Exploring a proxy to the CDS-Bond basis

Gustav Alfelt*

March 26, 2013

Abstract

This thesis investigates the behaviour of the CDS-bond basis during and after the 2008 financial crisis. It is found that the basis plunges deep into negative territory and that the theory of a zero CDS-bond basis is severly violated, not only during the crisis, but also in the subsequent years. Neither does it seem to return to the positive and less volatile levels observed before the crisis. Further it is investigated how an investor would have fared using CDS prices as bond risk proxy during the period, revealing an evident underestimation of risk in the proxy. Last a new proxy is constructed using linear regression models. The purpose of these model is to enable an investor to assess the riskyness of a bond that lacks reliable historical data by estimating the risk of the proxy in terms of Value at risk-quantities.

*I would like to thank professor Holger Rootzen, my advisor at the Department of Mathematical Sciences at Chalmers University, for his guidance and for sharing his knowledge. I am also thankful to my advisors at Nordeas office in Copenhagen, Jens Bisgaard and Allan Holm , for providing me with data as well as valuable comments and discussions. All errors in this presentation are my own.

Contents

1 Introduction

The CDS-bond basis is defined as the difference between the CDS price and bond spread with the same maturity written on the same financial entity. The relative stability of this quantity have laid ground for trading opportunities of arbitrageurs and, since there is often a lack of reliable bond data due to illiquidity, have also lead to the use of CDS prices to estimate bond risk.

During the recent financial crisis 2007/2008, financial markets experienced turmoil on a new scale. One of the most severely affected sectors was the credit market, where prices on lending and borrowing money is determined. Since the prices of CDSs and bonds are based on credit risk, these were naturally also greatly affected during the crisis. While the basis levels before the crisis fluctuated slighly above the zero-mark, these values would drop deep into negative territory. This could turn what at first seemed an arbitrage bargain into a great loss, or lead to major risk miscalculations.

This thesis will focus neither on the reasons behind this deviating basis behaviour, nor how basis trading was affected, but rather on two things: first it will investigate the impact of using CDS prices to measure bond risk during the crisis and its following years using a Value at risk-quantitiy and second; to search for a new, more reliable, way of proxying bond risk, using both CDS prices and several market indices. This proxy can then be used to estimate riskyness of bonds that lacks reliable pricing data by measuring the fluctuations of the proxy.

This report is organized as follows. Chapter 2 gives a basic introduction to the mechanics of bonds and CDSs, as well a quick review of the basis behaviour before and after the crisis years; chapter 3 explains what data have been used and how different quantities have been calculated; chapter 4 contains the findings of the data analysis, including Value at risk-measures using different methods and chapter 5 is a concluding chapter discussing the results, evaluating the methods and suggesting improvements of data set and methodology. Chapter A contains an Appendix.

For further reading on GPD and VaR-estimates, [4], [5] and [6] will be useful.[7], [8] and [9] sheds more light on potential factors used to model bond spreads and CDS-bond basis. [10], [11] and [12] further explores the dynamics of the CDS-bond basis, both post and pre-crisis. [13], [14] and [15] investigates the dynamics, liquidity and performance of trading basis.

2 CDS-bond relationship

2.1 Bonds

When a financial entity (corporation, government, etc) is in need of credit, it can issue a *bond*. A bond is a debt security, in which the issuer is obligated to pay the buyer or holder of the bond a principal amount, called the *par value*, at a pre-determined future date, called the *maturity date*. It usually also includes a periodic interest payment, called the *coupon*. Hence a bond is a loan from the bond holder to the issuer from the purchase date to the maturity date, with a recurring (for example annually or semi-annually) interest rate at the amount of the coupon.

Naturally, it is in the interest of an investor to know how much he or she would earn holding a bond until maturity, how large the *yield-to-maturity* (YTM) for the bond is. Suppose an investor buys a bond at time t_0, at price P, with annual coupon payments of amount C at dates $t_1, t_2, ..., t_{j-1}$ and $t_j = T$, which is the maturity date, and par value Bp. We then assume that the price is the present value of the future cash payments with each payment reinvested at the rate By, hence

$$P = C(1+By)^{-t_1} + C(1+By)^{-t_2} + ... + (C+Bp)(1+By)^{-t_j}. \qquad (1)$$

Solving this for By, we have determined what the yearly yield of the bond until its maturity is. Yield-to-maturity can be thought of as the interest rate the investor receives if he or she re-invests all coupon payments at this rate until the maturity date. It can be compared to investing the amount P into a money market account for T years, and will from now on be referred to as *bond yield*. Also note that as the bond price decrease, the bond yield will increase, and vice versa.

2.2 Credit Default Swaps

A *Credit Default Swap*, or CDS, is a bond derivative security, in which the seller of the CDS is obligated to compensate the buyer of the CDS in case of a default (because of bankrupcy, other failure to pay, restructuring, etc) of the underlying bond. In case of no default until the maturity of the CDS contract, the seller pays nothing. For this protection, the buyer has to pay the seller a periodic fee, usually quarterly, until either a default event or maturity, whichever occurs first. This fee, annualized, will be referred to as the *CDS price*. One can think of a CDS as an insurance on a bond.

2.3 Basis

There are numerous approaches to pricing credit risk, upon which bond and CDS prices depend. Most common are structural models based on the value of the firm or intensity-based models that assumes that the time of default is determined by a hazard rate. This thesis however, will rather focus on the theoretical no-arbitrage relation between CDS prices and bond yields.

Suppose an investor buys a bond, with yield By and maturity in T years. He borrows money to this investment from a lender at rate r. The investor will now enjoy a yield of $Bs = By - r$, called the *bond spread*, that is the premia earned from being exposed to the risk of the bond over the riskless rate r. Further, our investor proceeds to buy a T-year CDS on this bond from the CDS market at the cost CDS_p. Now, for there to be no arbitrage, the price of this CDS should be $CDSp = Bs$. This makes intuitive sense, since the investor have invested no capital (he or she borrowed the money to buy the bond) and bears no risk of the bond defaulting, and hence should not earn a profit. If $CDSp > Bs$ were true, the investor could write the CDS contract, short the bond and invest in the risk-free instrument r, and receive a risk-less profit. On the other hand if $CDSp < Bs$, the investor could borrow at r, invest in By and buy protection at $CDSp$ and again receive a profit at no risk.

The difference $CDSp - Bs$ is called the *CDS-bond basis* and as we have seen, its theoretical no-arbitrage value should be 0 *basis points* (bp). If any deviations should occur, arbitrageurs should quickly push the basis back to zero.

2.4 CDS as proxy for Bond risk

As we have seen, monitoring the CDS-bond basis can lead to profitable trading opportunities, but this relationship has other uses as well. Many bonds often have low liquidity and sometimes even lack pricing data. This makes estimating risks associated with the bond difficult and in itself risky. CDS contracts, however, are synthetic instruments which usually have higher liquidity than their bond counterparts. Hence they have more pricing data available and, as suggested by Blanco [1], much faster price discovery than corresponding bond spread (meaning that the bond spread tend to adjust to the CDS price rather than the other way round).

These properties, together with the theoretical CDS-bond basis value zero, enables an investor to estimate the risk of a bond using its CDS counterpart. This should make intuitive sense; if the CDS price and the bond spread are identical or closely follows, the CDS price and bond spread risk should be approximately equal. So assuming zero or small enough basis, an investor could use the time series of the CDS price to model the risk (with for example a Value at risk-measure) of corresponding bond spread. Note that in this case an actual investment in the CDS contract is not made, it is just used to estimate the risk of the bond.

2.5 Basis before and after 2007

In Blanco [1], the CDS-bond basis of 33 European and US firms during the period January 2, 2001 to June 20, 2002, is investigated. On a whole, the CDS-bond basis seems to revolve relatively close around zero, with an average of 14.6 basis points. Overall, the basis tend to be slightly positive, and Blanco suggests that CDS prices on average are an upper bound for bond spreads. This partly because exploiting a positive basis involves the difficulty of short selling a corporate bond.

The same relationship is investigated in Fontana [2], for 37 US firms in the period January 3, 2005 to April 1, 2009 (more than half a year after the "Lehman crash" on September 15, 2008). This period of 4 years and 4 months is divided in to three time spans. The average basis for the firms in these time spans are: 2.2 bp in the first, -17.2 bp in the second and -147.5 bp (-308.4 bp for the financial firms) in the third. Finger [3] extends this investigation on six financial bonds to September 2011, and finds that although the basis reverts from the levels seen during the financial crisis, they are still persistently negative and much more volatile than before 2007.

Clearly the zero basis theory that seemed to hold reasonably well before 2007 have been violated and the basis levels does not seem to return to low and stable levels either. Fontana [2] suggests a number of explanations to the basis behaviour during the crisis, among them a rising counterparty risk in the CDS contract, which during stable times is almost insignificant.

This thesis, however, aims not to explain the origin and persistance of the negative basis, but rather to investigate two other things. First, what the risks involved with using CDS prices to model bond risk post-2007 were. That is, to explore the model basis, and what additional risk it involved. Second, to search for a proxy for the basis itself, with the aim of constructing a more accurate post-2007 proxy for bond risk. The instruments investigated for this new proxy will, in addition to CDS prices, be EURIBOR, EUNIA, EURGovt, and the VIX. The purpose of this proxy is to be able to estimate the risk of bonds that lack reliable pricing data.

3 Data description

For the purpose of investigating the basis and model risk, a portfolio of 20 equally weighted bonds from European and US issuers is constructed. The portfolio have been selected to roughly represent the size of each sector on the bond market. The bonds are quoted in clean prices, that is, the market prices have been subtracted by the accrued interest $C * t/(t_{n+1} - t_n)$ (where t is the settlement date, t_n is the last coupon date and t_{n+1} is the next coupon date), to avoid the "mountain chain" that appear on the price graph from prices falling after coupon payments. Most bonds are from the banking sector, otherwise a cross section of bonds from different sectors have been made. The time period investigated is January 2, 2008 to March 23, 2012, and it consists of 1120 price observations. All bonds are quoted in euro and of fixed coupon type with annual coupon payments and par value 100 euro. Their complete description can be found in Table 1. Figure 1 and 2 contains box-plots of each bond spread and corresponding CDS price. It is easy to recognise that the data contains both high variation and a great deal of outliers. For most coporations the bond spread has a wider range of values than the CDS price. From each bond price, a daily yield to maturity By have been calculated[1]using equation 1.

As a interest rate benchmark, the Euro swap rate, as suggested by Blanco [1], of 9 different time lengths (1 month, 3 months, 6 months, 9 months, 1 year, 2 years, 5 years, 10 years and 30 years) have been used. Since the interest rate time length should match the lifetime of the bond, i.e. an investor must be able to borrow money to buy the bond, then hold both the bond and the loan until maturity, a synthetic rate have been made by interpolating rates of longer and shorter maturity than the bond. For example, a bond having 7.5 years to maturity have been matched by a rate $r_{7.5}$ constructed as $r_{7.5} = \frac{1}{2}r_{10} + \frac{1}{2}r_5$, where r_{10} and r_5 is the 10- and 5-year rate, respectively. For each bond, a daily bond spread Bs has been calculated as $Bs = By - r$ for the approriate rate r.

For each bond, 10 daily CDS prices with different lifetimes have been used: 6 months, 1 year, 2 years, 3 years, 5 years, 7 years, 10 years, 15 years, 20 years and 30 years. The CDS prices have been interpolated in the same way as the interest rates, in order to match their corresponding bonds' time to maturity. These constructed CDS prices can now be matched against their bonds' spreads Bs in order to determine the CDS-bond basis for the firm.

For 17 dates, bond data was present, but CDS and interest rate data absent.

[1]The YTM value have been calculated by computing the price of a wide range of YTM values, then interpolating these prices to match the given price. These values match the yield calculator on the Frankfurt Stock Exchange web site [16] with at least four digits.

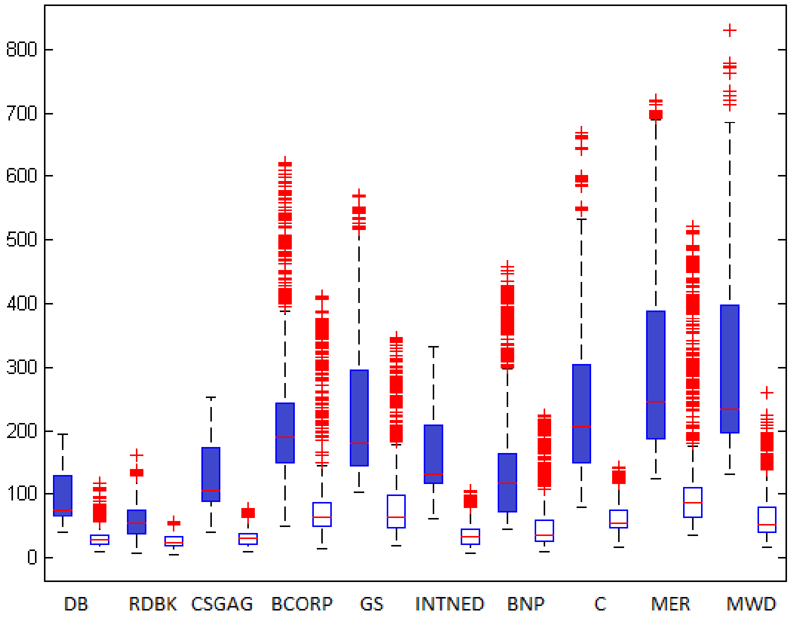

Figure 1: **Bond spread and CDS box plot.** *Box-plot of corporation bond and CDS. The filled boxes are bond spreads and the empty boxes are CDS prices. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. From left to right Deutsche Bank AG (DB), Rabobank Nederland (RDBK), Credit Suisse/London (CSGAG), Bank of America (BCORP), Goldaman Sachs Group (GS), ING Groep Nv (INTNED), BNP Paribas (BNP), Citigroup Inc (C), Merril Lynch & Co (MER) and Morgan Stanley (MWD).*

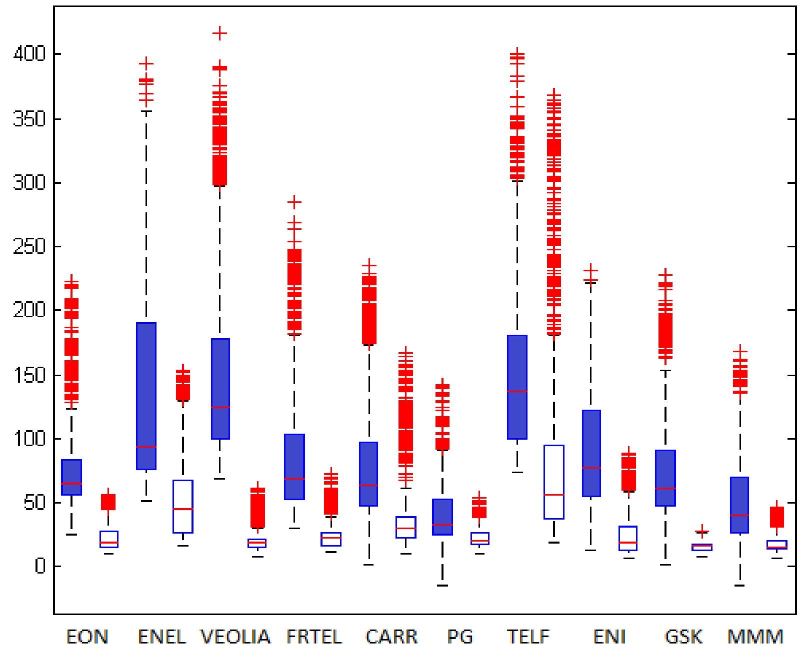

Figure 2: **Bond spread and CDS box plot.** *Box-plot of corporation bond and CDS. The filled boxes are bond spreads and the empty boxes are CDS prices. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. From left to right E.O.N International Finance (EON), Enel-Societa Per Azioni (ENEL), Veolia Enviroment (VEOLIA), France telecom (FRTEL), Carrefour (CARR), Procter & Gamble Co (PG), Telefonica Emisiones (TELF), ENI SpA (ENI), GlaxoSmithKline Capital (GSK) and 3M Co (MMM).*

Table 1: **Bond description.** *Names of bond issuer, industrial sector, bond coupon rate in percentage and time to maturity in years, from January 2, 2008.*

	Sector	Coupon rate	Time to maturity
Deutsche Bank AG	Banking	5.13	9.67
Rabobank Nederland	Banking	4.25	9.05
Credit Suisse/London	Banking	5.125	9.72
Bank of America	Banking	5.125	6.74
Goldman Sachs Group	Banking	5.125	6.79
ING Groep Nv	Insurance	4.75	9.42
BNP Paribas	Banking	5.431	9.69
Citigroup Inc	Banking	4.375	9.08
Merrill Lynch & Co	Banking	4.875	6.41
Morgan Stanley	Banking	5.5	9.76
E.ON International Finance	Utility	5.5	9.76
Enel-Societa Per Azioni	Utility	5.25	9.47
Veolia Environnement	Utility	5.125	14.4
France Telecom	Telecommunications	4.75	9.15
Carrefour	Consumer cyclical	5.125	6.78
Procter & Gamble Co	Consumer non-cyclical	4.5	6.36
Telefonica Emisiones	Telecommunications	4.674	6.10
ENI SpA	Energy	4.75	9.87
GlaxoSmithKline Capital	Consumer non-cyclical	5.625	9.95
3M Co	Capital goods	5	6.53

These data have been filled in using interpolationg between the previous and next trading date.

The indices to be used for the regression are EUNIA (3 month), EURIBOR (3 month), EURGovt (3 month) and VIX. For EUONIA and EURIBOR, 40 of the bond data dates were missing, for the EURGovt 17 and VIX 16, respectively. These dates have been filled in as the CDS and interest rate data, by interpolating values between the previous and next trading date.

4 Empirical analysis

This chapter will first investigate and make a statistical description of the relatiohship between the CDS prices $CDSp$ and bond spreads Bs, second what the discrepancies in risk measurement an investor would face using CDS prices as a proxy for bond risk and third how a new proxy for bond risk can be constructed by using regression models. The goal is to proxy the behaviour of bond yields such that an investor can use this proxy to estimate the risk of a bond that lacks reliable historical price data, a problem that is not uncommon among corporate bonds. For the analysis, the following definitions will be needed.

Definition 1. *Let $By_{i,n}$ and $Bs_{i,n}$ be the closing time bond yield and bond spread (calculated from equation 1), respectively, of bond i at trading day n, k the number of bonds in the portfolio, $CDSp_{i,n}$ be the closing time CDS price corresponding to bond i at trading day n and $r_{i,n}$ be the appropriate interpolated swap rate matching the maturity of bond i at trading day n. Then for every variable in this presentation, let*

$$\begin{aligned}
\Delta variable_n &= variable_{n+1} - variable_n \\
L1variable_n &= variable_{n-1} \\
Vvariable_n &= \frac{1}{k}\sum_{i=1}^{k} variable_{i,n}.
\end{aligned}$$

and

$$Vbasis_n = VCDSp_n - VBs_n,$$

So, for example, $\Delta By_{i,n}$ is the change in bond yield from day n to $n+1$ and VBs_n, $VCDSp_n$ and $Vbasis_n$ are the average bond spread, CDS price and basis across the portfolio. The subscripts i and n will sometimes be dropped from these variables for simplicity.

4.1 Statistical description

Figure 3 displays the average bond yield and swap rate of the bond in the portfolio. In Figure 4 graphs of the portfolio bond spread VBs_n, portfolio CDS price $VCDSp_n$ and portfolio basis $Vbasis_n$ can be seen. Notice two ”ravines” in the basis graph, one during the height of the financial crisis (mid 2008 to mid 2009), and one in a more recent period of distress (mid 2011 to march 2012). Also take note of how much smaller and less volatile

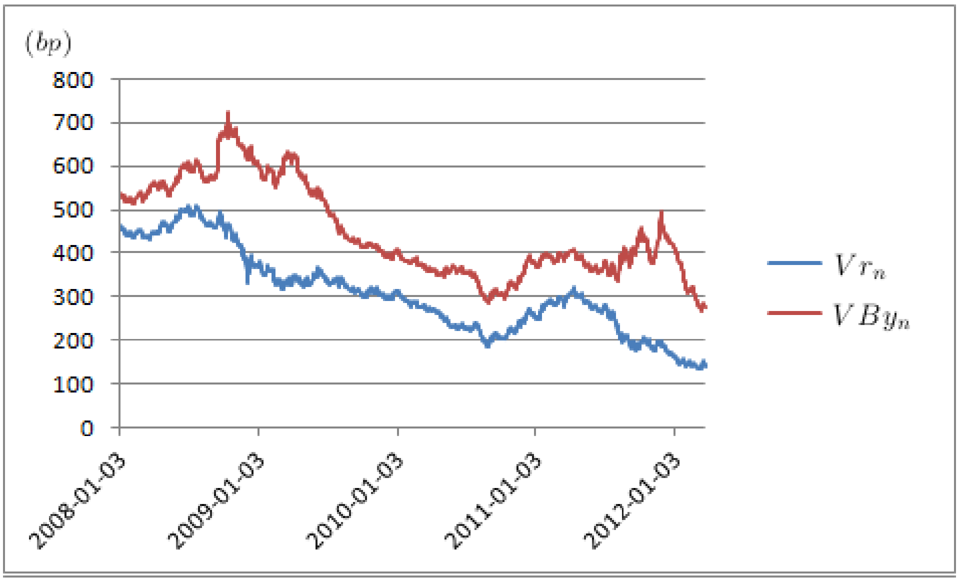

Figure 3: **Portfolio yield and rate.** *The average portfolio bond yield VBy_n and corresponding swap rate Vr_n for the 20 bonds in the portfolio.*

the average CDS price is, compared with the average bond spread.

Table 2 lists the average basis, basis volatility (namely the standard deviation of the daily changes in bp), $CDSp$ volatility, Bs volatility and correlation between $CDSp$ and Bs for each firm as well as the equally weighted portfolio. First, it is noticable that the zero basis theory is violated for all firms as well as the portfolio average. Neither does it agree with Blancos observations that basis on average is slightly positive. The values does however coincide with the findings of Fontana for the post 2007-period, namely that basis is largely negative. According to Figure 4 the basis is also persistently negative and does not return to positive levels during the period, also as concluded by both Fontana and Finger.

Furthermore, the difference in volatility of the portfolios $CDSp$ and Bs is rather large. The correlation does however suggest that for most firms as well as for the portfolio, the CDS price and bond spread moves in the same direction, but that is quite far away from 1, which would be the ideal value for a risk proxy. Still, the CDS prices does a poor job in capturing the volatile movements of the bond spread, which decidedly suggest against using CDS prices alone as a proxy to estimating bond risk.

Table 2: **Statistics on the firms basis, bonds and CDS prices.** *Each firms average basis, basis volatility σ_{basis} (standard deviation of the daily difference in bp), CDS price volatity σ_{CDSp}, bond spread volatility σ_{Bs} and correlation of CDS price and bond spread as well as statistics on the equally weighted portfolio.*

	Avg. basis	σ_{basis}	σ_{CDSp}	Corr($CDSp$, Bs)	σ_{Bs}
Deutsche Bank AG	-77.9	4.96	1.03	0.36	5.01
Rabobank Nederland	-46.22	4.36	0.54	0.85	4.28
Credit Suisse/London	-111.16	5.48	0.85	0.63	5.4
Bank of America	-174.89	10.77	3.36	0.84	11.21
Goldman Sachs Group	-190.2	16.18	3.88	0.59	17.19
ING Groep Nv	-140.25	8	1.55	0.56	7.83
BNP Paribas	-115.75	8.52	1.84	0.89	8.71
Citigroup Inc	-206.72	11.09	2.16	0.75	11.48
Merrill Lynch & Co	-247.96	13.66	4.26	0.61	14.41
Morgan Stanley	-279.78	23.87	3.91	0.73	25.05
E.ON International Finance	-69.39	4.65	0.38	-0.15	4.61
Enel-Societa Per Azioni	-107.69	5.92	1.51	0.9	6.01
Veolia Environnement	-146.08	5.67	0.46	0.2	5.69
France Telecom	-75.55	5.41	0.47	-0.01	5.43
Carrefour	-59.64	5.56	1.25	0.66	5.49
Procter & Gamble Co	-27.45	4	0.42	0.54	4
Telefonica Emisiones	-110.83	8.11	3.23	0.81	8.2
ENI SpA	-76.24	4.72	0.6	0.76	4.71
GlaxoSmithKline Capital	-71.51	4.95	0.23	0	4.94
3M Co	-41.53	5.85	0.38	0.51	5.83
Portfolio	-118.84	4.11	1.16	0.62	4.35

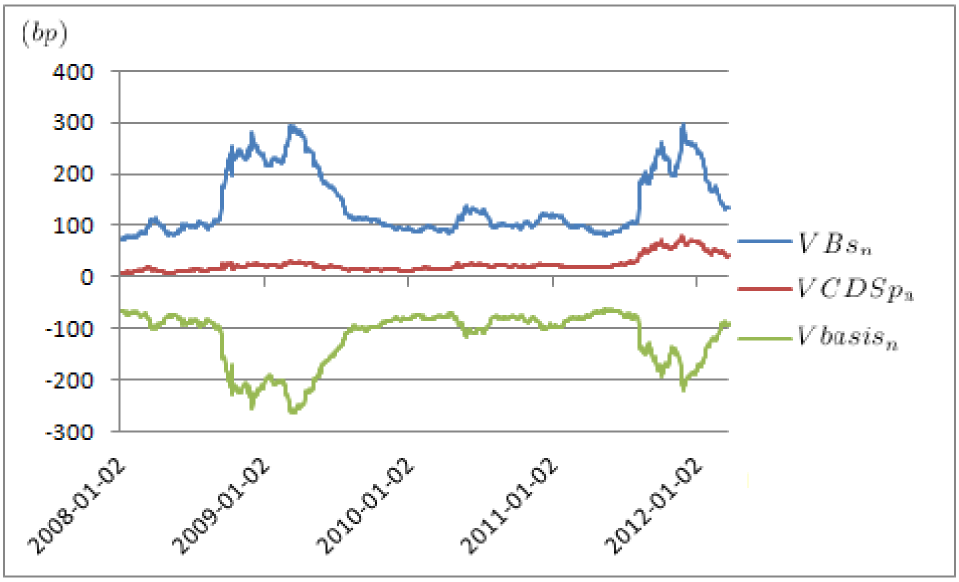

Figure 4: **Portfolio spread graphs.** *The average portfolio bond spread VBs_n, CDS price $VCDSp_n$ and basis $Vbasis_n$ for the 20 bonds in the portfolio.*

4.2 Value at risk

To estimate the risk associated with the bond, a *Value at Risk*-measure (VaR) is used. It is a commonly used risk measurement that tells what value is at risk of losing at a certain probability for a certain time span t, e.g. a single trading day. This is done by first observing the changes of a value V over each time period t in a time series, then fit these changes (which most commonly are return values of an asset, in percentage) to a distribution. From this fitted distribution, it is then possible to see what changes in V one can at least expect over any time span t, for any given probability, by examening the quantiles of corresponding percentage. For example, if an investor have fitted the daily returns of a stock to a distribution, the investor can then see what loss the stock is *at least* expected to experience in a single day over a 100 day period by looking at the 1%-quantile of said distribution.

In Table 3 the different VaR-estimates for the portfolios bond spread and CDS price values can be seen. It refers to first daily changes, ΔVBs_n and $\Delta VCDSp_n$ over the time period, then changes over three trading days, in bp. This is to capture the events when bond spreads increase substantially, i.e. the value of the bonds drop, and further how much the CDS price increases, at the same probability. In the upper table the changes have been fitted to a normal distribution, which is the most common approach when modelling return values (this is however, not a return value, as will be explained further down). In the lower table the changes have been fitted to a General Pareto distribution (GPD)[2]. This distribution models the tails of series, and hence is much more accurate in catching extreme events, such as large losses. This makes the GPD-approach highly suitable when estimating VaR. Statistics about the fit of the distributions can be found in section A. As one can expect, the GPD does a better job in modelling the more extreme increases of the bond spread and the CDS price. The statistics and QQ-plots of these distributions can be found in Table 9 and 10 Figure 8 to 11 and Figure 24 to 27 in section A.

One conclusion that can be drawn from Table 3 is for example that according to the GPD-model, there is a 1 in a 1000 risk that the bond spread will increase with at least 36.24 bp over a single trading day. Likewise it shows that at the same probability the CDS price will increase with at least 7.56 bp over a single day. Since bond spreads determine the bonds value over the corresponding rate $r_{i,n}$, it is comparable to making an investment at this rate, for example in a money market account. Hence fluctuations in Bs tells us how the risk premia for investing in the bond changes, but does not provide complete information about the price. Because of this, one can not solely look at Bs while estimating risks, since a change in the rate also can

[2]For further reading on General Pareto distribution, see [4], [5] and [6].

Table 3: **Value at Risk-estimates of bond spreads and CDS prices.** *VaR-estimates for the 95%-, 99%- and 99.9% quantile for the 1-day and 3-days increases (in bp) in portfolio bond spread and CDS price, respectively. In the upper table the VaR have been estimated with a normal assumption and in the lower with a General Pareto approach, where u is the threshold.*

Normal distribution	95%	99%	99.9%
ΔVBs_n	7.11	10.08	13.40
$\Delta VCDSp_n$	1.88	2.67	3.56
$VBs_{n+3} - VBs_n$	15.62	22.18	29.53
$VCDSp_{n+3} - VCDSp_n$	4.55	6.49	8.66

General Pareto distribution	u	95%	99%	99.9%
ΔVBs_n	7	5.97	14.48	36.24
$\Delta VCDSp_n$	2	1.63	3.52	7.56
$VBs_{n+3} - VBs_n$	6	16.26	35.65	73.44
$VCDSp_{n+3} - VCDSp_n$	2	4.66	9.30	17.38

result in a bond price change, Bs staying at the same level. To capture the full movement of a bond price, one has to look at the bond yield By, which is done further down in this presentation, as well as coupon rates, payment intensity and time to maturity (see equation 1). The estimates in Table 3 is rather to give a intuitive sense of how the bond spreads and CDS prices fluctuate and differ.

To make an actual estimation of how the portfolio value changes from day n to day $n+1$, one has to insert the changes of the bond spreads, CDS prices and rates from day n to day $n+1$ into the actual bond pricing formula, equation 1, for each bond in the portfolio. Naturally, since the observed bond yields generate the observed bond prices, these values are already known. Define the following variables

Definition 2. *Let $PBy_{i,n}$ be the closing price of bond i at trading day n. Then*

$$\begin{aligned} retBy_{i,n} &= \tfrac{PBy_{i,n+1} - PBy_{i,n}}{PBy_{i,n}} \\ vBy_n &= \tfrac{1}{k} \sum_{i=1}^{k} retBy_{i,n}. \end{aligned} \tag{2}$$

Hence $retBy_{i,n}$ is the percentage value change, or return, from day n to day $n+1$ for bond i, and vBy_n is the return of the equally weighted portfolio of all the k bonds from day n to day $n+1$. Since the bond yields will generate the market prices, equation 2 produces the actual return of the bond portfolio. To estimate a bond price using CDS prices, calculate the bond price for day $n+1$ that would have been realised if $\Delta Bs_{i,n}$ had been identical to the corresponding change in the CDS price, $\Delta CDSp_{i,n}$. From this the portfolio returns of the CDS estimation can be calculated.

Definition 3. *Let*

$$\begin{aligned} PCDSp_{i,n+1} &= \sum_{m=1}^{j-1} C(1 + By_{i,n} + \Delta CDSp_{i,n} + \Delta r_{n,i})^{-t_m} + \\ &\quad +(C + Bp)(1 + By_{i,n} + \Delta CDSp_{i,n} + \Delta r_{n,i})^{-t_j}. \\ retCDSp_{i,n} &= \frac{PCDSp_{i,n+1} - PBs_{i,n}}{PBs_{i,n}} \\ vCDSp_n &= \frac{1}{k}\sum_{i=1}^{k} retCDSp_{i,n}. \end{aligned}$$

Thus $PCDSp_{i,n+1}$ is the price of bond i if $\Delta Bs_{i,n}$ would move precisely as $\Delta CDSp_{i,n}$, i.e. the bond spread changes exactly like the CDS price does. Since the change in the rate $\Delta r_{n,i}$ also affects the bond price, this has to be included as well. Remember that $\Delta By_{i,n} = \Delta Bs_{i,n} + \Delta r_{i,n}$, which determines how the price of the bond will change to trading day $n+1$. In the above equation $\Delta Bs_{i,n}$ has simply been replaced by $\Delta CDS_{i,n}$. One can now compare vBy_n, the actual return of the portfolio, to $vCDSp_n$, the return of the portfolio as predicted by the CDS price changes. Note that this thesis focuses on 1-day portfolio returns.

The graphs of the actual return, vBy_n, and the return as estimated by $\Delta CDSp_{i,n}$, $vCDSp_n$, can be seen in Figure 5. They hint that the CDS estimated returns are less volatile then the actual. QQ-plots of ΔVBy_n against $\Delta V(CDSp_n + r_n)$ and vBy against $vCDSp$ can be found in Figure 12 and 13 in Appendix A. The VaR-estimates of the returns and volatilities can be seen in Table 4. Both estimates are fitted to a GPD, since these proved significantly more accurate then their normal counterparts. The CDS-proxy method underestimates the true risk of the portfolio decreasing for each percentile, and e.g. predicts a -1.31% loss in 1 day out of 1000, while the model fitted on actual data expects loss of -1.64%. This makes sense in connection to Table 3; it is easy to see that if the tails of the bond spread and the CDS price is largely differently distributed, so will the tail distribution of the portfolio retuns be. One might expect an even larger difference in return distributions here, but remember that the change in rates, which is present in both estimates, also affects the bond price, and determines a large part of the price changes. The GPD fit diagnostics of the

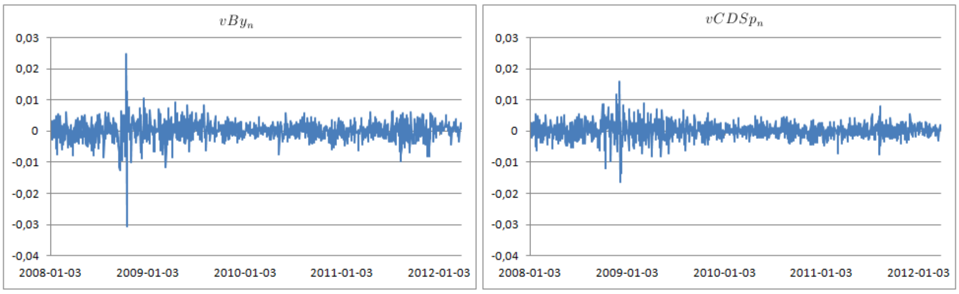

Figure 5: **Portfolio return graphs.** *The actual returns in the left graph and the CDS estimated returns in the right graph. The CDS estimated return are less volatile.*

Table 4: **Value at Risk-estimates of portfolio returns 1.** *VaR-estimates for the 95%-, 99%- and 99.9% quantile for the 1-day decreases (in %) of the portfolios actual and CDS-estimated returns together with the respective volatility (standard dev. of the daily returns). The data have been fitted to a GPD.*

General Pareto distribution					
	u	95%	99%	99.9%	σ
vBy_n	-0.5%	-0.48%	-0.87%	-1.64%	0.32%
$vCDSp_n$	-0.47%	-0.36%	-0.67%	-1.31%	0.26%

series can be found in Figure 28 and 29 in Appendix A. In these figures it is notable that the models fail to capture some of the most extreme values in both fits.

4.3 Estimation with regression

The previous sections shows that the theoretical relationship $Bs_{i,n} = CDSp_{i,n}$ is not true for the given data set, either in nominal terms, fluctuation or in what risk estimates they produce. This section will attempt to re-model this relationship with the addition of several market indices. One model will aim to replicate the actual bond spread curve in the data, and two others will attempt to model the daily changes of each bond yield, in order to produce a portfolio risk estimate that lies closer to the estimate fitted to the actual data.

4.3.1 Fitting the bond spread curve

The first model will try to explain the movement of the average portfolio bond spread, VBs_n. This is not an attempt to try and predict the next trading days price in terms of historic market data, but rather an investigation in possible explanations of the recent bond spread behaviour. Since it will also model a benchmark bond portfolio, it looks for general trends in for the whole bond market and does not take consideration for specific sectors. One approach is to keep the expression $Bs_{i,n} = CDSp_{i,n}$ and then add market indices to fit the discrepnacy between the variables with a linear regression. However, doing that will lock the 1-to-1 relationship between the bond spread and CDS price. Rejecting the aforementioned relationship, it is quite possible that the variables could be modeled as $Bs_{i,n} = \alpha CDSp_{i,n}$, $Bs_{i,n} = \alpha CDSp_{i,n}^{\beta}$ or similar. This thesis will stick to the linear case, and with market factors described further down, together with the $L1$-values of each variable the regression model will be set up as

$$\begin{aligned} VBs_n &= a_0 + \sum_{l=1}^{h} a_l * I_{l,n} + \sum_{l=1}^{h} b_l * L1I_{l,n} + a_{h+1} * VCDSp_n \\ &\quad + b_{h+1} * L1VCDSp_n + \epsilon_n \qquad (3) \end{aligned}$$

where $I_{1,n}, ..., I_{h,n}$ is the h different indices at day n, a_0 is a constant, $a_1, ..., a_{h+1}$ are the $h+1$ different regression coefficients for the nominal values, $b_1, ..., b_{h+1}$ the coefficients for the $L1$-values and ϵ_n is the error term at day n.

The indices that will, together with $VCDSp_n$, be used as regressors are: EUONIA (3 month), EURIBOR (3 month), EURGovt (3 month) and VIX.

EURIBOR-EUONIA: This is the Euro equivalence to LIBOR-OIS, which Fontana describes as an indicator of counterparty and funding liquidity risk,

Table 5: **Regression statistics 1.** *Regression statistics for estimation of equation 3.*

	Coeffiicent	Std. Err.	t	$p > \|t\|$	[95% Conf. int.]
VIX	2.153435	.1082442	19.89	0.000	[1.941049, 2.365821]
$EURIBOR$	256.8879	56.46038	4.55	0.000	[146.1069, 367.6689]
$VCDSp$	-3.031075	.5827179	-5.20	0.000	[-4.174427, -1.887723]
$L1VCDSp$	5.1574	.5916893	8.72	0.000	[3.996445, 6.318355]
$L1EUONIA$	-168.3564	7.113707	-23.67	0.000	[-182.3143, -154.3986]
$L1EURIBOR$	-181.6981	54.47376	-3.34	0.001	[-288.5811, -74.81501]
$L1EURGovt$	84.13324	5.034519	16.71	0.000	[74.255, 94.01147]
Constant	16.48483	2.579258	6.39	0.000	[11.42406, 21.54559]

Statistic	
Dependent var	VBs
Observations	1119
$F(7, 1111)$	1106.85
$p > F$	0.000
R^2	0.875
Adj. R^2	0.874
Root MSE	22.00
Correl. with d.v.	0.935

which pre-2007 was almost insignificant. Before the 2008 crisis this index fluctuated close around 10 bp, but jumped to over 206.9 bp during the crisis. The EURIBOR is short for Euro Interbank Offered Rate. It is based on the average rates at which about 50 european banks borrow from each other, at different maturities. EUONIA stands for The Euro Overnight Index Average, and is an interest rate swap index calculated on a panel of prime European banks. In the following regression, these indices will be used as separate regressors for maximum flexibility.

EURGovt: A fund of numerous different European government bonds and is somewhat an equivalence to the US T-bill. Fontana mentions the OIS-T-bill index as capturing "flight to quality", i.e. how investors tend to move their capital to less risky assets in stressful times.

VIX: Volatility Index, a marker of how stressed the market is, and is according to Fontana supposed to capture risk premium as well as funding costs due to deterioration of bond value in a negative basis trade.

The statistics of regression-equation 3 of the above regressors can be found in Table 5. A regression fit has been run several times, first with all the variables in equation 3. Then the variable with highest $p > |t|$ above 0.05 has been removed, and the regression has been re-run. This is repeated until all coefficients in the model have a t-statistic that is significant for the α-level 0.05. This thesis will not elaborate on the influences of each variable, but it is interesting to for example note that the portfolio bond spread VBs_n is positively influenced by yesterdays CDS price $L1VCDSp_n$, or $VCDSp_{n-1}$, but negatively and less substitially influenced by todays CDS price $VCDSp_n$. This somewhat supports the idea of faster price discovery for CDS prices, as suggested by Blanco. Overall, most of the significant variables are lagged values, hinting that bond spreads adjust somewhat slow to market factors.

Further, recall that the correlation between the portfolio bond spread and CDS price was 0.62, compared to the bond spread-regression correlation of 0.935. The portfolio bond spread, the regression of equation 3 and its ϵ_n-term can be seen in Figure 6. The regression graph manages *fairly* well to follow the bond spread graph, and somewhat capture most of the significant trends. From the curve of ϵ_n it is however clear that the financial crisis present in the first third of the series is rough terrain to model using these factors. A regression fitted to VBs_n from to the period 2010-01-01 to 2012-03-23 was made, with a R^2 measure of 0.9491, emphasizing that during the crisis there were explanatory factors unknown to this model that greatly affected the spread, but that had much less impact on the spread after the worst crisis period. The correlation of the residuals of day n and day $n+1$ was concluded to be 0.93, which strongly suggest that the error in the model is systematic and not random noise. This suggestion further indicates that there are additional, unknown, factors that impact the behaviour of bond spreads. QQ-plot of the error term ϵ_n against its fitted normal distribution and plot over time can be found in Figures 18 and 19 in Appendix A.

4.3.2 Fit and prediction of ΔBy

This section will try to model the daily changes of the bond yields in the portfolio. This model will then be used to make an risk-estimate of the bond portfolio much like the one made from CDS prices in section 4.2. However, the dependent variable will now be $\Delta By_{i,n}$ and the regressors will be the market indices used in section 4.3.1 together with $CDSp_{i,n}$. Hence this model will consider each bond individually, not in the aggregated portfolio sence as in last subsection. The regression equation will be

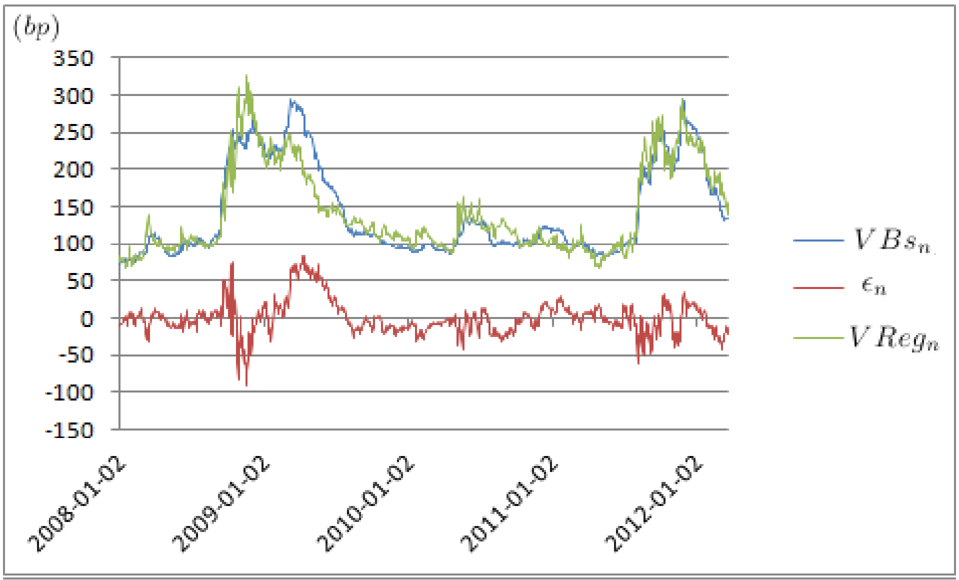

Figure 6: **Regression graph.** *Portfolio bond spread VBs_n, regression $VReg_n$ and error term ϵ_n of equation 3. The regression curve follows the bond spread curve fairly close.*

$$\begin{aligned}\Delta By_{i,n} &= a_0 + \sum_{l=1}^{h} a_l * I_{l,n} + \sum_{l=1}^{h} b_l * \Delta I_{l,n} + \sum_{l=1}^{h} c_l * L1I_{l,n} \\ &\quad + a_{h+1} * CDSp_{i,n} + b_{h+1} * \Delta CDSp_{i,n} \\ &\quad + c_{h+1} * L1CDSp_{i,n} + a_{h+2} * r_{i,n} + b_{h+2} * \Delta r_{i,n} \\ &\quad + c_{h+2} * L1r_{i,n} + \epsilon_{i,n}, \qquad (4)\end{aligned}$$

with the same notation as in equation 3 except that the Δ-value of each variable have been added, that each bonds CDS price is used to model each corresponding difference in bond yield, instead of only using portfolio values as dependent variable and regressors and that each relevant rate is included. The purpose of this is of course to capture each bond yields movement in respect to its CDS price and swap rate.

The statistics of the regression can be seen in Table 6. The variables have been eliminated in the same manner as in last subsection, eliminating the least significant variable then re-estimating the model until all regressors are significant at the α-level of 0.05. Not surprisingly, the variables for CDS price and rate are the most significant (in terms of $|t|$), most prominently $\Delta CDS_{i,n}$ and $\Delta r_{i,n}$, which makes intuitive sense. The R^2-statistic is only .226, telling that the fit is not very accurate. Neither is the error term

$\epsilon_{i,n}$ normally distributed, again indicating that there are factors outside the model affecting the bond yield movement. QQ-plot of $\epsilon_{i,n}$ against its fitted normal distribution and plot over time and bond can be found in Figure 20 and 21 in AppendixA.

In order to use this model to make a risk-estimation of the bond portfolio, the following definition is needed.

Definition 4. *Define*

$$
\begin{aligned}
\Delta Reg_{i,n} &= \Delta By_{i,n} - \epsilon_{i,n} \\
PReg_{n+1} &= \sum_{m=1}^{j-1} C(1 + By + \Delta Reg_{i,n})^{-t_m} + \\
&\quad +(C + Bp)(1 + By + \Delta Reg_{i,n})^{-t_j} \\
retReg_{i,n} &= \frac{PReg_{n+1} - PBs_{i,n}}{PBs_{i,n}} \\
vReg_n &= \frac{1}{k} \sum_{i=1}^{k} retReg_n. \qquad (5)
\end{aligned}
$$

Hence $\Delta Reg_{i,n}$ is the regressions fitted value to the bond yield value of bond i at trading day n, $PReg_{n+1}$ is the price of the bond at day $n+1$ if the bond yield moves identically to the regression value and $vReg_n$ is the portfolio return based on those daily price changes.

The actual return, vBy_n, the CDS-estimated return, $vCDSp_n$ and the return estimated by equation 4, $vReg_n$, can be seen in Figure 7. The volatility and VaR-estimates of the time series can be seen in Table 8. In terms of volatility, the distance between vBy_n and $vReg_n$ is only half that of vBy_n and $vCDSp_n$, and $vReg_n$ lies closer to vBy_n in all of the tail percentiles of the distribution, so the regression is able to improve the VaR-estimation somewhat. There is however some discrepancy, stemming from the difficulty in modelling the bond yield behaviour which, as mentioned before, clearly is influnced by additional factors not present in the model. Furthermore, all of the GPD-models contain outliers, which makes the most extreme tail fitting somewhat unreliable. QQ-plots of ΔVBy_n against ΔReg_n and vBy against $vReg$ can be found in Figures 14 and 15 together with fit diagnostics of the GPDs in Figures 28 and 30 in section A.

Finally, this thesis will investigate the regressors predictive power on ΔBy. Namely looking into how much of the bond yield change between trading day n and $n+1$ that can be explained by the data available at day n. This regression will use the same variables as in previous model, except that the

Table 6: **Regression statistics 2.** *Regression statistics for estimation of equation 4.*

	Coeffiicent	Std. Err.	t	$p > \|t\|$	[95% Conf. int.]
$EUONIA$	.2228673	.0266471	8.36	0.000	[.1706372, .2750975]
$EURIBOR$	-.0833007	.0223299	-3.73	0.000	[-.1270688, -.0395326]
$EURGovt$	-.1070137	.0149442	-7.16	0.000	[-.1363055, -.077722]
$CDS_{i,n}$	.6244625	.0281657	22.17	0.000	[.5692558, .6796692]
$r_{i,n}$	.050169	.0122362	4.10	0.000	[.0261852, .0741527]
$\Delta EUONIA$	.1679764	.0275874	6.09	0.000	[.1139032, .2220496]
ΔVIX	-.0020766	.0002814	-7.38	0.000	[-.0026282, -.0015249]
$\Delta EURIBOR$	.327449	.0525091	6.24	0.000	[.2245276, .4303705]
$\Delta EURGovt$	-.0328348	.0147889	-2.22	0.026	[-.0618221, -.0038476]
$\Delta CDS_{i,n}$	1.17591	.0308974	38.06	0.000	[1.115349, 1.236471]
$\Delta r_{i,n}$	.8014416	.0130102	61.60	0.000	[.7759406, .8269426]
$L1EUONIA$	-.1873719	.0257531	-7.28	0.000	[-.2378497, -.136894]
$L1VIX$	.0009884	.0001009	9.80	0.000	[.0007907, .0011861]
$L1EURIBOR$	.0681029	.0221608	3.07	0.002	[.0246662, .1115396]
$L1EURGovt$	.0909453	.0148487	6.12	0.000	[.0618409, .1200498]
$L1CDS_{i,n}$	-.6214556	.0281807	-22.05	0.000	[-.6766917, -.5662194]
$L1r_{i,n}$	-.0514401	.0122461	-4.20	0.000	[-.0754433, -.027437]
Constant	-.0222604	.0038603	-5.77	0.000	[-.0298268, -.0146939]

Statistic	
Dependent var	$\Delta By_{i,n}$
Observations	22360
$F(17, 22342)$	384.41
$p > F$	0.000
R^2	0.226
Adj. R^2	0.226
Root MSE	.09211
Correl. with d.v.	0.734

Table 7: **Regression statistics 3.** *Regression statistics for estimation of equation 6.*

	Coeffiicent	Std. Err.	t	$p > \|t\|$	[95% Conf. int.]
$EUONIA$	.037991	.0044279	8.58	0.000	[.029312 .04667]
$EURIBOR$	-.0293789	.0038602	-7.61	0.000	[-.0369452 -.0218126]
VIX	.0012334	.0001106	11.15	0.000	[.0010166 .0014501]
$r_{i,n}$	-.0077441	.0011976	-6.47	0.000	[-.0100914 -.0053968]
$L1\Delta EUONIA$	.3009719	.0294937	10.20	0.000	[.2431623 .3587815]
$L1\Delta EURIBOR$	.1882576	.0549686	3.42	0.001	[.0805153 .2959999]
$L1\Delta EURGovt$	-.1641736	.0159716	-10.28	0.000	[-.195479 -.1328681]
$L1\Delta CDS_{i,n}$	.8982653	.033507	26.81	0.000	[.8325892 .9639413]
$L1\Delta r_{i,n}$	.0447796	.0139649	3.21	0.001	[.0174073 .0721518]
Constant	-.0064847	.0031281	-2.07	0.038	[-.012616 -.0003534]

Statistic	
Dependent var	$\Delta By_{i,n+1}$
Observations	22360
$F(9, 22350)$	384.41
$p > F$	0.000
R^2	0.047
Adj. R^2	0.047
Root MSE	.1022
Correl. with d.v.	0.217

Δ-variables will be replaced by $L1\Delta$-variables since data for day $n+1$ is not available at day n. As a consequence the $L1$-variables will be dropped, since they have perfect colinearity with the variables and their $L1\Delta$-versions. The regression equation will then be

$$\begin{aligned}\Delta By_{i,n+1} &= a_0 + \sum_{l=1}^{h} a_l * I_{l,n} + \sum_{l=1}^{h} b_l * \Delta I_{l,n} + \sum_{l=1}^{h} c_l * L1\Delta I_{l,n} \\ &\quad + a_{h+1} * CDSp_{i,n} + b_{h+1} * \Delta CDSp_{i,n} \\ &\quad + c_{h+1} * L1\Delta CDSp_{i,n} + a_{h+2} * r_{i,n} + b_{h+2} * \Delta r_{i,n} \\ &\quad + c_{h+2} * L1\Delta r_{i,n} + \epsilon_{i,n}. \qquad (6)\end{aligned}$$

The results of estimating the coefficients in equation 6 can be seen in table 7. Now make the following definition.

Definition 5. *Define*

$$\begin{aligned}
\Delta Pred_{i,n+1} &= \Delta By_{i,n+1} - \epsilon_{i,n} \\
PPred_{n+1} &= \sum_{m=1}^{j-1} C(1 + By + \Delta Pred_{i,n+1})^{-t_m} + \\
&\quad +(C + Bp)(1 + By + \Delta Pred_{i,n+1})^{-t_j} \\
retPred_{i,n} &= \frac{PPred_{n+1} - PBs_{i,n}}{PBs_{i,n}} \\
vPred_n &= \frac{1}{k}\sum_{i=1}^{k} retPred_n. \qquad (7)
\end{aligned}$$

Hence $vPred_n$ is the portfolio return for day n calculated as if each $\Delta By_{i,n+1}$ had moved identically to the value $\Delta Pred_{i,n+1}$, which is to be interpreted as an prediction of $\Delta By_{i,n+1}$ conditioned on the known data at day n. A key difference between this regression and the previous two is that this one is making a claim about future values, while the others only tried to explain the present values. They are however all estimed using the full data set.

The plotted graph of $vPred_n$ can be seen in Figure 7, and it's GPD estimated VaR in Table 8. A QQ-plot and plot over time and bond for error term of the estimation of equation 6 and GPD fit diagnostics can be seen in Figure 22, 23 and 31 in Appendix A. Interestingly the predicted portfolio returns come out much more volatile than the original and previously estimated ones. Also notice that the GPD model fails to capture the most extreme outliers.

The idea now is that an investor could use an estimation of equation 6 and together with the present price (and yield) of some bond make an estimation of the probability distribution of tomorrows price, without any knowledge of how the price have behaved previously. Only inserting the value $\Delta Pred_{i,n+1}$ into By_n will however not be enough, since there is also an error term to be considered. Hence, to receive a probability distribution of tomorrows bond price, the investor would have to fit the error term in equation 6 to a distribution and let a variable, say ξ, have this distribution, and then inject a realization of this variable into the pricing formula along with $\Delta Pred_{i,n+1}$. This distribution must be identical and independent, and for the purpose of the regressions significance, also normal. Repeating this simulation a large number of times with different realizations of ξ, the probability distribution of tomorrows bond price, and hence the one day VaR, could be estimated. The equation to be simulated is defined as

$$
\begin{aligned}
SIMp_{i,n+1} \quad = & \quad \sum_{m=1}^{j-1} C(1 + By_{i,n} + \Delta Pred_{i,n+1} + \xi)^{-t_m} + \\
& +(C + Bp)(1 + By_{i,n} + \Delta Pred_{i,n+1} + \xi)^{-t_j}. \qquad (8)
\end{aligned}
$$

However, the error term of this regression, as well as the previous two regressions, suffer from heteroscedasticity and autocorrelation problems, as well as non-normality. This violates the basic Gauss-Markov assumptions of linear regressions and undermines the significance tests of the coefficients, greatly weakening the usefulness of these estimations. As mentioned before, these violations likely stem from other factors, not included in the model, influencing the spread or yield behaviour. It could also hint of endogeneity, correlation between variables and the error term, or that the linear approach simply is not feasible for the data.

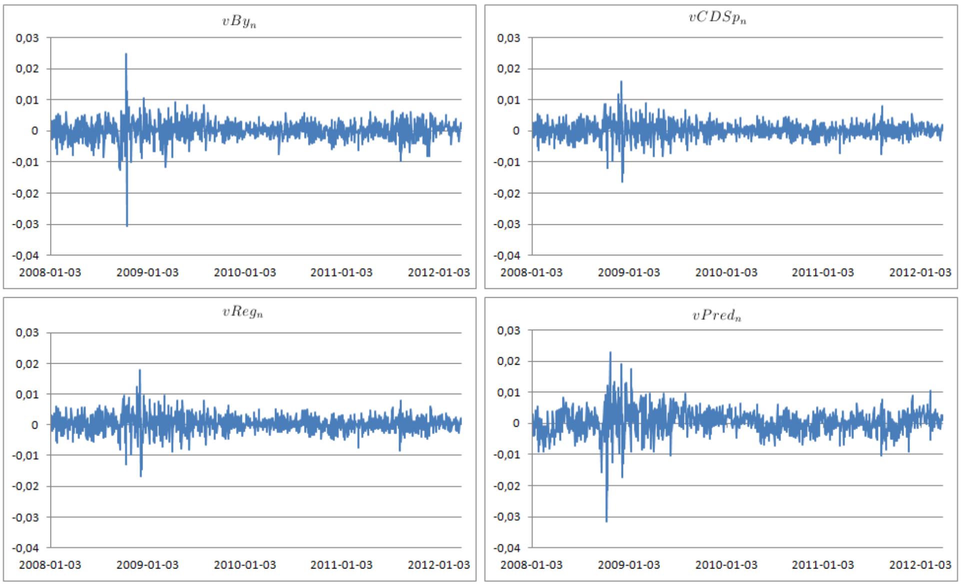

Figure 7: **Portfolio return graphs 2.** *From top left to bottom right: the acctual portfolio return* vBy_n*, CDS estimated portfolio return* $vCDSp_n$*, return from* $vReg_n$*, and return caulculated by* $vPred_n$*.* vBy_n *and* $vReg_n$ *are most similar in volatility.*

Table 8: **Value at Risk-estimates of portfolio returns 2.** *VaR-estimates for the 95%-, 99%- and 99.9% quantile for the 1-day decreases (in %) of the portfolios actual return, CDS-estimated returns and returns assessed with the two regressions together with the respective volatility. The data have been fitted to a GPD with threshold u.*

General Pareto distribution					
	u	95%	99%	99.9%	σ
vBy_n	-.5%	-.48%	-.87%	-1.64%	.32%
$vCDSp_n$	-.47%	-.36%	-.67%	-1.31%	.26%
$vReg_n$	-.49%	-.37%	-.70%	-1.38%	.29%
$vPred_n$	-.7%	-.62%	-.92%	-2.25%	.4%

5 Concluding discussion

The purposes of this thesis was to investigate: how the CDS-bond basis behaved during and after the financial crisis 2008; how an investor lacking cruicial bond data would have fared during this period using CDS prices to proxy bond risk; and if this proxy was unsuitable, to look for a better way to construct such a risk proxy. The purpose of this developed proxy is to enable an investor to estimate the riskyness of a bond even if it lacks reliable historical data. By measuring the fluctuations of the proxy, the investor would able to assess the bonds risk. The idea was also to make it general in such a way that it would be applied to any bond or portfolio of bonds.

The data set consisted of a portfolio constructed as a benchmark of US and European bonds, weighted in such a way that it would represent the bond market as a whole. To each of these bonds corresponding CDS prices and swap rates was matched with the bonds diminishing maturity. For the regression several market indices associated with bond mechanics was added.

First off, it was obvious from figure 4 together with the findings in Fontana and Finger that the zero difference basis was severly violated, both during and post crisis. It was largely negative during the crisis years, and however becoming less negative during the following years, did not return to the pre-crisis slighly positive basis level.

Second, this thesis aimed not to explicitly explain why this discrepancy occured, but rather to investigate what impact an investor would face using CDS prices to proxy bond risk with the data from this period. Figure 5 as well as Table 3 and 4 reveals that CDS prices during this period do not experience the same volatility as corresponding bond spreads, and using CDS prices alone underestimates the risks of holding a bond portfolio.

At this point it is obvious that the previously stable relationship $CDSp - Bs = 0$ had to be reinvestigated. This is done with a set of three regressions.

The first one aims simply to reconstruct the portfolio average bond spread curve of the time period using the portfolio average CDS price and a set of market indices. This is mainly an exercise to investigate the complexity of the bond markets mechanics. As it turns out, the estimated model reveals no simple and satisfying solution. Table 5 shows that a fit with R^2 of 0.875 was achieved, with the regression curve having a correlation of 0.935 with the dependent variable VBs. However, as revealed in Figure 18 and 19 in Appendix A, the estimation suffers from non-normality as well as severe autocorrelation and heteroscedasticity, greatly undermining the significance of the model.

A second regression is constructed to fit the observed $\Delta By_{i,n}$ in the data. Key differences from the first model is that it models the bond yield, hence including the swap rate, and that it takes each individual bond and CDS into account. The target is here to find a general pattern of how a bond fluctuates in relation to rate, CDS price and the mentioned indices. Table 6 reveals a fit with R^2 .226, a very weak fit. Figures 20 and 21 unveils that the error term of this model also suffers from non-normality, heteroscedasticity and autocorrelation. Again the significance of the model is undermined. The resulting values of this regression is injected into the bond pricing formula for relevant time and bond, producing portfolio return values that can then be compared with the return of the actual portfolio and the returns estimated by the CDS fluctuations. A discussion about these follows further down.

The third and last regression tries to make a claim about how the data set can be used to predict future bond price movements, specificly movements to tomorrows closing price. The model is set up to predict $By_{i,n+1}$, with only the data available at day n. In short, to forsee tomorrows bond yield with the data available today. The idea is then that an investor can use this model, together with information about CDS prices, rates, and indices, to make a forecast about tomorrows price. By also simulating the behaviour of the models error term (which have be independently, identicaly distributed for the simulation to be relevant, and normally distributed for the model to be significant), the investor can asses tomorrows bond price is distributed, and hence the 1-day-VaR. The strength of this application would be that it could be estimated without any knowledge of past bond prices, and as such could be used in assessing risks of bonds lacking reliable pricing data, one of the concerns this thesis aimed to address. However, as observed in Figure 22 and 23, also this error term suffer from non-normality, heteroskedasticity and autocorrelation, undermining the significance of the model.

Nontheless the resulting values of the prediction are tried out on the bond pricing formula as above, and the resulting daily return values of the portfolio can be seen in together with the results from the other exercises in Table 8. The results are quite intuitive. The tail of return values assessed with CDS prices, $vCDSp_n$ underestimates the true tail returns, simply as an result of CDS pricing being less erratic than bond spreads in the data. The tail of return values assessed by the regression fitting $\Delta By_{i,n}$, $vReg_n$, lies closer to the tail values than $vCDSp_n$ and about twice as close in terms of volatility, which is also to be expected, since the CDS values and additional other variables are used to fit the behaviour of the bond yield. Finally the tail of returns calculated with the prediction model, $vPred_n$, are more volatile than the actual returns, and quite dramatically overestimate the tail values. On another notice, the tails turns out to be difficult to capture in a model, even with the GPD approach. As seen in the fit diagnostics in Figure 24 to 31, all models struggle with a few extreme values. However, removing these outlies hardly changes the differences between return tails in Table 8.

A tiresome line of regression model error terms points to the conclusion that bond yield and spread behaviour over the time period is quite hard to explain with the help of only CDS prices and market indices. As mentioned earlier, a obvious culprit in this is the financial crisis itself, which naturally makes for troublesome terrain also in modelling. Excluding this part of the time series makes for better estimation of equation 3. It is however, as also mentioned, in these times robust models are of most importance. One conclusion drawn from the models non-conformist error terms could be that there simply are additional, or other, market factors affecting the bond price, just that they have not been incorporated in this model. One approach to improve the modelling could very well be to include additional factors, of macro nature for example. A very sound first step would however be to begin with broadening the data set to such a size that bonds could be divided by industrial sectors. This would be a highly relevant approach, since bond basis obviously behaves very different depending on their origin. Modelling in divided sectors could very well reveal much more useful results. Another notion to consider is the assumption of linearity as briefly mentioned in section 4.3. The addition of exponentials in the model can capture nuances that linearity will miss.

Furthermore, before using a model of this kind in application it first needs extensive out-of-sample testing, which in turn requires a considerably larger data set as well as a much more diversified portfolio to ensure robustness in the model.

A Distribution statistics

Table 9: **Normal distribution parameters:** *The parameters of the fitted normal distributions of* ΔVBs_n, $\Delta VCDSp_n$, $V_{Bs,n+3} - VBs_n$ *and* $V_{CDSp,n+3} - VCDSp_n$

	μ(mean)	σ(standard dev.)
ΔVBs_n	0.0543	4.3544
$\Delta VCDSp_n$	0.0328	1.1633
$V_{Bs,n+3} - VBs_n$	0.2169	9.6263
$V_{CDSp,n+3} - VCDSp_n$	0.1309	2.8462

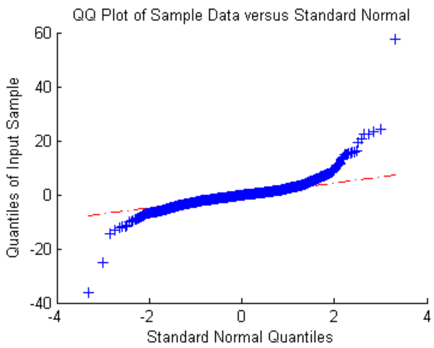

Figure 8: *Normal QQ-plot of daily changes* ΔVBs_n

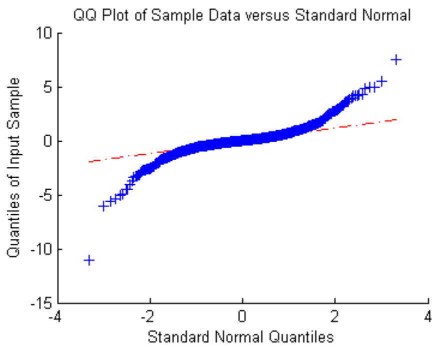

Figure 9: *Normal QQ-plot of daily changes* $\Delta VCDSp_n$

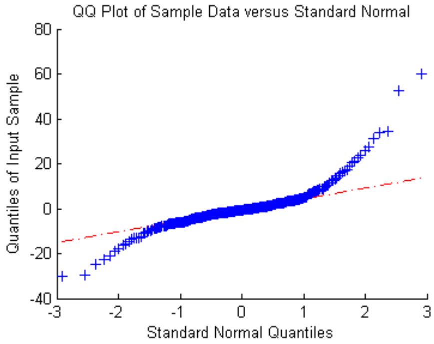

Figure 10: *Normal QQ-plot of* VBs_{n+3}

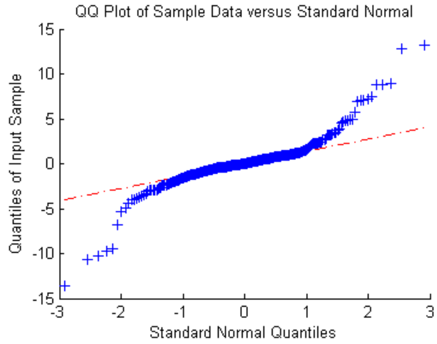

Figure 11: *Normal QQ-plot of* $VCDSp_{n+3}$

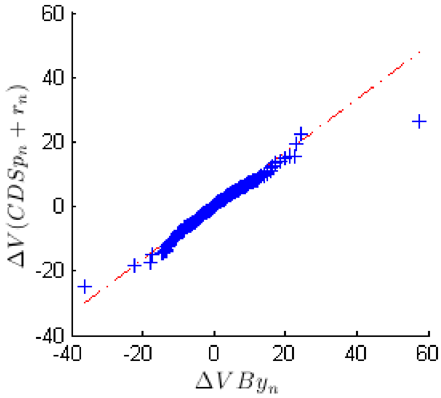

Figure 12: *QQ-plot of daily changes* ΔVBy_n *and* $\Delta V(CDSp_n + r_n)$

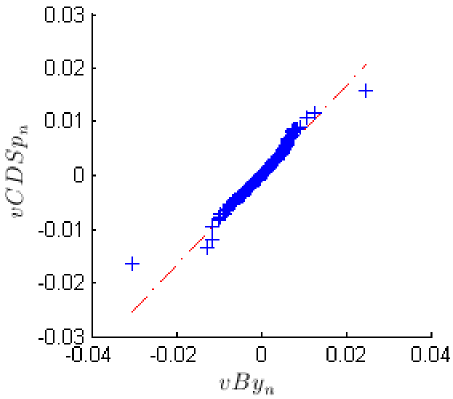

Figure 13: *QQ-plot of portfolio returns* vBy_n *and* $vCDSp_n$.

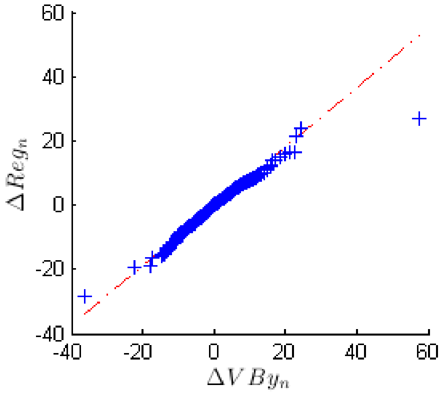

Figure 14: *QQ-plot of daiyl changes* ΔVBy_n *and* $\Delta VReg_n$.

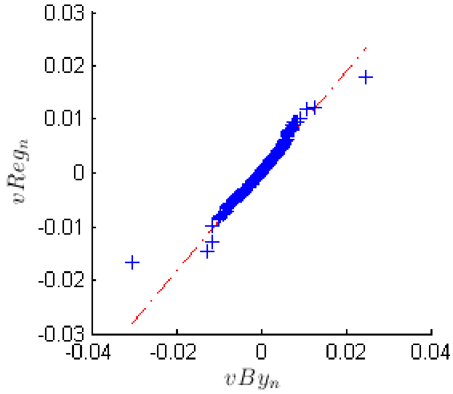

Figure 15: *QQ-plot of portfolio returns* vBy_n *and* $vReg_n$.

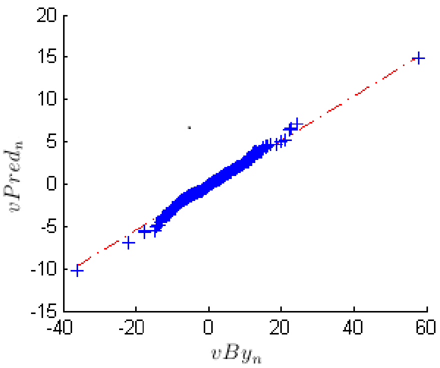

Figure 16: *QQ-plot of daily changes* ΔVBy_n *and* $\Delta VPred_n$.

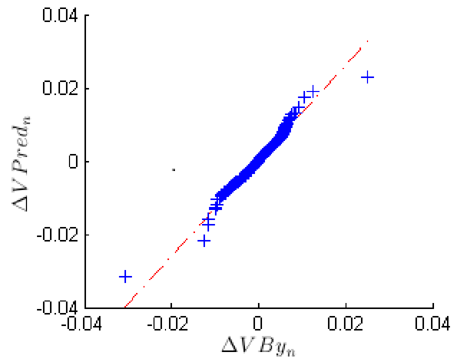

Figure 17: *QQ-plot of portfolio returns* vBy_n *and* $vPred_n$.

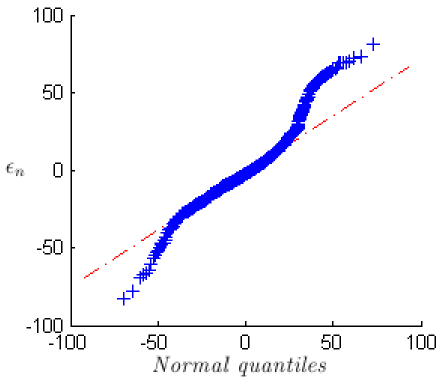

Figure 18: *QQ-plot of the error term* ϵ_n *in equation 3 against its fitted normal distribution.*

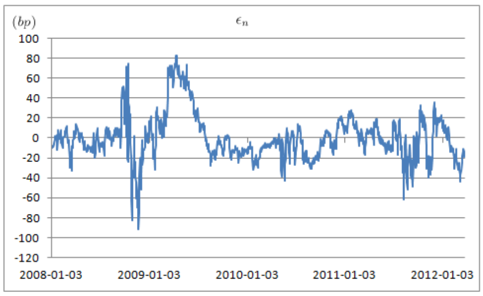

Figure 19: *The error term* $\epsilon_{i,n}$ *in equation 3 plotted over time.*

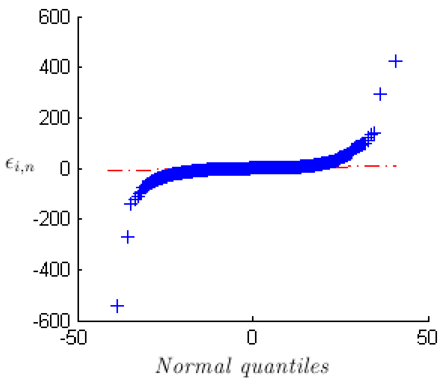

Figure 20: *QQ-plot of the error term* $\epsilon_{i,n}$ *in equation 4 against its fitted normal distribution.*

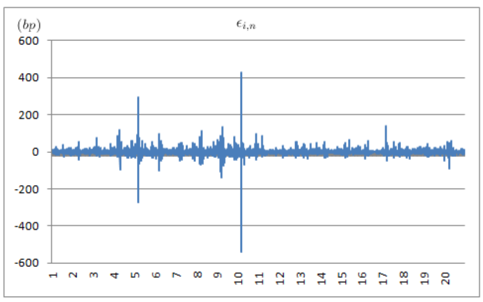

Figure 21: *The error term* $\epsilon_{i,n}$ *in equation 4 plotted over time and the 20 bonds.*

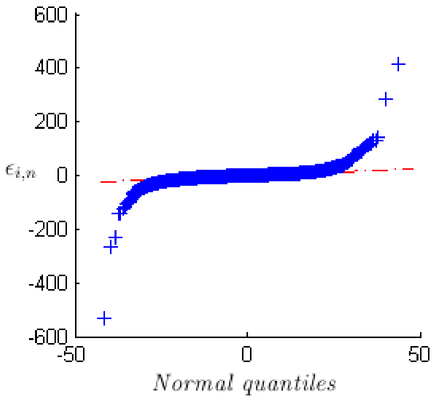

Figure 22: *QQ-plot of the error term $\epsilon_{i,n}$ in equation 6 against its fitted normal distribution.*

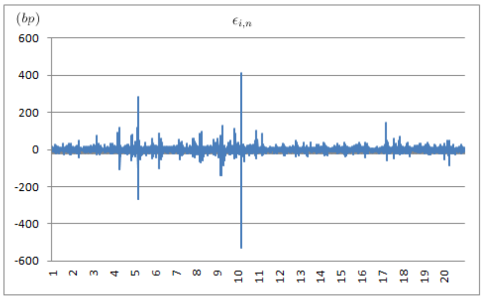

Figure 23: *The error term $\epsilon_{i,n}$ in equation 6 plotted over time and the 20 bonds.*

Table 10: **General Pareto distribution parameters:** *The parameters of the fitted GPD of ΔVBs_n, $\Delta VCDSp_n$, $VBs_{n+3} - VBs_n$, $VCDSp_{n+3} - VCDSp_n$, vBs_n, $vCDSp_n$, $vReg_n$ and $vPred_n$. Theshold level u have been chosen using a mean excess plot.*

	u	Exceedence rate	σ(scale)	ξ(shape)
ΔVBs_n	7 bp	3.93%	4.4198	.2945
$\Delta VCDSp_n$	2 bp	3.49%	1.0679	.2030
$VBs_{n+3} - VBs_n$	6 bp	14.29%	8.9881	.1566
$VCDSp_{n+3} - VCDSp_n$	2 bp	14.29%	2.4000	.0991
vBy_n	-.5%	4.47%	.0022	.1600
$vCDSp_n$	-.47%	2.68%	.0018	.1990
$vReg_n$	-.7%	2.68%	.0019	.1854
$vPred_n$	-.7%	2.95%	.0018	.4817

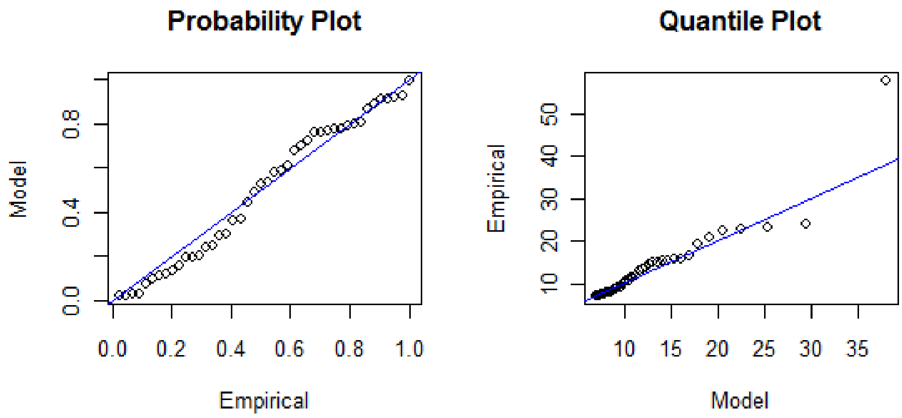

Figure 24: *GPD fit diagnostic of* ΔVBs_n.

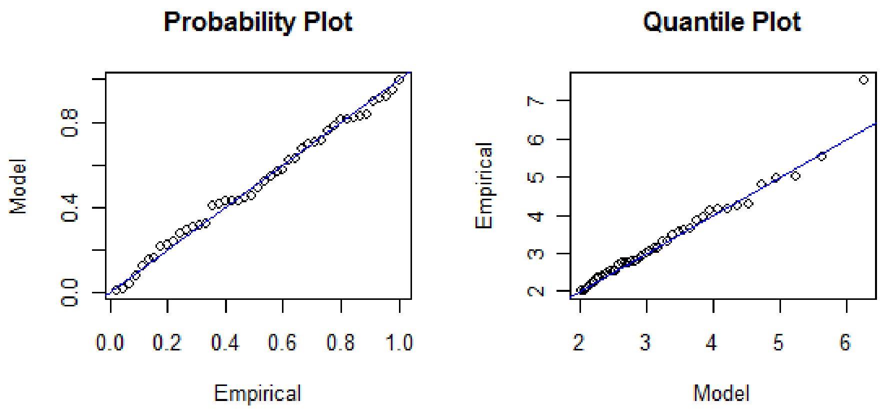

Figure 25: *GPD fit diagnostic of* $\Delta VCDSp_n$.

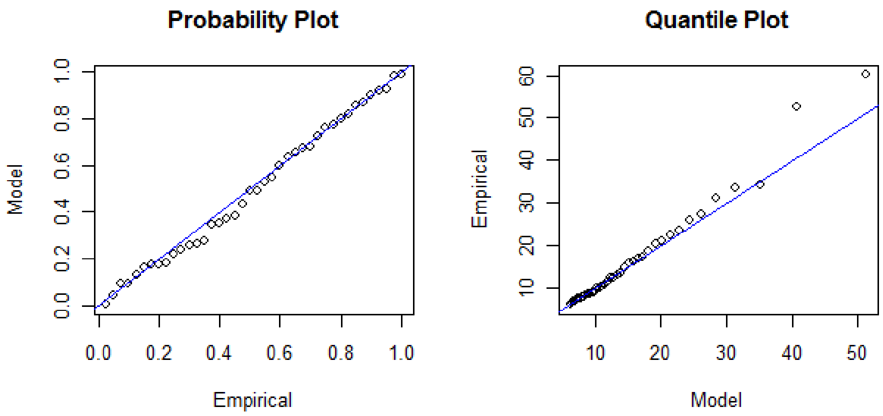

Figure 26: *GPD fit diagnostic of* $VBs_{n+3} - VBs_n$.

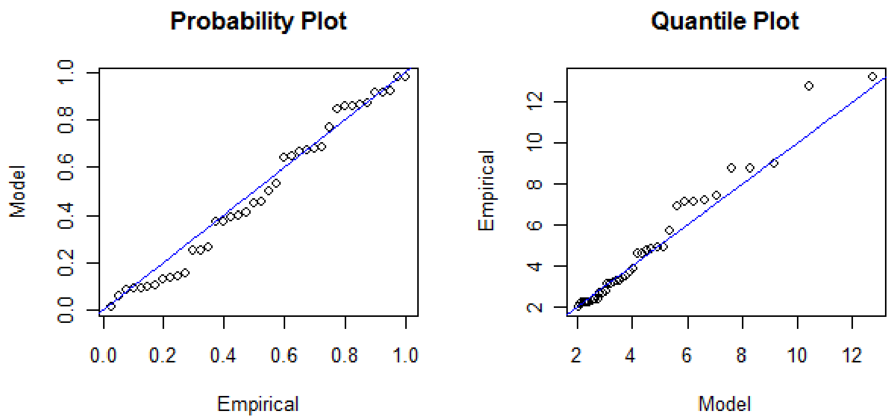

Figure 27: *GPD fit diagnostic of* $VCDSp_{n+3} - VCDSp_n$.

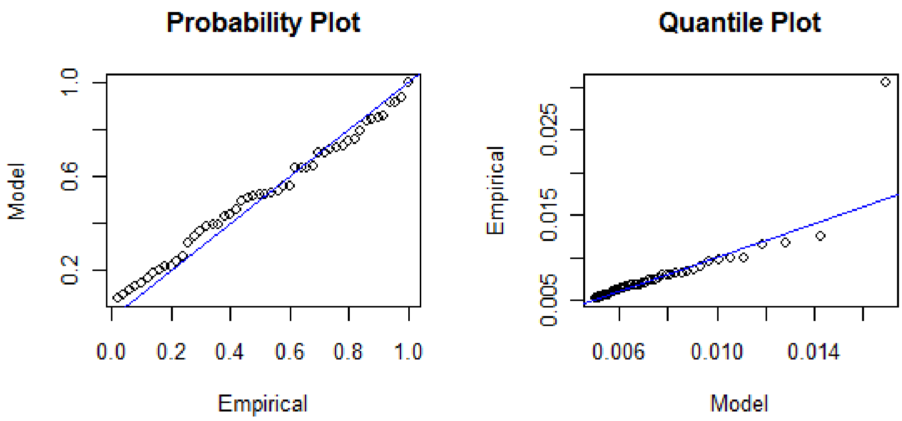

Figure 28: *GPD fit diagnostic of vBy.*

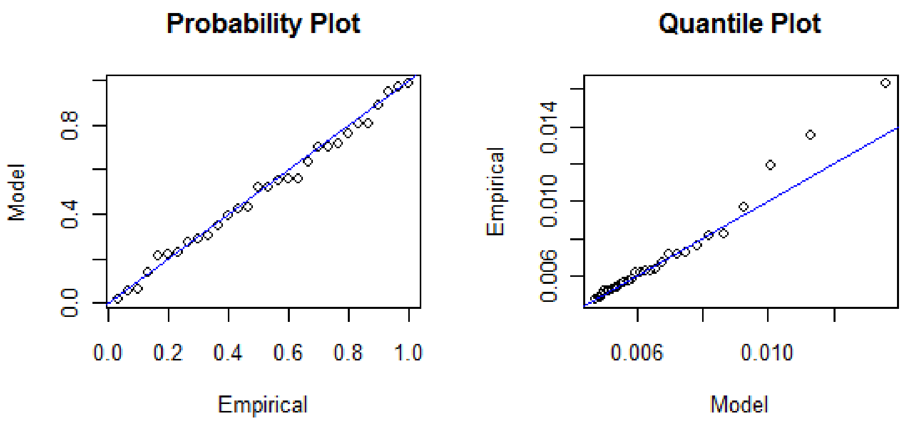

Figure 29: *GPD fit diagnostic of vCDSp.*

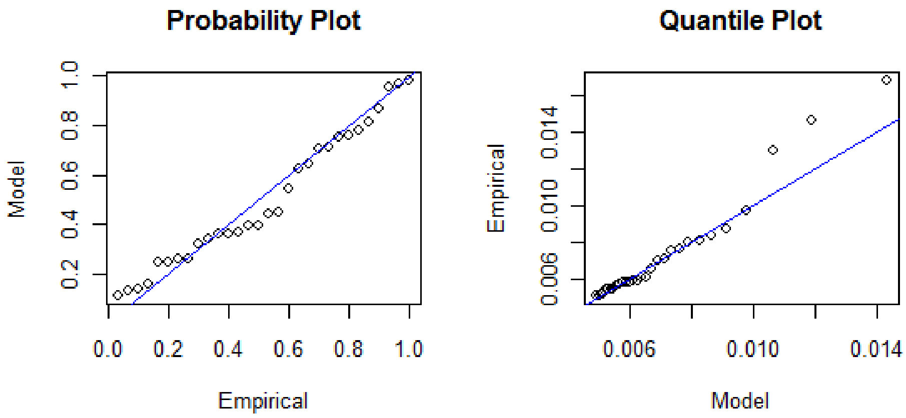

Figure 30: *GPD fit diagnostic of vReg.*

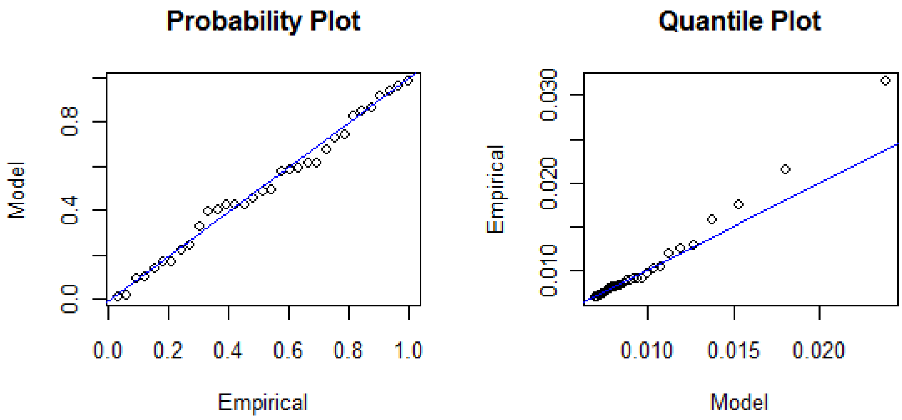

Figure 31: *GPD fit diagnostic of vPred.*

References

[1] Roberto Blanco, Simon Brennan and Ian W. Marsh, *An Empirical Analysis of the Dynamic Relation between Investment-Grade Bonds and Credit Default Swaps*. The Journal of Finance, Vol. 60, No.5, October 2005.

[2] Alessandro Fontana, *The persistent negative CDS-bond basis during the 2007/08 financial crisis*. Department of Economics, University of Ca' Foscari Venice, May 1, 2010.

[3] Christoffer Finger, *Missing the same risk twice*. Barra, MSCI, 2011.

[4] Vladimir O. Andreev, Oleg B. Okunev, Sergey Eu. Tinyakov, *Extreme value theory and peaks over threshold model: An application to the Russian stock market*. New York Science Journal, 2010;3(6).

[5] Stuart Coles, *An Introduction to Statistical Modeling of Extreme Values*. Springer London Ltd, 1st edition, 2001.

[6] Henrik Hult and Filip Lindskog, *Mathematical Modeling and Statistical Methods for Risk Management*. 2007.

[7] Jennie Bai, Pierre Collin-Dufresne, *The Determinants of the CDS-Bond Basis During the Financial Crisis of 2007-2009*. June 30, 2011.

[8] Pierre Collin-Dufresne, Robert S. Goldstein, J. Spencer Martin, *The Determinants of Credit Spread Changes*. The Journal of Finance, Vol. 56, No.6, December 2001. 2001.

[9] Patrick Houweling, Albert Mentink, Ton Vorst, *Comparing Possible Proxies of Corporate Bond Liquidity*. October 29, 2003.

[10] Wolfgang Bhler, Monika Trapp, *Explaining the Bond-CDS Basis The Role of Credit Risk and Liquidity*. CFR-working paper No 09-12, Center for Financial research, University of Cologne, 2009.

[11] Jan De Wit, *Exploring the CDS-Bond Basis*. NBB working paper No. 104, National Bank of Belgium, 2006.

[12] Haitao Li, Weina Zhang, Gi Hyun Kim, *The CDS-Bond Basis and the Cross Section of Corporate Bond Returns*. November 18, 2011.

[13] Qi Shang, *Limited Arbitrage Analysis of CDS-Basis Trading*. November, 2011.

[14] Monika Trapp, *Trading the Bond-CDS Basis - The Role of Credit Risk and Liquidity*. CFR-working paper No 09-16, Center for Financial research, University of Cologne, November, 2009.

[15] Amrut Nashikkar, Marti G. Subrahmanyam, Sriketan Mahanti, *Liquidity and Arbitrage in the Market for Credit Risk*. August 15, 2010.

[16] http://www.boerse-frankfurt.de/en/bonds/yield+calculator