This difference at the beginning is mostly due to the fact that for some of the variables, not all the data is available as for instance for the implied volatilities.

The very large peaks arise with the beginning of 2008. Especially in the out-of-sample case with 2008 Q3. The model seems to describe the movements at the end of the period quite well. It appears that all the three graphs tend to agree, that in the period before 2008 Q3 the actual spread is too high. In 2008 Q3, the model displays that the actual spread is perhaps too low.

5.3.4. Rank-Order Predictability

According to hedge fund managers and CDS traders, it is more important to get the relative ranking of the CDS spreads correct, rather than getting point estimates (Das et al 2007). To assess the rank-order predictability of the chosen model a cumulative accuracy profile (CAP) curve has been set up from which also the associated accuracy ratio (AR) statistic can be derived.

The CAP is set up by comparing the ranking of actual spread with fitted spreads. The fitted spreads and their corresponding actual spreads are ranked individually from highest to lowest. After this, 100 bins have been created to represent an increasing sample size. The first bin contains the top 1% of the fitted spreads. This bin is compared to the same bin with the top 1% of the actual spreads. The comparison is then made on how many of the fitted spreads in this bin have their corresponding actual spreads also in that same bin. This concept is then consequently applied to a second bin that represents the top 2% of spreads and so forth till the 100th bin, representing the total sample of spreads. If all things are set up in a correct manner, the 100th bin should result in a 100% overlap. Plotting the percentage for each bin results in a graph representing the cumulative accuracy profile as shown in figure 8. The AR is defined by Duffie, Saita and Wang (2005), as twice the area between the 45 degree line and the curve. The larger the area between these two lines, the higher the accuracy ratio of the model. An accuracy ratio above 50% is considered acceptable in general.

From several research articles the expectation was that the curve would start at a very low percentage and then would rapidly increase to eventually flatten at the end. In this case however, the model appears to achieve high rank-order predictability in the very first bin. Overall the graph seems to increase quite stable, however at the beginning and at the end there is a twist. Around the 10th bin, after first rapidly increasing the CAP decreases a bit before starting its upward straight line. Around the 90th bin, the CAP, it decreases before joining the 45 degree line. These two effects correlate with each other. In the total data set of actual spreads, there is some data missing from several companies. This is caused by the fact that not all CDSs have started trading from January 1st 2004. In ranking the total data set of actual values from highest to lowest, these missing points are set as the lowest values. Due to these missing parts, some values from the fitted spreads can't be compared correctly to their corresponding actual spread value. This effect mainly causes the two remarkable twists in the CAP.

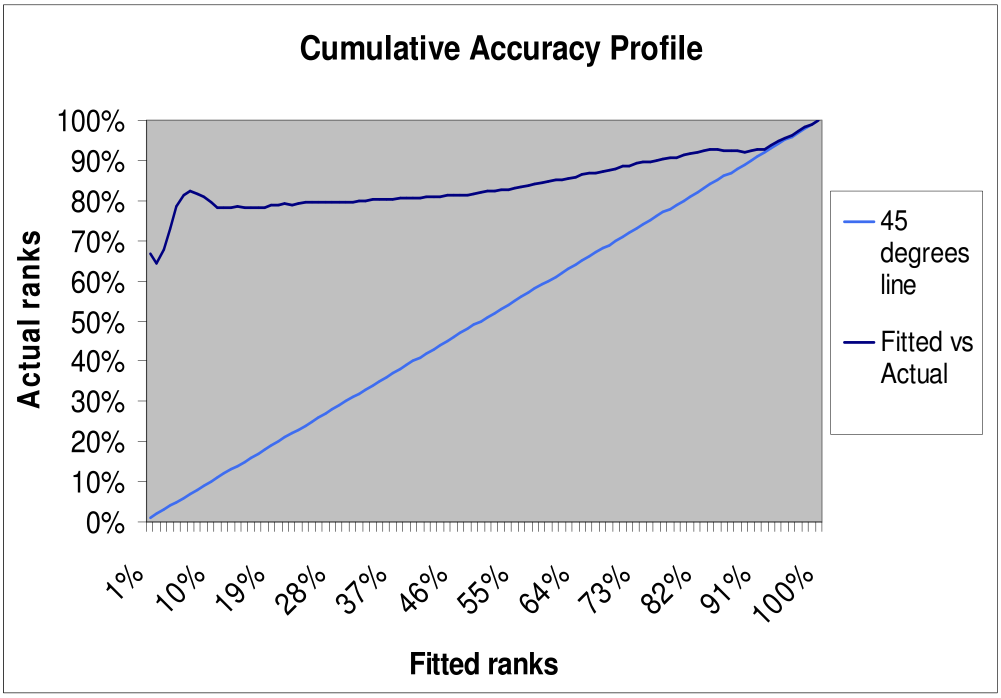


Figure 8 – The Cumulative Accuracy Profile graphs displays the rank-order probability of the data of the chosen model compared to the actual data. Twice the area between the 45 degree line and the curve is the accuracy ratio. The larger the area, the higher the accuracy.

The AR of the chosen model is 67.58%. As this is above 50%, this seems an acceptable model. Das et al (2007) describe the accuracy ratios of their accounting-based, market-based and comprehensive model. These results are respectively 56.7%, 56.5% and 61.6%. In this perspective comparing the result of the model from this research with theirs, the AR seems to be performing quite well. If some of the spread data weren't missing, the AR would most likely be even higher.

With the results from this chapter, the formula to calculate the describing CDS spread has been set up. However this formula is not allowed to be displayed in this public thesis.

6. FORECAST MODEL

6.1. Introduction

In addition to the describing model, a model has been tried out with a more predictive character. This model should attempt to give an indication on where the CDS spreads in the future might be and what their movements in future time are.

For this purpose the same variables as used for the describing model are taken into account, with the exception of Beta. Beta has been removed from the variables list because in the first testing phase, it appeared that the information that Beta contains, would create too much bias in predicting CDS spreads.

In addition to the list of variables, the supervisors at KCM asked if three additional variables could be taken into account. These variables are the volatility of Earnings per Share (EPS) over previous periods (EPSvol), an EPS dummy variable (Edummy) and a dummy variable called Ratings. EPSvol together with Edummy were chosen to replace the previous variables Evol and EBITdummy. Since Evol did not work according to plan, the idea was to determine the earnings stability by using EPSvol instead. EPSvol would not have the side effects as Evol did, acting as a size indicator instead of a stability indicator. Therefore the expectation is that the relation of EPSvol with CDS spreads will be positive. In case the EPSvol is high, this should indicate an unstable company which in turn should lead to an increase in CDS spreads.

The EPS dummy could either have the value 0 or 1. 1 represents a decrease in EPS compared to previous periods. This variable should act as an indicator of bad performance. In case there is a bad performance this should lead to an increase in the CDS spread. Therefore it is expected that the Edummy will have positive sign for its coefficient in the regressions.

The variable Ratings is based on the rating scheme of Moody's. Several ratings give a score for the model as is displayed in table 6. As an influence on CDS spreads, it is assumed that the difference between Aa1, Aa2 and Aa3 is not significant enough to give each rating a different score. Also the ratings from Ba3 and below are considered as a group and receive a lower score. Those with no rating will get zero as a score. Ratings is a reflection of how the rating agency Moody's considers the credit worthiness of companies. It is expected that a high rating should result in a relatively low CDS spread.

Table 6 – The scoring scheme used for the variable Ratings.

Moody's rating	Aaa	Aa1	Aa2	Aa3	A1	A2	A3	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Not Rated
Score for model	11	10	10	10	9	8	7	6	5	4	3	2	1	1	1	1	0

For testing the variables in their predictive value three different forecasting horizons have been chosen namely 2 weeks, 1 month and 2 months. For testing, the independent variables are set up in the same situation as in the descriptive model. The dependent variable CDS spread in this case is considering the forecast horizons brought forward by 2 weeks, 4 weeks or 8 weeks. This way the independent variables are put in a regression with CDS spreads of the forecasting horizon in the future.

6.2. Results

6.2.1. Univariate Analysis

All the variables went through the similar univariate and multivariate tests as in chapter 5. In the univariate regressions, there weren't that many differences in coefficients and significances of the earlier used variables. The univariate regression results of the three additional variables with a 2 weeks, 1 month and 2 months horizon can't be displayed due to confidentiality.

From the univariate regressions on the additional variables it appears that the Edummy is not significant in any of the forecasting horizons. Also the sign of its coefficient is in conflict with our sign expectation. The EPSvol only seems to work according to expectations in the 2 months forecast. However its results in the adjusted R^2 are quite moderate. The only additional variable that achieves satisfactory results is the Ratings variable. On its own the variable already achieves an adjusted R^2 above 0.5 in all the forecasting horizons. As a consequence of these results, only the variable Ratings is taken into account in the multivariate regressions.

6.2.2. Multivariate analysis

After reviewing the previous results, it was decided to only apply several variables in a multivariate regression. These variables would be applied to all the forecast horizons. For the fundamentals this meant that only NDE, ROA and Size would be taken into the regressions. For market, the implied volatility was the only variable, for macro the Sdiff, ISM and Slope were used and as explained earlier the variable Ratings was tested. The following combinations were used in the models:

M1: NDE, ROA, Size, Ivol3M, Sdiff
M2: NDE, ROA, Size, Ivol3M, ISM
M3: NDE, ROA, Size, Ivol3M, Sdiff, Ratings
M4: NDE, ROA, Size, Ivol3M, ISM, Ratings
M5: NDE, ROA, Size, Ivol3M, Sdiff, Slope, Ratings
M6: NDE, ROA, Size, Ivol3M, ISM, Slope, Ratings
M7: NDE, ROA, Size, Ivol3M, Ratings
M8: ROA, Size, Ivol3M, Sdiff, Ratings

The models where the variable Ratings is involved appear to achieve higher results than the other models according to the adjusted R^2. When including Ratings in the regressions, the fundamental variables are affected. Especially NDE and Size are found insignificant when Ratings is used in a regression. The addition of the variable Slope is found not powerful enough to be kept in the final model.

The results of the 2 weeks, 1 month and 2 months forecasting horizon models can't be shown due to confidentiality.

In the case of 1 month forecasting, the results are quite similar to those of the 2 weeks forecast. One remarkable change can be found in the coefficients of ROA. The sign is still negative, but the value is close to zero, where at the two weeks forecast the coefficient was higher than zero. This however should not affect the predictability of the models. Since the values of ROA of the companies are itself not larger than 1, the effect in calculating the lognormal CDS spread is not too large.

For the 2 months the coefficients of the ROA look more like the 2 weeks forecast than the 1 month forecast. In general however, the results for the three forecasting horizons are performing in a similar fashion. In comparison to the describing models, the coefficients of Ivol3M and Sdiff are a bit different. The coefficient of the Ivol3M is a bit higher, the coefficient of Sdiff is a bit lower. A possible explanation for this is that the forecasting periods are looking even further in 2008, from which we now know to be a very volatile period. More weight is given to the implied volatility to take these circumstances into account. This increase in value is compensated by a decrease in value of Sdiff.

For further testing, the decision was made to only test the models M1 and M8. Although the adjusted R^2 of M1 is lower than the models with Ratings, the positive results achieved by the variables of M1 in the describing model in 2008 Q3 contributed to make the decision in taking M1 into account. This choice proved profitable. When applying both the models to a selection of companies including their 2008 Q3 data, M8 is outperformed by M1. As an example, in figure 9 the difference is displayed between M1 and M8 applied on a company. The name of this company is not allowed to be published due to confidentiality.

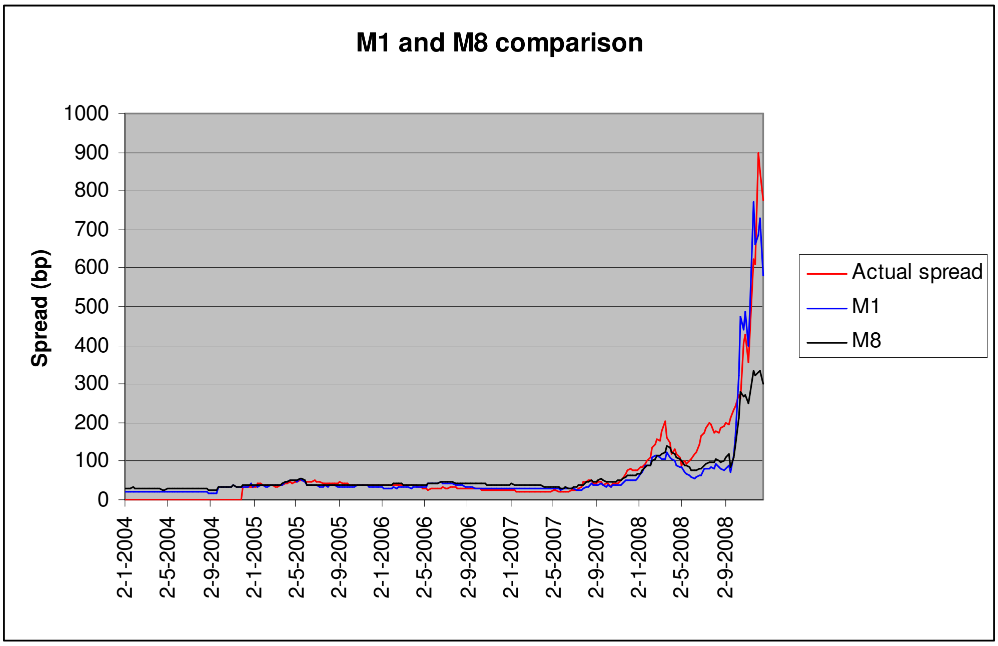


Figure 9 – This graph displays the difference between the actual spread of a company and the results from the models M1 and M8. Especially the period starting from 2008 shows quite some differences between the models.

The large difference between the two models is clearly made in the last period. Where the actual spread approaches 700 bp, the M1 tends to follow, but the M8 does not seem to achieve the same trend. This result is not only specific for this company, but for all tested companies. M8 can't keep up with the volatile market that especially is displayed in the latter part of 2008. When applying these two models in a regression including the data only of 2007 Q1 to 2008 Q3 for all time horizons, the M1 decreases just a little bit to an adjusted R^2 of 0.47 and the adjusted R^2 of M8 drops drastically to 0.38. Taking the regression results and the applications to a selection of companies into account, the decision has fallen to appoint M1 as the model to use for forecasting.

This decision results in formulas for calculating the CDS spreads that can't be displayed here due to confidentiality.

6.2.3. Applying the forecast model

Interesting is to see if the forecasting models could have some predictive power and could result in a positive result when using it as a short/long guide. A test has been set up to use the models on the period from 2007 Q1 to 2008 Q2. The trading rule set up to go either short when the model predicts that the CDS spread will increase or long when the model estimates a decrease in the CDS spread. This was tested on a weekly basis and only the top 10 and bottom 10 largest differences in spreads would be set to take a position. It was taken into account that once a position had been taken on a CDS, this position would have to be held for the forecasted period. So for instance when the 1 month forecast model indicates to go short on company X, this CDS could not be traded in the following 4 weeks and so would also be taken out of the top 10 for the coming 4 weeks. When applying all the 146 companies to every forecasting model, all applications resulted in a cumulative positive end result. Graphical results are not displayed due to confidentiality.

What draws the attention are mainly two things. The first is that in the year 2007 there is a loss when applying the model. There is very little profit from the shorts that year and the long positions achieve quite some loss. The losses due to longs after July 2007, is mainly due to that the model takes on long positions in companies that are classified as high yield. Companies that are called high yield have a credit rating below Baa3 (for Moody's) or BBB- (for S&P and Fitch). Companies with a rating equal or higher to BBB- are called investment grade companies.
The second part that draws attention is at the beginning of 2008. Here the short positioning is really paying off. After 2008 Q1 the long positions are also starting to contribute to the profit. The aspect that the shorts are achieving positive results is mainly due to the very volatile market and the increase in CDS spreads on average.

In case only the investment grade companies are taken into account, the three forecasting model all have a cumulative profit during the whole period starting from 2007 Q1. When extending the testing period by adding 2008 Q3 the profits are even a bit higher. Display of these results is not allowed due to confidentiality.

The total cumulative profit and loss displays stable and positive results for the whole period. In 2008 Q3 the longs are increasing in losses, however the shorts are even increasing more in profits. This is as explained earlier, mainly due to the increase in CDS spreads and that mostly long positions did not contribute to profits in reality.

In addition to this, we tested if it would matter if the trading rule would be changed to using the top 5 and bottom 5 of spread differences. This however did not result in any remarkable changes.

Comparing the cumulative profit and loss results of the three forecasting horizons, the 2 weeks is the most volatile. The longer the forecast horizon is, the more stable the results are. The total cumulative profit is also for the 2 weeks forecast the lowest, and for the 2 months the highest. The profit achieved by the 2 weeks forecast is also a bit too low if it were applied in reality. For every trade several costs like transaction costs are charged. This way there would not be much left of the profits from the 2 weeks forecast model.

Although both the 1 month and 2 months forecasting model achieve a high cumulative profit of respectively 2195.19 bp and 3213.9 bp, the decision for which model should be recommended in usage turned to the 1 month forecasting model. Although the 2 months forecasting model achieves more stable results, it is questionable if a forecast period of 2 months is not too long. The profit achieved by 1 month forecast is very satisfactory and gives the people at KCM more confidence in usage. Therefore the final forecasting model is set up in a 1 month forecasting horizon with the exclusion of high yield companies.

7. CONCLUSION AND FINAL THOUGHTS

7.1. Conclusion

As there are not many research articles on exploring the contents of CDS spreads, there is still a lot to explore. This thesis is a further step in this direction. With the usage of 146 non-financial European firms and a time period of 2004 Q1 to 2008 Q2, this research attempted to try to find variables that could describe or even forecast a CDS spread. It appeared that the combination of fundamental, market and macro variables are of influence to the CDS spread models. All the variables were selected either by results from other research articles or by advice of KCM. Most of these variables are already used in the Kempen Credit Score model. In this research, it appeared that some of the variables of Kempen Credit Score model were useful in describing and predicting the CDS spreads. Complemented with some additional variables the choice for the describing model after several OLS regressions, resulted in the following factors: NDE, ROA, Size, Ivol3M, Sdiff and Beta. Where the Sdiff and Beta give a good indication on how the market is performing, the other 4 variables give an indication on the firm-specific aspects required to describe a CDS spread.
With the tests for the forecasting models and the requirements of KCM, the choice had been made that the same variables with the exception of Beta were most suitable as a model. To avoid having a too long forecasting horizon, but also trying to keep stable results, the advised forecasting horizon is 1 month. The forecasting model appeared to achieve the best results when only applied on investment grade companies.

With the describing model, KCM will be able to get an indication on where the spread of a company's CDS should be and if the current spread is either too high or too low. The forecasting model has shown that it can be a successful model to use in predicting the CDS spreads. Tested over the period of 2007 Q1 to 2008 Q3, the forecasting model achieves a profit. One has to be careful however in putting all his faith in this forecasting model, since it is still a model. Based on the predictions of this forecasting model and the experience and common sense of the fixed income team, this model however could help setting up proper trading rules for dealing in CDS spreads.

7.2. Final thoughts

The satisfaction of how the model is performing in stable and more volatile times is very important. Both describing and forecasting models give a good insight in the position and movements of a CDS spread for non-financial firms. To find a model for describing the CDS spreads of financial firms like banks and insurance companies, the research should be set up with different sector specific variables. The variables in the current model are appropriate to use for non-financials. Firms, like banks require a different approach and a different definition of what their risk premium is.

The current describing and forecasting model are set up for non-financial firms covering different kind of sectors. Although the model is performing very well for these firms, more accuracy could perhaps be achieved by setting up models per sector. Some of these models will have overlap with the models from this research and some in addition will have overlap with each other, but several sectors will do distinct from each other. This distinction can perhaps be achieved by new coefficients for the variables or the implementation of more sector specific variables.

It will be interesting to see how the describing and forecasting model would work in the aftermath of the current credit crisis. I have a lot of confidence that the use of the current fundamental variables, the implied volatility and the yield difference between AAA corporates and BBB corporates can continue to give a proper indication on how CDS spreads are performing or what to expect of their movements.

With the proper usage and the common sense of a human being, the describing and forecasting model can be an effective tool in the world of CDS spreads. Unfortunately it is still no crystal ball.

REFERENCE LIST

- Altman E.I. (September 1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *Journal of Finance Vol. 23, No. 4, pp 589-609*

- Avramov D., G. Jostova (CFA), and A. Philipov (2007), "Understanding Changes in Corporate Credit Spreads", *Financial Analysts Journal, Vol. 63, No. 2*

- Benos A., G. Papanastasopoulos (June 2005), "Extending the Merton Model: A Hybrid Approach to Assessing Credit Quality", *EconWPA Paper Series, No. 0505020*

- Collin-Dufresne P., R.S. Goldstein and J. Spencer Martin (December 2001), "The Determinants of Credit Spread Changes", *The Journal of Finance, Vol. LVI, No.6*

- Das S.R., P. Hanouna, and A. Sarin (November 2007), "Accounting-Based versus Market-Based Cross-Sectional Models of CDS Spreads", *Working Paper Series, Villanova University and Santa Clara University*

- Duffie D. and K.J. Singleton (1999), "Modeling Term Structures of Defaultable Bonds", *The Review of Financial Studies, Vol. 12, Issue 4, pp 687-720*

- Duffie D., L. Saita and K. Wang (2005), "Multi-period Corporate Default Probabilities with Stochastic Covariates", *FDIC Center For Financial Research Working Paper No. 2006-05*

- Elton E.J., M.J. Gruber, D. Agrawal, and C. Mann (February 2001), "Explaining the Rate Spread on Corporate Bonds", *The Journal of Finance, Vol. LVI, No. 1*

- Fraas J.W. and I. Newman (2003), "Ordinary Least Squares Regression, Discriminant Analysis, and Logistic Regression: Questions Researchers and Practitioners should address when selecting an Analytic Technique", *Working Paper, Ashland University Library*

- Hartog R. den (2007), "Master Thesis: Constructing a European Credit Score Model", *Kempen Capital Management (Working paper)*

- Greatrex C.A. (March 2008), “The Credit Default Swap Market’s Determinants”, *Fordham University, Department of Economics Discussion Paper Series, No. 2008-05*

- Hull J., I. Nelken, and A. White (September 2004), “Merton’s Model, Credit Risk, and Volatility Skews”, *Journal of Credit Risk Vol. 1, No. 1*

- Jakovlev M (April 2007), “Determinants of Credit Default Swap Spread: Evidence from European Credit Derivatives Market”, *Master’s Thesis Lappeenranta University of Technology*

- Jarrow R. and S. Turnbull (1995), “Pricing options of financial securities subject to default risk”, *Journal of Finance 50, pp 53-86*

- Jarrow R., D. Lando, and S. Turnbull (1997), “A Markov Model for the Term Structure of Credit Risk Spreads”, *Review of Financial Studies, Vol. 10 , Issue 2, pp 481 - 523*

- Larsen R.J. and M.L. Marx (2001), “An Introduction to Mathematical Statistics and Its Applications”, *Prentice Hall International, Inc., Third edition*

- Litterman R. and T. Iben (1991), “Corporate bond valuation and the term structure of credit spreads”, *The Journal of Portfolio Management, Spring, pp 52 -64*

- Longstaff F.A., S. Mithal, and E. Neis (August 2003), “The Credit Default Swap Market: Is Credit Protection Priced Correctly”, *Working Paper Anderson School at UCLA*

- Merton R.C. (1974), “On the pricing of corporate debt: The risk structure of interest rate”, *Journal of Finance 29, pp 449 - 470*

- Morgan S.P. and J.D. Teachman (Nov 1988), “Logistic regression: Description, Examples, and Comparisons, *Journal of Marriage and the Family, Vol. 50, No. 4, pp 929 - 936*

- Ohlson, J.A. (1980), “Financial Ratios and the Probabilistic Prediction of Bankruptcy”, *Journal of Accounting Research, Vol. 18, No. 1, pp 109 - 131*

- Pindyck R.S. and D.L. Rubinfeld (July 1997), “Econometric models and economic forecasts”, *McGraw-Hill/Irwin, fourth edition*

- Sobehart J.R. and R.M. Stein (March 2000), "Moody's Public Firm Risk Model: A Hybrid Approach To Modeling Short Term Default Risk", *Moody's Investors Service Global Credit Research*

- Tudela M. and G. Young (2003), "A Merton-model approach to assessing the default risk of UK public companies", *Bank of England Working Paper No. 194*

APPENDIX A1.

Probability-Probability (P-P) plots of the CDS spreads fitted with a normal distribution and a lognormal distribution.

The data of the CDS spreads are tested on the probability scale by plotting the cumulative probabilities of the data under the assumed distribution against their expected probabilities.