Credit Curve Estimation and Corporate Bonds in the South African Market

by

Peter-John Clift

Assignment presented in partial fulfilment of the requirements for the degree of Master of Commerce in Financial Risk Management in the Faculty of Economics and Management Sciences at Stellenbosch University

Supervisor: Prof. W.J. Conradie

March 2020

PLAGIARISM DECLARATION

1. Plagiarism is the use of ideas, material and other intellectual property of another's work and to present it as my own.

2. I agree that plagiarism is a punishable offence because it constitutes theft.

3. I also understand that direct translations are plagiarism.

4. Accordingly, all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.

5. I declare that the work contained in this assignment, except otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

Acknowledgements

Prof. W.J. Conradie for his patience and guidance throughout. In my almost decade long relationship with Stellenbosch University he has been instrumental in influencing and shaping my future.

Mr Carel van der Merwe (almost Dr), who provided the initial topic, as well as valuable and insightful guidance around numerous technical aspects.

The Department of Statistics and Actuarial Science, for the valuable education I received, and excellence exhibited by everyone involved in this department.

Kobus, Divan and Jacques (the master's group), making the time spent together enjoyable.

My parents for giving me the opportunity to pursue my studies.

My wife, Michelle, for her love and support.

God, who among many other things, provides a purpose to what I do.

Abstract

Accurate fair value measurement of financial instruments serves as one of many mechanisms to enhance the integrity of financial institutions, particularly as it relates to counterparty credit risk. In this study, specific reference is made to credit spreads and the information that can be inferred from it for the purpose of fair value measurement. Market observable information, such as traded corporate bonds, together with accounting and share price information related to the issuers of these bonds, are used in order to construct credit spread curves. These credit curves are used as an input to calculate the value of corporate bonds, but can also be used in the calculation of measures related to counterparty credit risk management like the probability of default and loss given default parameters.

Currently there is no market standard model that can generate these credit curves. In this study, several models are introduced that may be appropriate to model credit spreads, as well as considerations for their application across a range of possible issuers. The accuracy of each model is tested by using these models to price newly issued corporate bonds and evaluating the resulting price difference from what is observed in the market.

Opsomming

Die akkurate billike waarde waardering van finansiële instrumente dien as een van vele meganismes om die integriteit van finansiële instellings te verbeter, veral ten opsige van teenparty kredietrisiko. In hierdie studie word spesifiek verwys na die krediet premie op korporatiewe effekte en die inligting wat daaruit afgelei kan word vir die doel van billike waarde bepaling. Markwaarneembare inligting, soos verhandelde korporatiewe effekte, sowel as rekenkundige- en aandeelprysinligting wat met die onderskrywers van hierdie effekte verband hou, word gebruik om krediet spreiding kurwes te op te stel. Hierdie krediet kurwes kan gebruik word om die waarde van korporatiewe effekte te bepaal, sowel as om parameters wat verband hou met die bestuur van teenparty kredietrisiko, soos waarskynlikheid van wanbetaling en die verlies gegewe wanbetaling, te bereken.

Daar is tans geen standaard model in die mark wat hierdie krediet kurwes kan genereer nie. In hierdie studie word verskillende modelle wat moontlik toepaslik kan wees om krediet premies te modelleer, asook oorwegings vir die toepassing daarvan vir 'n verskeidenheid moontlike onderskrywers van korporatiewe effekte voorgestel. Die akkuraatheid van elke model word getoets deur van hierdie modelle gebruik te maak om nuut uitgereikte korporatiewe effekte te prys en die gevolglike prysverskil te evalueer teenoor wat in die mark waargeneem word.

Table of contents

List of tables

List of figures

List of appendices

List of abbreviations and/or acronyms

AIP	All in price
ASW-spread	Asset swap spread
BESA	Bond Exchange of South Africa
Bps	Basis points
CCR	Counterparty credit risk
CDS	Credit default swap
CIB	Corporate investment bank
CVA	Credit value adjustment
DVA	Debt value adjustment
EAD	Exposure at default
EBA	European Banking Authority
EL	Expected loss
FRA	Forward rate agreement
G-spread	Government bond curve spread
IFRS	International financial reporting standards
JSE	Johannesburg Stock Exchange
LDA	Linear discriminant analysis
LGD	Loss given default
MTM	Marked-to-market
NFC	Non-financial corporate
PD	Probability of default
PV	Present value
YTM	Yield to maturity
Z-spread	Zero-coupon swap curve spread

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND TO PROBLEM

The financial crisis of 2007/2008 highlighted the importance of proper counterparty credit risk (CCR) management. The risk of a counterparty defaulting on their debt, or failing to meet their payment obligations, were prevalent during those years and the years preceding them. Consequently, there emerged a need for increased accuracy in terms of calculating the fair value of financial instruments.

Globally there has been a deliberate effort to enhance the transparency and integrity of banks and financial institutions by financial regulators. Of particular interest to this study is the framework for fair value measurements, namely IFRS (International Financial Reporting Standards) 13. In this specific standard a hierarchy for input data is provided with the aim of improved consistency and comparability in fair value measurements. A distinction is made between the quality of inputs and the related disclosures that should accompany them. Three different levels are defined in this hierarchy, with highest priority given to observed and unadjusted data, and lowest priority to unobservable inputs.

These levels are summarised as follows:

i.) Level 1 inputs can be actively observed in the market and is the most reliable source of data;

ii.) Level 2 inputs are inferred from level 1 inputs and include data points such as the implied volatility or credit spreads;

iii.) Level 3 inputs are generally unobservable and require a detailed disclosure to justify its use in calculating a fair value. These should only be used in instances where previous levels fail to provide adequate inputs.

In this study specific reference is made to credit spreads and the information that can be inferred from it for the purpose of fair value measurement. In short, as it relates to corporate bonds, the credit spread is the difference between the yield of a corporate bond and a corresponding risk-free rate. Alternatively, the credit spread is the additional spread that is added to the risk-free rates in order to obtain the market value in present value terms for a corporate bond, with the use of a discounted cash flow method. Formal definitions and details on credit spreads are discussed in subsequent chapters.

Essentially, credit spreads are useful inputs from which credit-risk related information can be inferred, predominantly the probability of default. An alternative source to credit spreads on corporate bonds from which to infer credit-risk related information is credit default swaps (CDS). Since CDS spreads are traded and directly market observable it will have a high priority in terms of the input data hierarchy recommended by IFRS 13. Since CDS spreads are actively traded, it is ideal to use. As South Africa does not currently have an active CDS market from which quoted CDS spreads can be inferred for a range of companies and sectors, alternative data should be considered. Therefore, bond yield data is the next best alternative, with the corresponding credit spreads inferred from them.

The calculation of credit spreads and corresponding credit curve is particularly useful for financial institutions, specifically relating to the pricing of risky loans and strategic decision-making regarding credit risk. With the credit curve, the ability to price newly issued risky bonds is enhanced. Furthermore, the credit spread can be used as input for many measures that relate to credit risk, such as default probabilities and credit valuation adjustments (CVA). Calculating these measures at variable valuation dates allows for the active management of CCR.

1.2 PROJECT OBJECTIVE

Currently there is no market standard model that can generate these credit curves. The objective of this study is to construct an accurate credit curve which can be used to improve the accuracy of the valuation of risky loans. This curve will further serve to enhance strategic decision-making with respect to risk, as it is an input for various measures that simplify the active management of CCR.

In this study, some of the following questions are addressed:

- What is the appropriate yield curve model for the construction of a credit curve in the South African market?
- Does this model remain valid given settings where the data is sparse?
- Is this model accurate when compared to the implied credit spreads of newly issued risky bonds?
- How could this model be useful for the calculation of measures that relate to credit risk, such as default probabilities or CVA?

The primary aim of this study is to develop a model to generate credit spreads of listed companies for a range of maturities.

1.3 CHAPTER OUTLINE

In chapter 2 a detailed discussion on default probabilities and the important difference between risk-neutral and real-world estimation thereof are provided. Credit spreads are discussed, including literature that relate to the various models and market information used to quantify it.

In chapter 3 risk-free rates are discussed, as well as the specific rates used in the South African market. The formal definition and notations related to credit spreads and term structures is given. The models that are applied in the subsequent analysis are introduced, being the Nelson-Siegel and Gaussian kernel.

In chapter 4 the outline of the research methodology applied, as well as a detailed discussion on the data that is used are provided. Here some preliminary results as it relates to the training data is given.

In chapter 5 the summarised results from the validation and test data is given.

Finally, in chapter 6, a summary of the main findings is provided as well as the proposal of some open questions for possible further research.

1.4 CONCLUSION

The value that this study could add to the existing literature is the construction of an accurate and feasible credit curve. This “final” credit curve is subject to many constraints, such as the intuitive notion that spreads of a higher credit rating should not be above that of a lower credit rating, for corresponding maturities. The key indicator of the appropriateness of using this curve in a South African context is to test how well it does in pricing newly issued risky loans.

CHAPTER 2

LITERATURE REVIEW: DEFAULT PROBABILITIES AND CREDIT SPREADS

2.1 INTRODUCTION

Credit risk arises when there is a non-zero probability that an obligor that issued debt or a counterparty in a derivative transaction may default. Credit risk has two components: probability of default (PD) and loss given default (LGD). PD refers to the probability that an obligor or counterparty fails to make their payments relating to the specific underlying instrument. LGD refers to the specific amount or proportion of the loss that is incurred when a default occurs. These two components, as well as the exposure at default (EAD) are used to calculate the expected loss (EL):

$$EL = PD \times LGD \times EAD. \tag{2.1}$$

EAD measures the amount of exposure in the event of default. EL is the decrease in market value resulting from the credit risk (De Laurentis et al., 2010).

In this chapter the estimation of default probabilities and how it relates to the fair value measurement of financial instruments is discussed. Details on the concepts of hazard rates and default probabilities are introduced, in particular as it relates to the use of credit spreads in the estimation thereof. Credit spreads and sources of market information to determine this are investigated, as well as elaborating on what is known as the “credit spread puzzle”.

2.2 RISK-NEUTRAL VS REAL-WORLD

Real-world default probabilities are typically estimated from historical data, whereas risk-neutral default probabilities are derived from market data using instruments such as credit default swaps (CDS) or bonds. Unknown parameters in pricing models, that are assumed to be correct, are calibrated such that the model price is equal to the observed market price. As an example, equity volatilities can be calculated from historical equity prices, or observed from what is implied by options that trade on these equities. Implied volatilities are determined by applying the Black-Scholes-Merton model and using the observed market price of the option. The estimate of volatility with the use of historical data may be entirely unrelated to the implied volatility observed at that point in time in the market. In this case, the implied volatility would be the risk-neutral version of volatility and the historical estimate of volatility the real-world.

A measure, also referred to as a probability measure, defines the market price of risk (Hull, 2012). Risk-neutral valuations are applied under the Q-measure and real-world under the P-measure. In pricing and valuation of financial instruments market practitioners generally aim to calibrate to risk-neutral parameters where possible, and resort only to real-world parameters in cases where the relevant market data is unobservable. The use of risk-neutral parameters is reasonable as hedging can only be performed with market observable instruments (Gregory, 2012). Real-world parameters are typically applied in instances such as scenario generation for risk management purposes.

Risk-neutral default probabilities are therefore estimated by solving the relevant model parameters so that the model implied price is equal to the market price. Risk-neutral default probabilities typically overstate the actual probability of default, since the actual probability of default is not the only factor that determines it. Other factors may include liquidity or a default risk premium. It should be expected that the risk-neutral default probabilities are generally higher than real-world default probabilities since investors are risk-averse and therefore price in a premium for accepting the default risk. Alternatively, real-world default probabilities are estimated from historical data.

If there were no expected excess return between risk-free bonds and risky corporate bonds, then real-world default probabilities and risk-neutral default probabilities, as estimated from bond prices, would be the same. This is not the case, however, as is illustrated in numerous academic studies. In particular, Altman (1989) tracked the performance of a portfolio of corporate bonds, across various credit ratings, and found that the returns outperformed the risk-free benchmark of treasury bonds. This indicates that investors could expect higher returns from investing in corporate bonds compared to investing in the risk-free treasury bonds.

Hull et al. (2005) provides an empirical comparison between real-world and risk-neutral probabilities of default. The real-world 1-year default probabilities were estimated from average cumulative default rates published by Moody's, between 1970 and 2003. The risk-neutral probabilities of default were estimated from Merrill Lynch bond indices, with the approximation for the 1-year risk-neutral default probability a bond given by:

$$\frac{y-r}{1-R} \tag{2.2}$$

where y is the bond's yield, r is the yield on a risk-free bond that pays off the same cash flows as the bond, and R is the recovery rate. This approximation is discussed in detail in the subsequent section, with the results of the comparison given in Table 2.1 below.

Table 2.1 Comparison between real-world and risk-neutral default probabilities, in basis points

Rating	Real-world	Risk-neutral	Ratio	Difference
Aaa	4	67	16.8	63
Aa	6	78	13	72
A	13	128	9.8	115
Baa	47	238	5.1	191
Ba	240	507	2.1	267
B	749	902	1.2	153
Caa and Lower	1690	2130	1.3	440

Source: Hull et al. (2005)

The results indicate that the difference between the default probabilities increases as credit quality declines, although the ratio of the risk-neutral to real-world default probabilities decreases. As a simple numerical example to illustrate the implications of these results, consider a one year zero coupon bond that is Ba rated and pays off $R\ 100$ at maturity. Further, ignore the time value of money and assume that in the case of default there is no recovery on the underlying asset. The real-world implied price of this bond would be $R\ 97.6, \big(100 \times (1 - 2.40\%)\big)$, but on average markets price this bond to be valued at $R\ 94.93, \big(100 \times (1 - 5.07\%)\big)$. The size of the difference between the two default probability estimates is often referred to as the credit spread puzzle. This, in part, refers to the expected excess return of corporate bonds over the risk-free bonds, a concept which is investigated further in subsequent sections.

2.3 PROBABILITY OF DEFAULT FOR FAIR VALUE ESTIMATION

Following the discussion above, the next step is to determine the appropriate method for estimating probability of default. In 2013, the IFRS 13 accounting guidelines were introduced. IFRS 13 provides a single framework for guidance around fair value measurement for financial derivatives, which subsequently encouraged convergence in approaches among market practitioners (Gregory, 2012). The following extract provides some clarity on the concept of fair value:

"*IFRS 13 defines fair value as the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date (an exit price). When measuring fair value, an entity uses the assumptions that market participants would use when pricing the asset or the liability under current market conditions, including assumptions about risk. As a result, an entity's intention to hold an asset or to settle or otherwise fulfil a liability is not relevant when measuring fair value*." (IFRS 13: Fair Value Measurement, 2017)

This definition of fair value proves to highlight the contrast in the appropriateness of estimation approaches for probability of default between risk-neutral and real-world. If the "exit price" is to be replicated as close as reasonably possible, then observable market data should determine the fair value of a financial derivative and not historical data. This is done to the extent that the relevant market data is reliably observable. This would lead to the conclusion that a risk-neutral framework is the appropriate framework for the purpose of determining the fair value of a financial instrument.

In addition, the "exit price" notion introduces the concept of debt value adjustment (DVA): the credit value adjustment (CVA) charge of a replacement counterparty when exiting a transaction (Gregory, 2012). This leads to the counterintuitive notion of a counterparty's own credit risk being treated as a benefit in the fair value calculation of a financial derivative. In essence, DVA would serve to decrease the fair value of a liability or increase the fair value of an asset. This concept and framework, including CVA and DVA, are discussed in more detail in section 3.2.2.

2.3.1 Hazard and Survival Functions

Typically, definitions that pertain to survival analysis refer to the probability of a specific event's occurrence. The event that is relevant in this case is the default event; therefore, subsequent definitions are given in light of this.

Definition 2.1 Survival Function: Let T be a non-negative continuous random variable with the cumulative density function $F(t)$ on $[0,\infty)$ and probability density function $f(t)$. The probability that a default event has occurred by time t is given by $F(t) = P(T < t)$. Then the survival function, or the probability that a default event has not occurred, is given by:

$$S(t) = P(T > t) = 1 - F(t). \tag{2.3}$$

Alternatively, the distribution of T can be characterised by the hazard function:

Definition 2.2 Hazard Function: The hazard function is defined as the instantaneous probability of default, given by:

$$\lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt | T > t)}{dt}. \tag{2.4}$$

The numerator of equation (2.4) is the conditional probability that a default event will occur in the interval $[t, t + dt)$ provided that it has not occurred before. The denominator is the length of the time interval. The instantaneous probability of default is then obtained by taking the limit as the width of the time interval approaches zero.

Given that the probability density function of T is given by:

$$f(t) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt)}{dt}.$$

With the use of Bayes' rule, equation (2.4) can be written as follows:

$$\lambda(t) = \lim_{dt\to 0} \frac{P(t \le T < t + dt|T > t)}{dt}$$

$$\lambda(t) = \lim_{dt\to 0} \frac{P(t \le T < t + dt)}{P(T > t)dt}$$

$$\lambda(t) = \frac{f(t)}{S(t)}.$$

Furthermore, given that:

$$\frac{dS(t)}{dt} = \frac{d\big(1 - F(t)\big)}{dt}$$

$$= -f(t),$$

it follows that:

$$-\frac{d(\log{(S(t)})}{dt} = \frac{-\frac{dS(t)}{dt}}{S(t)}$$

$$= -\frac{f(t)}{S(t)}$$

$$= \lambda(t).$$

Therefore:

$$-\frac{d(\log{(S(t)})}{dt} = \lambda(t).$$

Taking the integral on both sides gives:

$$-\log\big(S(t)\big) = \int_0^t \lambda(s)ds$$

$$S(t) = \exp\left[-\int_0^t \lambda(s)ds\right]. \tag{2.5}$$

The integral in the square bracket is referred to as the cumulative hazard rate. Using the simplifying assumption that risk is constant through time, $\lambda(t) = \lambda$, equation (2.5) can be rewritten as:

$$S(t) = \exp(-\lambda t). \tag{2.6}$$

It is useful to redefine some of the previous functions in terms of the cumulative hazard rate and the parameters in equation (2.6).

The cumulative density function, being the probability of default between time zero and time t, is given by:

$$P[T < t] = F(t) = 1 - e^{-\lambda t}. \tag{2.7}$$

This is also referred to as the cumulative probability of default. Then the survival function, or the probability of no default between time zero and time t, is given by:

$$P[T \geq t] = 1 - P[T < t] = 1 - F(t) = e^{-\lambda t}. \tag{2.9}$$

This is also referred to as the probability of survival. From equations (2.7) and (2.8), the cumulative probability of default converges to 1 as t grows large and the probability of survival converges to 0 as t grows large. The sum of the cumulative probability of default and the survival probability is also equal to 1 at every point in time.

Definition 2.3 Marginal Default Probability: The marginal default probability follows from the cumulative probability of default, i.e.:

$$dF(t) = \lambda e^{-\lambda t} dt. \tag{2.10}$$

Intuitively, this represents the marginal increase in the cumulative probability of default. The marginal default probability is strictly a positive number, since the cumulative probability of default is monotone increasing. Since the cumulative probability of default converges to 1 as t grows large, the marginal probability of default would converge to 0 as t grows large. The rate of this convergence is determined by the size of the hazard rate parameter. Intuitively, the marginal default probability between any two sequential dates can be interpreted as the difference between the cumulative default probabilities of the later and the first date. This is often the definition applied in practice, with the marginal default probability then given by:

$$PD(t_1, t_2) = F(t_2) - F(t_1),$$

with $PD(t_1, t_2)$ denoting the probability of default between time t_1 and t_2, and $t_1 \leq t_2$.

Alternatively, the probability of survival decreases over time:

$$\frac{dS(t)}{dt} = -\lambda e^{-\lambda t} < 0.$$

Figures 2.1 and 2.2 below illustrate the marginal and cumulative probabilities of default for various values of λ across a range of maturities.

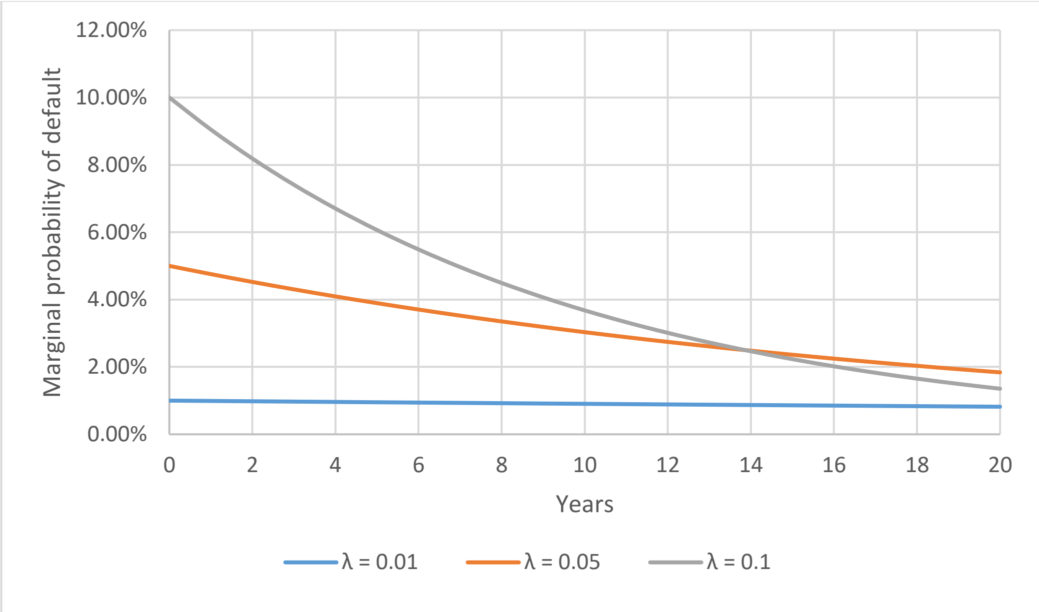

Figure 2.1: Marginal probability of default

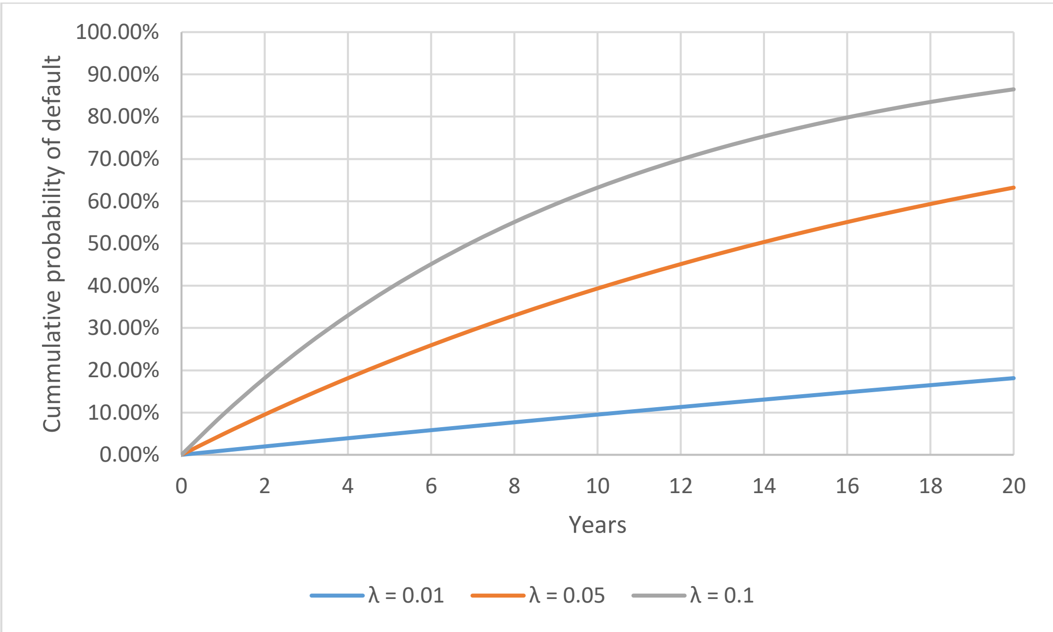

Figure 2.2: Cumulative probability of default

Definition 2.4 Conditional Default Probability: The conditional default probability is the probability of default over some time horizon $(t, t+\tau)$ given that there has been no default up to time t. This is given by:

$$P(T < t + \tau | T > t) = \frac{P(T > t \cap T < t + \tau)}{P(T > t)}. \tag{2.11}$$

Effectively, this is the ratio of the joint probability of survival up to point t and default occurring in the interval $(t, t+\tau)$ over the survival probability up to point t. The joint probability of survival up to point t and default in the interval $(t, t+\tau)$ is the event of defaulting between time $t+\tau$ and time t, therefore $P(T > t \cap T < t + \tau) = F(t+\tau) - F(t)$. Hence, equation (2.11) can be rewritten as:

$$P(T < t + \tau | T > t) = \frac{F(t+\tau) - F(t)}{S(t)},$$

which is the difference in the unconditional probability of default up to time $t+\tau$ and time t, divided by the probability of survival up to time t.

2.3.2 Risk-Neutral Estimates of Default Probabilities

In this section, the derivation of default probabilities from market prices are shown, with the majority of the definitions and notation drawn from Malz (2017). Default probabilities drawn from market prices are known as being risk-neutral. The alternative to risk-neutral default probabilities are real-world default probabilities and the difference between these are discussed in a subsequent section.

2.3.2.1 Constant Hazard Rates

Credit default swaps and bonds are the main instruments that lend themselves to default probability estimation, with the simplest of these being a zero-coupon corporate bond. This instrument is used in this section to illustrate the basic analytics, with more sophisticated extensions shown in subsequent sections. The following notation is applied:

$p_{0,\tau}$ = Present value of a default-free, or government, τ-year zero coupon bond,

$p_{0,\tau}^{corp}$ = Present value of a defaultable, or corporate, τ-year zero coupon bond,

r_τ = Continuously compounded discount rate on the default-free bond,

z_τ = Continuously compounded credit spread on the defaultable bond,

R = Recovery rate,

$\hat{\lambda}_\tau$ = τ-year risk-neutral hazard rate,

$1 - e^{-\hat{\lambda}_\tau}$ = Annualized risk-neutral default probability.

For both zero-coupon bonds it is assumed that they pay one unit of the given currency at maturity, such that the present value of a default-free τ-year zero coupon bond is given by:

$$p_{0,\tau} = e^{-r_\tau \tau}.$$

The credit spread z_τ is added to the discount rate r_τ to obtain the present value of the defaultable τ-year zero coupon bond, given by:

$$p_{0,\tau}^{corp} = e^{-(r_\tau + z_\tau)\tau} = p_{0,\tau} e^{-z_\tau}.$$

By definition the defaultable bond is more risky than the default-free bond, thus the defaultable bond would be cheaper since both have the same payoff at maturity. Therefore:

$$p_{0,\tau}^{corp} \leq p_{0,\tau}.$$

This then implies that $z_\tau \geq 0$.

To estimate the hazard rates it is necessary to assume that the issuer of the defaultable bond can experience a default at any time in the τ-year horizon and that in the event of default the holder of the bond will receive a deterministic and known recovery amount at the maturity of the bond. The recovery rate, denoted by R, is a fraction of the par amount of the bond, being one unit in this case. The recovery amount is received at the maturity date irrespective of when the bond defaults.

The risk-neutral τ-year probability of default is given by $1 - e^{-\hat{\lambda}_\tau \tau}$, with the estimated hazard rate assumed to be constant initially. If the recovery rate is assumed to be zero, $R = 0$, then the expected risk-neutral payoff of a defaultable bond that receives either one unit of the given currency or nothing at maturity, is given by:

$$E\left(p_{\tau,\tau}\right) = e^{-\hat{\lambda}_\tau \tau}.1 + \left(1 - e^{-\hat{\lambda}_\tau \tau}\right).0.$$

The expected present value of the payoffs is given by:

$$E\left(p_{0,\tau}\right) = e^{-r_\tau \tau}\left(e^{-\hat{\lambda}_\tau \tau}.1 + \left(1 - e^{-\hat{\lambda}_\tau \tau}\right).0\right).$$

Discounting this payoff with the risk-free rate is justified since the defaultable bond price and credit spread reflects the risk premium as well as an estimate of the true probability of default, which are both embedded in $\hat{\lambda}_\tau$. Therefore, with the risk-neutral default probabilities, the present value of the payoffs are set to be equal to the price of the defaultable bond. From this the risk-neutral hazard rate is estimated:

$$e^{-(r_\tau + z_\tau)\tau} = e^{-r_\tau \tau}\left(e^{-\hat{\lambda}_\tau \tau}.1 + \left(1 - e^{-\hat{\lambda}_\tau \tau}\right).0\right)$$

$$=> \qquad e^{-(r_\tau + z_\tau)\tau} = e^{-(r_\tau + \hat{\lambda}_\tau)\tau}$$

$$=> \qquad \hat{\lambda}_\tau = z_\tau$$

Thus, the risk-neutral hazard rate is equal to the credit spread. Assuming that the recovery rate is a positive value on $(0,1)$ then:

$$e^{-(r_\tau+z_\tau)\tau} = e^{-r_\tau\tau}\left(e^{-\hat{\lambda}_\tau\tau}.1 + \left(1-e^{-\hat{\lambda}_\tau\tau}\right).R\right) \qquad (2.12.a)$$

$$=> \qquad e^{-z_\tau\tau} = e^{-\hat{\lambda}_\tau\tau} + \left(1-e^{-\hat{\lambda}_\tau\tau}\right)R$$

$$=> \qquad e^{-z_\tau\tau} = 1-\left(1-e^{-\hat{\lambda}_\tau\tau}\right)(1-R)$$

$$=> \qquad 1-e^{-z_\tau\tau} = \left(1-e^{-\hat{\lambda}_\tau\tau}\right)(1-R)$$

$$=> \qquad \frac{1-e^{-z_\tau\tau}}{1-R} = 1-e^{-\hat{\lambda}_\tau\tau} = P(T<\tau).$$

The above equation implies that the τ-year probability of default is equal to the additional credit-spread discount on the defaultable bond, divided by the LGD. Alternatively:

$$1-e^{-z_\tau\tau} = P(T<\tau)(1-R)$$

$$= P(T<\tau)LGD,$$

i.e. the additional credit-spread discount on the defaultable bond is equal to the product of the τ-year probability of default and the LGD. Furthermore, taking the logs of equation (2.12.a) gives:

$$-(r_\tau+z_\tau)\tau = -r_\tau\tau + \log\left(e^{-\hat{\lambda}_\tau\tau} + \left(1-e^{-\hat{\lambda}_\tau\tau}\right)R\right),$$

$$z_\tau\tau = -\log\left(e^{-\hat{\lambda}_\tau\tau} + \left(1-e^{-\hat{\lambda}_\tau\tau}\right)R\right). \qquad (2.12.b)$$

With the approximation $e^x \approx 1+x$ and $\log(1+x) \approx x$, and taking exponents, then the right-hand side of equation (2.12.b) can be written as:

$$e^{-\hat{\lambda}_\tau\tau} + \left(1-e^{-\hat{\lambda}_\tau\tau}\right)R \approx 1-\hat{\lambda}_\tau\tau + \hat{\lambda}_\tau\tau R$$

$$=> \qquad 1-\hat{\lambda}_\tau\tau + \hat{\lambda}_\tau\tau R = 1-\hat{\lambda}_\tau\tau(1-R).$$

Hence,

$$\log\left(e^{-\hat{\lambda}_\tau\tau} + \left(1-e^{-\hat{\lambda}_\tau\tau}\right)R\right) = \log\left(1-\hat{\lambda}_\tau\tau(1-R)\right)$$

$$\approx -\hat{\lambda}_\tau\tau(1-R).$$

Thus, from equation (2.12.b) it follows that:

$$z_\tau\tau \approx \hat{\lambda}_\tau\tau(1-R)$$

$$=> \qquad z_\tau \approx \hat{\lambda}_\tau(1-R) \qquad (2.13.a)$$

$$=> \qquad \hat{\lambda}_\tau \approx \frac{z_\tau}{1-R}. \qquad (2.13.b)$$

The credit spread is approximately equal to the product of the risk-neutral hazard rate and the LGD. Furthermore, with the implications of the same approximation, $e^x \approx 1 + x$, and $1 - e^{-\hat{\lambda}_\tau}$ being the annualised probability of default:

$$\hat{\lambda}_\tau \approx 1 - e^{-\hat{\lambda}_\tau}.$$

Therefore, equation (2.13.a) implies that the credit spread is approximately equal to the product of the probability of default and the LGD. These approximations work well when both the spreads and risk-neutral default probabilities are not large.

2.3.2.2 Time Varying Hazard Rates

In this subsection the extension of the hazard rates and probability of defaults that accommodates hazard rates that vary over time is provided. Let $F(t)$ denote the t-year probability of default, then it follows from equation (2.3) and (2.5) that:

$$F(t) = 1 - e^{\int_0^t \lambda(s)ds}. \tag{2.14}$$

If the hazard rates are assumed to be constant (equation (2.7)) then $\lambda(t) = \lambda$, for $t \in [0, \infty)$. Equation (2.14) is then reduced to $F(t) = 1 - e^{-\lambda t}$. Suppose that the hazard rates are derived from CDS spreads that have three traded maturities, t_1, t_2 and t_3, then three piecewise constant hazard rates are derived from the dates as:

$$\lambda(t) = \begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{Bmatrix} for \begin{Bmatrix} 0 < t \leq t_1 \\ t_1 < t \leq t_2 \\ t > t_2 \end{Bmatrix}.$$

The integral from equation (2.14) is then given by:

$$\int_0^t \lambda(s)ds = \begin{Bmatrix} \lambda_1 t \\ \lambda_1 t_1 + (t - t_1)\lambda_2 \\ \lambda_1 t_1 + (t_2 - t_1)\lambda_2 + (t - t_2)\lambda_3 \end{Bmatrix} for \begin{Bmatrix} 0 < t \leq t_1 \\ t_1 < t \leq t_2 \\ t_2 < t \end{Bmatrix}.$$

Hence, the probability of default distribution is estimated by observing the relevant spreads in the market and backing out $F(t)$ through a bootstrapping procedure by solving the relevant market instrument across the specified points in time. This would be similar to the analysis shown in the previous section.

2.4 CREDIT SPREAD ANALYSIS

In this section, various sources of information as well as models that relate it to credit spreads are considered. This follows the mathematical derivations given in section 2.3 and relates it to market observable information. Additionally, some variations regarding the use of market observable information and credit spreads are discussed, as well as suggestions in cases of a clear lack of relevant information.

2.4.1 Models and information

A CDS is a derivative that offers protection against an obligor defaulting on its debt. Therefore, CDS spreads are considered to be a reliable measure of default risk since they provide the compensation that market participants require for bearing that specific risk. Given that many countries, including South Africa, lack a readily available and liquid CDS market, there is a need for a general approach with which to determine the appropriate credit spread proxy for counterparties. Currently no such method or model, that has been standardised, exists (Gregory, 2012). This is not surprising given the amount of subjectivity involved, in particular with respect to the estimation of a credit curve for a counterparty that lacks the relevant traded data from which to infer this.

In cases where credit spreads are not liquid, or market observable, institutions are required to proxy the credit spreads based on their rating, region, and sector. Even though these categories are broad, there may still be instances where data constraints exist. The intersection method (or bucketing method), proposed by the European Banking Authority (EBA), averages data of illiquid names across the relevant sub-categories to determine the implied proxy spread (EBA, 2013). With this methodology the proxy spread of obligor i is defined as:

$$S_i^{proxy} = \frac{1}{N}\sum_{j=1}^{N} S(j),$$

where $N \geq 1$ is the number of liquid names in the same rating, region and sector sub-categories as obligor i and $S(j)$ their corresponding spreads.

In Choudrakis et al. (2013) some of the shortcomings of the practical implementation of the intersection method are discussed. These include problems such as instances where there is simply not enough data points for a chosen bucket with some sectors, regions, or rating intersections. These buckets contain few or no liquid obligors, which results in an undefined spread. This constraint may lead to choosing a much broader definition for each sub-category, such as grouping all South-American obligors together. This may lead to unique information to each obligor being lost, such as differentiating between Brazil and Argentina in the abovementioned example. Data constraints often lead to historical instability, with cases of rating migrations in sparsely populated buckets causing the average spread to move significantly when the given spread either enters or exits the bucket.

Finally, with any model that attempts to calculate proxy spreads it should be expected that the spreads produced are monotonic by rating. This implies that worse ratings have an equal or larger spread than the ratings that are superior. With the intersection method, the rating is the strongest indicator of CDS spreads. Choudrakis et al. (2013) found that the intersection method frequently produces proxy spreads that are not monotonic by rating.

An alternative to the intersection methodology, introduced by Choudrakis et al. (2013), is the cross-section methodology. With this methodology, the proxy spread of obligor i is given by:

$$S_i^{proxy} = M_{glob} M_{sctr(i)} M_{rgn(i)} M_{rtg(i)} M_{snty(i)},$$

with $scrt(i), rgn(i)$, $rtg(i)$ and $snty(i)$ denoting the sector, region, rating, and seniority of obligor i respectively. A fundamental assumption in this methodology is that there is a single multiplicative factor, such as the regional factor, that is independent of the sector, rating, or seniority of all the obligors that fall within the chosen region.

The calibration of the cross-section factors to market data is done by minimising the total squared difference in log spreads. This is applied on equation (2.15):

$$y_i = \sum_{j=1}^{n} A_{ij} x_j, \tag{2.15}$$

where $y_i = \log(S_i^{proxy})$, $x_j = \log(M_j)$, and n is the total number of factors (total number of sectors, regions, ratings, seniorities, and one global factor). A is an indicator matrix that equals 1 for a corresponding factor of an obligor and 0 otherwise.

Compared to the intersection method, Choudrakis et al. (2013) found that the cross-section method results on more stable historical spreads as well as producing spreads that are monotonic by rating substantially more often. Additionally, the trade-off between choosing how broad the categories for each bucket is in the intersection method does not exist in the cross-section method. This effectively negates the possibility of losing information unique to a particular obligor.

A limitation that is present in both the intersection and cross-section methodologies is the prevalence of CDS quotes for obligors with similar ratings, regions and sectors that have significant deviations. Sourabh et al. (2018) propose that to achieve a substantial increase in accuracy it is necessary to incorporate additional information such as equity data. Perhaps the most prominent justification for the use of equity prices comes from Merton (1974). The model proposed by Merton, an example of a structural approach, is based on the premise that the event of default occurs when the structure of the company is no longer considered worthwhile. This model assumes that the firm's only debt issue is a zero-coupon bond with a face value of F, maturing at time T. If the firm is unable to pay the principal at time T, then the firm is considered to be defaulted and the equity worthless. Alternatively, if the firm's asset value at time T, denoted by V_T, is more than the debt principal the firm is able to repay the debt. Then the equity value is given by $V_T - F$. These two possible payoffs are similar to a call option on the assets of the firm, with the outstanding debt as the strike price, and can be modelled with the use of the Black-Scholes-Merton option pricing formula. The value of the firm's equity is given by:

$$E_T = \max(V_T - F, 0).$$

In this framework the relevant variables include interest rates, the principal value of the debt, the company's asset value and the volatility of the company's assets. The Black-Scholes-Merton formula gives the present value of the equity, E_0, by:

$$E_0 = V_0 N(d_1) - Fe^{-rT} N(d_2),$$

with

$d_1 = \frac{\ln\left(\frac{V_0}{D}\right)+\left(r+\frac{\sigma_V^2}{2}\right)T}{\sigma_V\sqrt{T}}$ and $d_2 = d_1 - \sigma_V\sqrt{T}$

and

σ_V = the volatility of the assets, and

r = the risk-free rate corresponding to time T.

The risk-neutral probability of default is given by:

$$P(V_T < F) = N(-d_2).$$

This framework implies that an important driver of default is volatility, leverage, and market interest rates. The asset volatility, σ_V, is not directly observable, but can be estimated from equity data provided the company is publicly traded. Although probability of default is not exactly the same as credit spreads, given the links derived in previous sections it is relevant to look at what information drives each, as these should be considered to be related.

Given the links between equity data and credit spreads, Sourabh et al. (2018) propose amending the model provided in the cross-section methodology. These amendments include simply adding the additional factors of equity returns and volatility of equity returns to the current cross-section model. The addition of these factors led to increased accuracy of proxy spreads compared to both the intersection and original cross-section methodology. Sourabh et al. (2018:480) goes on to state that this methodology may provide a solution in cases where financial market participants revert to using historical probabilities of default due to a lack of data.

In determining the probability of default for companies, an alternative source to market information is accounting-based information. This is notably discussed in Altman (1968), which proposed what is known as Altman's Z-score. This is a scoring methodology that classifies companies as either performing or in default. Linear discriminant analysis (LDA) is applied to determine the relevant parameters for predicting default. The variables are chosen based on their estimated contribution to the probability of default and come from wide range of accounting ratios. The resulting model is given by:

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.00999X_5,$$

where

X_1= Working capital / Total assets,

X_2= Retained assets / Total assets,

X_3= Earnings before interest and taxes / Total assets,

X_4= Market value of equity / Book value of total debt,

X_5= Sales / Total assets, and

Z = Overall index.

The output in the form of the overall index level, Z-score, effectively provides a method to evaluate or rank a company's relative likelihood to default. The higher the Z-score, the more likely a firm is to be performing. This method has several drawbacks, including the possibility of companies having the same Z-score but being in different states, performing and default.

Das et al. (2008) investigated the relative performance of models that use either accounting-based information or market-based information to explain CDS spreads. They found that the models that apply accounting-based information explain CDS spreads at least as well as structural models that make use of market-based information. An additional advantage that models using accounting-based information have is the ability to quantify credit risk for companies that are not publicly traded. Das et al. (2008) concludes, however, that accounting-based and market-based information should be considered as complimentary since models that make use of both sets of information tend to explain a substantially larger portion of CDS spreads.

Additional research on the use of accounting data to explain credit spreads include Demirovic et al. (2015), where the authors employ a sample of credit spreads on vanilla corporate bonds issued by non-financial companies, consisting of 349 firms and a total of 11,632 quarterly observations. The initial idea is to evaluate the assumption that underlies the efficient market hypothesis: that market prices should reflect all available information. This has the implication that a structural model, such as the Merton (1974) model, which uses market information, would outperform models that use accounting data in explaining variations in the credit spread on corporate bonds. The conclusion reached by the authors is that although equity volatility and Merton's distance-to-default outperform accounting variables in explaining variations in the credit spread, accounting variables are incrementally informative when considered in conjunction with market-based measures. This implies that both sets of information is required to obtain the highest explanatory power for the underlying model.

2.4.2 Credit Spread Puzzle

The spreads of corporate bonds are typically much wider than the spreads implied by expected default losses. This characteristic of financial markets is initially discussed in a previous section with reference to the difference between risk-neutral and real-world estimation methodology for probability of default. Similarly, the gap between market observed spreads and expected default losses, referred to as the “credit spread puzzle” (Amato et al., 2003), is discussed in this section.

Amato et al. (2003) investigates the determinants of credit spreads, with the use of bond indices covering the period from January 1997 to August 2003. It is important to note that the derived spreads are compared to expected loss (EL). Estimates of expected loss are calculated in this case with the use of one-year ratings transition matrix, which includes the rating migrations as well as defaults. The expected losses are then calculated as the average of an issue defaulting within the next T years times loss given default. This is similar to a real-world estimation of expected losses, calculated from historical data. The results from their comparison is given in Table 2.2 below.

Table 2.2 Spreads and expected default losses, in basis points

Rating	**Maturity**							
	1–3 years		**3–5 years**		**5–7 years**		**7–10 years**	
	Spread	**Expected loss**	**Spread**	**Expected loss**	**Spread**	**Expected loss**	**Spread**	**Expected loss**
AAA	49.50	0.06	63.86	0.18	70.47	0.33	73.95	0.61
AA	58.97	1.24	71.22	1.44	82.36	1.86	88.57	2.70
A	88.82	1.12	102.91	2.78	110.71	4.71	117.52	7.32
BBB	168.99	12.48	170.89	20.12	185.34	27.17	179.63	34.56
BB	421.20	103.09	364.55	126.74	345.37	140.52	322.32	148.05
B	760.84	426.16	691.81	400.52	571.94	368.38	512.43	329.40

Source: Amato et al. (2003)

The results in Table 2.2 indicate that for higher credit ratings, the difference in spreads and expected losses are smaller than for lower credit ratings across maturities. Conversely, the ratio of spread to expected loss is larger for higher credit ratings than for lower credit ratings across maturities. An interesting observation is that the term structures of spreads have different shapes across the rating grades. The term structures are upward-sloping for the higher rated investment grade bonds, hump-shaped for BBB bonds and downward sloping for the non-investment grade bonds.

As is illustrated in Table 2.2, across all rating categories and maturities, expected loss accounts for only a minor portion of spreads. Amato et al. (2003) notes further that additional factors such as taxes, risk premium, liquidity and risk of unexpected losses contribute to explain the spreads on corporate bonds. It is important to note that the expected loss numbers are not comparable to default rates, since these are reduced through the multiplication of the loss given default parameter.

Longstaff et al. (2005) did a similar analysis and determined that the default component accounts for the majority of the credit spreads of corporate bonds, across all credit ratings. In particular, they noted that the default component represents the following percentages of the spreads across credit ratings: 51% for AAA/AA-rated bonds, 56% for A-rated bonds, 71% for BBB-rated bonds and 83% for BB-rated bonds. The non-default components were found to be strongly related to bond specific components, such as liquidity and the outstanding principal amount. In addition to these, Longstaff et al. (2005) determined that the non-default component related to taxes is comparatively insignificant.

Giesecke et al. (2010) used an extensive dataset of corporate bonds, spanning 1866 to 2008, and found that, on average, credit spreads are roughly twice as large as default losses. This implies that the ratio of risk-neutral to actual, or real-world, default losses is roughly two. This is on average across time and credit ratings, with the particular sample containing bonds issued by firms in the United States of America from the non-financial sector.

Liquidity is another key component that contributes to credit spreads on corporate bonds. It is comparatively difficult to quantify the extent that liquidity, or illiquidity, contributes to the observed credit spreads. Houweling et al. (2005) investigated various proxies that can be used to measure corporate bond liquidity. Several factors, including issued amount, if a firm's equity is listed, and age, were identified as viable proxies. Firstly, the larger the issued amount of a bond, the more frequently the bond is assumed to be traded, which would reduce illiquidity. Secondly, listed companies have more information publicly available, which would potentially reduce cost of making a market for bonds of listed companies compared to those of unlisted companies. This would then have the effect of reducing illiquidity. Thirdly, the age of a bond, which is the time since issuance, is found to be positively related to illiquidity. The underlying reasoning for this is that as the bond gets older an increasing percentage thereof is absorbed into portfolios with buy-hold strategies. This implies that less trading on these bonds occur, which would increase illiquidity.

2.4.3 CDS vs Bond Spread Basis Adjustment

Given that a CDS can be used to hedge a position in a corporate bond, it may be informative to consider how CDS and bond spreads relate to one another. The difference in spread between a CDS and a corporate bond is referred to as the CDS-bond basis, defined as:

$$CDS\_bond\ basis = CDS\ spread - Excess\ of\ bond\ yield\ over\ risk\ free\ rate. \qquad (2.16)$$

This is most clearly illustrated through an example, similar to how it is discussed in Hull (2012). Suppose that an investor buys a 5-year corporate bond that pays an annual coupon of 8% at par value of R100 per R100 notional. The investor then enters into a 5-year CDS that has a spread of 2%, or 200 basis points, as protection on the issuer of the bond defaulting. If the bond issuer does not default, the investor earns a net of 6% per year. If the bond issuer does default, the investor earns a net of 6% per year up to the point of default. At the point of default, the investor then receives the notional value of the bond through the CDS, which can then be invested at the risk-free rate for the remainder of the 5-year period. Effectively, the investor has converted his risky bond yielding 8% into a risk-free bond yielding 6%.

The above example serves to illustrate that an n-year CDS spread should be approximately equal to the n-year credit spread the CDS reference entity's bonds have over the risk-free rate. The difference between these two spreads is referred to as the CDS bond basis. If the CDS bond basis is non-zero, it implies that a theoretical arbitrage opportunity exists.

This example is simplified and ignores much of the exact market dynamics surrounding the actual trading conventions of these instruments. Much research has gone into quantifying and understanding the drivers of the CDS bond basis. Zhu (2004) confirmed the theoretical prediction that the CDS bond basis is zero on average in the long run, although there are short run discrepancies. Possible reasons for the short run discrepancies are due to the different responses to changes in credit conditions. The existence of transaction costs and illiquidity enable small arbitrage opportunities between the two markets to exist. Furthermore, the author notes that the impracticality of short selling bonds does not enable traders to dynamically take advantage of such arbitrage opportunities.

De Wit (2006) performed a cointegration analysis, as proposed by Engle et al. (1987), to investigate the long run relationship between CDS spreads and bond spreads. The conclusion was that these variables are in fact cointegrated, although for the period 2004-2005 the median CDS bond basis was positive (7.5 basis points). Furthermore, the CDS bond basis for emerging market sovereign entities is significantly higher than for corporate issuers, while USD issuers have significantly higher CDS bond basis compared to issuers in the European markets. This emphasises the authors assertion that the CDS bond basis tends to be market specific and numerous factors, in particular liquidity, affect the results.

More recent research indicates a much larger and positive CDS bond basis than what is suggested by Zhu (2004) and De Wit (2006). In particular, Gyntelber et al. (2017) found that there is an equilibrium CDS bond basis at a certain point that indicates arbitrageurs' step in and carry out basis trades only when the expected gain from the arbitrage trade is greater than the trading costs. This resulting threshold would represent the costs of trading the specific instruments. The data for their analysis consisted of CDS contracts and government bonds for France, Germany, Greece, Ireland, Italy, Portugal, and Spain from January 2008 to December 2011. The persistence of a positive basis between sovereign CDS and sovereign bond spreads indicates that the theoretical no-arbitrage condition of a zero basis is prohibited through these transaction costs. The authors find that this transaction costs, or equilibrium CDS bond basis, is on average 190 basis points during the Euro sovereign credit crisis in 2010-2011, compared to an average of 80 basis points before the crisis. This sharp increase likely reflects the increased risks of engaging in such trades during the crisis, resulting in higher thresholds.

In contrast to Gyntelber et al. (2017), other research has found that there is a negative CDS bond basis. Kim et al. (2016) uses data of senior unsecured U.S. dollar denominated CDS spreads, and senior unsecured, fixed-rate, straight bonds with semi-annual coupon payments that has credit rating information readily available. Bonds with embedded options, floating coupons, and less than one year to maturity, are removed. The sample period is between 2 January 2001 and 31 December 2008. The CDS spreads are interpolated for a given firm to match the maturity for the corresponding corporate bond spreads, such that the CDS bond basis can be calculated. The average CDS bond basis from this sample, across all maturities and ratings, is -40 basis points. This non-zero CDS bond basis is explained by a set of market frictions and risks involved in the basis arbitrage, such as liquidity and counterparty credit risk.

Bühler et al. (2009) performed a similar analysis, with a sample of CDSs and bonds, denominated in Euro, from 1 June 2001 to 30 June 2007. Selected results of their analysis are given in Table 2.3 below. It is important to note that the CDS bond basis is defined the other way around in their methodology, with the CDS spreads being subtracted from the bond spreads to give the basis. Therefore, a positive basis would imply a negative basis based on equation (2.16) as well as the research mentioned previously. The results indicate an average negative CDS bond basis of 48.75 basis points, across all ratings and sectors. It is interesting to note that the absolute CDS bond basis is on average smaller for lower credit ratings than for higher credit ratings.

Table 2.3 CDS bond basis comparison

		AAA-BBB	BB-CCC	Financial	Non-Financial	All
Bond Spreads	Mean	93.05	327.69	59.81	122.58	91.92
	Std. Dev.	123.83	311.91	126.63	151.66	126.74
	Min.	-199.60	-20.56	-195.54	-199.60	-199.60
	Max.	1,603.09	2,288.17	1,380.78	2,288.17	2,288.17
CDS Spreads	Mean	42.69	309.46	19.03	70.30	55.44
	Std. Dev.	57.10	246.82	18.25	110.15	96.19
	Min.	3.00	38.50	3.00	5.33	3.00
	Max.	1,393.75	1,874.88	310.00	1,874.88	1,874.88
Basis	Mean	50.29	18.32	40.74	52.03	48.75
	Std. Dev.	115.99	195.84	127.71	118.30	121.22
	Min.	-516.48	-726.41	-217.97	-726.41	-726.41
	Max.	1,573.93	1,462.31	1,375.71	1,573.93	1,573.93

Bühler et al. (2009).

It is important to note that there are some differing methodologies applied between a few of the aforementioned research. In particular, Gyntelber et al. (2017) used the asset-swap spread as the corporate bond credit spread, while Kim et al. (2016) used the Z-spread. The differences in these definitions would influence the results, and a more detailed discussion on these are provided in chapter 3. From Table 2.3 it is also noted that there are what appears to be outliers in the observations, such as negative bond spreads and maximum bond spreads of more than 2000 basis points. Outliers such as these may or may not be excluded based on some filtering rule applied, which could vary based on the specific researcher's approach.

Given the differences of methodology and results, it may be an interesting research project to reperform either set of results from Gyntelber et al. (2017) and Kim et al. (2016) with consistent definitions, to see whether the results would change, and if so in which direction.

2.4.4 Credit Spread Mapping

Banks would generally have numerous counterparties for derivative trades and risky loans. Therefore, credit curve estimation is an important input when pricing these. Often these counterparties may not have liquid CDSs or bonds that are readily observable in the market from which the relevant credit-risk related information can be observed. This serves to emphasise the importance of a general credit curve estimation framework, as well as the subjectivity in the construction thereof.

Regulators generally propose mapping to be based on rating, region and sector, as discussed in section 2.4. Other potential issues banks face when constructing credit curves are the necessity to consider the tenor, seniority, liquidity, and reference instrument characteristics when inferring credit-risk related information from them. In light of this, Gregory (2012) has suggested the following decision tree framework to determine the appropriate counterparty credit spread:

i.) Is there a liquid CDS? If yes, then this can be used directly to determine the credit spread.

ii.) Is there is no liquid CDS, but some other relevant liquid instrument like a bond? If yes, then this can be used to derive the credit spread, along with a basis adjustment.

iii.) If neither set of information are available, then a suitable company that has the relevant information available can be used as proxy. This may be a parent company or the sovereign, which would then require some adjustment to the derived credit spread to reflect the increased riskiness.

iv.) If all three previous conditions are not able to be met, then generic mapping can be done. This would of course be substantially more subjective but given the lack of observable and relevant information the only reasonable resolution. This would include considerations proposed by regulators such as rating, region and sector.

2.5 CONCLUSION

This chapter forms the first part of the literature study. Herein the framework of probability of default estimation is discussed, in particular the importance and relevance of the difference between risk-neutral and real-world is highlighted. Specifics on the credit spreads are provided, with reference to the credit spread puzzle. The difference between sources of information to quantify credit spreads are discussed, which is important throughout this study. Finally, it is noted that there is no standardised way to produce a range of credit curves for each counterparty required, and that the available market information should be utilised as far as possible. A suggested credit curve mapping framework is then provided.

In the following chapter detail on the modelling of credit spreads is provided, which builds on the theoretical framework of credit spreads as well as the market information related to it. This is done by discussing the appropriate risk-free rate to use, as well as detailing several credit spread term structure models.

CHAPTER 3
LITERATURE REVIEW: CREDIT SPREAD MODELING

3.1 INTRODUCTION

"Why would an investor hold a security with an expected loss? Because he believes the credit spread more than compensates for the expected loss" - Malz (2011:203). In other words, investors readily invest in riskier instruments due to the comparatively higher yield that accompanies it.

As an example, suppose there exists a risk-free bond that pays a 6% annual coupon with one year to maturity and a similar defaultable bond that pays an 8% annual coupon with one year to maturity. If both these bonds trade at par, being R100 per R100 notional, then the market believes that the additional credit spread inherent to the defaultable bond's discount factor adequately captures or represents the additional risk inherent to it. This is derived from implied difference in discount factors that would be applied on both cash flows to get a present value of R100 in each case. Comparatively the risk-free bond would have no additional spread added on the risk-free rate when discounting the final payment. This credit spread, including the modelling thereof, and risk-free rates are discussed in this chapter.

3.2 RISK-FREE RATES

3.2.1 Curves used in South Africa

For a corporate bond, the credit spread is one of the most important measures investors use for credit security selection. This provides a way to estimate the credit-related risk that will be assumed by investing in the specific bond and serves as a key input in the pricing thereof.

Several types of credit spread measures exist, with the relevance of each determined by the data availability in the specific market as well as the corresponding instrument's inherent characteristics. To inform the choice of spread measure, it is therefore necessary to consider the choice of the specific risk-free rate with which it is possible to infer a credit spread. In South Africa there are three prominent risk-free rates applied in the pricing of fixed income debt: bond curve, swap curve, and real curve (JSE, 2012).

These curves represent the term structure of the specified interest rates across a continuum of maturities. The interest rates that comprise the curves are derived from each curve's underlying instruments and calculated such that the curve prices each underlying instrument as close as possible to its traded market price. Table 3.1 provides the relevant instruments and their corresponding marked-to-market (MTM) yield to maturities (YTM) for the bond curve, as at 31 December 2018.

Table 3.1 Bond curve inputs

Input	**MTM YTM**
SAFEX Overnight	6.590
TB91D	7.610
TB182D	7.760
TB273D	7.820
TB364D	7.720
R208	6.995
R2023	8.100
R186	8.875
R2030	9.385
R213	9.490
R209	9.705
R2037	9.830
R2044	9.925
R2048	9.900

Source: JSE

The inputs include an overnight rate, numerous short dated treasury bills, and several government bonds. Each input has a corresponding yield, which is the traded YTM that reproduces the current market value of the instrument. As an example, the YTM of a fixed coupon bond is derived by solving for y_{YTM} in the below equation:

$$PV_{Bond} = \sum_{i=1}^{n} C_i \mathrm{e}^{-y_{YTM} \times T_i} + m\,\mathrm{e}^{-y_{YTM} \times T_n},$$

with

C_i = The periodic coupon payment,

T_i = The corresponding time to maturity of payment i,

m = The final repayment of the principal or face value,

PV_{Bond} = The current market value of the relevant bond.

The bootstrapping and interpolation methodology applied in the curve construction is not discussed here and the interested reader is referred to the work of Hagan and West (2006), and Hagan and West (2008) for further details. Ultimately, a curve is constructed such that it will price back its inputs as close as possible, termed a “perfect-fit” curve. For example, each future cash flow from the R2023 bond is discounted at the rate that corresponds to the future date on the curve which should closely reproduce the current value of the R2023 bond. An example of this curve is provided in Figure 3.1.

Whereas the bond curve represents the zero-coupon yields at which the South African government can obtain funding, across a range of maturities, the swap curve represents the zero-coupon yields at which interbank funding can be obtained. The inputs for this curve are given in Table 3.2:

Table 3.2 Swap curve inputs

Point	**Instrument**
Overnight	SAFEX Overnight
1 Month	1 Month JIBAR
3 Month	3 Month JIBAR
4 - 24 Months	FRAs
2 - 30 Years	Swaps

Source: JSE

The inputs in Table 3.2 are denoted differently to those in Table 3.1, since there are a range of maturities with corresponding yields for the swaps and forward rate agreements (FRAs) that go into the construction of the curve. The swap curve is constructed in a similar way such to be a "perfect-fit" curve, by pricing back its inputs as close as possible.

Finally, the real curve represents the real zero-coupon yields which the South African government can obtain funding. This curve is constructed with inflation linked money market instruments, and inflation linked government bonds. The inputs for this curve are given in Table 3.3:

Table 3.3 Real curve inputs

Input	**MTM YTM**
R212	2.890
R197	2.900
I2025	3.140
R210	3.230
I2029	3.240
I2033	3.290
R202	3.280
I2038	3.305
I2046	3.335
I2050	3.340

Source: JSE

The inputs to the real curve have substantially lower yields compared to those of the bond curve. This is due to these yields being provided in real terms and not nominal terms. The general approximation of nominal interest rates in terms of real interest rates and inflation is given by:

$$Nominal\ Interest\ Rate \approx Real\ Interest\ Rate + Inflation.$$

Again, this curve is constructed such that it will price back its inputs as close as possible, being a "perfect-fit" curve. The bonds from which this curve is constructed have been issued such that both their coupons and principal are linked to the South African Consumer Price Index (CPI) as published by Statistics South Africa (Stats SA). The historical CPI information can be found at www.statssa.gov.za.

In Figure 3.1 below the resulting curves as at 31 December 2018 are given.

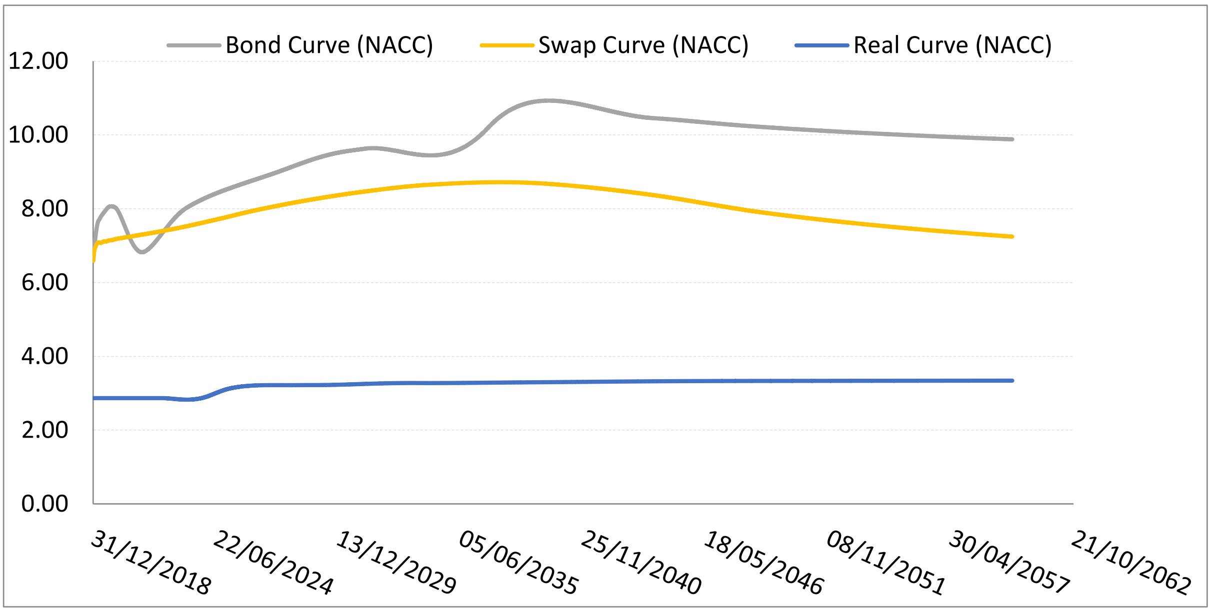

Figure 3.1: Bond, Swap, and Real curves, as at 31 December 2018

Source: JSE

Financial institutions, such as banks, would construct these curves on a continuous basis for pricing and risk management purposes. These would then inform their pricing of specific derivatives and fixed income instruments. In a functioning capital market, yield curves serve numerous purposes (JSE 2012), including:

i.) The valuation of any future cash flow or series of cash flows,

ii.) They provide a reliable indication of the current expectation of future interest rates,

iii.) They inform the analysis of fixed income trading strategies, and iv.) They are used, in conjunction with other market information such as swaption volatility surfaces, to calibrate no-arbitrage term structure models. Examples of these models include the models of Ho and Lee (1986) and Hull and White (1990).

3.2.2 Overnight Indexed Swap Curve

In the South African market, the swap curve is generally considered to be the appropriate risk-free rate used in the pricing of financial instruments such as derivatives and corporate bonds that are not inflation linked. The bond curve is primarily used to price government bonds, and the real curve to price inflation linked bonds. Some banks in South Africa also consider the OIS curve (overnight indexed swap curve) as the appropriate risk-free rate. This is a curve constructed in a similar way to that of the swap curve, but has different inputs. The inputs to an OIS curve are overnight indexed swaps, interest rate swaps in which a fixed rate of interest is exchanged for a floating rate that corresponds to the daily overnight rates.

Currently some banks in South Africa construct OIS curves, generally through an approximation thereof since the market for the relevant instruments are not yet adequately available. More developed markets, such as the United States (US) and European market, do in fact make use of the OIS curve since the relevant instruments are available.

Hull and White (2013) provides an interesting discussion on the appropriate risk-free rate to use, contrasting the US Libor (London interbank offered rate, this term is used interchangeably between the United Kingdom and US) curve with the OIS curve. The Libor curve is similar to the swap curve in South Africa. Therein the authors note that, although there is no “perfect” risk-free rate, the OIS rate is currently the best proxy available. The three-month LIBOR-OIS spread, the spread between three-month Libor and the three-month OIS swap rate, is observed to be on average 10 basis points during normal market conditions. In stressed market conditions, such as the financial crisis of 2008, this spread rose to a high of 364 basis points. This dramatic increase demonstrates that the Libor rates are a poor proxy for the risk-free rate.

Hull and White (2013) note that a potential counterargument may be that Libor rates are often a reflection of the credit risk of the two parties in a derivatives transaction, which implies that it may be the appropriate choice for a risk-free rate proxy. This argument, however, leads to an incorrect calculation of the no-default value of a derivative or derivative portfolio. The credit risk for both parties in a derivative transaction can be taken into account through CVA and DVA. CVA represents the reduction in the value of a derivatives portfolio to allow for a possible default by counterparty and DVA represents the increase in the value of a derivatives portfolio to allow for the possibility of one’s own default. The resulting value of the portfolio is then given by:

$$f = f_{nd} - CVA + DVA,$$

with

f_{nd} = The no-default value of the portfolio.

With the use of Libor rates in discounting it has the effect of incorporating at least some of the credit risk into the discount rate. This then obscures the typical calculation of CVA and DVA, since there may be an element of double counting for credit risk. Furthermore, in the process of calculating CVA or DVA it is necessary to estimate future exposures, which would also require the correct no-default value. Therefore, the OIS rate (or curve) should be viewed as the best proxy of the risk-free rate for the purposes of applying risk-neutral valuation and calculating no-default derivative values.

3.3 CREDIT SPREAD MODELING

3.3.1 General definitions

The credit spread may be the most important measure used by investors for credit security selection. The credit spread provides a relatively intuitive way to evaluate the return compensation that an investor will receive for assuming the credit-related risks associated with the specific credit security. As a simple example, suppose that two fixed rate bonds that have similar maturities and coupon payment dates, both trading at par value, but one has a higher coupon rate than the other does. For these bonds to have an equal value, the one with the higher coupon rate would have a higher discount rate for each future payment and therefore a higher credit spread.

A bond's credit spread is a function of several factors, including default risk and market liquidity risk. A number of definitions of credit spread measures exist, including the G-spread (Government bond curve spread), ASW-spread (asset swap spread) and Z-spread (zero-coupon swap curve spread). The G-spread is the additional spread over the bond curve that would need to be added such that the discounted future cash flows of a particular bond are equal to the current market value of the bond. The ASW-spread is slightly more intricate as it is a package that combines an interest rate swap with a bond that has the effect of changing the basis of the bond. A fixed rate bond is combined with an interest rate swap wherein the bondholder pays a fixed rate and receives a floating rate. The floating coupon on the interest rate swap will be the reference JIBAR rate (Johannesburg Interbank Agreed Rate) plus a spread. The net effect is that the bondholder who previously would have received fixed cash flows, now receives floating cash flows. The ASW-Spread is the resulting spread added on the JIBAR rates in this asset swap which ensures the current market value to the bondholder remains unchanged.

The Z-spread is the additional spread over the interest rate swap curve which would need to be added such that the discounted future cash flows of a particular bond are equal to the current market value of the bond. This is illustrated in equation (3.1) below:

$$PV_{Bond} = C_1 e^{-(r_1+Z)T_1} + C_2 e^{-(r_2+Z)T_2} + \cdots + (C_n + m) e^{-(r_n+Z)T_n}$$

$$PV_{Bond} = \sum_{i=1}^{n} C_i \mathrm{e}^{-(r_i+Z)T_i} + m\, e^{-(r_n+Z)T_n}, \tag{3.1}$$

with

Z = The constant Z-spread,

$r_1, \dots, r_n$ = The zero rates,

C_i = The periodic coupon payment,

T_i = The corresponding time to maturity of payment i,

m = The final repayment of the principal or face value, and

PV_{Bond} = The current market value of the relevant bond.

The Z-spread for shorter dated and higher credit quality bonds are very similar to the ASW-spread. The Z-spread is typically the higher of the two (Choudry, 2005).

With the constant Z-spread calculated, the Zero-Z-spread can be inferred from it. The constant Z-spread and the Zero-Z-spread would be the same when a single bond is considered. A constant Z-spread, in principle, is similar to that of YTM since it is a constant spread added to each future cash flow of a bond. The Zero-Z-spread is, in principle, similar to zero rates in that it is an additional discount spread for a specific point in the future and not applicable to all the future cash flows of a bond. It requires a sample of bonds from which it is derived. The application of Zero-Z-spreads are illustrated in the present value of a bond equation below:

$$PV_{Bond} = ce^{-T_1(r_1+z_1)} + \cdots + ce^{-T_{n-1}(r_{n-1}+z_{n-1})} + (m+c)e^{-T_n(r_n+z_n)},$$

with

z_i = The Zero-Z-spread, corresponding to time T_i.

Therefore, if there is only one coupon remaining on a given bond, the Zero-Z-spread is the same as the constant Z-spread:

$$PV_{Bond} = (m+c)e^{-T_n(r_n+z_n)} = (m+c)e^{-T_n(r_n+Z)}.$$

Implying that: $z_n = Z$.

If, however, there is more than one coupon payment on a bond remaining, each subsequent Zero-Z-spread rate, z_i, can be solved, given the previous rates are known. This implies that each subsequent Zero-Z-spread is solved using the previously solved Zero-Z-spread and by stepping through time. The sample of bonds are ordered according to maturity to facilitate this process. The Zero-Z-spread rates are interpolated to correspond with the relevant maturity dates of each coupon payment, since the chosen sample of bonds would likely have different payment dates. The final Zero-Z-spread rate is solved with the following equation, with the use of bond n:

$$PV_{Bond_n} = \sum_{i=1}^{n-1} ce^{-T_i(r_i+z_{I,i})} + (m+c)e^{-T_n(r_n+z_n)} \tag{3.2}$$

$$=> \qquad PV_{Bond_n} - \sum_{i=1}^{n-1} ce^{-T_i(r_i+z_{I,i})} = (m+c)e^{-T_n(r_n+z_n)}$$

$$=> \qquad \ln\left(PV_{Bond_n} - \sum_{i=1}^{n-1} ce^{-T_i(r_i+z_{I,i})}\right) = -T_n(r_n+z_n)\ln(m+c)$$

$$=> \qquad T_n(r_n+z_n) = -\ln\left(\frac{PV_{Bond_n} - \sum_{i=1}^{n-1} ce^{-T_i(r_i+z_{I,i})}}{m+c}\right)$$

$$=> \qquad z_n = \frac{-1}{T_n}\ln\left(\frac{PV_{Bond_n} - \sum_{i=1}^{n-1} ce^{-T_i(r_i+z_{I,i})}}{m+c}\right) - r_n,$$

where $z_{I,i}$ is the interpolated Zero-Z-spread, corresponding to the relevant point in time to match the respective cash flows.

3.3.2 Modelling Credit Spread Term Structures

The term structure of credit spreads describes the relationship between credit spreads and their maturities. This is similar in nature to that of the interest rate term structure. A properly estimated term structure of both interest rates and credit spreads are essential for obtaining intrinsic values to instruments such as corporate bonds. In this section, several methods for obtaining a term structure of credit spreads are discussed. Applications of term structure models include interest rates, credit spreads as well as PDs. The majority of the literature on term structure models pertain to interest rates, and as such the related formulas and discussions are given in this same context.

3.3.2.1 General notation

The term structure of interest rates can be presented in three theoretically equivalent ways: zero rates, forward rates, and discount rates. This section discussed the relationship between these concepts. All rates are denoted under a continuous compounding convention.

The discount rate is most clearly illustrated in the price of a zero-coupon bond, with a face value of R100, and term to maturity of t years:

$$P(t) = 100\, e^{-r(t)t}$$

$$= 100\, d(t),$$

with $r(t)$ the continuously compounded zero rate corresponding to time t. The discount function $d(t)$ is obtained from $e^{-r(t)t}$. Typically, the discount curve starts at one and decreases with increasing maturities, although it can be larger than 1 when interest rates are negative, as is the case in many European and Asian market at the time of writing.

Zero-coupon rates are spot rates, since these are interest rates applicable to specific points in the future. The spot rate curve, also referred to as the yield curve, would then describe these rates as they relate to various points in the future. Forward rates, however, can be derived from the spot rate curve. The forward rate between the future dates t_1 and t_2 is the annualised interest rate that can be locked in today. The relation between the forward rates and spot rates are given by:

$$e^{t_2 r(t_2)} = e^{t_1 r(t_1)+(t_2-t_1)f(t_1,t_2)},$$

with the rate $f(t_1, t_2)$ defined as the continuously compounded annualized forward rate between future dates t_1 and t_2. The rate $f(t_1, t_2)$ can be calculated as follows:

$$\begin{aligned} f(t_1,t_2) &= \frac{\mathrm{t}_2\mathrm{r}(\mathrm{t}_2) - t_1 r(t_1)}{t_2 - t_1} \\ &= r(t_2) + \frac{\mathrm{r}(\mathrm{t}_2) - r(t_1)}{t_2 - t_1} t_1. \end{aligned}$$

This implies that if the spot rate curve is upward sloping, then the forward rates will be higher than the spot rates. If the spot rate curve is flat, then the forward rates would be identical to the spot rates.

The instantaneous forward rate, $f(t)$, is the annualised forward rate at time t, for an infinitesimally small interval. From this, the annualised spot rate at time t, $r(t)$, can be calculated as the equally weighted average of instantaneous forward rates between time zero and time t, given by:

$$r(t) = \frac{1}{t}\int_0^t f(x)dx.$$

This implies that the spot rate at time t is equal to the average of the instantaneous forward rates from time 0 to time t.

3.3.2.2 Nelson-Siegel

Nelson and Siegel (1987) introduced a simple and parsimonious model with the purpose of being flexible enough to represent a range of shapes generally associated with the yield curve, or equivalently the zero curve. They assumed that the instantaneous forward rate at maturity t, $f(t)$, is given by the solution to a second order partial differential equation with real roots for spot rates.

This is from the notion that if spot rates are generated by a differential equation, then the forward rates should be the solution to it, since the forward rate curve is derived from it.

This approach fits the forward curve of instantaneous forward rates with:

$$f(t) = \beta_0 + \beta_1 \exp\left(-\frac{t}{\tau}\right) + \beta_2 \exp\left(-\frac{t}{\tau}\right), \tag{3.3}$$

where β_0, β_1 and β_2 are determined by initial conditions, and τ a constant associated with the equation, $\tau > 0$. These parameters are discussed in more detail below. With this equation a range of forward rate curves can be generated with different shapes, depending on the values of β_1 and β_2.

Let $r(t)$ denote the zero rates, given by the average of the forward rates:

$$r(t) = \frac{1}{t}\int_0^t f(x)dx$$

$$=> \; r(t) = \frac{1}{t}\int_0^t \left[\beta_0 + \beta_1 \exp\left(-\frac{x}{\tau}\right) + \beta_2 \exp\left(-\frac{x}{\tau}\right)\right] dx.$$

Then:

$$r(t) = \beta_0 + (\beta_1 + \beta_2)\frac{1 - \exp\left(-\frac{t}{\tau}\right)}{\frac{t}{\tau}} - \beta_2 \; \exp\left(-\frac{t}{\tau}\right).$$

Or equivalently:

$$r(t) = \beta_0 + \beta_1 \frac{1 - \exp\left(-\frac{t}{\tau}\right)}{\frac{t}{\tau}} + \beta_2\left(\frac{1 - \exp\left(-\frac{t}{\tau}\right)}{\frac{t}{\tau}} - \exp\left(-\frac{t}{\tau}\right)\right). \tag{3.4}$$

The components of equation (3.4) are:

- β_0 is the long-term component for interest rates,
- β_1 is the short-term component,
- β_2 is the medium-term component, and
- τ is the decay factor.

β_0 is essentially multiplied by one for all maturities and therefore the constant long-term component. The weighting function for β_1 starts at one, if $t = 0$, and decays to zero as t increases. This points to the short-term component, as it diminishes with increasing maturity. The weighting function for β_2 has a humped shape, starting at zero and increasing thereafter, with an eventual decay to zero as maturity increases. This describes what is expected of a medium-term component, since it starts at zero, implying it is not short-term, and end with zero, implying that it is not long-term. These components are illustrated with Figure 3.2 below, with the notation of t being replaced by m.

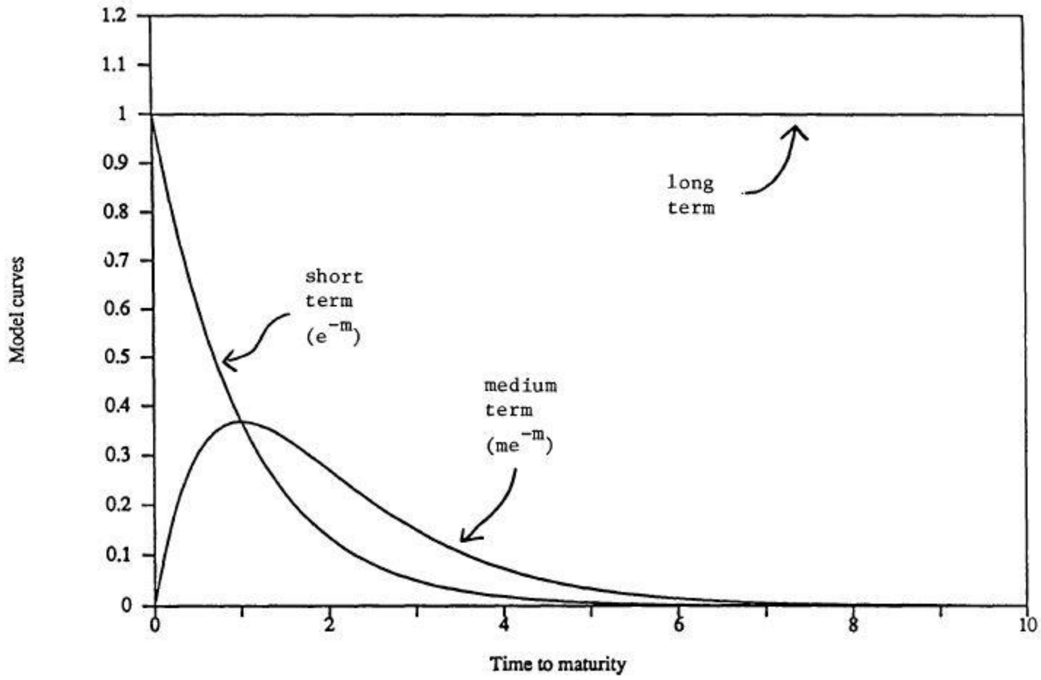

Figure 3.2: Components of the yield curve

Source: Nelson and Siegel, 1987:447

From equation (3.3) and Figure 3.2 it is observed that as the time to maturity increases the zero rates approaches the value β_0. The value of the zero rates at time zero is given by $(\beta_0 + \beta_1)$. The magnitude and sign of β_2 determines the shape of the medium-term component, often referred to as the "hump".

The estimation process for the parameters of the Nelson-Siegel model is complicated by the fact that it is non-linear due to the parameter τ. A possible solution to this is to perform the calibration of the β's over a grid of values for τ with linear least squares to determine the best fit (Nelson and Siegel, 1987). This approach simplifies the calibration process, as the resulting equation is then a linear equation, for each given value of τ. This is applied by selecting a vector of possible values for τ, say $(1, 2, \ldots, 10)$. Then for each value of τ in the vector the corresponding parameters β_0, β_1 and β_2 are estimated using the method of least squares. Finally, the resulting parameters are selected based on the τ-value that produces the most optimal fit.

There are many extensions to the Nelson and Siegel model, including that of Svensson (1994). Svensson proposed an extension with the aim of increasing flexibility and improving the fit. This was done by adding a fourth term for the second "hump", with two additional parameters β_3 and τ_2. The function for the forward curve is then given as:

$$f(t) = \beta_0 + \beta_1 \exp\left(-\frac{t}{\tau_1}\right) + \beta_2 \exp\left(-\frac{t}{\tau_1}\right) + \beta_3 \frac{t}{\tau_2} \exp\left(-\frac{t}{\tau_2}\right).$$

The zero rates are derived in a similar way to equation (3.3) by integrating the forward rate, which results in:

$$r(t) = \beta_0 + \beta_1 \frac{1-\exp\left(-\frac{t}{\tau_1}\right)}{\frac{t}{\tau_1}} + \beta_2 \left(\frac{1-\exp\left(-\frac{t}{\tau_1}\right)}{\frac{t}{\tau_1}} - exp\left(-\frac{t}{\tau_1}\right)\right)$$
$$+ \beta_3 \left(\frac{1-\exp\left(-\frac{t}{\tau_2}\right)}{\frac{t}{\tau_2}} - exp\left(-\frac{t}{\tau_2}\right)\right).$$

3.3.2.3 Gaussian kernel weighting

The methodology for using the Gaussian kernel weighting is based on the manner in which the Reserve Bank of Australia estimated the Australian non-financial corporate (NFC) bond credit spread curve (Arsov at al., 2013). This weighting method only uses market data to determine the spread at specific maturities. This method is in contrast to that of parametric models like the Nelson-Siegel model, since it does not impose a particular functional form on the credit spread curve, but rather allows the observed data to determine the shape (Arsov at al., 2013). Weights are assigned to all the relevant spreads, based on a Gaussian normal distribution. This Gaussian normal distribution has an expected value equal to the specific maturity of the spread in question, implying that market data of spreads close to this point will be more relevant than observations of the spread that are further away.

Following the notation presented in Arsov et al. (2013), let $S(t)$ denote the spread at maturity t, with t as the target maturity. Then the Gaussian kernel spread estimator, $S(t)$, is given by:

$$S(t) = \sum_{i=1}^{n} w_i(t;\sigma)\, S_i. \qquad (3.5)$$

Here $w_i(t,\sigma)$ denotes the weight assigned to the i^{th} bond in the sub-sample and S_i the spread of the corresponding bond. The parameter σ is the standard deviation of the normal distribution and is measured in years. This parameter determines the weights of each observed spread based on the distance between the corresponding bond's remaining time to maturity and time t. Larger values for σ lead to a wider effective window of relevant data, which will produce a smoother fit for the curve. The trade-off, however, is that data which should in fact be less relevant, will play a more significant role in determining the estimated spread at a specific maturity. Alternatively, with sparse data and a small σ value, certain points could effectively determine a large portion of the estimated curve. Therefore, σ is also interpreted as the smoothing parameter for this method.

Very large values of σ results in the estimated spread being equal to a weighted average of all the observed spreads in the sub-sample.

The general form of the weighting function is given by:

$$w_i(t;\sigma) = \frac{K(t_i - t;\sigma)F_i}{\sum_{j=1}^{n} K\left(t_j - t;\sigma\right)F_j}. \tag{3.6}$$

Here, $K(t_i - t;\sigma)$ is Gaussian kernel function. F_i is the face value of the i^{th} bond in the sub-sample, since the assumption is made that bonds with higher face value are priced with a higher degree of accuracy, therefore their corresponding spread should be given a larger weight. This method assigns positive weights to spreads of bonds in the same sub-sample, with greater weight assigned for those observations that are closer to the specific maturity for which an estimate is obtained.

The Gaussian kernel is given by:

$$K(t_i - t;\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(t_i - t)^2}{2\sigma^2}\right]. \tag{3.7}$$

The kernel function is a symmetric, continuous, and bounded real-valued function that integrates to one (Arsov at al., 2013). The Gaussian is the most popular of the many possible candidate kernels, with the particular choice of the kernel having a minor effect on the eventual estimates (Linton et al., 2001).

Finally, determining the optimal smoothing parameter, σ, plays a major role in the eventual credit curve. There exists a natural trade-off between the smoothness of the resulting credit curve and the goodness-of-fit of the estimates, measured as the sum of squared residuals between the observed spreads and the estimated spreads. Low values for σ produce estimates with smaller residuals by assigning higher weights to bonds closest to the target tenor. This would then result in the estimated credit curve reflecting more of the noise in the observations, and furthermore may not be representative of the true credit spread for that specific target tenor. Alternatively, low values for σ may produce estimates with larger residuals.

Although the optimal choice of the smoothing parameter is largely guided by the economic plausibility of the resulting credit spread estimates, it is also possible to determine the optimal value of the smoothing parameter by both minimising the sum of squared residuals of the credit spread estimates and penalising excessive irregularity of the estimated credit spread curve:

$$\min_{\sigma}\left\{(1-\lambda)\sum_{i=1}^{N}[S(t_i,\sigma) - S_i]^2 + \lambda\int\left[\frac{d^2S(u,\sigma)}{du^2}\right]^2 du\right\}. \tag{3.8}$$

In equation (3.8) above the first term measures the goodness-of-fit of the Gaussian kernel estimate, while the second term measures the curvature of the estimated spread curve. The parameter λ, $(0 \leq \lambda \leq 1)$, determines the trade-off between the fit and the curvature terms in the objective function.

3.4 CONCLUSION

This chapter concludes the literature review and introduces the risk-free curves and the practical definitions of the credit curves to be applied in this study. The swap curve, which is the risk-free curve used in the South African market, is discussed as well as what should appropriately be considered when concluding on a curve that is representative of the notion of what is “risk-free”. Finally, the modelling of credit curves is discussed, with the introduction of the Nelson-Siegel and Gaussian kernel models to be applied in subsequent analysis.

In the context of this study, the important practical considerations and definition of terms related to the application of credit spread modelling is discussed. This builds on the theoretical framework of credit spreads discussed in chapter 2 and provides the important practical considerations and definition of terms related to the application of credit spread modelling, as it is applied in chapters 4 and 5.

CHAPTER 4
RESEARCH METHODOLOGY

4.1 INTRODUCTION

In this chapter the data as well as the models applied are discussed in detail. The various stages of data application, training data, validation data, and test data is described, as well as aspects related to data pre-processing and filtering applied to use the raw data. Preliminary results related to the training data are provided, with the rest of the results provided in chapter 5, as well as in the appendix.

4.2 DATA DESCRIPTION

The primary sets of data used in this research are mark-to-market bond data, risk-free rates, and company specific accounting variables together with corresponding share price data. The next section outlines the details of each set of data.

4.2.1 Bond data

Mark-to-market bond data, published by the Bond Exchange of South Africa (BESA), a subsidiary of JSE Limited, is used from which to calculate the credit spreads of ZAR denominated corporate bonds. This dataset covers the relevant market information covering the spectrum of listed bonds and is compiled at the end of each business day. The relevant fields, or bond specific details, used in subsequent calculations are provided in appendix A. The dates at which these datasets are used is discussed in the sections below. The evolution of the bond market over the past several years is illustrated in Figure 4.1. This figure comprises year-end data from 2011 to 2018 and includes all issuers and instrument types.

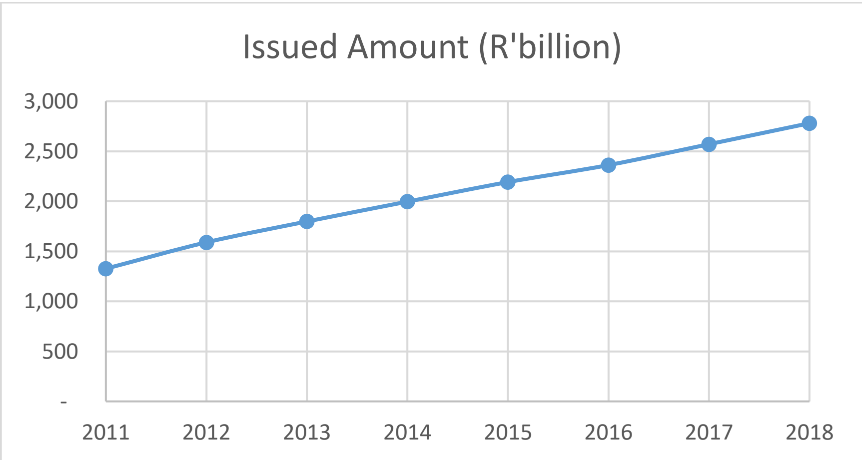

Figure 4.1: Total issued amount for listed bonds in South-Africa

Source: JSE

In Figure 4.1 it is indicated that the total issue amount has substantially increased from 2011 to 2018. It may also be the case that as the listed debt market matures in South Africa, an increasing amount corporates and government institutions issue listed debt on the South African exchange as opposed to other sources of funding or exchanges, which may explain some of the increasing trend in Figure 4.1.

Extensive filtering and pre-processing of the bond data is required in order to be used in the subsequent calculations. In particular, the following filtering rules are applied to arrive at the subset of data used in subsequent calculations:

i.) Exclude all bonds that are not vanilla fixed coupon paying.

ii.) Exclude all bonds with callable features.

iii.) Exclude all bonds that are issued by governments or government related entities.

The first rule is applied on the “Pricing Class Code” variable, which is specified for each bond. This rule results in the exclusion of instrument types such as floating rate notes, inflation-linked bonds, credit linked notes and amortizing instruments. The reasoning for the exclusion is to obtain a dataset that has homogenous bonds, as close as possible, so that differences in spread calculations do not fluctuate due to inherent features such as an amortising notional amount or being linked to the credit quality of an entity other than the issuer. This is also the reasoning behind the exclusion of callable bonds, where the issuer may redeem the bond before it reaches the stated maturity date.

The third rule results in the exclusion of all bonds that are issued by governments or government related entities. Examples of issuers related to this exclusion include: Republic of South Africa (which is the set of government bonds from which the government bond curve is constructed), Transnet, City of Cape Town Metropolitan Municipality, City of Tshwane Municipality and the Trans-Caledon Tunnel Authority. The reasoning for this exclusion is that in practice government bonds are priced from the government bond curve, without a spread. Furthermore, government related issuers may have implicit or explicit guarantees from the government which would affect the spreads as observed from market prices.

The resulting subset of bond data would include vanilla fixed coupon paying corporate bonds, with no callable features and no assumed government support. Further filtering is also applied in an ad hoc basis, where misspecification of certain information, such as instrument type, have been observed.

4.2.2 Risk-free rates

The zero-coupon swap curve, as described in section 3.2, is used as the risk-free rate throughout the subsequent calculations. These swap curves are sourced from BESA, as provided by the JSE Limited. How the swap curves are used is discussed in the sections below.

4.2.3 Share price data and accounting variables

Share price data and corresponding accounting variables related to numerous companies, which are both listed on the JSE and have listed bonds in issue, are sourced from INET BFA. The relevant accounting variables used in subsequent calculations are provided in Table 4.2. How the share price data and accounting variables are used is discussed in the sections below.

4.3 DATA ANALYSIS

The data analysis and modelling consist of three phases: training data, validation data, and test data. The training data phase uses the data to gain information and calibrate certain parameters. Thereafter, the validation phase applies the information learned to fit the various models and tune parameters. Finally, the test data phase is used to evaluate model performance. The purpose of these are to determine whether market observable information can be used to effectively quantify credit related parameters inherent in bond instruments.

4.3.1 Training data

The initial phase in the data analysis is to gain information and calibrate certain parameters. Only market observable information is used. Bond data at the end of each year, from 2011 to 2015, is used along with corresponding accounting variables and share price data. Linear regression is applied to determine which variables best predicts the Z-spreads calculated from the bond data and risk-free rates. The steps and calculations are detailed below.

4.3.1.1 Z-spread calculation

The first step is to calculate the constant Z-spread (referred to as the Z-spread) for each bond in the subsample identified. The Z-spread is the additional spread over the interest rate swap curve that would need to be added such that the discounted future cash flows of a particular bond are equal to the current market value of the bond. This is illustrated in equation (4.1):

$$PV_{Bond} = \sum_{i=1}^{n} C_i \mathrm{e}^{-(r_i+Z)T_i} + m\, e^{-(r_n+Z)T_n}. \tag{4.1}$$

with

Z = The Z-spread,

$r_1, \dots, r_n$ = The zero rates,

C_i = The periodic coupon payment,

T_i = The corresponding time to maturity of payment i,

m = The final repayment of the principal or face value, and

PV_{Bond} = The current market value of the relevant bond.

The PV_{Bond} is obtained from the All in Prices (AIP) provided in the dataset, which is the current market value of the particular bond, on the specified valuation date, which includes any accrued interest. This is also referred to as the dirty value. Conversely, the clean value of the bond does not include any accrued interest. All bonds are priced based on an AIP per R100 nominal.

The following methodology is used to determine the Z-spread for each bond:

i.) Generate a schedule of coupon payment dates.

ii.) Determine the risk-free rate corresponding to each coupon date.

iii.) Discount each coupon and principal back to the valuation date, taking into account a Z-spread variable that is added to each risk-free rate.

iv.) Solve for the specific Z-spread that sets the sum of the discounted future cash flows equal to the current AIP.

This methodology is followed for each respective bond in the subsample. An extract of the results is given in Table 4.1 below.

Table 4.1 Sample of bond information used, and results obtained

Bond Code	**ISIN No**	**Issuer**	**All in Price**	**Issue Date**	**Coupon Rate %**	**Coupon Frequency**	**Maturity Date**	**Z-spread**
ABN12	ZAG000 042128	ABSA BANK LIMITED	117.27	23/07/2007	13.5	2	22/07/2014	0.0215
CBL06	ZAG000 072786	CAPITEC BANK	112.81	02/11/2009	13	2	02/11/2014	0.0236
FRBN05	ZAG000 042169	FIRSTRAND BANK LIMITED	117.82	24/07/2007	13.5	2	23/07/2014	0.0193

Source: JSE

In Table 4.1 bond specific information for three bonds form the subsample, as at 30 December 2011, with the corresponding Z-spreads are given. From this table the relevant AIP, coupon rate, coupon frequency (which are all semi-annually) and maturity date can be observed. These are used, along with the corresponding risk-free rates at each coupon date, to solve the Z-spread. The results indicate that for these bonds with relatively similar time to maturity, the bond issued by FirstRand Bank Limited has the lowest Z-spread.

In Figure 4.2 the solved Z-spreads for the entire subsample of bonds, as at 30 December 2011 are graphed.

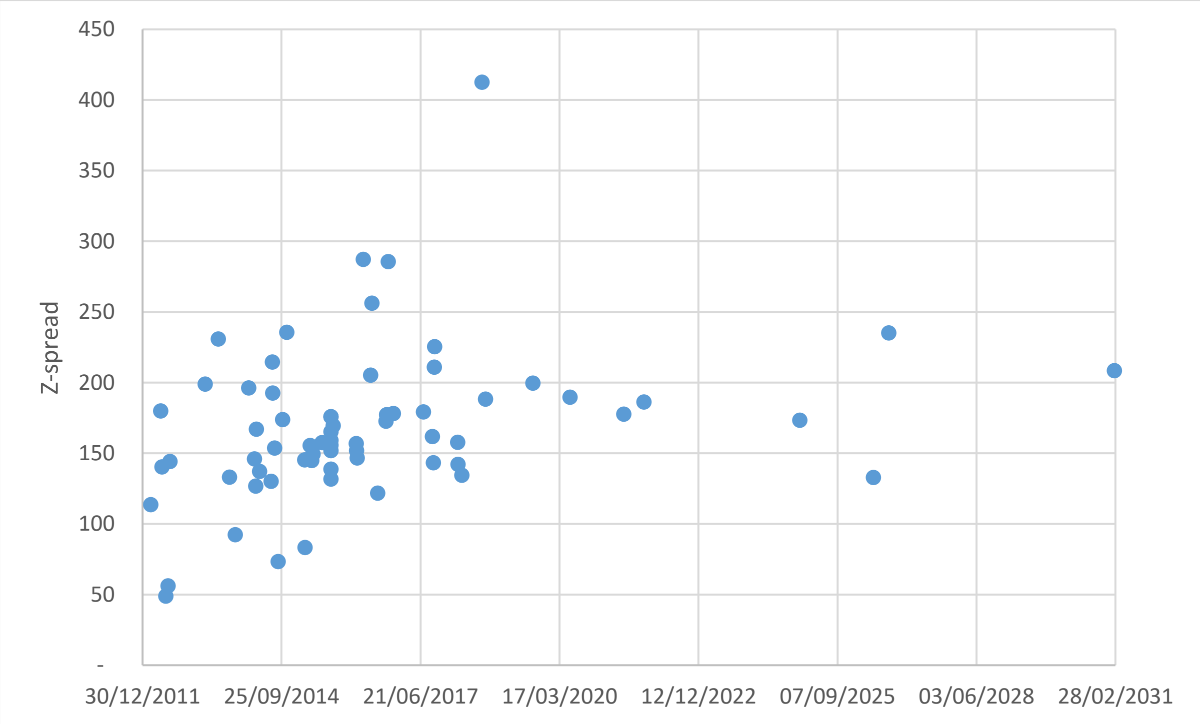

Figure 4.2: Z-spread results as at 30 December 2011

Figure 4.2 has the maturity date on the x-axis and the Z-spread on the y-axis (in basis points). From this graph, although there are a lot of dispersion in the data, an increasing trend in the Z-spreads across maturity dates can be observed which appears to flatten out after approximately 10 years.

The process of solving Z-spreads is done for each respective year end, from 2011 to 2015. Subsamples are determined by only including bonds from companies that have bonds listed in each year as well as share price data and accounting variables available. Some of these issuers have different names across the data, but altogether 14 unique issuers are identified. In total 512 bonds were identified (68 for 2011, 69 for 2012, 106 for 2013, 123 for 2014 and 128 for 2015) which are to be used in subsequent regression analyses.

4.3.1.2 Regression Analysis: Model description

Linear regression is applied to determine what market observable information should be used to predict the Z-spreads calculated from the bond data and risk-free rates. The answer to what is a reliable predictor of spreads is not clear from the literature reviewed. The conclusion reached by both Demirovic at al. (2015) and Das et al. (2008) is that accounting-based and market-based information should be considered as complimentary since models that make use of both sets of information tend to explain a substantially larger portion of credit spreads.

A similar approach to Das et al. (2008) is followed, with the use of accounting variables and market information being considered. Accounting variables for the years 2011 to 2015 of each respective unique bond issuer are used. The market-based variable considered is the annualised share price volatility. In addition to these variables, bond specific information is also used in the regression analysis, in particular, the $\log(Maturity)$ and standardised variables related to the issuer class and guarantee type for each respective bond.

The issuer class variable is identified from a visual inspection as to what bond characteristics may contribute to explaining the Z-spread. Both the guarantee type and issuer class variables were considered. Each bond in the subsample has a corresponding issuer class, segmenting the issuers in various categories and guarantee types. For non-financial (referred to as non-banks) issuers there is no clear distinction to be made by guarantee type from the data, as there appears to be some overlap in the classification used. The relative seniority is not made clear and a thorough inspection of each bond contract is necessary to be certain. This is not in the scope of this study. Whereas the guarantee type varies across time, the issuer class variable is consistent and provides a clear distinction.

To apply these variables in the linear regression they are first standardised, i.e. for the issuer class (IC) variable:

$$\frac{IC_i - \mu_{IC}}{\sigma_{IC}},$$

with:

$IC_i =$ the mean Z-spread per distinct issuer class i,

$\mu_{IC} =$ the mean of the set of IC values, and

$\sigma_{IC} =$ the standard deviation of the set of IC values.

For the guarantee type (GT) variable:

$$\frac{GT_i - \mu_{GT}}{\sigma_{GT}},$$

with:

$GT_i =$ the mean Z-spread per distinct guarantee type i,

$\mu_{GT} =$ the mean of the set of GT values, and

$\sigma_{GT} =$ the standard deviation of the set of GT values.

All the possible explanatory variables are concatenated into a matrix and the corresponding Z-spreads in a vector. Assuming there are n observations and k predictor variables, the linear regression model can be written as:

$$y = X\beta + \epsilon,$$

with:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

the response vector of Z-spreads,

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix},$$

the matrix of the observations of the predictor variables and an intercept,

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix},$$

the vector of regression coefficients, and

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_k \end{bmatrix},$$

the vector of residual terms.

4.3.1.3 Coefficient Determination

This linear model is typically fitted with the use of least squares. In this case, however, alternative fitting procedures are used in an attempt to yield better prediction accuracy and model interpretability. James et al. (2013) notes that if the number of observations per variable is not much larger than the number of variables, then the use of least squares to fit the model can produce substantial variance, resulting in overfitting and consequently poor predictions on future observations.

Constraining or shrinking the estimated coefficients may lead to the reduction in the variance at the cost of an insignificant increase in bias. This may lead to substantial improvements in the accuracy for out of sample predictions. Additionally, model interpretability may improve as it is often the case that some or many of the variables used in a multiple regression model are somewhat unrelated to the response variable. By removing such variables that are potentially irrelevant, it would lead to a model that is easier to interpret. Least squares estimation is unlikely to yield any coefficient estimates that are exactly zero, therefore an alternative method which shrinks the coefficient estimates towards zero is applied. Shrinking the coefficient estimates towards zero can both increase model interpretability and significantly reduce their variance. The two best-known techniques for shrinking the regression coefficients towards zero are ridge regression and the lasso (James et al., 2013).

The lasso technique is applied in the subsequent regression analysis to aid the variable selection. The interested reader is referred to chapter 6 of James et al. (2013) for a thorough discussion on shrinkage methods and in particular the lasso technique. In addition to the lasso technique, judgemental exclusions are applied to eliminate certain variables that appear irrelevant. The principle of parsimony is held to, as a simpler model than the model with 22 explanatory variables, as determined by Das et al. (2008), is preferred.

4.3.1.4 Results

Following the variable selection process described above, linear regression is applied to determine the resulting regression coefficients. A description of all the variables derived is given in Table 4.2.

Table 4.2 Predictor variables descriptions

Variable Name	**Description**
Debt / Equity	Debt over equity, also referred to as the leverage ratio
Current Ratio	Current assets divided by current liabilities, and is popular metric used to assess a company's short-term liquidity
Price / EBITDA	Share price over earnings before interest, tax, depreciation and amortisation
Volatility	Annualised share price volatility
IClass	Standardised variable related to the issuer class for each respective bond
GType	Standardised variable related to the guarantee type for each respective bond
Log (Time)	Natural logarithm for the time to maturity from the specified valuation date, in years

Two different models are fitted on the entire sample of Z-spreads. The results are given in Tables 4.3 and 4.4.

Table 4.3 Regression output: fit 1

Observations	512			
F-statistic (7,504)	58.95			
R^2	0.4502			
Adjusted R^2	0.4425			
Coefficient	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	-0.13500	0.0316	-4.2720	< 0.001
Current Ratio	-0.00639	0.0010	-6.6450	< 0.001
Debt / Equity	-0.00059	0.0001	-6.5860	< 0.001
Volatility	0.04662	0.0048	9.7150	< 0.001
Price / EBITDA	0.00004	0.0000	2.5380	0.0115
IClass	0.00351	0.0005	7.5260	< 0.001
GType	0.00825	0.0006	14.8880	< 0.001
Log (Time)	0.01559	0.0032	4.8400	< 0.001

Table 4.4 Regression output: fit 2

Observations	512			
F-statistic (6,505)	66.98			
R^2	0.4431			
Adjusted R^2	0.4365			
Coefficient	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	-0.13460	0.0318	-4.2380	< 0.001
Current Ratio	-0.00592	0.0009	-6.2390	< 0.001
Debt / Equity	-0.00064	0.0001	-7.3650	< 0.001
Volatility	0.04550	0.0048	9.4700	< 0.001
IClass	0.00363	0.0005	7.7840	< 0.001
GType	0.00820	0.0006	14.7330	< 0.001
Log (Time)	0.01555	0.0032	4.8010	< 0.001

The resulting R^2 and $Adjusted\ R^2$ is not much different between the two models, where the second model, fit 2, has the Price / EBITDA variable removed. This is done to investigate the change in $Adjusted\ R^2$ for a slightly more parsimonious model through the removal of the predictor variable with the lowest degree of significance from fit 1.

It is observed that in the South-African market, banks typically have the lowest credit spreads on their bonds in issue, but do not necessarily have lower leverage ratios compared to other corporate bond issuers. This has the effect that the Debt / Equity variable's coefficient is negatively related to credit spreads. An intuitive understanding of credit risk would lead to the belief that a lower proportion of debt that a company has would lead to an increased likelihood of their ability to repay their debt, which is not the case here. This may be due to specific leverage ratio's that South African banks operate at or similar regulatory requirements. Therefore, in line with Demirovic at al. (2015), the sample is divided into two sections: banks and non-banks. Similar regression analysis is applied to these subsets, with the results given in Tables 4.5 and 4.6.

The results from fit 3 and 4 indicate an improved overall fit based on the increased R^2 and $Adjusted\ R^2$ values. In particular the R^2 value related to the regression applied on non-banks increased to 0.5615. The results appear to be in line with those obtained in Demirovic at al. (2015), with none of their models producing an R^2 value above 0.63. The main discernible difference between the fit on banks and non-banks appear to be the preference for near-term earnings liquidity to be more relevant to non-banks than banks in determining credit spreads. This is seen by the inclusion of the Current Ratio and Price / EBITDA variables being included, whereas with banks the Debt / Equity is included. Including the Debt / Equity appears to obscure the results if all the issuers are considered in a single set, whereas the sub-setting applied results in increased higher explanatory power.

Table 4.5 Regression output: fit 3 - Banks

Observations	432			
F-statistic (7,426)	74.74			
R²	0.4673			
Adjusted R²	0.4611			
Coefficient	**Estimate**	**Std. Error**	**t value**	**p-value**
(Intercept)	0.01701	0.0026	6.6700	< 0.001
Debt / Equity	-0.00105	0.0001	-7.8960	< 0.001
Volatility	0.04521	0.0051	8.7960	< 0.001
IClass	0.00629	0.0009	7.4000	< 0.001
GType	0.00753	0.0006	12.1660	< 0.001
Log (Time)	0.00070	0.0003	2.1540	0.0318

Table 4.6 Regression output: fit 4 - Non-banks

Observations	80			
F-statistic (5,74)	18.95			
R^2	0.5615			
Adjusted R^2	0.5319			
Coefficient	**Estimate**	**Std. Error**	**t value**	**p-value**
(Intercept)	-0.00025	0.0029	-0.0850	0.93232
Current Ratio	-0.00383	0.0014	-2.8350	0.00591
Volatility	0.03131	0.0070	4.4780	< 0.001
Price / EBITDA	-0.00028	0.0001	-2.0240	0.04654
IClass	0.00242	0.0005	4.8640	< 0.001
Log (Time)	0.00210	0.0004	5.0750	< 0.001

4.3.2 Validation Data

Typically research that investigates the links between accounting data, market data, and credit spreads conclude with the results of the regression analysis, as is the case with Das et al. (2008) and Demirovic et al. (2015). For this study, the regression is used as an input to the subsequent bond pricing. From literature reviewed, it is not clear whether there exists a particular market standard method to construct credit curves for the purpose of pricing newly issued bonds based on accounting and market related data. It is therefore noted that there are several elements of judgement applied in an attempt to arrive at a model that is sensible, as outlined in section 4.3.2.2.

4.3.2.1 Data Period

The validation data includes mark-to-market bond data from 1 January 2016 to 31 December 2016. The filtering applied is similar to what is described in the bond data section above. In addition to this filtering, exclusions of Z-spread outliers are made. These exclusions are in accordance with Xiao (2010) and removes outliers where the calculated Z-spread is not between zero and six percent.

4.3.2.2 Model application

The curve estimation and pricing steps are as follows:

i.) Identify the newly issued bonds that fit the same criteria as identified in section 4.3.1.1.

ii.) Import datasets corresponding to these new issues, including risk-free rates and accounting variables and market data (share price volatility). This would include the full set of listed bonds on each date that the newly issued bonds are investigated.

iii.) Filter data to get only vanilla bonds that are listed and has no callable features.

iv.) Remove the newly issued bonds from this data set, at each date.

v.) Determine the predicted Z-spread for each bond in issue at the respective dates, with the use of the output from the regression models.

vi.) Create a subset of bonds, at each respective new issue date, by excluding bonds where the predicted Z-spread without the $log\ (Time)$ component is outside the first or third quantile (less than 25% or more than 75%). The $log\ (Time)$ component is removed in an attempt to get the middle section of bonds across time and exclude outliers. Not removing the $log\ (Time)$ component may weigh the resulting quantiles towards those in the sub-sample that have longer times to maturity.

vii.) Solve the market implied Z-spreads for each of the bonds in the subset determined above.

viii.) Perform bootstrapping on the Z-spreads of the subset of bonds determined above to get the Zero-Z-spreads.

ix.) Once these are determined the various curves are fitted to the Zero-Z-spreads: Gaussian kernel, logarithmic and Nelson-Siegel.

Firstly, the Gaussian kernel, as outlined in chapter 3, is applied. Three different values for the σ parameter is tested, being 1, 2, and 5. The σ parameter is given as an annualised value. Secondly, a logarithmic curve is fitted with the use of least squares. Finally, the Nelson-Siegel, as outlined in chapter 3, is fitted. Therefore, the three models range from non-parametric with the Gaussian Kernel, to fully parametric both simple and relatively complex in the logarithmic and Nelson-Siegel models respectively.

x.) Shift each of the estimated curves, at each new issue date with a quantile methodology, based on the following binning probabilities:

   a. $Q_1 = (0.2, 0.4, 0.6, 0.8)$,
   b. $Q_2 = (0.1, 0.25, 0.75, 0.9)$.

The above vectors, Q_1 and Q_2, give the cut points for the the bins in which each of the bonds in the subset are divided into. This is done by first ordering all the predicted Z-spread without the $log\ (Time)$ components in ascending order, then dividing these into the bins as specified by Q_1 and Q_2 on a percentage level. All the scores in each bin are then averaged to get an average bin score. For example, with Q_1, the first bin would be the average of the lowest 20% of scores in the subsample.

The curve adjustment factors are then calculated as the difference between the average bin scores and the middle bin scores. Therefore, the resulting shifts would be increasing and negative for the first two, zero for the middle curve, and positive for the remaining two. The same method is followed with Q_2. The reason for the difference between Q_1 and Q_2 is to investigate the possible differences in results from this methodology, as well as to investigate the notion that bonds which correspond with the lowest or highest curve would potentially require a greater adjustment.

xi.) Determine five distinct curves based on the adjustments described above for each respective date of new issue and each curve estimation method.

xii.) Determine which curve the newly issued bond corresponds to, based on its predicted Z-spread without the $log\ (Time)$ component score. Price each of the new issues based on these curves determined above with a discounted cash flow method to get the model implied price. The formula in accordance with equation (4.1) is applied.

xiii.) Evaluate the results by considering the percentage difference ("$\%D$") and the $PV0x$.

The formula for the $\%D$ is given below, being the difference between the bond price as implied from the model and from that observed in the market, divided by the price observed in the market:

$$\%D = \frac{PV_{Bond:\,model\ implied} - PV_{Bond:\,market\ observed}}{PV_{Bond:\,market\ observed}}. \qquad (4.2)$$

The $PV0x$ value is an adaptation from the commonly known $PV01$ metric used in interest rate sensitivity calculations. The $PV01$ metric measures the difference in present value given a one basis point parallel shift in the discount curve applied. The formula to calculate the $PV01$ value is:

$$PV01 = PV_{Bond:\,no\ shift} - PV_{Bond:\,1\ Bps\ shift},$$

where

$$PV_{Bond:\,no\ shift} = PV_{Bond:\,model\ implied}$$

and, with similar notation used in equation (3.2):

$$PV_{Bond:\,1\ Bps\ shift} = \sum_{i=1}^{n-1} ce^{-T_i\left(r_i + z_{I,i} + 1\right)} + (m + c)e^{-T_n\left(r_n + z_n + 1\right)}.$$

The approximation for calculating the $PV0x$ is then given by:

$$PV0x \approx \frac{PV_{Bond:\,model\ implied} - PV_{Bond:\,market\ observed}}{PV01}. \qquad (4.3)$$

The $PV0x$ value gives the approximate amount of basis points (Bps) the curve needs to be shifted for the bond price as implied from the model to be equal to the bond price observed in the market. Therefore, from a bond pricing perspective, the $\%D$ may be considered the more informative metric, whereas from a curve estimation perspective it may be the $PV0x$.

xiv.) Finally, a further sensitivity is applied on the curves estimated above. An increased parallel basis point shift between the five curves described in point xi. This is done in order to investigate the use of increased discrimination between the credit quality of bond issuers, as well as to capture what was excluded due to omitting the $log\ (Time)$ components from the adjustment factors. From the initial results, without any additional shifts, it appears that additional shifts may result in increased accuracy, particularly for the non-bank related bond issuers. The shifts are applied across the five curves, with the number of basis points in each case given below:

a. $Shift\ 1 = (-20, -10, 0, 10, 20)$ – Adjustment of 10 Bps,
b. $Shift\ 2 = (-40, -20, 0, 20, 40)$ – Adjustment of 20 Bps,
c. $Shift\ 3 = (-60, -30, 0, 30, 60)$ – Adjustment of 30 Bps.

4.3.3 Test Data

The test data covers the period 1 January 2017 to 30 June 2018 and makes use of the same model application described for the validation data section above, as well as the same filtering and outlier exclusions. The results are provided and discussed in the next chapter.

4.4 CONCLUSION

In this chapter the data and models applied are discussed in detail. Firstly, the filtering and pre-processing of the raw bond data is discussed in order to arrive at the relevant sample. The training data stage identifies various market and accounting variables that can be used to predict the credit spreads related to corporate bonds. It is observed that sub-setting the data to distinguish between bonds issued by banks and those issued by non-banks add to the explanatory power of the prediction variables.

Subsequent to the training data step the validation data methodology is detailed. The models described in chapter 3 are used to construct credit curves to fit the observable Zero-Z-spreads and used in the pricing of newly issued bonds. The measures used to describe the results are also detailed and given in the next chapter.

CHAPTER 5
FINDINGS

5.1 INTRODUCTION

This chapter covers the validation data and test data results. The distinction is made in order to investigate the out-of-sample performance of the estimation techniques, as well as the fine tuning of the adjustments applied on the validation data.

5.2 NOTATION OF RESULTS

It is useful to indicate the notation of the various results obtained, following from the model application outlined in the previous chapter. This includes the 5 curves (point ix), 2 quantiles (point x) and 4 shifts (point xiv). In total each new bond issue would thus have 40 unique results for both the $\%D$ and $PV0x$ metric, with their related equations given by:

$$\%D = \frac{PV_{Bond:\,model\ implied} - PV_{Bond:\,market\ observed}}{PV_{Bond:\,market\ observed}},$$

and

$$PV0x \approx \frac{PV_{Bond:\,model\ implied} - PV_{Bond:\,market\ observed}}{PV01}.$$

The five models applied to estimate the credit curves are the Gaussian Kernel with the σ parameter being 1, 2, and 5, the logarithmic curve and the Nelson-Siegel curve. In the subsequent results this would be given in this order as mentioned, numbered one to five. The adjustments related to the basis point shifts, from point xiv, are stated clearly, along with the two different quantile methodologies applied.

5.3 VALIDATION DATA RESULTS

5.3.1 Results related to banks

Table 5.1 Validation data results: banks

		Q1					Q2				
	Model	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
No Adjustment	**Mean %D**	-0.55%	-0.49%	-0.44%	-0.72%	-0.12%	-0.54%	-0.48%	-0.43%	-0.70%	-0.11%
	Mean (|%D|)	1.27%	1.20%	1.17%	1.30%	1.20%	1.22%	1.15%	1.15%	1.25%	1.19%
	Mean PV0x	-16	-15	-16	-20	-4	-16	-15	-16	-20	-3
	Mean (|PV0x|)	34	32	33	35	32	32	31	33	33	31
Adjustment = 10 Bps	**Mean %D**	-0.76%	-0.70%	-0.65%	-0.92%	-0.33%	-0.64%	-0.59%	-0.53%	-0.81%	-0.21%
	Mean (|%D|)	1.45%	1.42%	1.37%	1.51%	1.30%	1.30%	1.26%	1.23%	1.36%	1.24%
	Mean PV0x	-21	-20	-21	-25	-8	-18	-18	-19	-23	-6
	Mean (|PV0x|)	37	37	38	39	34	34	34	35	36	33
Adjustment = 20 Bps	**Mean %D**	-0.96%	-0.91%	-0.86%	-1.13%	-0.53%	-0.74%	-0.69%	-0.64%	-0.91%	-0.31%
	Mean (|%D|)	1.73%	1.70%	1.67%	1.75%	1.60%	1.43%	1.39%	1.37%	1.49%	1.37%
	Mean PV0x	-26	-25	-26	-30	-13	-21	-20	-22	-25	-9
	Mean (|PV0x|)	43	43	44	44	40	37	36	37	38	35
Adjustment = 30 Bps	**Mean %D**	-1.16%	-1.10%	-1.06%	-1.32%	-0.73%	-0.84%	-0.79%	-0.74%	-1.01%	-0.41%
	Mean (|%D|)	2.06%	2.03%	2.03%	2.13%	1.96%	1.56%	1.53%	1.56%	1.72%	1.52%
	Mean PV0x	-31	-30	-31	-35	-18	-24	-23	-24	-28	-12
	Mean (|PV0x|)	50	49	51	52	47	40	39	42	44	38

5.3.2 Discussion of results

In Table 5.1 it is observed that the difference between the various credit curve models are minor in the case of applying no basis point adjustments to the curves. The $mean(|PV0x|)$ values range between 32 and 35 with the use of Q1 and 31 and 33 with Q2. A similar narrow range of differences is observed for the $mean(|\%D|)$ values. There also does not appear to be any significant change between the use of Q1 or Q2. Finally, the application of basis point adjustments of certain curves do not lead to more accurate results. In fact, the results indicate a worse fit across all adjustments compared to applying no adjustment.

5.3.3 Results related to non-banks

Table 5.2 Validation data results: non-banks

		Q 1					Q 2				
	Model	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
No Adjustment	**Mean %D**	1.64%	1.82%	2.54%	1.49%	2.31%	1.65%	1.83%	2.54%	1.50%	2.32%
	Mean (|%D|)	3.11%	3.36%	3.33%	3.55%	3.37%	3.05%	3.30%	3.28%	3.48%	3.34%
	Mean PV0x	41	45	61	38	56	41	46	61	38	56
	Mean (|PV0x|)	72	77	77	81	77	71	76	76	80	77
Adjustment = 10 Bps	**Mean %D**	2.00%	2.18%	2.90%	1.85%	2.67%	1.89%	2.07%	2.79%	1.74%	2.56%
	Mean (|%D|)	2.49%	2.86%	3.18%	3.09%	3.15%	2.48%	2.95%	3.02%	3.17%	3.08%
	Mean PV0x	49	54	69	46	64	47	51	67	44	62
	Mean (|PV0x|)	59	67	75	72	74	59	69	71	74	72
Adjustment = 20 Bps	**Mean %D**	2.37%	2.54%	3.27%	2.21%	3.03%	2.14%	2.32%	3.04%	1.99%	2.81%
	Mean (|%D|)	2.37%	2.67%	3.27%	2.72%	3.03%	2.30%	2.69%	3.04%	2.91%	2.90%
	Mean PV0x	57	61	77	54	72	52	57	72	49	67
	Mean (|PV0x|)	57	64	77	65	72	56	64	72	70	69
Adjustment = 30 Bps	**Mean %D**	2.74%	2.92%	3.64%	2.58%	3.41%	2.40%	2.58%	3.30%	2.25%	3.07%
	Mean (|%D|)	2.74%	2.92%	3.64%	2.76%	3.41%	2.56%	2.88%	3.30%	3.13%	3.16%
	Mean PV0x	65	69	85	62	80	58	62	77	55	73
	Mean (|PV0x|)	65	69	85	66	80	61	68	77	74	75

5.3.4 Discussion of results

In Table 5.2 it is observed that an increased disparity exists between models and adjustments applied for the results related to non-banks compared to the results related to banks. This is likely due to the departure from homogeneity of the issuers. In particular, the results with the use of the Gaussian Kernel, where $\sigma = 1$ and Q1 is used, appears to provide the closest fit.

The 10 Bps and 20 Bps adjustments applied to the curves does appear to improve the fit and overall the Gaussian Kernel, where $\sigma = 1$ or $\sigma = 2$, and logarithmic curve outperform the Gaussian Kernel where $\sigma = 5$ and Nelson-Siegel.

It is noted that there is a single bond, issued by Northam Platinum Limited, which is the main source of difference. The results for this specific bond are given in appendix A, having the bond reference of NB1. None of the models, including the adjustments appear to be able to price the bond back to the market price given the bond specific information. The coupon rate for this bond is 13.5%, which is towards the upper range of the underlying sample. This implies that the observable data would underestimate the credit spread required, to price back to the market observable price at issue date.

5.4 TEST DATA RESULTS

5.4.1 Results related to banks

Table 5.3 Test data results: banks

		Q 1					Q 2				
	Model	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
No Adjustment	**Mean %D**	-1.13%	-1.08%	-0.98%	-1.27%	-0.43%	-1.22%	-1.17%	-1.07%	-1.36%	-0.52%
	Mean (|%D|)	1.64%	1.67%	1.64%	1.98%	1.52%	1.68%	1.71%	1.73%	2.02%	1.61%
	Mean PV0x	-32	-31	-30	-33	-16	-34	-33	-32	-36	-18
	Mean (|PV0x|)	45	46	47	52	44	46	47	49	53	46
Adjustment = 10 Bps	**Mean %D**	-1.99%	-1.94%	-1.84%	-2.12%	-1.29%	-1.91%	-1.86%	-1.76%	-2.05%	-1.22%
	Mean (|%D|)	2.29%	2.27%	2.23%	2.54%	2.06%	2.19%	2.17%	2.12%	2.43%	1.96%
	Mean PV0x	-50	-50	-48	-52	-34	-50	-49	-48	-51	-33
	Mean (|PV0x|)	58	59	59	63	54	57	57	57	61	53
Adjustment = 20 Bps	**Mean %D**	-2.83%	-2.78%	-2.68%	-2.96%	-2.14%	-2.60%	-2.55%	-2.45%	-2.73%	-1.91%
	Mean (|%D|)	3.02%	3.00%	2.95%	3.24%	2.67%	2.76%	2.74%	2.69%	2.98%	2.37%
	Mean PV0x	-69	-69	-67	-71	-53	-65	-65	-63	-67	-49
	Mean (|PV0x|)	74	75	74	79	67	70	70	70	74	61
Adjustment = 30 Bps	**Mean %D**	-3.66%	-3.61%	-3.52%	-3.78%	-2.98%	-3.27%	-3.23%	-3.13%	-3.41%	-2.59%
	Mean (|%D|)	3.73%	3.71%	3.67%	3.95%	3.27%	3.32%	3.30%	3.25%	3.54%	2.82%
	Mean PV0x	-89	-88	-87	-91	-72	-81	-81	-79	-83	-65
	Mean (|PV0x|)	91	91	91	95	80	83	83	83	87	71

5.4.2 Discussion of results

The results are similar to the validation data set, where the use of Q1 and Q2 do not influence the results substantially and the basis point adjustments to the curves do not improve the overall results. Additionally, where no basis point adjustment is applied, the difference between the various credit curve models are minor.

5.4.3 Results related to non-banks

Table 5.4 Test data results: non-banks

		Q 1					Q 2				
	Model	1	2	3	4	5	1	2	3	4	5
No Adjustment	Mean %D	-0.93%	-0.83%	-0.50%	-1.37%	3.80%	-0.91%	-0.81%	-0.47%	-1.34%	3.83%
	Mean (|%D|)	0.93%	0.83%	0.50%	1.37%	4.32%	0.91%	0.81%	0.52%	1.34%	4.37%
	Mean PV0x	-22	-24	-18	-30	102	-21	-23	-17	-30	102
	Mean (|PV0x|)	22	24	18	30	112	21	23	18	30	113
Adjustment = 10 Bps	Mean %D	-0.96%	-0.86%	-0.52%	-1.39%	3.74%	-1.03%	-0.93%	-0.59%	-1.46%	3.66%
	Mean (|%D|)	0.98%	0.86%	0.77%	1.43%	3.85%	1.07%	0.98%	0.86%	1.51%	4.00%
	Mean PV0x	-20	-22	-16	-29	104	-21	-23	-17	-30	103
	Mean (|PV0x|)	23	22	21	33	107	25	24	22	35	110
Adjustment = 20 Bps	Mean %D	-0.98%	-0.87%	-0.53%	-1.40%	3.68%	-1.15%	-1.05%	-0.71%	-1.58%	3.50%
	Mean (|%D|)	1.28%	1.31%	1.40%	1.62%	3.68%	1.40%	1.33%	1.20%	1.85%	3.64%
	Mean PV0x	-18	-20	-14	-27	107	-21	-23	-17	-30	104
	Mean (|PV0x|)	33	29	31	41	107	35	30	27	46	106
Adjustment = 30 Bps	Mean %D	-0.99%	-0.88%	-0.54%	-1.41%	3.63%	-1.27%	-1.16%	-0.83%	-1.69%	3.35%
	Mean (|%D|)	1.99%	2.02%	2.10%	2.24%	3.63%	1.80%	1.73%	1.81%	2.25%	3.35%
	Mean PV0x	-17	-18	-13	-25	109	-22	-23	-18	-30	104
	Mean (|PV0x|)	51	48	47	58	109	48	42	42	58	104

5.4.4 Discussion of results

In Table 5.3 an improved fit is observed when the Gaussian Kernel is used, for all values of σ, compared to the logarithmic and Nelson-Siegel models. With the validation data an improved fit is observed when the 10 Bps and 20 Bps adjustments are applied, whereas in this case applying no adjustment or an adjustment of 10 Bps appear to improve the fit.

5.5 CONCLUSION

This chapter provides the results from the analysis detailed in chapter 4, as it relates to the validation data and test data. Detailed results for each specific bond considered in the abovementioned analysis is provided in appendix A. The results related to banks indicate that there is no significant difference between using different models and parameters for the estimation of credit curves. Shifting the curves also does not lead to improved results. These results remained relatively consistent between the validation data and test data.

For non-banks, where data is comparatively sparser, the non-parametric model, being the Gaussian Kernel, performed relatively better compared to the parametric model, especially the Nelson-Siegel model. A shift of 10 Bps also appears to improve the results.

CHAPTER 6
SUMMARY, CONCLUSION AND RECOMMENDATIONS

6.1 INTRODUCTION

This chapters provides a brief overview of the discussions of and the observations made from the various models and data applied throughout this study.

Initially the difference between what is considered risk-neutral and real-world estimates of default probabilities is discussed, as well as the importance of obtaining risk-neutral default probabilities for the purpose of fair value estimation. The different sources of data to obtain these estimates, in particular CDS and bond spreads, are discussed. The limitations of the South African market, as it pertains to the CDS data, lead to the conclusion that an alternative data source should be used, such as credit spreads on corporate bonds. These spreads, being the Z-spread, is defined as well as various sources of market and accounting information that could serve to quantify the related spread. Subsequently, two models are introduced, the Nelson-Siegel and Gaussian kernel models, that along with the logarithmic curve are used to fit a credit curve to the observed spreads. These, in conjunction with appropriate risk-free curves, are used to price back newly issued bonds in an attempt to verify the accuracy of the curves.

6.2 SUMMARY OF MAIN FINDINGS

The training data used led to the observation that sub-setting the data to distinguish between banks and non-banks increases the explanatory power of the market and accounting data as it relates to observable corporate bond spreads. Although the resulting R^2 and $adjusted\ R^2$ value are not particularly high, it is in line with related research, as set out in chapter 2. This serves to further underline the subjectivity of what is being done: attempting to accurately use market observable variables in explaining corporate bond credit spreads.

Subsequently, credit curves are constructed to fit market observable credit spreads, and in conjunction with the abovementioned analysis, newly issued bonds are priced. The accuracy of the models and corresponding analysis are tested by investigating the accuracy compared to the market prices. For the subset containing bonds issued by listed banks in South Africa, the results indicate that there is no significant difference between using different models and parameters for the estimation of credit curves. This may be due to various factors, including the relative homogeneity between South African banks and their listed corporate bonds.

For non-banks subset, where data is comparatively sparser, the non-parametric Gaussian kernel model performed relatively better compared to the parametric models, especially compared to the Nelson-Siegel model. A shift of 10 Bps also led to improved results.

There are several occasions where outliers in the results significantly influence the overall results, such as the case of the Northam Platinum Limited bond (bond reference of NB1 in appendix A). Care should be taken when evaluating these results, as outliers may be prevalent when only market observable information is considered in a market with limited information.

In general, market data related to banks are much more available compared to non-banks, especially in the South African market. It is therefore noted that the use of proxy data may in many cases be unnecessary, and bank specific credit curves would probably be better suited to price newly issued debt of the corresponding bank. For non-bank issuers of debt this is not the case, and often proxy data is the only reasonable source of information. In this regard, the curves applied produced reasonable results, except for the Nelson-Siegel curve. Where data is sparse or when several outliers obscure the data to which the Nelson-Siegel model is fitted, often the resulting shape produced is not sensible. Therefore, an element of judgement would have to be applied when the Nelson-Siegel model is used in this context.

6.3 RECOMMENDATIONS AND FURTHER RESEARCH

It is not clear what is an acceptable level of accuracy when considering the construction of credit curves and the pricing of newly issued bonds. The analysis presented above is, from the authors perspective, unique.

There are several additional considerations that may be applied that relates to the models and data used in this study as well as questions for further research. A few of these are given below:

i.) What level of accuracy is acceptable when considering the construction of credit curves as it relates to the pricing of newly issued bonds, particularly as it pertains to the $\%D$ and $PV0x$ measures?

ii.) What other measures should be considered informative for the evaluation of credit curves and the pricing of newly issued bonds, other than the $\%D$ and $PV0x$ measures?

iii.) What additional information should be considered in the analysis applied in section 4.3.1.3 to produce an improved fit?

iv.) Would additional sub-setting of the non-banks subset, such as distinguishing between mining companies and property companies, be informative and useful?

v.) Would the use of alternative curve fitting, such as cubic splines, improve the results?

Finally, it needs to be noted that the estimation of credit curves with the use of market data may require an element of judgement in order to be relevant and applicable to the South African context, as opposed to merely using a standardised model to be applied across all potential issuers of bonds.

REFERENCES

Altman, E. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), September: 589-609.

Amato, J. D. and Remolona, E.M. 2003. The Credit Spread Puzzle. *BIS Quarterly Review*, 5, December: 51-63.

Arsov, I., Brooks, M. and Kosev, M. 2013. New Measures of Australian Corporate Credit Spreads. *Reserve Bank Bulletin*, December.

Bühler, W. and Trapp, M. 2009. Explaining the Bond-CDS Basis – The Role of Credit Risk and Liquidity. *CFR Working paper*, No. 09-12.

Choudry, M. 2005. *Understanding the Z-spread*. [Online]. Available: http://www.yieldcurve.com/mktresearch/learningcurve/choudhry_zspread_aug2005_logo.pdf . [2019, October 11].

Das, S.R., Hanouna, P. and Sarin, A. 2009. Accounting-based versus market-based cross-sectional models of CDS spreads. *Journal of Banking and Finance*, 33 (4): 719-730.

Demirovic, A., Tucker, J. and Guermat, C. 2015. Accounting data and the credit spread: An empirical investigation. *Centre for Global Finance, Faculty of Business & Law, University of the West of England*. Coldharbour Lane, Bristol BS16 1QY.

De Laurentis, G., Maino, R. and Molteni, L. 2010. *Developing, validating, and using internal ratings*. John Wiley & Sons. West Sussex, United Kingdom.

De Wit, J. 2006. Exploring the CDS-Bond Basis. *NBB Working Paper*, No 104 - November.

EBA, Technical standards in relation with credit valuation adjustment. 2013. [Online] Available: https://www.eba.europa.eu/regulation-and-policy/credit-risk/draft-regulatory-technical-standards-in-relation-to-creditvaluation-adjustment-risk [2018, December 30].

Engle, R. and Granger, C. 1987. Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica*, 55.

Giesecke, K., Longstaff, F., Schaefer, S. and Strebulaev, I. 2010. Corporate Bond Default Risk: A 150-Year Perspective, *NBER Working Paper*, 15848.

Gregory, J. 2012. *Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets*. Croydon: John Wiley & Sons Ltd.

Gyntelberg, J., Hördahl, P., Ters, K. and Urban, J. 2017. Arbitrage costs and the persistent non-zero CDS-bond basis: Evidence from intraday euro area sovereign debt markets. *BIS Working Papers*, No 631. April.

Hagan, P.S. and West, G. 2006. Interpolation Methods for Curve Construction. *Applied Mathematical Finance,*13(2), June: 89-129.

Hagan, P.S. and West, G. 2008. Methods for constructing a yield curve. *Wilmott magazine*, May: 70-81.

Hastie, T., Tibshirani, R., Witten, D. and James, G. 2013. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics 103, New York.

Houweling, P., Mentink, A. and Vorst, T. 2005. Comparing possible proxies of corporate bond liquidity. *Journal of Banking and Finance,* 29: 1331-1358.

Ho, T. and Lee, S. 1986. Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, 41(5): 1011-1029.

Hull, J.C. 2012. *Options, Futures, and other Derivatives*. Pearson/Prentice Hall, New Jersey.

Hull, J. and White, A. 1990. Pricing interest rate derivative securities. *The Review of Financial Studies,* 3(4): 573-592.

Hull, J. and White, A. 2013. LIBOR vs. OIS: The Derivatives Discounting Dilemma. *Journal of Investment Management,* 11 (3): 14-27.

Hull, J., Predescu, M. and White, A. 2005. Bond Prices, Default Probabilities and Risk Premiums. *Journal of credit risk*, 1(2 Springs): 53-60.

IFRS 13: Fair Value Measurement. 2017. [Online] Available: https://www.ifrs.org/issued-standards/list-of-standards/ifrs-13-fair-value-measurement/ [2018, December 30].

JSE. 2012. The JSE Zero-Coupon Yield Curves: Methodology Document. September.

Kim, G., Li, H. and Zhang, W. 2016. The CDS-Bond Basis Arbitrage and the Cross Section of Corporate Bond Returns. *Journal of Futures Markets*, December 2016.

Linton, O., Mammen, E., Nielsen, J. and Tanggard, C. 2001. Yield Curve Estimation by Kernel Smoothing Methods. *Journal of Econometrics*, 105: 185-223.

Longstaff, F., Mithal, S. and Neis, E. 2005. Corporate yield spreads: default risk or liquidity? New evidence from the credit-default swap market. *Journal of Finance*, 60: 2213-2253.

Malz, A. 2011. *Financial Risk Management: Models, History and Institutions.* John Wiley & Sons, Hoboken, New Jersey.

Merton, R.C. 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance,* 29*,* May:449-470.

Nelson, C. R. and Siegel, A.F. 1987. Parsimonious Modelling of Yield Curves. *The Journal of Business*, 60(4), October:473-489.

Sourabh, S., Hofer, M. and Kandhai, D. 2018. Liquidity risk in derivatives valuation: an improved credit proxy method. *Quantitative Finance*, 18:3: 467-481

Svensson, L.E.O. 1992. Estimating and interpreting forward interest rates: Sweden 1992-1994. *NBER Working Paper*, 4871.

Xiao, J. 2010. Term Structure Estimation for U.S. Corporate Bond Yields. *RiskMetrics Journal*, 2(1): 19-34.

Zhu, H. 2004. An Empirical Comparison of Credit Spreads Between the Bond Market and the Credit Default Swap Market. *BIS Working Paper,* 160, Proceedings of the European Financial Management Association Basel Meeting, 13.

APPENDIX A:

A.1 BOND DETAILS FOR NEW ISSUES IDENTIFIED

A.1.1 Validation data

A.1.1.1 Banks

Table A 1 Validation data bond details: banks

Bond Ref	Bond Code	Issuer	Issue Date	Maturity Date	Coupon Rate %
B1	IBL79	INVESTEC BANK LIMITED	27/01/2016	27/01/2019	10.00%
B2	NBK25A	NEDBANK LIMITED	18/02/2016	17/02/2023	10.66%
B3	CBL26	CAPITEC BANK	06/05/2016	06/05/2021	11.11%
B4	NBK26A	NEDBANK LIMITED	10/05/2016	10/05/2023	10.68%
B5	NBK27A	NEDBANK LIMITED	10/05/2016	10/05/2026	11.15%
B6	IBL87	INVESTEC BANK LIMITED	20/05/2016	20/05/2019	9.75%
B7	IBL88	INVESTEC BANK LIMITED	20/05/2016	20/05/2021	10.32%
B8	SSN038	THE STANDARD BANK OF SOUTH AFRICA LIMITED	26/07/2016	26/01/2021	9.55%
B9	NBK28A	NEDBANK LIMITED	02/08/2016	02/08/2023	10.01%
B10	NBK29A	NEDBANK LIMITED	02/08/2016	31/07/2026	10.50%
B11	ABS18	ABSA BANK LIMITED	26/09/2016	26/09/2023	9.84%
B12	SSN041	THE STANDARD BANK OF SOUTH AFRICA LIMITED	13/10/2016	13/10/2023	9.80%
B13	IBL99	INVESTEC BANK LIMITED	21/10/2016	21/10/2019	9.00%
B14	IVC094	INVESTEC BANK LIMITED	06/12/2016	06/12/2021	10.80%
B15	IVC098	INVESTEC BANK LIMITED	14/12/2016	14/06/2018	8.97%

A.1.1.2 Non-Banks

Table A 2 Validation data bond details: non-banks

Bond Ref	Bond Code	Issuer	Issue Date	Maturity Date	Coupon Rate %
NB1	NHM002	NORTHAM PLATINUM LIMITED	13/05/2016	12/05/2021	13.50%
NB2	LGL06	THE LIBERTY GROUP LIMITED	04/10/2016	04/10/2022	10.20%
NB3	GRT17	GROWTHPOINT PROPERTIES LIMITED	17/10/2016	17/10/2023	10.15%
NB4	KAP007	KAP INDUSTRIAL HOLDINGS LIMITED	26/10/2016	26/10/2021	10.23%

A.1.2 Test data

A.1.2.1 Banks

Table A 3 Test data bond details: banks

Bond Ref	Bond Code	Issuer	Issue Date	Maturity Date	Coupon Rate %
B16	SBS50	THE STANDARD BANK OF SOUTH AFRICA LIMITED	31/01/2017	31/01/2022	9.46%
B17	SBS51	THE STANDARD BANK OF SOUTH AFRICA LIMITED	31/01/2017	31/01/2024	9.78%
B18	NBK30A	NEDBANK LIMITED	20/02/2017	20/02/2024	9.60%
B19	FRX27	FIRSTRAND BANK LIMITED	07/03/2017	07/03/2027	10.19%
B20	FRX32	FIRSTRAND BANK LIMITED	07/03/2017	31/03/2032	10.52%
B21	IVC110	INVESTEC BANK LIMITED	19/05/2017	19/05/2021	9.77%
B22	IBL101	INVESTEC BANK LIMITED	24/05/2017	24/05/2022	9.08%
B23	SBS56	THE STANDARD BANK OF SOUTH AFRICA LIMITED	12/06/2017	12/06/2022	8.95%
B24	SSN053	THE STANDARD BANK OF SOUTH AFRICA LIMITED	26/10/2017	26/10/2020	8.38%
B25	SSN059	THE STANDARD BANK OF SOUTH AFRICA LIMITED	02/03/2018	05/10/2022	8.40%
B26	FRC272	FIRSTRAND BANK LIMITED	24/05/2018	30/01/2023	9.58%
B27	FRC274	FIRSTRAND BANK LIMITED	05/06/2018	02/05/2028	9.80%
B28	IVC133	INVESTEC BANK LIMITED	06/06/2018	06/12/2019	8.02%

A.1.2.2 Non-Banks

Table A 4 Test data bond details: non-banks

Bond Ref	Bond Code	Issuer	Issue Date	Maturity Date	Coupon Rate %
NB5	TL24	TELKOM SA LIMITED	04/09/2017	05/09/2022	9.04%
NB6	TL25	TELKOM SA LIMITED	04/09/2017	04/09/2024	9.57%
NB7	CGR34	CALGRO M3 DEVELOPMENTS LIMITED	22/09/2017	21/09/2018	8.39%
NB8	DSY03	DISCOVERY LIMITED	21/11/2017	21/11/2024	10.46%
NB9	TL28	TELKOM SA LIMITED	24/04/2018	24/04/2025	9.28%

A.2 INDIVIDUAL BOND RESULTS

A.2.1 No Adjustment – Q1

Bond Ref	%D					PV0x				
	Models									
	1	2	3	4	5	1	2	3	4	5
B1	-0.73%	-0.90%	-1.24%	-1.15%	-0.64%	-28	-34	-47	-44	-24
B2	-0.39%	-0.17%	0.21%	-0.19%	0.85%	-8	-3	4	-4	17
B3	2.72%	2.65%	2.24%	1.95%	2.57%	67	65	55	48	63
B4	-0.46%	-0.74%	-0.72%	-1.26%	-0.57%	-9	-15	-14	-25	-11
B5	-1.13%	-0.59%	0.23%	-0.52%	0.59%	-19	-10	4	-9	9
B6	-1.28%	-0.91%	-1.17%	-1.11%	-0.98%	-49	-34	-44	-42	-37
B7	-0.49%	-0.64%	-0.97%	-1.26%	-1.02%	-12	-16	-24	-32	-26
B8	-1.34%	-1.36%	-1.53%	-1.98%	-1.93%	-36	-37	-42	-54	-53
B9	-0.68%	-0.64%	-0.36%	-1.09%	-0.58%	-13	-13	-7	-21	-11
B10	-1.27%	-0.83%	0.22%	-0.60%	0.44%	-20	-13	3	-9	7
B11	-1.28%	-1.33%	-1.27%	-1.27%	-1.15%	-25	-26	-25	-25	-22
B12	-2.32%	-2.38%	-2.26%	-2.61%	-1.97%	-46	-47	-45	-52	-39
B13	-1.65%	-1.59%	-1.81%	-1.75%	-1.07%	-63	-60	-69	-66	-40
B14	2.68%	2.66%	2.54%	2.46%	3.09%	65	65	62	60	75
B15	-0.61%	-0.65%	-0.78%	-0.36%	0.57%	-43	-46	-55	-25	40
B16	-1.14%	-1.21%	-1.36%	-1.36%	-0.54%	-28	-30	-34	-34	-13
B17	-1.92%	-1.71%	-1.49%	-1.98%	-0.97%	-38	-34	-29	-39	-19
B18	-2.22%	-2.06%	-1.71%	-2.12%	-0.62%	-44	-40	-34	-42	-12
B19	-1.88%	-1.52%	-0.94%	-2.15%	-0.88%	-30	-24	-15	-35	-14
B20	-0.35%	-0.34%	-0.03%	-1.99%	0.08%	-5	-5	-0	-28	1
B21	0.15%	0.20%	0.25%	0.33%	0.48%	4	6	7	10	14
B22	-2.38%	-2.22%	-2.20%	-2.24%	-1.50%	-59	-55	-55	-56	-37
B23	-2.00%	-1.92%	-2.02%	-2.39%	-2.13%	-50	-47	-50	-60	-53
B24	-3.49%	-3.69%	-3.88%	-3.38%	-2.54%	-134	-142	-150	-130	-96
B25	0.36%	0.64%	0.77%	0.88%	2.63%	9	17	20	23	68
B26	2.76%	2.95%	3.29%	3.36%	3.91%	72	76	85	87	100
B27	-1.40%	-1.87%	-2.05%	-2.12%	-1.98%	-22	-30	-33	-34	-32
B28	-1.21%	-1.35%	-1.39%	-1.38%	-1.55%	-86	-95	-98	-98	-110
NB1	9.50%	10.37%	10.98%	10.07%	11.17%	227	245	258	239	262
NB2	-0.27%	-0.56%	0.41%	-1.72%	-0.14%	-6	-12	9	-38	-3
NB3	-1.98%	-2.27%	-1.59%	-2.23%	-1.97%	-39	-45	-31	-45	-39
NB4	-0.69%	-0.25%	0.35%	-0.16%	0.20%	-17	-6	9	-4	5
NB5	-1.35%	-1.22%	-0.95%	-1.86%	7.49%	-33	-30	-23	-46	168
NB6	-0.64%	-0.58%	-0.18%	-1.88%	10.89%	-12	-11	-3	-37	184
NB7	-0.13%	-0.35%	-0.43%	-0.09%	1.75%	-13	-37	-44	-10	179
NB8	-1.71%	-1.47%	-0.93%	-2.20%	-1.30%	-34	-29	-19	-44	-26
NB9	-0.84%	-0.54%	0.01%	-0.80%	0.18%	-16	-10	0	-15	3

A.2.2 No Adjustment – Q2

Bond Ref	%D					PV0x				
	Models									
	1	2	3	4	5	1	2	3	4	5
B1	-0.65%	-0.82%	-1.16%	-1.08%	-0.56%	-25	-31	-44	-41	-21
B2	-0.38%	-0.16%	0.22%	-0.18%	0.86%	-8	-3	4	-4	17
B3	2.42%	2.36%	1.95%	1.66%	2.28%	60	58	48	41	56
B4	-0.40%	-0.67%	-0.66%	-1.20%	-0.50%	-8	-13	-13	-24	-10
B5	-1.05%	-0.50%	0.31%	-0.44%	0.67%	-17	-8	5	-7	11
B6	-1.28%	-0.91%	-1.17%	-1.11%	-0.98%	-49	-34	-44	-42	-37
B7	-0.49%	-0.64%	-0.97%	-1.26%	-1.02%	-12	-16	-24	-32	-26
B8	-1.23%	-1.25%	-1.43%	-1.88%	-1.83%	-33	-34	-39	-51	-50
B9	-0.76%	-0.72%	-0.44%	-1.17%	-0.66%	-15	-14	-9	-23	-13
B10	-1.09%	-0.65%	0.40%	-0.42%	0.62%	-18	-10	6	-7	10
B11	-1.38%	-1.44%	-1.37%	-1.37%	-1.26%	-27	-28	-27	-27	-25
B12	-2.11%	-2.17%	-2.05%	-2.40%	-1.76%	-42	-43	-41	-48	-35
B13	-1.73%	-1.66%	-1.88%	-1.82%	-1.14%	-66	-63	-71	-69	-43
B14	2.68%	2.66%	2.54%	2.46%	3.09%	65	65	62	60	75
B15	-0.61%	-0.65%	-0.78%	-0.36%	0.57%	-43	-46	-55	-25	40
B16	-1.41%	-1.48%	-1.63%	-1.64%	-0.82%	-35	-37	-41	-41	-20
B17	-2.27%	-2.05%	-1.84%	-2.32%	-1.31%	-45	-41	-36	-46	-26
B18	-2.56%	-2.40%	-2.06%	-2.46%	-0.96%	-51	-47	-40	-49	-19
B19	-1.60%	-1.25%	-0.67%	-1.88%	-0.60%	-25	-20	-10	-30	-9
B20	-0.03%	-0.02%	0.30%	-1.68%	0.40%	-0	-0	4	-23	5
B21	0.26%	0.31%	0.36%	0.44%	0.59%	8	9	11	13	17
B22	-2.25%	-2.09%	-2.07%	-2.11%	-1.37%	-56	-52	-51	-53	-34
B23	-2.00%	-1.91%	-2.01%	-2.39%	-2.13%	-50	-47	-50	-59	-53
B24	-3.62%	-3.82%	-4.02%	-3.52%	-2.67%	-139	-147	-155	-135	-102
B25	0.13%	0.41%	0.55%	0.65%	2.40%	3	11	15	17	62
B26	2.57%	2.76%	3.09%	3.17%	3.72%	67	72	80	82	95
B27	-1.86%	-2.32%	-2.50%	-2.56%	-2.43%	-29	-37	-40	-41	-39
B28	-1.21%	-1.35%	-1.39%	-1.38%	-1.55%	-86	-95	-98	-98	-110
NB1	9.39%	10.26%	10.87%	9.96%	11.05%	225	243	256	237	259
NB2	-0.31%	-0.60%	0.37%	-1.75%	-0.18%	-7	-13	8	-39	-4
NB3	-1.87%	-2.16%	-1.48%	-2.12%	-1.86%	-37	-43	-29	-42	-37
NB4	-0.62%	-0.18%	0.41%	-0.10%	0.26%	-16	-5	10	-3	6
NB5	-1.34%	-1.20%	-0.93%	-1.84%	7.51%	-33	-30	-23	-46	168
NB6	-0.62%	-0.56%	-0.16%	-1.86%	10.91%	-12	-11	-3	-36	184
NB7	-0.11%	-0.33%	-0.41%	-0.07%	1.77%	-11	-35	-42	-8	181
NB8	-1.76%	-1.52%	-0.98%	-2.25%	-1.35%	-35	-30	-20	-46	-27
NB9	-0.71%	-0.41%	0.14%	-0.67%	0.31%	-14	-8	3	-13	6

A.2.3 10 Bps Adjustment – Q1

Bond Ref	%D					PV0x				
	Models									
	1	2	3	4	5	1	2	3	4	5
B1	-1.25%	-1.43%	-1.76%	-1.68%	-1.17%	-48	-55	-68	-64	-45
B2	0.62%	0.84%	1.23%	0.82%	1.88%	12	16	24	16	36
B3	1.91%	1.84%	1.43%	1.15%	1.76%	47	46	36	29	44
B4	-0.97%	-1.24%	-1.23%	-1.76%	-1.07%	-19	-25	-25	-36	-21
B5	-1.74%	-1.20%	-0.39%	-1.13%	-0.03%	-29	-20	-6	-19	-1
B6	-1.28%	-0.91%	-1.17%	-1.11%	-0.98%	-49	-34	-44	-42	-37
B7	-0.49%	-0.64%	-0.97%	-1.26%	-1.02%	-12	-16	-24	-32	-26
B8	-2.07%	-2.09%	-2.27%	-2.71%	-2.66%	-57	-57	-62	-75	-73
B9	-0.17%	-0.13%	0.15%	-0.58%	-0.07%	-3	-3	3	-11	-1
B10	-2.50%	-2.07%	-1.05%	-1.85%	-0.83%	-41	-34	-17	-30	-13
B11	-0.76%	-0.82%	-0.76%	-0.75%	-0.64%	-15	-16	-15	-15	-12
B12	-3.33%	-3.38%	-3.27%	-3.61%	-2.97%	-67	-68	-66	-73	-60
B13	-1.39%	-1.33%	-1.54%	-1.48%	-0.81%	-53	-50	-58	-56	-30
B14	2.68%	2.66%	2.54%	2.46%	3.09%	65	65	62	60	75
B15	-0.61%	-0.65%	-0.78%	-0.36%	0.57%	-43	-46	-55	-25	40
B16	-1.94%	-2.01%	-2.16%	-2.16%	-1.35%	-49	-50	-54	-54	-34
B17	-2.93%	-2.72%	-2.51%	-2.98%	-1.98%	-59	-54	-50	-60	-39
B18	-3.22%	-3.06%	-2.73%	-3.12%	-1.64%	-64	-61	-54	-62	-32
B19	-3.12%	-2.78%	-2.21%	-3.39%	-2.14%	-51	-45	-35	-55	-34
B20	-1.81%	-1.80%	-1.49%	-3.40%	-1.39%	-25	-25	-21	-49	-19
B21	-0.53%	-0.48%	-0.43%	-0.35%	-0.20%	-16	-14	-13	-10	-6
B22	-3.18%	-3.01%	-3.00%	-3.04%	-2.31%	-80	-76	-75	-77	-58
B23	-2.80%	-2.72%	-2.82%	-3.19%	-2.93%	-70	-68	-71	-80	-74
B24	-4.01%	-4.20%	-4.40%	-3.90%	-3.06%	-155	-163	-171	-151	-117
B25	-0.39%	-0.11%	0.02%	0.12%	1.86%	-11	-3	1	3	49
B26	1.99%	2.18%	2.51%	2.59%	3.13%	52	57	65	67	81
B27	-2.66%	-3.12%	-3.29%	-3.36%	-3.22%	-43	-51	-53	-55	-52
B28	-1.21%	-1.35%	-1.39%	-1.38%	-1.55%	-86	-95	-98	-98	-110
NB1	8.67%	9.53%	10.13%	9.23%	10.32%	209	227	240	221	244
NB2	0.19%	-0.10%	0.87%	-1.27%	0.32%	4	-2	19	-28	7
NB3	-0.97%	-1.27%	-0.57%	-1.22%	-0.96%	-19	-25	-11	-24	-19
NB4	0.11%	0.56%	1.16%	0.64%	1.01%	3	14	29	16	25
NB5	-2.16%	-2.03%	-1.76%	-2.66%	6.60%	-54	-50	-44	-67	149
NB6	-1.67%	-1.61%	-1.22%	-2.90%	9.71%	-33	-31	-24	-57	166
NB7	0.06%	-0.16%	-0.23%	0.10%	1.95%	7	-17	-24	10	198
NB8	-0.71%	-0.46%	0.08%	-1.21%	-0.29%	-14	-9	2	-24	-6
NB9	-0.32%	-0.01%	0.53%	-0.27%	0.71%	-6	0	10	-5	13

A.2.4 10 Bps Adjustment – Q2

Bond Ref	%D					PV0x				
	Models									
	1	2	3	4	5	1	2	3	4	5
B1	-0.91%	-1.09%	-1.42%	-1.34%	-0.83%	-35	-41	-54	-51	-31
B2	0.63%	0.85%	1.24%	0.83%	1.89%	12	17	24	16	37
B3	1.62%	1.55%	1.14%	0.86%	1.47%	40	39	29	22	37
B4	-0.40%	-0.67%	-0.66%	-1.20%	-0.50%	-8	-13	-13	-24	-10
B5	-1.05%	-0.50%	0.31%	-0.44%	0.67%	-17	-8	5	-7	11
B6	-1.28%	-0.91%	-1.17%	-1.11%	-0.98%	-49	-34	-44	-42	-37
B7	-0.49%	-0.64%	-0.97%	-1.26%	-1.02%	-12	-16	-24	-32	-26
B8	-1.60%	-1.62%	-1.79%	-2.25%	-2.20%	-43	-44	-49	-61	-60
B9	-0.76%	-0.72%	-0.44%	-1.17%	-0.66%	-15	-14	-9	-23	-13
B10	-1.71%	-1.28%	-0.24%	-1.05%	-0.02%	-28	-21	-4	-17	-0
B11	-1.38%	-1.44%	-1.37%	-1.37%	-1.26%	-27	-28	-27	-27	-25
B12	-2.62%	-2.67%	-2.56%	-2.90%	-2.26%	-52	-53	-51	-58	-45
B13	-1.73%	-1.66%	-1.88%	-1.82%	-1.14%	-66	-63	-71	-69	-43
B14	2.68%	2.66%	2.54%	2.46%	3.09%	65	65	62	60	75
B15	-0.61%	-0.65%	-0.78%	-0.36%	0.57%	-43	-46	-55	-25	40
B16	-2.21%	-2.28%	-2.43%	-2.43%	-1.62%	-56	-57	-61	-61	-41
B17	-3.27%	-3.06%	-2.85%	-3.32%	-2.33%	-66	-61	-57	-67	-46
B18	-3.56%	-3.40%	-3.07%	-3.46%	-1.98%	-72	-68	-61	-69	-39
B19	-2.23%	-1.88%	-1.30%	-2.51%	-1.24%	-36	-30	-21	-40	-20
B20	-0.77%	-0.76%	-0.44%	-2.39%	-0.34%	-11	-10	-6	-34	-5
B21	-0.08%	-0.03%	0.02%	0.10%	0.25%	-2	-1	1	3	7
B22	-2.65%	-2.49%	-2.47%	-2.51%	-1.78%	-66	-62	-62	-63	-44
B23	-2.80%	-2.72%	-2.82%	-3.18%	-2.93%	-70	-68	-70	-80	-73
B24	-4.14%	-4.34%	-4.53%	-4.04%	-3.20%	-160	-168	-176	-156	-122
B25	-0.62%	-0.34%	-0.20%	-0.10%	1.63%	-17	-9	-5	-3	43
B26	1.80%	1.99%	2.32%	2.40%	2.94%	47	52	61	62	76
B27	-3.10%	-3.56%	-3.73%	-3.80%	-3.66%	-50	-58	-61	-62	-60
B28	-1.21%	-1.35%	-1.39%	-1.38%	-1.55%	-86	-95	-98	-98	-110
NB1	8.56%	9.42%	10.02%	9.12%	10.20%	207	225	238	219	242
NB2	-0.31%	-0.60%	0.37%	-1.75%	-0.18%	-7	-13	8	-39	-4
NB3	-0.86%	-1.15%	-0.46%	-1.11%	-0.85%	-17	-23	-9	-22	-17
NB4	0.18%	0.62%	1.23%	0.71%	1.07%	4	15	30	17	26
NB5	-2.14%	-2.01%	-1.75%	-2.64%	6.62%	-53	-50	-43	-66	149
NB6	-1.65%	-1.59%	-1.20%	-2.88%	9.74%	-32	-31	-23	-57	167
NB7	0.08%	-0.14%	-0.21%	0.12%	1.97%	9	-15	-22	12	200
NB8	-1.26%	-1.02%	-0.48%	-1.76%	-0.85%	-25	-20	-9	-35	-17
NB9	-0.19%	0.12%	0.67%	-0.14%	0.84%	-4	2	13	-3	16

A.2.5 20 Bps Adjustment – Q1

Bond Ref	%D					PV0x				
	Models									
	1	2	3	4	5	1	2	3	4	5
B1	-1.77%	-1.95%	-2.28%	-2.20%	-1.69%	-68	-75	-88	-85	-65
B2	1.65%	1.87%	2.27%	1.85%	2.92%	32	36	44	36	56
B3	1.10%	1.04%	0.63%	0.35%	0.96%	28	26	16	9	24
B4	-1.47%	-1.74%	-1.72%	-2.25%	-1.57%	-29	-35	-35	-46	-32
B5	-2.33%	-1.80%	-1.00%	-1.74%	-0.65%	-39	-30	-16	-29	-11
B6	-1.28%	-0.91%	-1.17%	-1.11%	-0.98%	-49	-34	-44	-42	-37
B7	-0.49%	-0.64%	-0.97%	-1.26%	-1.02%	-12	-16	-24	-32	-26
B8	-2.80%	-2.82%	-2.99%	-3.44%	-3.39%	-77	-78	-83	-95	-94
B9	0.35%	0.39%	0.67%	-0.06%	0.45%	7	7	13	-1	9
B10	-3.72%	-3.30%	-2.29%	-3.08%	-2.09%	-62	-54	-37	-50	-34
B11	-0.25%	-0.30%	-0.24%	-0.24%	-0.12%	-5	-6	-5	-5	-2
B12	-4.32%	-4.37%	-4.26%	-4.60%	-3.97%	-88	-89	-86	-94	-80
B13	-1.13%	-1.06%	-1.28%	-1.22%	-0.54%	-42	-40	-48	-46	-20
B14	2.68%	2.66%	2.54%	2.46%	3.09%	65	65	62	60	75
B15	-0.61%	-0.65%	-0.78%	-0.36%	0.57%	-43	-46	-55	-25	40
B16	-2.74%	-2.80%	-2.95%	-2.95%	-2.15%	-69	-71	-75	-75	-54
B17	-3.93%	-3.72%	-3.51%	-3.98%	-2.99%	-79	-75	-71	-81	-60
B18	-4.22%	-4.06%	-3.73%	-4.12%	-2.65%	-85	-82	-75	-83	-53
B19	-4.34%	-4.01%	-3.45%	-4.61%	-3.38%	-72	-66	-56	-76	-55
B20	-3.24%	-3.23%	-2.93%	-4.78%	-2.82%	-46	-46	-41	-70	-40
B21	-1.20%	-1.15%	-1.10%	-1.02%	-0.87%	-36	-34	-33	-30	-26
B22	-3.97%	-3.81%	-3.79%	-3.83%	-3.10%	-101	-97	-96	-97	-78
B23	-3.60%	-3.51%	-3.61%	-3.98%	-3.73%	-91	-89	-91	-101	-94
B24	-4.52%	-4.72%	-4.91%	-4.42%	-3.58%	-176	-184	-192	-171	-138
B25	-1.13%	-0.86%	-0.73%	-0.62%	1.10%	-31	-23	-20	-17	29
B26	1.23%	1.42%	1.75%	1.82%	2.36%	33	37	46	48	62
B27	-3.89%	-4.35%	-4.51%	-4.58%	-4.45%	-64	-72	-74	-76	-73
B28	-1.21%	-1.35%	-1.39%	-1.38%	-1.55%	-86	-95	-98	-98	-110
NB1	7.84%	8.69%	9.29%	8.40%	9.47%	191	209	222	203	226
NB2	0.65%	0.36%	1.33%	-0.81%	0.78%	14	8	29	-18	17
NB3	0.05%	-0.24%	0.46%	-0.20%	0.06%	1	-5	9	-4	1
NB4	0.92%	1.37%	1.98%	1.46%	1.82%	23	34	48	36	44
NB5	-2.96%	-2.83%	-2.57%	-3.45%	5.72%	-74	-71	-64	-87	130
NB6	-2.69%	-2.63%	-2.24%	-3.90%	8.55%	-53	-52	-44	-78	148
NB7	0.26%	0.03%	-0.04%	0.29%	2.15%	27	3	-4	30	218
NB8	0.31%	0.56%	1.11%	-0.20%	0.73%	6	11	21	-4	14
NB9	0.20%	0.51%	1.07%	0.25%	1.24%	4	10	20	5	23

A.2.6 20 Bps Adjustment – Q2

Bond Ref	%D					PV0x				
	Models									
	1	2	3	4	5	1	2	3	4	5
B1	-1.17%	-1.35%	-1.68%	-1.60%	-1.09%	-45	-52	-65	-61	-42
B2	1.66%	1.88%	2.28%	1.86%	2.93%	32	36	44	36	56
B3	0.81%	0.75%	0.35%	0.07%	0.67%	20	19	9	2	17
B4	-0.40%	-0.67%	-0.66%	-1.20%	-0.50%	-8	-13	-13	-24	-10
B5	-1.05%	-0.50%	0.31%	-0.44%	0.67%	-17	-8	5	-7	11
B6	-1.28%	-0.91%	-1.17%	-1.11%	-0.98%	-49	-34	-44	-42	-37
B7	-0.49%	-0.64%	-0.97%	-1.26%	-1.02%	-12	-16	-24	-32	-26
B8	-1.97%	-1.99%	-2.16%	-2.61%	-2.56%	-54	-54	-59	-72	-70
B9	-0.76%	-0.72%	-0.44%	-1.17%	-0.66%	-15	-14	-9	-23	-13
B10	-2.33%	-1.90%	-0.87%	-1.67%	-0.66%	-38	-31	-14	-27	-10
B11	-1.38%	-1.44%	-1.37%	-1.37%	-1.26%	-27	-28	-27	-27	-25
B12	-3.12%	-3.18%	-3.06%	-3.40%	-2.76%	-62	-64	-61	-68	-55
B13	-1.73%	-1.66%	-1.88%	-1.82%	-1.14%	-66	-63	-71	-69	-43
B14	2.68%	2.66%	2.54%	2.46%	3.09%	65	65	62	60	75
B15	-0.61%	-0.65%	-0.78%	-0.36%	0.57%	-43	-46	-55	-25	40
B16	-3.00%	-3.07%	-3.22%	-3.22%	-2.42%	-76	-78	-82	-82	-61
B17	-4.26%	-4.05%	-3.85%	-4.31%	-3.33%	-87	-82	-78	-88	-67
B18	-4.55%	-4.40%	-4.07%	-4.45%	-2.99%	-93	-89	-82	-90	-60
B19	-2.85%	-2.51%	-1.94%	-3.12%	-1.87%	-46	-40	-31	-51	-30
B20	-1.50%	-1.49%	-1.18%	-3.10%	-1.07%	-21	-21	-16	-44	-15
B21	-0.42%	-0.37%	-0.32%	-0.24%	-0.09%	-12	-11	-9	-7	-3
B22	-3.05%	-2.89%	-2.87%	-2.91%	-2.18%	-77	-73	-72	-73	-54
B23	-3.60%	-3.51%	-3.61%	-3.98%	-3.72%	-91	-89	-91	-101	-94
B24	-4.66%	-4.85%	-5.05%	-4.55%	-3.72%	-181	-189	-197	-177	-143
B25	-1.36%	-1.08%	-0.95%	-0.84%	0.87%	-37	-29	-26	-23	23
B26	1.04%	1.23%	1.56%	1.63%	2.17%	28	33	41	43	57
B27	-4.33%	-4.78%	-4.95%	-5.01%	-4.88%	-71	-79	-82	-83	-81
B28	-1.21%	-1.35%	-1.39%	-1.38%	-1.55%	-86	-95	-98	-98	-110
NB1	7.74%	8.58%	9.18%	8.29%	9.36%	188	207	220	201	224
NB2	-0.31%	-0.60%	0.37%	-1.75%	-0.18%	-7	-13	8	-39	-4
NB3	0.17%	-0.13%	0.57%	-0.09%	0.18%	3	-3	11	-2	3
NB4	0.98%	1.43%	2.04%	1.52%	1.89%	24	35	50	37	46
NB5	-2.94%	-2.81%	-2.55%	-3.44%	5.74%	-74	-71	-64	-87	131
NB6	-2.67%	-2.61%	-2.22%	-3.88%	8.57%	-53	-52	-44	-78	149
NB7	0.28%	0.05%	-0.02%	0.31%	2.17%	29	5	-2	32	220
NB8	-0.76%	-0.51%	0.03%	-1.26%	-0.35%	-15	-10	1	-25	-7
NB9	0.34%	0.65%	1.20%	0.38%	1.38%	6	12	22	7	26

A.2.7 30 Bps Adjustment – Q1

Bond Ref	%D					PV0x				
	Models									
	1	2	3	4	5	1	2	3	4	5
B1	-2.29%	-2.47%	-2.80%	-2.71%	-2.21%	-89	-95	-109	-105	-85
B2	2.69%	2.91%	3.31%	2.89%	3.98%	51	56	63	55	75
B3	0.31%	0.24%	-0.16%	-0.43%	0.17%	8	6	-4	-11	4
B4	-1.96%	-2.23%	-2.22%	-2.75%	-2.06%	-40	-45	-45	-56	-42
B5	-2.93%	-2.40%	-1.61%	-2.34%	-1.26%	-49	-40	-27	-39	-21
B6	-1.28%	-0.91%	-1.17%	-1.11%	-0.98%	-49	-34	-44	-42	-37
B7	-0.49%	-0.64%	-0.97%	-1.26%	-1.02%	-12	-16	-24	-32	-26
B8	-3.53%	-3.55%	-3.72%	-4.16%	-4.11%	-98	-98	-103	-116	-115
B9	0.87%	0.91%	1.20%	0.45%	0.97%	17	17	23	9	19
B10	-4.91%	-4.50%	-3.51%	-4.29%	-3.32%	-83	-75	-58	-71	-54
B11	0.27%	0.21%	0.28%	0.28%	0.40%	5	4	5	5	8
B12	-5.29%	-5.35%	-5.24%	-5.57%	-4.95%	-109	-110	-108	-115	-102
B13	-0.86%	-0.80%	-1.01%	-0.96%	-0.27%	-32	-30	-38	-36	-10
B14	2.68%	2.66%	2.54%	2.46%	3.09%	65	65	62	60	75
B15	-0.61%	-0.65%	-0.78%	-0.36%	0.57%	-43	-46	-55	-25	40
B16	-3.52%	-3.59%	-3.74%	-3.74%	-2.94%	-90	-92	-96	-96	-75
B17	-4.91%	-4.70%	-4.50%	-4.96%	-3.98%	-101	-96	-92	-102	-81
B18	-5.20%	-5.05%	-4.72%	-5.10%	-3.65%	-107	-103	-96	-104	-74
B19	-5.55%	-5.22%	-4.67%	-5.81%	-4.60%	-93	-87	-77	-98	-76
B20	-4.63%	-4.62%	-4.33%	-6.13%	-4.22%	-67	-67	-62	-92	-61
B21	-1.87%	-1.82%	-1.77%	-1.69%	-1.54%	-56	-55	-53	-51	-46
B22	-4.75%	-4.59%	-4.58%	-4.62%	-3.89%	-122	-118	-117	-118	-99
B23	-4.39%	-4.30%	-4.40%	-4.76%	-4.51%	-112	-110	-112	-122	-115
B24	-5.04%	-5.23%	-5.42%	-4.93%	-4.10%	-197	-205	-213	-192	-159
B25	-1.87%	-1.60%	-1.46%	-1.36%	0.34%	-51	-43	-40	-37	9
B26	0.48%	0.66%	0.99%	1.06%	1.60%	13	18	26	28	42
B27	-5.11%	-5.55%	-5.72%	-5.78%	-5.65%	-85	-93	-96	-97	-95
B28	-1.21%	-1.35%	-1.39%	-1.38%	-1.55%	-86	-95	-98	-98	-110
NB1	7.03%	7.87%	8.46%	7.58%	8.64%	173	191	204	185	208
NB2	1.11%	0.82%	1.80%	-0.36%	1.24%	24	18	38	-8	27
NB3	1.09%	0.79%	1.50%	0.83%	1.10%	21	15	29	16	21
NB4	1.74%	2.19%	2.81%	2.28%	2.65%	42	53	68	55	64
NB5	-3.75%	-3.62%	-3.36%	-4.24%	4.85%	-95	-92	-85	-108	112
NB6	-3.69%	-3.64%	-3.25%	-4.89%	7.40%	-74	-73	-65	-99	130
NB7	0.45%	0.22%	0.15%	0.49%	2.34%	47	23	16	50	237
NB8	1.34%	1.59%	2.15%	0.83%	1.76%	26	31	41	16	34
NB9	0.73%	1.04%	1.60%	0.78%	1.78%	14	20	30	15	33

A.2.8 30 Bps Adjustment – Q2

Bond Ref	%D					PV0x				
	Models									
	1	2	3	4	5	1	2	3	4	5
B1	-1.44%	-1.61%	-1.94%	-1.86%	-1.35%	-55	-62	-75	-72	-52
B2	2.70%	2.92%	3.33%	2.90%	3.99%	52	56	63	55	75
B3	0.02%	-0.04%	-0.44%	-0.72%	-0.12%	1	-1	-11	-18	-3
B4	-0.40%	-0.67%	-0.66%	-1.20%	-0.50%	-8	-13	-13	-24	-10
B5	-1.05%	-0.50%	0.31%	-0.44%	0.67%	-17	-8	5	-7	11
B6	-1.28%	-0.91%	-1.17%	-1.11%	-0.98%	-49	-34	-44	-42	-37
B7	-0.49%	-0.64%	-0.97%	-1.26%	-1.02%	-12	-16	-24	-32	-26
B8	-2.33%	-2.36%	-2.53%	-2.97%	-2.92%	-64	-65	-69	-82	-81
B9	-0.76%	-0.72%	-0.44%	-1.17%	-0.66%	-15	-14	-9	-23	-13
B10	-2.94%	-2.52%	-1.50%	-2.29%	-1.29%	-48	-41	-24	-37	-21
B11	-1.38%	-1.44%	-1.37%	-1.37%	-1.26%	-27	-28	-27	-27	-25
B12	-3.62%	-3.67%	-3.56%	-3.90%	-3.26%	-73	-74	-72	-79	-66
B13	-1.73%	-1.66%	-1.88%	-1.82%	-1.14%	-66	-63	-71	-69	-43
B14	2.68%	2.66%	2.54%	2.46%	3.09%	65	65	62	60	75
B15	-0.61%	-0.65%	-0.78%	-0.36%	0.57%	-43	-46	-55	-25	40
B16	-3.79%	-3.85%	-4.00%	-4.00%	-3.21%	-97	-99	-103	-103	-82
B17	-5.24%	-5.03%	-4.83%	-5.29%	-4.32%	-108	-103	-99	-109	-88
B18	-5.53%	-5.38%	-5.05%	-5.43%	-3.99%	-114	-110	-103	-112	-81
B19	-3.47%	-3.13%	-2.56%	-3.74%	-2.49%	-57	-51	-41	-61	-40
B20	-2.22%	-2.21%	-1.90%	-3.79%	-1.79%	-31	-31	-26	-55	-25
B21	-0.76%	-0.71%	-0.65%	-0.58%	-0.43%	-22	-21	-19	-17	-13
B22	-3.44%	-3.29%	-3.27%	-3.31%	-2.58%	-87	-83	-82	-84	-65
B23	-4.38%	-4.30%	-4.40%	-4.76%	-4.51%	-112	-110	-112	-122	-115
B24	-5.17%	-5.37%	-5.56%	-5.07%	-4.24%	-202	-210	-218	-198	-164
B25	-2.09%	-1.82%	-1.69%	-1.58%	0.12%	-57	-50	-46	-43	3
B26	0.29%	0.48%	0.80%	0.87%	1.41%	8	13	21	23	37
B27	-5.54%	-5.98%	-6.14%	-6.21%	-6.08%	-93	-101	-104	-105	-102
B28	-1.21%	-1.35%	-1.39%	-1.38%	-1.55%	-86	-95	-98	-98	-110
NB1	6.92%	7.76%	8.35%	7.47%	8.53%	170	189	202	182	206
NB2	-0.31%	-0.60%	0.37%	-1.75%	-0.18%	-7	-13	8	-39	-4
NB3	1.20%	0.90%	1.61%	0.95%	1.21%	23	17	31	18	23
NB4	1.80%	2.26%	2.87%	2.34%	2.71%	44	55	69	57	65
NB5	-3.74%	-3.61%	-3.35%	-4.23%	4.87%	-95	-91	-84	-108	112
NB6	-3.67%	-3.62%	-3.23%	-4.87%	7.43%	-73	-72	-64	-99	130
NB7	0.47%	0.24%	0.17%	0.51%	2.36%	49	25	18	52	239
NB8	-0.25%	-0.01%	0.54%	-0.75%	0.16%	-5	0	11	-15	3
NB9	0.87%	1.18%	1.74%	0.91%	1.91%	16	22	32	17	36

APPENDIX B: EXCERPTS OF SELECTED R CODE USED

Appendix B provides certain excerpts of code used in this study. This is not an A-Z segmentation of all the code applied, as there are numerous sections that pertain to parts such as data pre-processing. The full code is available upon request.

B.1 Constant Z-spread estimation

```
####### Constant Z-Spread estimation ########

Run appropriate data before the relevant constant Z-Spread can be calculated

###################################################################
Names_ZR_Loop = as.character(paste("ZR_", substr(Names_ZRates, 10, 17), sep =
""))
Names_VData_Loop = as.character(paste("V_", Names_Filter, sep = ""))

for (p in 1:length(AnchorDates))
{
  # Set anchor date (ad), which is the start date or evaluation date
  ad1 = ymd(AnchorDates[p])
  # Specify which data set should be used:
  Data_V1 = get(Names_VData_Loop[p])
  # Specify appropriate Zero_Rates:
  Zero_Rates = get(Names_ZR_Loop[p])

  Z_Spread = rep(0, nrow(Data_V1))

  for (i in 1:nrow(Data_V1))
  {
    Get_C_ZS = function(Z)
    {
      # Get vector of first interest dates:
      FirstInterestDate = as.Date(as.numeric(Data_V1[i,"First Interest
Date"]), origin = "1899-12-30")
      MaturityDate = as.Date(as.numeric(Data_V1[i,"Maturity Date"]), origin =
"1899-12-30")

      C_Dates = rep(0, 1)
      if (as.numeric(Data_V1[i,"Coupon Frequency"])==2)
      {
        C_Dates = seq(as.Date(FirstInterestDate), as.Date(MaturityDate), "6
months")
      }else {
        C_Dates = seq(as.Date(FirstInterestDate), as.Date(MaturityDate),
"quarters")
      }

      # The above code section assumes that the only possible coupon
frequency is 2 or 4 (times per year)

      if(C_Dates[length(C_Dates)] <(MaturityDate - 80))
      {
        C_Dates = c(C_Dates, MaturityDate)
      }

      # Change weekend dates to following Monday:
      for (j in 1:length(C_Dates))
      {
```

```
        if(weekdays(C_Dates[[j]]) == "Saturday")
        {C_Dates[[j]] = C_Dates[[j]]+days(2)
        }else if (weekdays(C_Dates[[j]]) == "Sunday")
        {C_Dates[[j]] = C_Dates[[j]]+days(1)}
        else
        {C_Dates[[j]] = C_Dates[[j]]
        }
      }

      # C_Dates can only begin after anchor date, ad1:
      C_Dates = C_Dates[C_Dates>ad1]

      # Get Coupon Frequency:
      C_Freq = as.numeric(Data_V1[i,"Coupon Frequency"])

      # Coupon amount payable on the coupon dates:
      Coupon_Rate = as.numeric(Data_V1[i,"Coupon Rate %"])/C_Freq

      # Get vector of corresponding Zero-Rates:
      Zero_Rates_on_C = rep(0, length(C_Dates))
      Zero_Rates = as.data.frame(Zero_Rates)
      X = ymd(Zero_Rates[,1])
      Y = as.numeric(Zero_Rates[,2])
      Zero_Rates_on_C = Y[match(C_Dates, X)]/100

      # Get vector of times to discount back to Ad1:
      T_1 = (C_Dates - ad1)/365
      T_2 = as.numeric(T_1)

      AIP_T = sum(Coupon_Rate * exp(- T_2*(Z + Zero_Rates_on_C))) +
        100 * exp(-T_2[[length(C_Dates)]]*(Z +
Zero_Rates_on_C[[length(C_Dates)]]))

      AIP = as.numeric(Data_V1[i, "All In Price"])

      AIP - AIP_T
    }

    Z <- uniroot(Get_C_ZS,c(-100, 100))
    Z_Spread[[i]] = Z$root
  }

  assign(paste("CZ_", Names_Filter[p], sep = ""), Z_Spread)
}

# Concatenate to Data:

Names_Filter_1 = as.character(paste("V_", Names_Filter, sep = ""))
Names_Filter_2 = as.character(paste("CZ_", Names_Filter, sep = ""))
```

B.2 Application of regression model

```
###################################################################
# Application Of regression model before bootstrapping is done:
###################################################################

# Data Preparation:

# Rename data, such that banks and non-banks data sets are separate:
```

```
New_Issue_Data_Final_Banks      =
New_Issue_Data_Final[!is.na(match(New_Issue_Data_Final[, "Issuer"],
Unique_Issuers_Banks)), ]
New_Issue_Data_Final_Non_Banks  =
New_Issue_Data_Final[is.na(match(New_Issue_Data_Final[, "Issuer"],
Unique_Issuers_Banks)), ]

Issue_Dates_Banks     =
as.Date(as.numeric(unlist(New_Issue_Data_Final_Banks[,"Issue Date"])), origin
= "1899-12-30")
Issue_Dates_Non_Banks =
as.Date(as.numeric(unlist(New_Issue_Data_Final_Non_Banks[,"Issue Date"])),
origin = "1899-12-30")

Names_Final_Reg_Q_Banks     = paste("Final_Data_Banks_", gsub("-", "",
Issue_Dates_Banks), sep = "")
Names_Final_Reg_Q_Non_Banks = paste("Final_Data_Non_Banks_", gsub("-", "",
Issue_Dates_Non_Banks), sep = "")

# Remove duplicates:
Names_Final_Reg_Q_Banks = unique(Names_Final_Reg_Q_Banks)
Names_Final_Reg_Q_Non_Banks = unique(Names_Final_Reg_Q_Non_Banks)

###################################################################
# Regression Model Fit:
###################################################################

Coefs_FFit_B_2 = fit_B_2$coefficients
Coefs_FFit_B_2 = as.numeric(Coefs_FFit_B_2)

Coefs_FFit_NB_3 = fit_NB_3$coefficients
Coefs_FFit_NB_3 = as.numeric(Coefs_FFit_NB_3)

# Banks:
for(i in 1:length(Names_Final_Reg_Q_Banks))
{
  Data = get(Names_Final_Reg_Q_Banks[i])
  Cred_Without_LogTime_B = rep(0, nrow(Data))

  for(j in 1:nrow(Data))
  {
    Cred_Without_LogTime_B[j] = Coefs_FFit_B_2[1] + Coefs_FFit_B_2[2] *
Data[j, "Debt_Equity"]  +
      Coefs_FFit_B_2[3] * Data[j, "Volatility"]   + Coefs_FFit_B_2[4] *
Data[j, "IVec_B_NB"] +
      Coefs_FFit_B_2[5] * Data[j, "GTVec_B_NB"]
  }

  # Sort the results of the credit calculation using quantiles:
  Y_2 = quantile(Cred_Without_LogTime_B)
  Z_2 = rep(0, nrow(Data))
  X_2 = Cred_Without_LogTime_B
  Z_2[X_2<Y_2[2]] = 1
  Z_2[X_2>=Y_2[2] & X_2<=Y_2[4]] = 2
  Z_2[X_2>Y_2[4]] = 3

  Cred_Qt_B = Z_2
  X_B = X_2

  Data = cbind(Data, X_B, Cred_Qt_B)
  assign(Names_Final_Reg_Q_Banks[i], Data)
```

```
}

# Non - Banks:
for(i in 1:length(Names_Final_Reg_Q_Non_Banks))
{
  Data = get(Names_Final_Reg_Q_Non_Banks[i])
  Cred_Without_LogTime_NB = rep(0, nrow(Data))


  for(j in 1:nrow(Data))
  {
    Cred_Without_LogTime_NB[j] = Coefs_FFit_NB_3[1] + Coefs_FFit_NB_3[2] *
Data[j, "Current_Ratio"]  +
      Coefs_FFit_NB_3[3] * Data[j, "Volatility"]    + Coefs_FFit_NB_3[4] *
Data[j, "Price_EBITDA"] +
      Coefs_FFit_NB_3[5] * Data[j, "IVec_B_NB"]
  }

  # Sort the results of the credit calculation using quantiles:
  Y_2 = quantile(Cred_Without_LogTime_NB)
  Z_2 = rep(0, nrow(Data))
  X_2 = Cred_Without_LogTime_NB
  Z_2[X_2<Y_2[2]] = 1
  Z_2[X_2>=Y_2[2] & X_2<=Y_2[4]] = 2
  Z_2[X_2>Y_2[4]] = 3

  Cred_Qt_NB_F3 = Z_2
  X_NB_F3 = X_2

  Data = cbind(Data, X_NB_F3, Cred_Qt_NB_F3)
  assign(Names_Final_Reg_Q_Non_Banks[i], Data)

}
```

B.3 Zero-Z-spread estimation

```
###################################################################
# Zero-Z-Spread estimation is done:
###################################################################

For (q in 1:length(Names_Final_Reg_Q_Banks))
{
  # Specify Data:
  Data = get(Names_Final_Reg_Q_Banks[q])
  col_ref = match(ymd(substr(Names_Final_Reg_Q_Banks[q], 18, 25)),
AnchorDates_New)

  # Specify Anchordate:
  ad1 = AnchorDates_New[col_ref]
  # Specify corresponding zero rates:
  Zero_Rates = as.numeric(unlist(New_ZeroRates[, col_ref+1]))

  # 1. Filter and sort exclude certain bonds:
  Mat = subset(Data, Data$Cred_Qt_B == 2)
  Mat = subset(Mat, Mat$Logs_Vec_New > 0) # Remove Matured Bonds
  Mat = Mat[order(Mat$`Maturity Date`), ]

 # 2. Identify when coupon payments are going to be made:
  # Create an empty matrix in which coupon dates are stored:
```

```
  Pay_Mat = matrix(0, nrow(Mat), 120)  # The 120 is chosen as the estimated
  # maximum of the amount of potential
  # future coupon payments
  Pay_Mat = as.data.frame(Pay_Mat)

  # Create an empty matrix in which the zero rates that corresponds to
  # each coupon date are stored:
  Zero_P_Mat = matrix(0, nrow(Mat), 120)
  Zero_P_Mat = as.data.frame(Zero_P_Mat)
  i
  for (i in 1:nrow(Pay_Mat))
  {
    # Get vector of first interest dates:
    FirstInterestDate = as.Date(as.numeric(Mat[i,"First Interest Date"]),
origin = "1899-12-30")
    MaturityDate = as.Date(as.numeric(Mat[i,"Maturity Date"]), origin =
"1899-12-30")

    C_Dates = rep(0, 1)
    if (as.numeric(Mat[i,"Coupon Frequency"])==2)
    {
      C_Dates = seq(as.Date(FirstInterestDate), as.Date(MaturityDate), "6
months")
    }else
    {
      C_Dates = seq(as.Date(FirstInterestDate), as.Date(MaturityDate),
"quarters")
    }

    # Change weekend dates to following Monday:
    for (j in 1:length(C_Dates))
    {
      if(weekdays(C_Dates[[j]]) == "Saturday")
      {C_Dates[[j]] = C_Dates[[j]]+days(2)
      }else if (weekdays(C_Dates[[j]]) == "Sunday")
      {C_Dates[[j]] = C_Dates[[j]]+days(1)
      }else
      {C_Dates[[j]] = C_Dates[[j]]
      }
    }

    # C_Dates can only begin after anchor date:
    C_Dates = C_Dates[C_Dates>ad1]

    # Get vector of corresponding Zero-Rates:
    # (Amend if coupon dates exceed the corresponding dates of
    # the zero-rates)
    Zero_Rates_on_C = rep(0, length(C_Dates))

    TT_1 = as.numeric(C_Dates - ad1)
    Z_TT_1 = Zero_Rates[TT_1]
    Zero_Rates_on_C = as.numeric(unlist(Z_TT_1))/100

    # Input the zero-rates in Zero_P_Mat:
    for (j in 1:length(C_Dates))
    {
      Zero_P_Mat[i, j] = Zero_Rates_on_C[[j]]
    }

    # Input the coupon payments in Pay_Mat:
    for (j in 1:length(C_Dates))
```

```
    {
      Pay_Mat[i, j] = C_Dates[[j]]
    }
  }

  Pay_Mat = Pay_Mat - as.numeric(ad1)
  Pay_Mat[Pay_Mat < 0] <- 0
  # Now Pay_Mat represents the number of days after the evaluation date

  # 3. Create vector of coupon payments, that correspond to the frequency:
  Coupons = rep(0, nrow(Mat))
  for (i in 1:nrow(Mat))
  {
    # Get Coupon Frequency:
    C_Freq = as.numeric(Mat[i,"Coupon Frequency"])

    # Coupon amount payable on the coupon dates:
    Coupon_Rate = as.numeric(Mat[i,"Coupon Rate %"])/C_Freq
    Coupons[[i]] = Coupon_Rate
  }

  # 4. Get vector of All in Prices:
  AIP = as.numeric(Mat[, "All In Price"])

  # 5. Get vector of constant Z-Spreads:
  CZSpread = as.numeric(Mat[, "Z_Spread"])

  # 6.

  # This code shows the interpolation method that estimates the Zero-Z-
Spread:

  # A matrix is created that has all the maturity dates of each bond in the
one column and the corresponding zero Z-Spread next to it.

  # The first value of the matrix is (1, c_z_1), where c_z_1 is the estimated
constant zero rate of the bond that matures first. This is necessary for the
interpolation.

  Z_Z_S_Mat = matrix(1, nrow = nrow(Zero_P_Mat)+1, ncol = 2)
  Z_Z_S_Mat[-1,2] = CZSpread
  # Set day 1 value equal to the first Constant Z-Spread
  Z_Z_S_Mat[1,2]  = Z_Z_S_Mat[2,2]

  number = rep(1, nrow(Pay_Mat))
  for (i in 1:nrow(Pay_Mat)) {
    number[[i]] = sum(Pay_Mat[i,]>0)
  }

  # Loop that interpolates and constructs the Zero-Z-Spread:
  for (i in 1:nrow(Pay_Mat))
  {
    Z_Z_S_Mat[i+1,1] = max(Pay_Mat[i,])

    if (number[[i]] == 1)
    {
      Z_Z_S_Mat[i+1,2] = CZSpread[[i]]
    } else {

      I = rep(0, number[[i]]-1)
      D = rep(0, number[[i]]-1)
```

```
      for (j in 1:length(I))
      {
        # Get the discount factors through interpolation of
        # Zero-Z-Spreads that are already calculated:

        I[[j]] = approx(Z_Z_S_Mat[1:(i+1),1], Z_Z_S_Mat[1:(i+1),2] , xout =
Pay_Mat[i,j], method = "linear")$y

        # Calculate the "discount factors", that should be summed in next
section:
        D[[j]] = Coupons[[i]]*exp(-(Pay_Mat[i,j]/365)*(Zero_P_Mat[i,j] +
I[[j]]))
      }
      Z_Z_S_Mat[i+1,2] =(((-1/(Pay_Mat[i, number[[i]]]/365))*
                            (log((AIP[[i]]-sum(D, na.rm=TRUE)) / (100 +
Coupons[[i]]))))
                         - Zero_P_Mat[i, number[[i]]])
    }
  }

  assign(Names_ZZMats_Banks[q], Z_Z_S_Mat[2:nrow(Z_Z_S_Mat), ])

}
```

B.4 Curve estimation

```
# Gaussian Kernel function:
GaussianKernelYield = function(Sigma, X, Y, Z)
{
  Z = length(Z)
  GKYield = rep(0, Z)

  for(x in 1:Z)
  {
    GK = rep(0, length(X))
    for(i in 1:length(X))
    {
      # The Kernel function:
      GK[i] = (1/((sqrt(2*pi)*Sigma))) * exp(-0.5*(((X[i] - x)/365)/Sigma)^2)
    }

    Weight = rep(0,length(X))
    for(i in 1:length(X))
    {
      Weight[i] = GK[i]*100/(sum(GK*100))
    }             # Can insert true Face value of bonds to adjust weights
    GKYield[x] = sum(Y * Weight)
  }
  return(GKYield)
}

for(q in 1:length(Names_ZZMats_Non_Banks))
{
  # Specify Data:
  Data = get(Names_ZZMats_Non_Banks[q])
  X = Data[ ,1]
  Y = Data[ ,2]
  Y = Y * 10000
  Z = 1:6000
```

```
  # Gaussian Kernel:
  Y_GK_1 = GaussianKernelYield(1, X = X, Y = Y, Z = Z)
  Y_GK_2 = GaussianKernelYield(2, X = X, Y = Y, Z = Z)
  Y_GK_5 = GaussianKernelYield(5, X = X, Y = Y, Z = Z)

  # Logarithmic Regression:
  fit_Log = lm(Y ~ log(X))
  logCoefs =  fit_Log$coefficients
  Y_Log = logCoefs[1] + log(Z) * logCoefs[2]

  # Nelson Siegel
  NSParameters = Nelson.Siegel(rate = Y, maturity = X)
  Y_NS = rep(0, 6000)
  B0 = NSParameters[1,1]
  B1 = NSParameters[1,2]
  B2 = NSParameters[1,3]
  Lam1 = NSParameters[1,4]

  for(i in 1:length(Y_NS))
  {
    Y_NS[i] = NelsonSiegel_Output(B0, B1, B2, Lam1,  i)
  }

  ZZS_Fits = cbind(Y_GK_1, Y_GK_2, Y_GK_5, Y_Log, Y_NS)
  assign(Names_ZZMats_Non_Banks_Fits[q], ZZS_Fits)

}
```

B.5 Quantile adjustments

```
###################################################################
# Quantile Adjustment Estimation
###################################################################

Adj_Q_1_Banks = matrix(0, nrow = length(Names_ZZMats_Banks), ncol = 5)
Adj_Q_2_Banks = matrix(0, nrow = length(Names_ZZMats_Banks), ncol = 5)

# Banks:
for(i in 1:length(Names_ZZMats_Banks))
{

  # Specify Data:
  Data = get(Names_Final_Reg_Q_Banks[i])

  X_Vec_2 = Data[, "X_B"] # X = Cred_Without_LogTime
  Q_Vec = X_Vec_2

  Q_1 = quantile(Q_Vec, probs = c(0.2, 0.4, 0.6, 0.8))
  Q_2 = quantile(Q_Vec, probs = c(0.1, 0.25, 0.75, 0.9))

  Mean_1 =   mean(Q_Vec[Q_Vec <  Q_1[1]])
  Mean_2 =   mean(Q_Vec[Q_Vec >= Q_1[1] & Q_Vec < Q_1[2]])
  Mean_3 =   mean(Q_Vec[Q_Vec >= Q_1[2] & Q_Vec < Q_1[3]])
  Mean_4 =   mean(Q_Vec[Q_Vec >= Q_1[3] & Q_Vec < Q_1[4]])
  Mean_5 =   mean(Q_Vec[Q_Vec >= Q_1[4]])

  Adj_Q_1_Banks[i, 1] = Mean_1 - Mean_3
  Adj_Q_1_Banks[i, 2] = Mean_2 - Mean_3
  Adj_Q_1_Banks[i, 3] = Mean_3 - Mean_3
```

```
  Adj_Q_1_Banks[i, 4] = Mean_4 - Mean_3
  Adj_Q_1_Banks[i, 5] = Mean_5 - Mean_3

  Mean_1 =   mean(Q_Vec[Q_Vec <  Q_2[1]])
  Mean_2 =   mean(Q_Vec[Q_Vec >= Q_2[1] & Q_Vec < Q_2[2]])
  Mean_3 =   mean(Q_Vec[Q_Vec >= Q_2[2] & Q_Vec < Q_2[3]])
  Mean_4 =   mean(Q_Vec[Q_Vec >= Q_2[3] & Q_Vec < Q_2[4]])
  Mean_5 =   mean(Q_Vec[Q_Vec >= Q_2[4]])

  Adj_Q_2_Banks[i, 1] = Mean_1 - Mean_3
  Adj_Q_2_Banks[i, 2] = Mean_2 - Mean_3
  Adj_Q_2_Banks[i, 3] = Mean_3 - Mean_3
  Adj_Q_2_Banks[i, 4] = Mean_4 - Mean_3
  Adj_Q_2_Banks[i, 5] = Mean_5 - Mean_3

}
```

B.6 Pricing newly issued bonds

```
###################################################################
# Here all new issues are priced, using all the estimated curves
###################################################################

# Set Data:
Data = New_Issue_Data_Final_Banks

Results_Q_1_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127))
Results_Q_2_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127))

Diff_Q_1_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127))
Diff_Q_2_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127))

For (p in 1:nrow(Data))
{
  # Set anchor date (ad), which is the start date or evaluation date
  Date_ref = as.Date(as.numeric(unlist(Data[p, "Issue Date"])), origin =
"1899-12-30")
  col_ref = match(Date_ref, AnchorDates_New)

  Data_Ref_Get = as.Date(as.numeric(unlist(Data[p, "Issue Date"])), origin =
"1899-12-30")
  Dates_Dataset = ymd(substr(Names_Final_Reg_Q_Banks, 18, 25))
  Data_Ref_Get_Final = match(Data_Ref_Get, Dates_Dataset)

  ad1 = AnchorDates_New[col_ref]

  # Specify which data set should be used:
  Data_V1 = Data[p, ]
  Zero_Rates = as.numeric(unlist(New_ZeroRates[, col_ref+1]))

  # Get Curves:
  temp_Mat_Q_1_Banks = get(Names_Curves_Banks_Q_1[Data_Ref_Get_Final])
  temp_Mat_Q_2_Banks = get(Names_Curves_Banks_Q_2[Data_Ref_Get_Final])

  for (i in 1:ncol(temp_Mat_Q_1_Banks))
  {
```

```
    FirstInterestDate = as.Date(as.numeric(Data_V1[ 1,"First Interest
Date"]), origin = "1899-12-30")
    MaturityDate =      as.Date(as.numeric(Data_V1[ 1,"Maturity Date"]),
origin = "1899-12-30")

    C_Dates = rep(0, 1)
    if (as.numeric(Data_V1[1 ,"Coupon Frequency"])==2)
    {
      C_Dates = seq(as.Date(FirstInterestDate), as.Date(MaturityDate), "6
months")
    }else {
      C_Dates = seq(as.Date(FirstInterestDate), as.Date(MaturityDate),
"quarters")
    }
    if(C_Dates[length(C_Dates)] < (MaturityDate - 80)) # To capture maturity
date if seq function fails
    {
      C_Dates = c(C_Dates, MaturityDate)
    }

    # Change weekend dates to following Monday:
    for (j in 1:length(C_Dates))
    {
      if(weekdays(C_Dates[[j]]) == "Saturday")
      {C_Dates[[j]] = C_Dates[[j]]+days(2)
      }else if (weekdays(C_Dates[[j]]) == "Sunday")
      {C_Dates[[j]] = C_Dates[[j]]+days(1)}
      else
      {C_Dates[[j]] = C_Dates[[j]]
      }
    }
    # C_Dates can only begin after ancor date, ad1:
    C_Dates = C_Dates[C_Dates>ad1]

    # Get Coupon Frequency:
    C_Freq = as.numeric(Data_V1[1,"Coupon Frequency"])

    # Coupon amount payable on the coupon dates:
    Coupon_Rate = as.numeric(Data_V1[1,"Coupon Rate %"])/C_Freq

    # Get vector of corresponding Zero-Rates + Credit Spreads:
    TT_1 = as.numeric(C_Dates - ad1)

    Disc_Rates_Q_1  = temp_Mat_Q_1_Banks[, i]/10000 + Zero_Rates[1:6000]/100
    Disc_Rates_Q_2  = temp_Mat_Q_2_Banks[, i]/10000 + Zero_Rates[1:6000]/100

    Disc_Rates_on_C_Q_1 = Disc_Rates_Q_1[TT_1]
    Disc_Rates_on_C_Q_2 = Disc_Rates_Q_2[TT_1]

    # Get vector of times to discount back to Ad1:
    T_1 = (C_Dates - ad1)/365
    T_2 = as.numeric(T_1)

    AIP_Q_1 = sum(Coupon_Rate * exp(- T_2*(Disc_Rates_on_C_Q_1))) + 100 *
exp(-T_2[[length(C_Dates)]]*(Disc_Rates_on_C_Q_1[[length(C_Dates)]]))
    AIP_Q_2 = sum(Coupon_Rate * exp(- T_2*(Disc_Rates_on_C_Q_2))) + 100 *
exp(-T_2[[length(C_Dates)]]*(Disc_Rates_on_C_Q_2[[length(C_Dates)]]))

    AIP_From_Data = as.numeric(Data_V1[1, "All In Price"])

    # Assign to results_matrix:
```

```
    Results_Q_1_Banks[p, i] = AIP_Q_1
    Results_Q_2_Banks[p, i] = AIP_Q_2

    Diff_Q_1_Banks[p, i]    = (AIP_Q_1 - AIP_From_Data)/AIP_From_Data # MAD
    Diff_Q_2_Banks[p, i]    = (AIP_Q_2 - AIP_From_Data)/AIP_From_Data

  }
}
```

B.7 Quantile determination for newly issued bonds

```
####################################################################
# Here all quantiles from the regression model for the new issues are
determined:
####################################################################

# Results_Q_1_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127))
# Results_Q_2_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127))

# Diff_Q_1_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127))
# Diff_Q_2_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127))

####################################################################

# Coefs_FFit2 = c("Intercept", "Debt / Equity", "Vol", "IssueClassVec_B",
"GT_Vec_B", "Logs_Vec_B")
Coefs_FFit_B_2 = fit_B_2$coefficients
Coefs_FFit_B_2 = as.numeric(Coefs_FFit_B_2)

Data = New_Issue_Data_Final_Banks
Pricing_Error_Q_1_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127)/5)
Pricing_Error_Q_2_Banks = matrix(0, nrow = nrow(Data), ncol =
ncol(Curve_Mat_Banks_Q1_20160127)/5)

Q_1_Quantiles_Banks = rep(0, nrow(Data))
Q_2_Quantiles_Banks = rep(0, nrow(Data))

for(j in 1:nrow(Data))
{
  Data_Ref_Get = as.Date(as.numeric(unlist(Data[j, "Issue Date"])), origin =
"1899-12-30")
  Dates_Dataset = ymd(substr(Names_Final_Reg_Q_Banks, 18, 25))
  Data_Ref_Get_Final = match(Data_Ref_Get, Dates_Dataset)

  Data_Quantiles_X = get(Names_Final_Reg_Q_Banks[Data_Ref_Get_Final])
  Data_Quantiles_X_2 = Data_Quantiles_X[, "X_B"] # X = Cred_Without_LogTime

  Q_Vec = Data_Quantiles_X_2

  Q_1_X = quantile(Q_Vec, probs = c(0.2, 0.4, 0.6, 0.8))
  Q_2_X = quantile(Q_Vec, probs = c(0.1, 0.25, 0.75, 0.9))

  New_Cred_Without_LogTime = Coefs_FFit_B_2[1] + Coefs_FFit_B_2[2] * Data[j,
"Debt_Equity"]  +
```

```
    Coefs_FFit_B_2[3] * Data[j, "Volatility"]   + Coefs_FFit_B_2[4] * Data[j,
"IVec_B_NB"] +
    Coefs_FFit_B_2[5] * Data[j, "GTVec_B_NB"]

  # Determine quantile interval of new issue:
  if (New_Cred_Without_LogTime < Q_1_X[1])
  {     Q_1_NewIssue = 1
  } else if (Q_1_X[1] < New_Cred_Without_LogTime & New_Cred_Without_LogTime
<= Q_1_X[2])
  {     Q_1_NewIssue = 2
  } else if (Q_1_X[2] < New_Cred_Without_LogTime & New_Cred_Without_LogTime
<= Q_1_X[3])
  {     Q_1_NewIssue = 3
  } else if (Q_1_X[3] < New_Cred_Without_LogTime & New_Cred_Without_LogTime
<= Q_1_X[4])
  {     Q_1_NewIssue = 4
  } else if (Q_1_X[4] < New_Cred_Without_LogTime )
  {     Q_1_NewIssue = 5
  }

  if (New_Cred_Without_LogTime < Q_2_X[1])
  {     Q_2_NewIssue = 1
  } else if (Q_2_X[1] < New_Cred_Without_LogTime & New_Cred_Without_LogTime
<= Q_2_X[2])
  {     Q_2_NewIssue = 2
  } else if (Q_2_X[2] < New_Cred_Without_LogTime & New_Cred_Without_LogTime
<= Q_2_X[3])
  {     Q_2_NewIssue = 3
  } else if (Q_2_X[3] < New_Cred_Without_LogTime & New_Cred_Without_LogTime
<= Q_2_X[4])
  {     Q_2_NewIssue = 4
  } else if (Q_2_X[4] < New_Cred_Without_LogTime )
  {     Q_2_NewIssue = 5
  }

  # Get sequence of indices for various models:
  Q_1_s_NewIssue = Q_1_NewIssue + seq(0,20,5)
  Q_2_s_NewIssue = Q_2_NewIssue + seq(0,20,5)

  Q_1_Quantiles_Banks[j] = Q_1_NewIssue
  Q_2_Quantiles_Banks[j] = Q_2_NewIssue

  # Assign to matrix:
  Pricing_Error_Q_1_Banks[j, ] = Diff_Q_1_Banks[j, Q_1_s_NewIssue]
  Pricing_Error_Q_2_Banks[j, ] = Diff_Q_2_Banks[j, Q_2_s_NewIssue]

}
```