Modelling Exposure At Default Without Using Conversion Factors

Macquarie University

STAT825: Statistics Project
Master of Applied Statistics
Semester Two, 2015

Mark Thackham (41411889)

supervised by:
A/Prof Jun Ma

November 10, 2015

Dedication

谨以此文献给佳。

Acknowledgements

I would like to thank Associate Professor Jun Ma who supervised my project.

I would also like to thank three anonymous reviewers for providing feedback and suggestions.

Contents

Chapter 1

Executive summary

1.1 Abstract

Banks accredited by their regulator to use the Advanced Internal Ratings Based (A-IRB) approach are required to provide their own estimates for calculating their minimum credit capital; these estimates rely on statistical and analytical models to predict Probability of Default (PD), Loss Given Default (LGD) and Exposure at Default (EAD). This thesis focusses on estimating EAD for banks granting revolving loans to large corporates and leverages the Global Credit Data (GCD) database.

This thesis briefly discusses why risk management, particularly credit risk management, is important for banks and we survey the existing EAD modelling literature which to date has had less focus than PD and LGD modelling.

Our prosed methodology models both loan balance at default (EAD) and changes in loan limit at default as random variables, modelling their joint dynamics via a two stage model – the first stage estimates the probability that limits decrease while the second stage estimates EAD conditional on changing limits. To the best of our knowledge, our approach is the first to estimate EAD and changes in loan limit directly for large corporate revolving facilities using the GCD database.

Our model suggests that the key drivers of EAD include: limit; balance; utilisation; risk rating; and time to maturity. We also find evidence that banks actively manage limits in the lead up to default, and that these changes in limits have substantial effects on the outcomes of realised EAD.

1.2 Project Goals

This project trains a statistical model for estimating Exposure at Default (EAD) for revolving loan facilities to large corporates. Our approach directly models the joint dynamics $log_{10}(EAD)$ and changes in limits, as opposed to the more common approach in literature and in industry to model "Credit Conversion Factor" (CCF). Our model is trained using loss data from the Global Credit Data (GCD) consortium. Our model could be implemented by a large internationally active bank that advances revolving facilities to large corporate counterparties. The model estimates could be used for risk management purposes such as: pricing; provisioning; limit management; economic capital; stress testing; and regulatory capital (subject to the approval by the bank's regulator).

1.3 Structure of the Report

This project proceeds as follows. Chapter 2 introduces the concepts of risk, return and how risk management can help strike the desired balance between the two. We also provide some background to the "Basel Accords", which are regulatory risk management frameworks to help to provide controls for the risks that banks face. We focus our discussion on the "Basel II" accord, in particular how it requires statistical and analytical estimation of key risk management inputs and that one of these is exposure at default (EAD), the focus of this thesis. Finally, we end by introducing the Global Credit Data (GCD) reference data set used to train our model.

Chapter 3 contains a literature review and begins first by briefly retouching at a high level on the regulatory setting for credit risk capital estimation by banks that are accredited to use the Advanced Internal Ratings Based (A-IRB) approach. We outline the key terminology used for EAD estimation. This is followed by a summarisation of relevant available literature that discusses EAD estimation for banks.

Chapter 4, which comprises the bulk of the report, contains our modelling methodology and results. We discuss the modelling data before and after applying filters, introduce our proposed methodology, and then present univariate and multivariate analysis that helps select the important covariates and allows parameter estimation. We next present our final model and analyse the results by comparing predicted and observed values of $log_{10}(EAD)$. We also outline the sensitivity of our selected covariates for predicting changes in limits and estimates of $log_{10}(EAD)$. We end the section by discussing the knowledge discovered while also suggesting avenues for further research.

Finally, the appendices contain details of the univariate, bivariate and multivariate analysis. We also provide the computer code, in both SAS and R, to fit our final model.

Chapter 2

Background to Credit Risk

2.1 Introduction

While banks offer many products and services, with individual institutions specialising in different parts of the value chain (for example, investment banks, commercial banks, and retail banks), the granting of loans to customers is one of the core services that they provide (Apostolik et al., 2009). Some examples include issuing of a consumer credit card to a private individual, or a trade-finance facility for a large corporate client.

The granting of these loans exposes the bank to **credit risk**, which is defined as:

> ***"the potential loss a bank would suffer if a borrower fails to meets its obligations"*** (Apostolik et al., 2009).

Risks that build-up unchecked within either an individual institution or the wider financial system may cause insolvency of an institution and/or instability in the financial system. As credit risk is typically the largest risk class a bank faces, it thus attracts significant attention from bank both: management who wish to run profitable institutions for their stakeholders; and regulators who aim to achieve over a stable and functioning banking system so that institutions can fulfil the financial promises they make.

To help incentivise prudently run banks and admonish excessive risk taking, banking regulators stipulate a minimum level of capital that banks must hold, called "regulatory capital", and is designed to be risk sensitive. That is, the more risk that a bank chooses to take on, the higher the minimum level of capital they will be required to hold. Safer banks whose management decide they should operate more prudently are allowed to hold additional capital beyond the stipulated minimum if they wish and this is common in practice. For an example of a regulatory view, the Australian Prudential Regulation Authority (APRA) who regulate banks in Australia require banks to *"maintain adequate capital ... to act as a buffer against the risk associated with its activities"* (APRA, 2015).

In this chapter, we introduce the concept of risk, demonstrating how it is intrinsically linked to economic returns and how banks make use of the process of risk managemt to identify, quantify and control risk to reduce their risk profile to within its targeted risk appetite. We also discuss the regulatory risk management control framework for banks, which is commonly known as the "Basel Accords". Our comments largely centre on the second of the accords ("Basel II"), focussing on the statistical quantification inputs to calculate of credit risk-weighted assets. It is estimation of one of these components, exposure at default (EAD), that is the focus of this thesis. Finally, we also introduce the Global Credit Data (GCD) reference data set used to train our model.

2.2 Risk Management

2.2.1 What is Risk

There are a wide variety of slightly differing definitions of "risk", with no single definition suitable in all settings. The Oxford English Dictionary (2015) defines risk as ***"exposure to the possibility of loss, injury, or other adverse or unwelcome circumstance"***.

For a view specific to finance, McNeil et al. (2005) provide two definitions in their book "Quantitative Risk Management" which are: ***"any event or action that may adversely affect an organization's ability to achieve its objectives and execute its strategies"***, or alternatively ***"the quantifiable likelihood of loss or less than expected returns"***.

In his book "Value at Risk", Jorion (2001) states that risk can generally be defined as the uncertainty of outcomes when compared to expectation. The author further says that the origins of the word "risk" can be traced from Latin, through the French "risqué" and the Italian "risco". The original sense of "risco" is to cut off like a rock, from the Latin "re-" (back) and "secare" (to cut). Hence the sense of peril to sailors who had to navigate around the dangers of sharp rocks.

Clearly these above definitions of risk could apply to and effect the outcomes and continued operation any organisation and not only a bank. Take for example:

- a non-financial corporation, such as a manufacturer, which might suffer a loss of market share and hence a fall in profits from the effects of entry of a new competitor to their market;
- a university, which might see its international student enrolments decrease due to a substantial appreciation of its country's currency that causes many potential students to prefer alternate universities in cheaper countries;
- a charity might see its charitable donations it receives decrease during an economic recession;
- a large cyclone in Brazil might largely destroy the country's sugar cane crop for one year causing world sugar prices to spike temporarily and resulting in Australian farmers enjoying higher than expected profits.

All of these examples are characterised by outcomes that have deviated to be either better or worse than expectations.

Given the focus of this thesis is on financial risk management for banks, we will now focus our discussion and examples from this point on specifically on banks. For instance, Apostolik et al. (2009) (who also agree that there are multiple definitions of risk) provide some concrete examples of various risks that a bank may encounter:

- Borrowers may submit payments late or fail altogether to make repayments;
- A depositor may wish their money returned faster than the bank anticipated;
- Market interest rates may change and decrease the value of the bank's loans;
- Investments made by the bank may unexpectedly lose value;
- Human input errors, frauds, computer systems or natural disasters may lead to losses.

2.2.2 What is Return, and How is it Related to Risk

Returns are the financial gains that accrue (in the main) from exposure to risk, with higher returns expected in exchange for taking on and being exposed to higher risk. In some sense, return is the reward or the incentive for taking on risk. In essence, these simply describe the adage of "no risk- no return", a trade-off which is widely accepted in the business world (Lam, 2003). Crouchy et al. (2006) refer to this as the "conflict of risk and reward", whereby in commercial activities, if one wants to achieve a higher rate of return on average, one often has to assume more risk. For instance, a bank would charge a higher interest rate for credit card than for a home loan because the home loan is secured by property and thus lower risk than the unsecured credit card.

In financial theory, risk and return are inextricably linked (Peirson et al., 2002). Lam (2003) says that both risk and return must always be jointly considered and balanced via the process of risk management, which entails the key steps of: risk identification, risk quantification and risk control.

2.2.3 Risk Identification, Quantification and Control

In order to manage risks, they need to be firstly identified, secondly quantified, and thirdly controlled.

Risk identification techniques (an entire topic in itself) can include the creation of a taxonomy of all the risks that a bank is exposed to, and result in the capturing these on a risk register. Banks are exposed to a myriad of individual risks, but the three main classes they are exposed to (in decreasing order of importance) as defined by Apostolik et al. (2009) are:

- Credit risk – loss from default of a borrower or counterparty (either in part or in full), whether due to inability or unwillingness;
- Operational risk – direct or indirect loss from either inadequate or failed internal processes, people, systems, or natural disasters;
- Market risk – losses from changes in market prices (typically relating to changes in: interest rates; foreign currency; commodities; or equities).

Assuming that all relevant risks have been identified, focus can shift to their quantification (the primary topic for this thesis). Jorion (2001) notes that probability theory can be (and is indeed often) used to help measure risk. Lam (2003) lists seven dimensions of risk quantification, noting that not all dimensions are relevant for all risks:

- Probability – how likely is the event to occur?
- Exposure – what does the bank stand to lose?
- Severity – what loss is likely to be suffered?
- Volatility – how uncertain is the future?
- Time Horizon – how long will the bank be exposed to the risk?
- Correlation – how are the individual risks related to each other?
- Capital – how much safety margin should a bank put aside to cover unexpected losses?

Once risks have been identified and quantified, appropriate controls need to be implemented to reduce the risks to a level that is within the bank's risk appetite. Chapman (2006) lists four methods for controlling identified risks, noting that a particular risk can have more than control applied:

- Avoid – cease the activity and remove the exposure to the risk;
- Reduce – reduce the exposure to the risk;
- Retain – decide to accept the exposure to the risk;
- Transfer – obtain indemnity to reduce severity in the event that the risk occurs.

Placing controls on risks that have been successful identified and quantified will typically result in the reduction rather than the complete elimination in risk. Any left over risk is known as "residual risk" and a primary outcome of risk management is to reduce the level of this "residual risk' to within a tolerable level, known as the bank's "risk appetite", allowing the bank to target and obtain the desired level of return in a controlled manner. For example, if a bank had identified that the level of risk for a credit card portfolio was too high for its risk appetite, it could enact one or more of the following controls:

- Avoid the risk, by selling the portfolio to a competitor bank;
- Reduce the risk, by selling less of the product;
- Retain the risk, deciding that it remains well within the bank's overall risk appetite;
- Transfer the risk, by purchasing a insurance against credit losses from an investment bank to hedge the risk.

For instance, in order to move this risk to be within the bank's "risk appetite", the control to reduce the monthly growth rate from 10% per month to 5% per month could be selected. The remaining risk would be the "residual risk".

2.2.4 The Benefits of Risk Management

The failure by banks to execute appropriate risk management has been brought into sharp focus after the credit crisis and "Global Financial Crisis" (GFC) of 2007 and 2008, and showed that when banks mismanage their risks the spill-over can affect not only the banking industry, but entire economies. The crisis clearly showed that a lack of focus by banks on risk management, and in particular on credit risk management (the primary risk that a bank takes on) can ultimately lead to a their financial demise, as evidenced by the failure and nationalisation of many household-name banks around the world.

While the causes and ramifications of the GFC have been – and will no doubt continue to be – debated and discussed at length, this thesis will not seek to do so. Instead, it focusses on estimation using statistical methods of a key quantitative component of credit risk, the main class of risk a bank faces.

From a theoretical point of view McNeil et al. (2005) state that an important reason for undertaking risk management is to balance the competing expectations of the bank's many stakeholders, including: shareholders; customers; management; regulators; politicians; and the public at large. They reason that for banks, there is a societal viewpoint and clear expectation that the banking system should be run smoothly stating that the regulatory process of Basel Accords has been strongly motivated from the fear of systematic risk that may spill over (and actually did spill-over during the Global Financial Crisis of 2007 and 2008) from one bank to another bank, industry or even country.

2.3 The Basel Accords

The genesis of the Basel Accords can be traced back to 1974, when the Central Banks of the then 10 largest economies of the world represented by the G10 established the Basel Committee of Banking Supervision (BCBS) with an aim of *"setting minimum standards for the regulation and supervision of banks"* (Bank of International Settlements, 2014). The committee does not create legal regulations, but instead devises supervisory standards and guidelines that its now nearly 30 member countries are expected to implement in their local jurisdictions.

2.3.1 Basel I

The BCBS published "Basel I: the Basel Accord" in 1988 which set out the minimum capital requirement for banks. It introduced the calculation of risk-weighted assets (RWA), which was designed to embed a risk-sensitivity into the capital that a bank must hold. That is, the more a bank increases its risk, the more capital that it will need to hold to reflect that higher risk that it faces. It also stipulated a minimum capital to risk-weighted assets ratio of at least 8%.

To undertake these two calculations, assets (loans) are grouped into categories according to their risk and RWA is to calculated as the multiplication of a specified risk-weight and the loan size. Table 2.1 below shows an example of this calculation for 5 different loans. The total exposure is \$600, the RWA is calculated as \$370, and equation 2.1 shows that the minimum capital required is \$29.60. Risk sensitivity is embedded in RWA calculations, because a \$1 increase in a "loan to a AAA rated bank" will cause RWA to increase by only 20 cents, where as a \$1 increase in a "loan to a B+ rated corporate" will cause RWA to increase by \$1.50.

$$Minumum\ Capital = 8\% \times \$370 = \$29.60 \tag{2.1}$$

Asset	Risk Weight	Loan Amount	Risk-Weighted Asset
Loan to a AAA Rated Government	0%	\$100	\$0
Loan to a AAA Rated Bank	20%	\$100	\$20
Residential Mortgage	50%	\$100	\$50
Loan to a BBB Rated Corporate	100%	\$100	\$100
Loan to a B+ Rated Corporate	150%	\$100	\$150
Total		**\$600**	**\$370**

Table 2.1: Example Calculation of Risk Weighted Assets

Several amendments were made to "Basel I" over time, including consideration of capital adequacy for market risk in the 1996 "Market Risk Amendment (Bank of International Settlements, 2014).

2.3.2 Basel II

In 2004 the BCBS released the "Revised Capital Framework", which is widely known as "Basel II". While it does superscede "Basel I", it seeks to enhance and build upon it rather than starting completely afresh. It does this by introducing the concept of "three pillars" to the risk measurement and management of a bank's capital adequacy, namely:

1. Pillar I – minimum capital requirements for credit, operational and market risks, which strengthened rules from Basel I for calculating RWA.

2. Pillar II – supervisory review of an institution's internal capital adequacy and assessment process. It also provides a framework for dealing with other risks a bank faces, including (for example): systematic risk; liquidity risk; legal risk; pension risk; concentration risk; and strategic risk.

3. Pillar III – stipulating public disclosure of key risk and financial metrics so that market discipline can help encourage sound banking practices.

This thesis focusses on credit risk (the largest risk that a bank takes on), so we will now concentrate on methods for calculating credit risk-weighted assets for use in calculating capital for a bank as per first pillar of "Basel II". There are two alternate approaches:

- Standardised – this method largely follows the "Basel I" calculation, and stipulates exactly what risk weights are to be applied to loans for calculating risk-weighted assets.

- Internal Ratings Based (IRB) – introduced under "Basel II", it sets out the conditions a bank must satisfy in order for their local regulator to approve them to estimate risk-weighted assets using their own preferred statistical methods and their own data. IRB accredited banks can either adopt the Foundation (F-IRB) or Advanced (A-IRB) approach.

Banks anticipate a certain level of credit losses (known as "expected loss") which is seen as a cost of doing business. Estimation of expected loss helps inform loan pricing and loan provisioning. While clearly portfolios with different risk will likely encounter different levels of expected loss, the estimation and prediction of such losses forms the main basis for the application of statistical and analytical methods in credit risk modelling.

However this loss experience will likely vary from year to year, depending on the number and the severity of the losses encountered. Figure 2.1 below is a reproduction of figure 1 from an explanatory note form the Basel Committee for Banking Supervision (Bank of International Settlements, 2005). It shows a stylisation of annual credit losses over time that for each year deviate around the central tendency ("expected loss") of the probability distribution function of credit losses. Deviations beyond expected loss are called "unexpected loss" and banks are required to hold regulatory capital to absorb such losses.

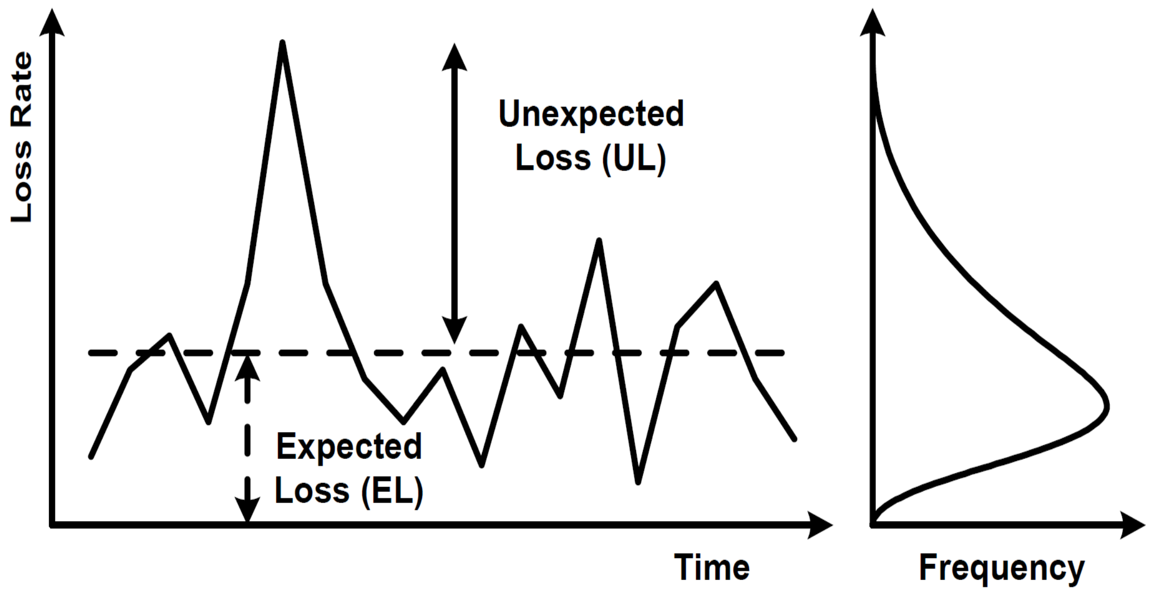


Figure 2.1: Stylisation of Annual Credit Losses. Source: Bank of International Settlements (2005)

Resources that are stored as capital are unable to be re-deployed by the bank for profit generating activities, so there is a clear incentive to strike a balance between the capital they hold verses the ability for the bank to withstand a period of large losses. This trade-off is struck by the selection of a very high confidence level (typically 99.9%) from the credit loss distribution and capital is held to absorb losses up to this quantile. This quantile is termed theoretically in the literature as the "Value at Risk" or "VaR" (see for example Jorion (2001)), and represents the largest loss with 99.9% confidence suffered by the credit portfolio over a 1-year period. Losses beyond this quantile (also called the VaR point) will lead to the bank's insolvency. The relationship between "expected loss", "unexpected loss" and the VaR point for the distribution of credit losses are displayed graphically in figure 2.2 below:

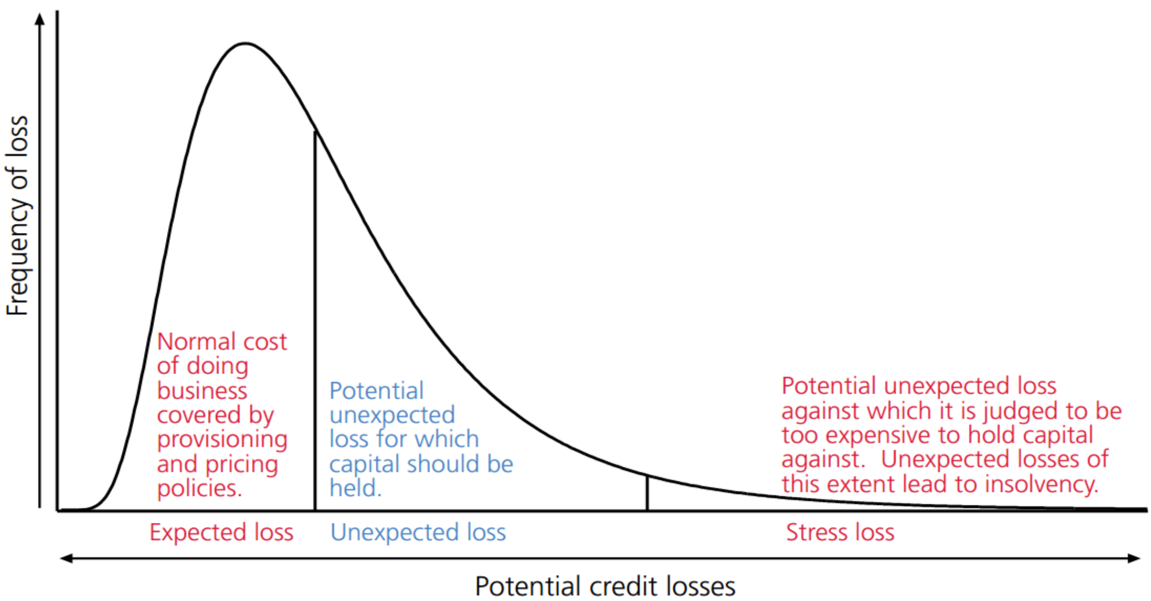


Figure 2.2: Credit Loss Distribution. Source: Yeh et al. (2005)

Basel II supplied a set of formulas detailing how banks are to estimate "expected loss" and "unexpected loss". Banks accredited under the advanced internal ratings based approach (A-IRB) are allowed to estimate these quantities using their own statistical methods and data to estimate credit risk components – the key inputs to the regulatory formulas. Apostolik et al. (2009) defines these risk components as follows:

- "Probability of Default" (PD) – The probability a customer will fail to make full and timely repayments;
- "Exposure At Default" (EAD) – The expected value of the loan at the time of default;
- "Loss Given Default" (LGD) – The amount of the loss as a percentage of EAD;
- "Maturity (M)" – The remaining life of the loan, in years.

The first moment of the credit loss distribution ("expected loss") is calcuated as the product of these three values (Apostolik et al., 2009):

$$EL = \underbrace{PD}_{\text{frequency}} \times \underbrace{EAD \times LGD}_{\text{severity}} \tag{2.2}$$

If one were to assume that individual credit defaults and losses within a portfolio were independent, then the distribution of the sum of these losses as the portfolio size goes to infinity would be well approximated by the central limit theorem (Vasicek, 2002). However credit defaults and losses within a portfolio are correlated meaning that the distribution of credit losses is characterised by a skewed distribution as demonstrated in figures 2.1 and 2.2 above.

To cater for this correlation, the mathematical construct used in Basel II has its initial groundings from the Merton model (Merton, 1974), which defines that a firm defaults if its asset value falls below some critical value (typically but not necessarily related to its debt and liabilities). This work was extended by Vasicek (2002) to construct the Asymptotic Single Risk Factor Model, using the following two assumptions:

- the banks holds a large homogeneous portfolio of loans with no excessive concentrations to any particular loan;
- each loan, conditional on a given realisation of the common (latent) systematic factor X, is independent to each other loan. Each loan is thus related to each other loan via a common correlation $(-\sqrt{R})$ to the single common latent factor (X).

Heitfield and Barger (2003) provide a simple explanation of the Vasicek (2002) model. Obliger i defaults if the (normalised) return on the firm's assets (signified by Y_i) falls below the standardised default threshold γ_i:

$$Y_i = \epsilon_i\sqrt{1-R} - X\sqrt{R} \leq \gamma_i = \Phi^{-1}(PD_i) \tag{2.3}$$

Where:

- $X := N(0,1)$ common (latent) systematic factor, independent of ϵ_i
- $\epsilon_i := N(0,1)$ idiosyncratic systematic factor, independent of X and ϵ_j, $j \neq i$
- $R :=$ common asset correlation
- $\gamma_i :=$default threshold

- PD_i :=probability of default for obliger i
- $\Phi^{-1}(.)$:= the inverse of the standard normal cumulative density function

While several authors beyond Merton (1974) and Vasicek (2002) (see for example Rutkowski and Tarca (2014)) discuss the Brownian motion construct underlying equation 2.3, it is worthwhile discussing its intuition. Due to the distributional and independence assumptions of X and ϵ_i, the following results are obtained:

- $Covar[Y_i, X] = -\sqrt{R}$ (each obliger is correlated to the single common latent factor)
- $Covar[Y_i, Y_i] = 1$
- $E[Y_i] = 0$
- $Covar[Y_i, Y_j] = 0$, $i \neq j$ (obligers are independent to each other)

Using equation 2.2, the conditional expected loss function for exposure i for a given realisation of X and for an $EAD = \$1$ is:

$$\begin{aligned}
c_i(x) &= P[Y_i \leq \gamma_i | X = x].LGD_i \\
&= P[\epsilon_i\sqrt{1-R} - X\sqrt{R} \leq \Phi^{-1}(PD_i)|X = x] \times LGD_i \\
&= P\left[\epsilon_i \leq \frac{\Phi^{-1}(PD_i) + X\sqrt{R}}{\sqrt{1-\rho}} \middle| X = x\right] \times LGD_i \\
&= \Phi\left[\frac{\Phi^{-1}(PD_i) + x\sqrt{R}}{\sqrt{1-R}}\right] \times LGD_i \\
c_i(\Phi^{-1}(0.999)) &= \Phi\left[\frac{\Phi^{-1}(PD_i) + \Phi^{-1}(0.999)\sqrt{R}}{\sqrt{1-R}}\right] \times LGD_i \qquad (2.4)
\end{aligned}$$

Equation 2.4 is the quantile that corresponds to the 99.9^{th} percentile of the credit loss distribution. That is, its the Value at Risk as discussed in figure 2.2. We can obtain an expression for the “unexpected loss” by taking away the ”expected loss” as per equation 2.2 from the Value at Risk in equation 2.4:

$$UL = \Phi\left[\frac{\Phi^{-1}(PD_i) + \Phi^{-1}(0.999)\sqrt{R}}{\sqrt{1-R}}\right].LGD_i - PD_i.LGD_i \qquad (2.5)$$

Finally, we present the entire set of capital formulas as per the Basel II accord. Heitfield and Barger (2003) note that correlation is “hard wired”, in that its parametrisation is stipulated to be a function of probability of default and that banks cannot alter its functional form. The Basel II capital formula also makes an explicit adjustment for maturities, which again is prescribed to be a function of probability of default.

Correlation (for corporates, banks and sovereign counterparties).

$$R = 0.12\left[\frac{1-e^{-50PD}}{1-e^{50}}\right] + 0.24\left[1 - \frac{1-e^{-50PD}}{1-e^{50}}\right] \qquad (2.6)$$

Maturity Adjustment

$$b = [0.11852 - 0.05478.ln(PD)]^2 \qquad (2.7)$$

Capital Requirement (we derived the portion in the first square brackets in equation 2.5)

$$K = \left[LGD.\Phi \left(\sqrt{\frac{1}{1-R}} \Phi^{-1}(PD) + \sqrt{\frac{R}{1-R}} \Phi^{-1}(0.999) \right) - PD.LGD \right] \times \left[\frac{1 + (M - 2.5)b}{1 - 1.5b} \right] \tag{2.8}$$

Risk-Weighted Assets

$$RWA = K.12.5.EAD \tag{2.9}$$

Before we leave this section, we quickly loop back and discuss how the Basel II capital framework aligns with Lam (2003)'s seven dimensions of risk quantification (see section 2.2.3). Table 2.2 below shows that the Basel II parameters cover off all seven dimensions.

Risk Dimension	**Lam's Descriptoin**	**Basel II Parameter**
Probability	how likely is the event to occur?	PD
Exposure	what does the bank stand to lose?	EAD
Severity	what loss is likely to be suffered?	LGD
Volatility	how uncertain is the future?	UL
Time Horizon	how long will the bank be exposed risk?	M
Correlatoin	how are the individual risks related?	R
Capital	safety margin to cover unexpected losses?	K

Table 2.2: Comparison of Seven Risk Qualification Dimensions of Lam (2003) to the Basel II Risk Parameters

2.3.3 Basel III

Finally, in 2010 the BCBS released "Basel III", which revises and strengthens the three pillars of "Basel II". Briefly, the three pillars will remain but it introduces several new ratios (such as a capital conservation buffer, countercyclical capital buffer, leverage ratio, and minimum liquidity ratio). These, and many other new measures, are to be implemented over a 5 to 10 year phase-in period beginning from 2013 onwards. As "Basel III" relates to accounting and financial ratios and not statistical estimation of risk, we will not consider it any further in this thesis.

2.4 The Global Credit Data (GCD)

To help counter the problem of data paucity for own-estimation of EAD and LGD, the Global Credit Data (GCD) Consortium was created as a credit data-pooling initiative. The initiative grew from approximately 10 founding European banks in 2004 to approximately 47 banks from Europe, Africa, Asia, North America and Australia. The entire database consists of over 100,000 defaulted facilities representing more than €200b across all the non-retail Basel Asset Classes spanning 20 year period up to 2015 (Global Credit Data, 2015).

Member banks who cede their own internal credit data to one more Basel Asset Classes are availed, via the principle of reciprocity, to receive access to the pooled member data in the asset classes they ceded to. This "give-to-get" philosophy allows each member bank to selectively participate in those Basel Asset Classes of most interest to them. All the data is anonomised, so neither customers nor contributing banks can be distinguished in the data returned to members. Member banks cede new or updated data for defaulted facilities together with a suite of related covariates twice annually (Global Credit Data, 2015).

This thesis will use data of one member bank's view of the GCD data to train a predictive model estimating exposure at default.

Chapter 3

Literature Review

3.1 Regulatory Setting

The Basel Accords, published in various iterations since 1988 by the Bank for International Settlements, outline the requirements Banks must satisfy to be accredited to use the Advanced Internal Ratings Based (A-IRB) method to calculate regulatory capital for credit risk. These requirements include appropriate quantitative estimation of credit risk components: Probability of Default (PD); Loss Given Default (LGD); and Exposure at Default (EAD) (BIS, 2006).

The majority of both academic and practitioner attention to date has focussed on the more readily estimable Probability of Default, and to a lesser extent Loss Given Default, with relatively less attention paid to Exposure at Default (see Jacobs and Bag (2011), Financial Conduct Authority (2007), and Brown (2011)). This may in-part be due to the contingent nature and resulting data paucity of EAD which is estimated using only defaulted obligors, where as PD uses both defaulted and non-defaulted obligors.

Despite this, quantitative EAD estimation is becoming a more active research area and is receiving larger focus by regulators and in industry, and this section briefly covers some of the related literature. We start with the definition from the Bank for International Settlements, who define EAD as:

> ***"... the expected gross exposure of the facility upon default of the obligor"*** (BIS, 2006).

3.2 Terminology

For accounting purposes, a loan granted to a customer is split into two parts. The balance drawn by the customer (that is, funds the customer has already withdrawn from their loan account) which is a receivable the bank is owed by the customer and is thus an asset on the bank's balance sheet. This is known as the "on balance sheet" exposure. The remaining limit that is not yet drawn down (that is, funds the customer is still entitled to withdraw later on but has not yet done so) is not yet a receivable the bank is owed by the customer. This is known as the "off balance sheet" exposure. As a quick concrete example, suppose a bank grants a loan to a customer for \$100 but the customer choose to draw only \$15. Here, the "on balance sheet" amount is \$15 and the "off balance sheet" amount is (\$100 - \$15) = \$85.

The Basel Accords introduce the concept of a Credit Conversion Factor (CCF), defined as the proportion of the remaining limit (ie: the "off balance sheet" exposure) that will likely be drawn-down in the event of a default (see paragraph paragraph 310 in BIS (2006)). Thus the Exposure at Default (EAD) can be estimated by summing the drawn balance plus the CCF multiplied by the remaining limit. This is demonstrated formulaically in figure 3.1 below:

$$EAD = \underbrace{B_t}_{\text{on balance sheet}} + \underbrace{\overbrace{CCF}^{\text{Proportion of remaining limit}} \times \overbrace{(L_t - B_t)}^{\text{remaining limit}}}_{\text{off balance sheet}} \tag{3.1}$$

This same relationship between EAD and CCF is graphically in 3.1 below:

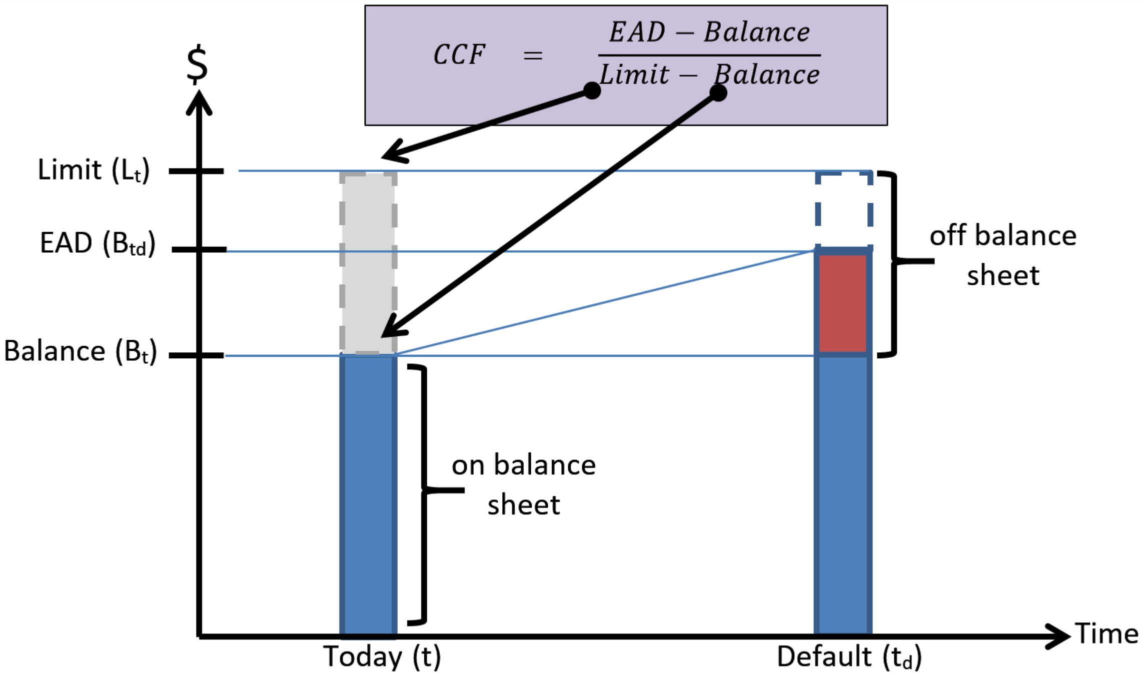


Figure 3.1: Stylisation of Credit Conversion Factor Calculation

The motivation for CCF stems from the hypothesis that a facility that defaults sometime in the future would be expected to have an EAD of at least the balance today, plus some "conversion" of the unused limit (Taplin et al., 2007). Note that A-IRB accredited banks are not required to use CCF in order to estimate EAD (Taplin et al., 2007). In addition, regulators in the UK (Financial Conduct Authority, 2014) and in Continental Europe (Committee of European Banking Supervisors, 2006) will allow, in appropriate circumstances, the use of own EAD estimates that do not rely on conversion factors. However the use of conversion factors for the Standardised Approach (BIS, 2006) means that the substantial majority of EAD literature to date focusses on its estimation.

At the very least, a key benefit of directly modelling EAD rather than indirectly via a conversion factor would be avoidance of strong non-unimodality and extreme values in empirical distributions of CCF. For example we reproduce in figure 3.2 below figure 1 from Brown (2011). The extreme values, as well (in this case) strong bi-modality are clearly visible.

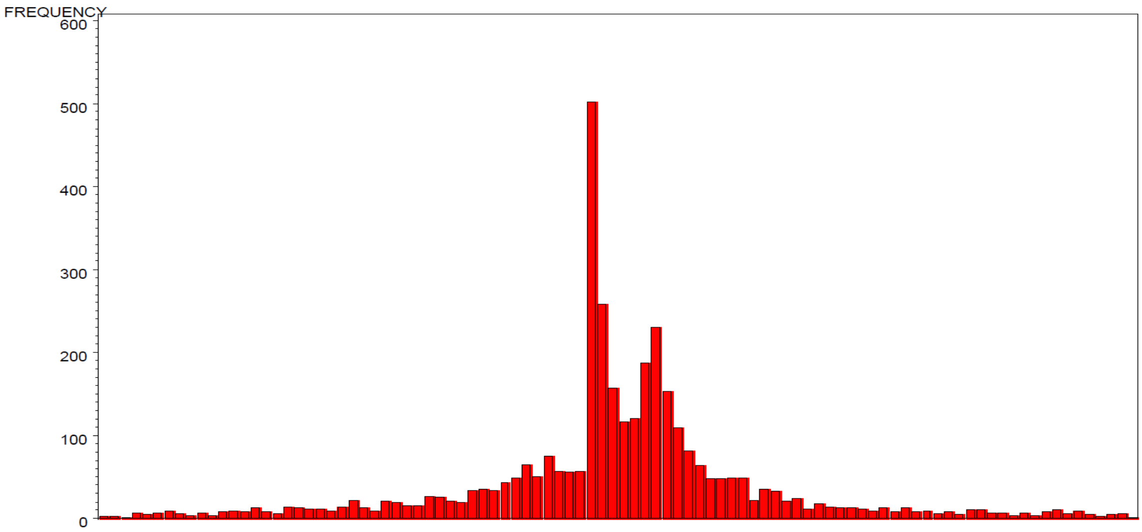


Figure 3.2: Distribution of or Credit Cards CCF. Source: Figure 1 From Brown (2011)

There also appears to be no universally agreed terminology for conversion factors, with some authors preferring to use the term "Loan Equivalent Factor" (LEQ) for what the Basel Accord defines as CCF. We prefer to use CCF to be inkeeping with Basel Accord, but highlight where we encounter inconsistency in terminology and provide details of the underlying calculation for clarity.

The definition of EAD embodies a facility level estimation, which may be why several authors choose to focus only one or a few facility types. Accordingly, this project focusses on estimating EAD for **"Continent Credit Lines"** or **"Revolving Facilities"**, which Investopedia (2015) defines as:

> ***"An arrangement between a financial institution, usually a bank, and a customer that establishes a maximum loan balance that the bank will permit the borrower to maintain. The borrower can draw down on the line of credit at any time, as long as he or she does not exceed the maximum set in the agreement"*.**

3.3 Literature Review

While the area of quantitative EAD estimation is a developing research area, there are several interesting and informative contributions by authors to date. A key recurring theme in almost all the literature surveyed is the preferred focus on modelling CCF (or some other conversion factor) as the response variable rather than choosing to model EAD as the response variable. There is also no clear agreement between different authors in how to deal with empirical values of CCF outside the sensible range of [0,1]. Authors typically also find that CCF are strongly bimodal with modes at 0 and 1. Common approaches to this problem are to either Winsorize or truncate the data.

A good place to start is with Taplin et al. (2007) who use the term CCF as per our definition in figure 3.1 above. They strongly argue that CCF is not universally appropriate, and demonstrate clear deficiencies both numerically for calculating CCF when balances are nearly fully drawn as well as conceptually for estimates of CCF greater than one.

The authors show that for $0 < CCF < 1$ implies that EAD cannot be greater than the limit, stating that this is both an empirically unreasonable restriction as many datasets (including the one they use in their paper) which have a substantial number of empirical CCFs outside this range. It also does not make intuitive sense, as an EAD greater than the limit is typically a condition of default.

The authors also show that $CCF > 1$ has several undesirable properties. The first is that for accounts that have $Balance > Limit$, EAD will be estimated to be smaller than both balance and the limit. For example, using equation 3.1:

- Suppose: $Balance = 1.5$, $Limit = 1$ and $CCF = 1.5$
- Then: $EAD = 1.5 + 1.5 \times (1 - 1.5) = 0.75$, which is smaller than the balance and the limit.

The second undesirable property is that as the balance increases, the estimate EAD (and thus the risk) decreases. For example, again using equation 3.1, suppose we have two loans, one undrawn with a balance of 0 and one fully drawn with the balance equal to the limit:

- Loan A is undrawn : $Balance_A = 0$, $Limit_A = 1$ and $CCF = 1.5$
- Loan B is fully drawn: $Balance_B = 1$, $Limit_B = 1$ and $CCF = 1.5$
- Then: $EAD_A = 1.5$ and $EAD_B = 1$, which suggests that the fully drawn loan is lower risk.

The third undesirable property is the numerical instability when the denominator in the CCF calculation is close to zero, and that CCF becomes undefined when the denominator equals zero. These situations occur when the balance and the limit are either approximately equal or equal respectively.

The paper suggest two alternate model parametrisations, both of which model EAD/Limit, but retain balance as the only explanatory covariate in the models. They fit their models using business credit card data from the Bank of Western Australia (BankWest, which is now wholly owned by Commonwealth Bank of Australia).

A critique by Moral (2006) from the Central Bank of Spain (Banco de España) provides a review of several methodologies. The author begins by defining two factors for estimating EAD: an alternate definition Credit Conversion Factor (CCF) and a definition of Loan Equivalent (LEQ) that aligns the Basel Accord definition of CCF as per figure 3.1 above.

The author goes on to critically observe that because CCF estimation typically adopt a regression-based approach that the inherently assumed symmetric loss function may not appropriately penalise uncertainty. They also re-iterate the common credit risk terminology regarding observational periods of: "Fixed Time Horizon", "Cohort Approach", and "Variable Time Horizon". They discusses the advantages and disadvantages of each, and state that banks typical use of the first two methods may not account for all relevant information due to their conventional focus on a reference date. Moral suggests that, at a minimum, the following risk drivers are considered for modelling: facility type; covenants; limit; balance; utilisation; time to default; rating class; facility status; and macroeconomic indicators.

With regard to truncation, Moral states "*[t]he common practice of [truncating] the realised CCF factors to [0, 1] is not justified and, in general, it is not possible to conclude ex ante if the associated CCF estimates are biased in a conservative manner*". We have seen several papers which we mention below that either truncate or Winsorize CCF values outside [0,1], and agree strongly that such treatment in inappropriate without also estimating the degree of bias induced.

A paper by Jacobs (2010) begins by setting out clearly the mathematical equations for Loan Equivalent (LEQ), Credit Conversion Factor (CCF) and EAD Factor (EADF). Again, the author's definition in their paper for LEQ is the one that actually aligns with the Basel Accord of CCF as per figure 3.1 above.

The paper details results of the modelling CCF using GLM regression techniques and finds that credit rating and utilisation have the strongest predictive power, while other important factors include: leverage; liquidity; debt cushion; along with mild evidence of counter-cyclicality. The author also discusses the lack of empirical evidence that CCF are bounded between 0% and 100%, and describes as "ad-hoc" the methods (such as Winsorizing, capping, and flooring) that are typically employed for dealing with values outside this range and suggests more "enlightened" methodologies could include robust statistics and quantile regression. Models are estimated using Moody's Ultimate Recovery Database (MURD) and make the key assumption that balance, limit and EAD are identically equal at the point of default.

Araten and Jacobs (2001) investigate a dataset of 408 defaulted revolving credit lines from JPMorganChase & Co. They define LEQ as "*...the portion of a credit line's undrawn commitment that is likely to be drawn down by the borrower in the event of default*", which is consistent with our definition of CCF in figure 3.1. For their analysis, they truncate CCF values at 0, and find that the resulting distribution is strongly bi-modal, with modes at 0% and 100%. They empirically find that the two main drivers of CCF is credit quality and time to maturity, but go on to opine that other common sense covariates could include: tenor of commitment; nature of obligor's business; access to commercial paper market; usage; size; commitment level; facility type; and borrower domicile. The final reported model is:

$$CCF = 48.36 - 3.49 \times FacityGrade + 10.87 \times TimeToDefault \tag{3.2}$$

Jimènez et al. (2009) obtain a census of 20 years of data for corporate revolving lines from the Spanish Credit Registry. This data provides a unique view for EAD estimation, as it includes both defaulted and non-defaulted facilities where all other studies (that we know of) estimate models only on defaulted facilities. Rather than using CCF or EAD, their analysis uses utilisation as the response (but their paper presents results in terms of LEQ, and again this aligns with the Basel Accord Definition of CCF):

$$Utilisation = \frac{Balance}{Limit} \tag{3.3}$$

Their model is able to detect utilisation increases up to 5 years prior to default, depending on whether firms have previously defaulted as an explanatory covariate in their regression. In addition, they also find that commitment size, collateralisation and maturity are key drivers of utilisation. Figure 3.3 below reproduces figure 2 from Jimènez et al. (2009) and shows that years to default is a very powerful covariate.

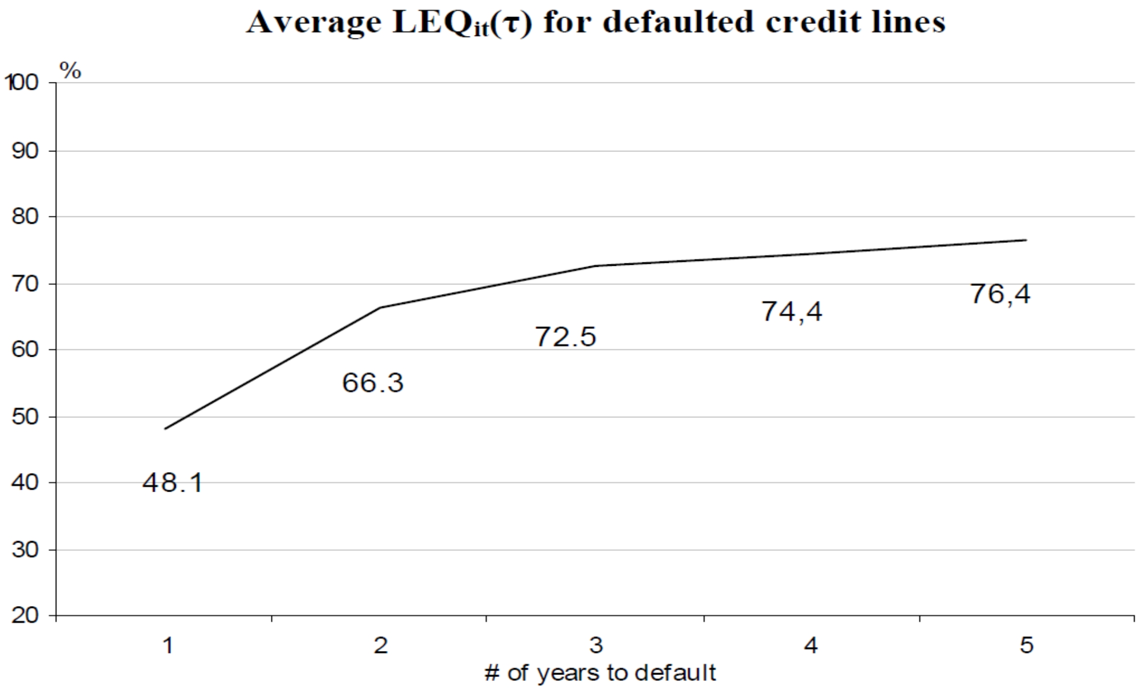


Figure 3.3: Average CCF for Defaulted Credit Lines. Source: Figure 2 from Jimènez et al. (2009)
.

Brown (2011) begins by saying that "*to date very little model development and validation has been reported on the estimation of EAD*". The paper estimates several competing models for CCF, which is defined in the paper identically as our definition of CCF in 3.1 above. Several models are built using credit data from a UK financial institution from 2001 to 2004, and include an OLS regression, and various logit and cumulative logit models that rely on partitioning CCF. The identified statistically significant variables include: utilisation; limit; limit minus balance; time-to-default; credit rating; average days delinquent in past six months. Another retail study by Qi (2009) focussing instead on United States retail credit card data finds that "*...borrowers are more active than lenders in the 'race to default'*". Finally, a study by Agarwal et al. (2006) on retail home equity lines of credit (HELOC) in the United States also find that borrowers with deteriorating credit quality increase their utilization.

A presentation by Leow and Crook (2013) at the 2013 Credit Scoring & Credit Control Conference provides a clear explanation of how risk components fit into regulatory capital calculation as well as noting that EAD is routinely estimated using LEQ and include its mathematical definition. They define LEQ as per the Basel Accord's definition of CCF (as per our definition in figure 3.1), but also define an alternate "CCF" as the ratio of EAD to balance (ie: the scaling factor applied to the observed balance today to obtain an EAD in the future). Their final model uses neither "CCF" (such defined by the authors) nor CCF from the Base Accord, but instead estimates a two-step mixture model to directly estimate customer-level EAD for UK retail credit card data as per equation 3.4 below. The author's model employs a repeated measures design using a random effect to to cater for hierarchical correlation and is thus general enough to not only estimate EAD but also loan balance at any time up to default.

The final estimate of EAD relies on 3 models: (1) a predicted probability that balance is grater than limit at the time of default; (2) a panel model with subject-level random-effect to estimate limit at default; and (3) a panel model with subject-level random-effect for the balance at default.

$$\tilde{B}_{it} = [P(B_{it} \geq L_{it}) \times \hat{L}_{it}] + [(1 - P(B_{it} \geq L_{it})) \times \hat{B}_{it}] \tag{3.4}$$

Where:
$\tilde{B}_{it}$ = model estimate for EAD for i$^{\text{th}}$ account
$\hat{B}_{it}$ = balance at time t for i$^{\text{th}}$ account
$\hat{L}_{it}$ = limit at time t for the i$^{\text{th}}$ account

The presentation contains a histogram of $\frac{Balance}{Limt}$ on slide 7, and we supply a copy in figure 3.4 below. While the ratio is trimmed at values of $\frac{Balance}{Limt} < 3$ and is not used for their model, our data (see figure 3.5) for revolving credit lines to large corporates bares a striking resemblance.

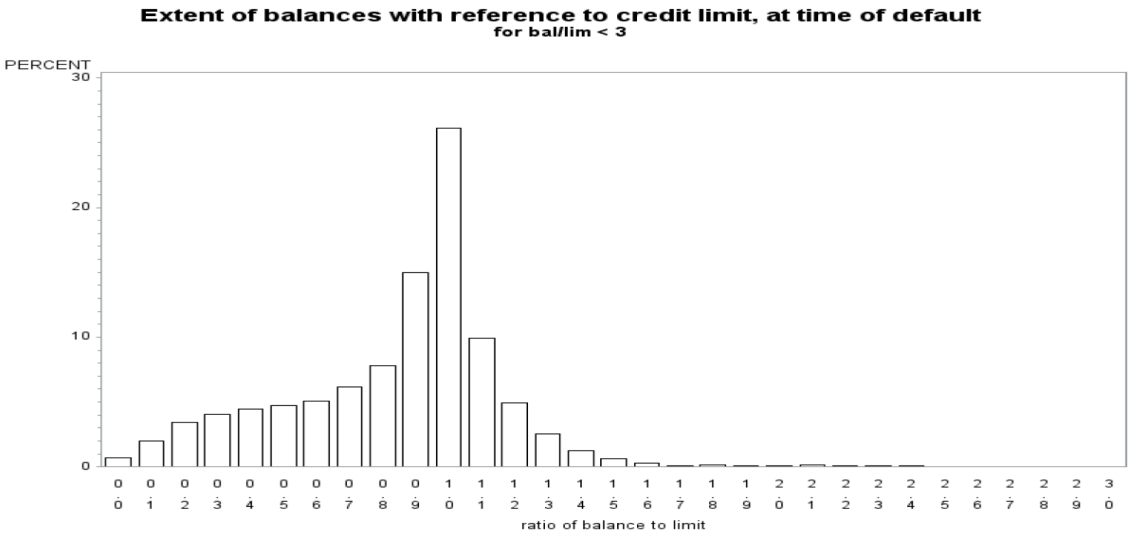


Figure 3.4: $\frac{Balance}{Limt}$ For Credit Cards. Source: Slide 7 Leow and Crook (2013)

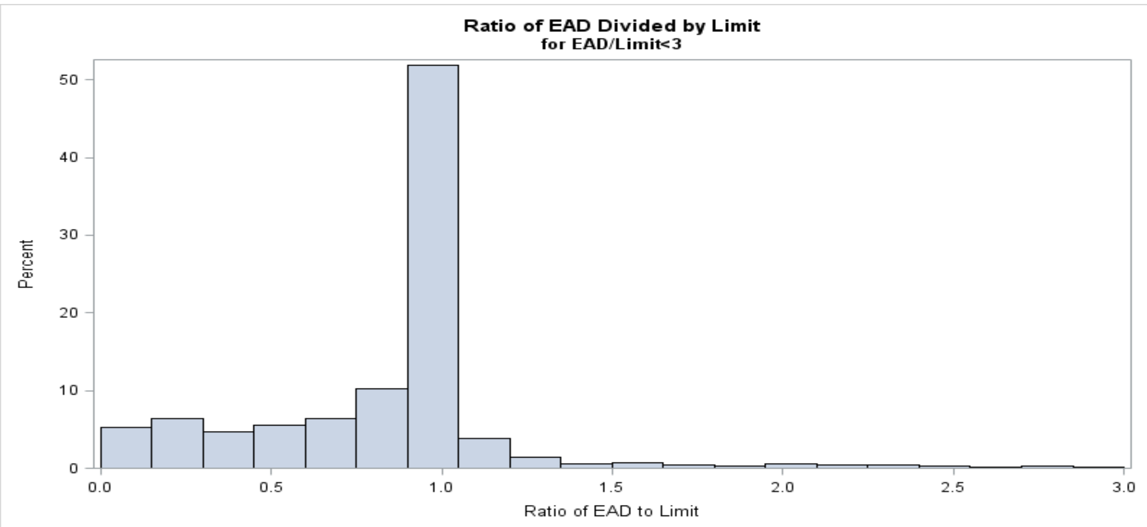


Figure 3.5: $\frac{Balance}{Limt}$ for GCD Large Corporate Revolving Credit Lines

A presentation by Tong et al. (2015) at the 2015 Credit Scoring & Credit Control Conference compares several models for estimating EAD for retail credit cards. The presentation shows that models that have CCF as the response variable perform better for accounts with usage below 90% while models that have CCF as the response variable perform better for usage above 90%. In order to compare their models, the authors truncate their sample data to have values of CCF in the range [0,1].

A Masters thesis by Mantel (2012) models CCF for facilities to large corporates using the GCD data. The analysis focusses on four facility types (revolvers, term loans, letters of credit, and working capital) for facilities with values of CCF less than 80%. For all four facility types the selected covariates in the OLS regression are size, utilisation and their interaction, however each model obtain different numerical parameter estimates. The author also finds (similar to Jacobs (2010)) some evidence of cylicality in CCF estimates.

For an alternate view, some authors (for example: Araten and Jacobs (2001), Loukoianova et al. (2013), Jacobs and Bag (2012)) recognise that obligors with a contingent credit line hold a put option to draw funds up to a specified limit from the bank, and in particular Jacobs and Bag (2012) outline a framework to price such an option. Sufi (2009) recognises this as a moral hazard problem, which is mitigated by banks posing strict covenants. Jacobs and Bag (2011) state that contingent facilities typically contain material adverse change (MAC) clause, which in effect means the factliy is unconditionally cancellable by the issuing institution. Finally, Witzany (2011) models EAD using default intensities by introducing a default density function g(s), and g(s)$\triangle$s is the probability that default happens during the time interval[s, s + $\triangle$s].

3.4 Proposed Model

Given the identified shortcomings of modelling exposure at default indirectly via CCF (both conceptually and numerically) our model will be designed directly estimate EAD. We will train our model using loss data from the Global Credit Data (GCD) consortium focussing on revolving facilitates issued to large corporates. With time constraints of only 15 weeks to compete this thesis, we will focus on building one model, and leave testing our chosen model against other competing models as an element of future research.

Credit loss data for large corporate bank loans is notoriously difficult to assemble, given that this type of lending by its nature is low risk and thus results in low levels of defaults and losses. It is for this precise reason that the 47 member banks of the GCD consortium compile and (anonymously) share their respective loss data to provide the basis for a critical mass of data to statistically estimate EAD. It is also for this same reason (of data paucity) that we will deliberate choose to use all available large corporate data for revolving facilities rather than retaining a holdout dataset to test our final chosen model against. We will however validate our model using a non-parametric bootstrap cross-validation technique to help give some comfort that the parameter estimates are stable. We also suggest a practical process to potentially validate the model on GCD consortium data that becomes available in the future, but we also leave this as a suggestion for future research.

Our model could be implemented by a large internationally active bank that advances revolving facilities to large corporate counterparties, and its estimates used for risk management purposes such as: pricing; provisioning; limit management; economic capital; stress testing; and regulatory capital (subject to the bank's regulator).

Chapter 4

Statistical Modelling

4.1 Data, Filtering and Sampling

4.1.1 Original Data

The data used in this thesis is that of one member bank of the GCD consortium, containing approximately 10,000 defaulted faculties. Due to the extremely sensitive nature of this data, we are only allowed to display some brief summary statistics of both the unfiltered and filtered data. Table 4.1 shows the reference dataset by facility type before the application of filters. The results show revolvers are the most common facility type.

		Exposure at Default (€Millions)					
Facility Type	N	Sum	Min	P25	Median	P75	Max
Revolver	3,415	19,933	0.0000	0.0193	0.5720	4.1667	573.2030
Term Loan	3,333	25,340	0.0000	0.1534	1.1345	5.5544	475.5688
Other	2,797	8,094	0.0000	0.0048	0.1020	0.9548	752.8469
OVERALL	**9,545**	**53,366**	**0.0000**	**0.0297**	**0.4383**	**3.5312**	**752.8469**

Table 4.1: Summary of Reference Data Set by Facility Type

We focus our analysis on revolvers, and table 4.2 shows the distribution by geography.

		Exposure at Default (€Millions)					
Geography	N	Sum	Min	P25	Median	P75	Max
Europe	2,026	9,037	0.0000	0.0050	0.3947	2.9064	329.9214
North America	693	6,791	0.0000	0.0347	0.8087	9.7087	573.2030
Asia	349	1,193	0.0000	0.1891	0.7648	3.8265	100.4712
Australia/NZ	183	1,257	0.0000	0.0247	0.2395	3.2744	95.7402
Other	164	1,655	0.0000	0.1709	2.3137	7.1778	184.9302
OVERALL	**3,415**	**19,933**	**0.0000**	**0.0193**	**0.5720**	**4.1667**	**573.2030**

Table 4.2: Summary of Revolvers in Reference Data Set by Geography

4.1.2 Data Filtering and Preprocessing

Araten and Jacobs (2001) state that "*[t]he importance of carefully screening and cleaning data cannot be overemphasized*". With this in mind, we apply some brief filters as per table 4.3 that remove some observations. These relate to removal of facilities that are not to large corporates (for example: banks, sovereigns and specialised lending) as well as the removal of small limits at and small EAD.

Filter	Reason	Count	Percent (%)
Exclude	Not large Corporate	398	11.63
Exclude	Limit <€500	640	18.74
Exclude	EAD <€500	233	6.82
Include	Modelling Dataset	2,144	62.81
TOTAL		**3,415**	**100.00**

Table 4.3: Filters Applied

The final modelling dataset consists of 2,144 defaulted revolving facilities to large corporates that pass the identified filters.

4.1.3 The Decision to Not Retain a Holdout Sample

For construction of this model, we have decided not to retain a hold-out sample. Given that our methodology (as outlined in section 4.2) requires splitting the modelling dataset of 2,144 observations into two segments of 1,445 and 699 respectively, retaining a holdout sample may have lead to some explanatory covariates not having sufficient volume. For example, some categorical covariates have less than 100 observations in a given category, and thining this further may have caused such a covariate to not be selected in our model. This data paucity for large corporate credit risk modelling is a key challenge for such modelling problems and is indeed one of the motivating factors behind the formation of the GCD.

Further, while such measure are out-of-scope for our modelling exercise, there are alternate means that one could undertake to back test our model. GCD data is updated semi-annually and there are three sources from which additional data is added to the database. (Note that our dataset includes resolved defaults up to September 2014).

1. additional resolved defaults from ***prior to*** September 2014 that existing members provide beyond those already resolved defaults that they have previously ceded;
2. additional resolved defaults from ***after*** September 2014 that existing members will provide in the future.
3. additional resolved defaults from ***brand new*** members joining GCD for the first time.

In our experience, credit risk model development activities – which can include modelling, documentation, policy alignment and internal/external approval – can take up to 12 months to complete. This time frame would conceivably allow enough additional data from an updated GCD release as outlined above to undertake both in-time and out-of-time back testing for this model.

4.2 Methodology

4.2.1 Model Methodology

Several authors, (including Araten and Jacobs (2001), Jacobs (2010) , Qi (2009), Agarwal et al. (2006), Mantel (2012))recognise there are two separate counter-acting dynamics driving EAD in the lead-up to a customer default:

1. Banks will seek to manage (up or down) the available limit for a financially distressed customer
2. A financially distressed customer will seek to draw up the remaining funds to the limit to attempt to stave off insolvency

This "arm-wrestle" is depicted in a highly stylised example below in figure 4.1 below. At some time "t" prior to default, the loan balance B_t is below the limit L_t. For a customer that does eventually default at time "t_d", as this time approaches the customer begins to draw up their balance towards the limit while the bank, which has been monitoring its customer closely noticing the increase risk of default, manages the limit (typically) down. At the time of default the loan balance has increased from B_t to B_{td} while the limit has decreased from L_t to L_{td} .

Jacobs (2010) identifies this as an adverse selection problem in the context of revolving facilities, where if a borrower's fortunes improve their ability to pay-down or negotiate better pricing increases; however if a borrower's fortunes decline, there is an incentive to draw down the unused proportion of the commitment.

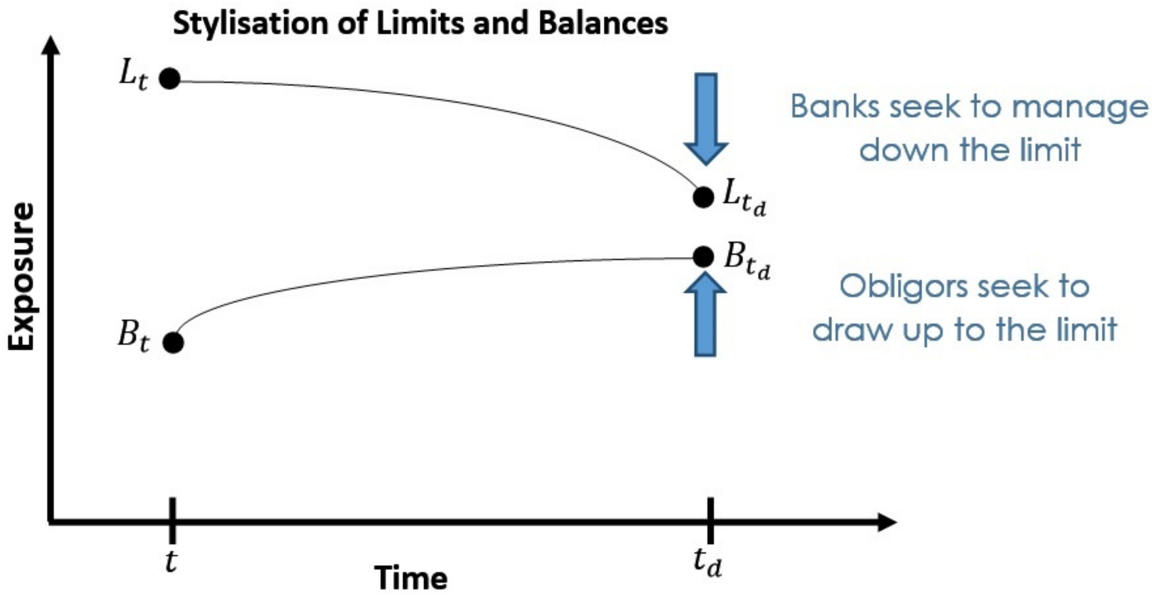


Figure 4.1: Stylisation of Limits and Balances for a Distressed Bank Customer

Where:

t = one year prior to time of default (t_d)

B_t =: balance at time t

L_t =: limit at time t

t_d =: time of default

B_{td} =: balance at default =: EAD

L_{td} =: limit at default

Note that this is a stylised example to highlight the joint dynamics of balance and limit, and it is not necessary that:

- Limits will always decrease in the lead-up to default;
- Balances will always increase in the lead-up to default; and
- Balances are always less than the limits.

In any event, there are clearly two dynamics at play so to model this duality we propose a two stage model. The first stage will capture whether or not the bank has decreased the limit, while the second stage will model the EAD conditional on whether the bank lowers the limit or not. The schematic in figure 4.2 below shows the construct of our model, detailing how the "stage one" logistic model is trained using all the observations while the "stage two" models are trained using observations for where there is a decrease in limits (finite mixture model) or observations for where the limit stays the same or increase (ordinary least squares model). Table 4.9 in the next section details how the model is applied for either testing or implementation purposes.

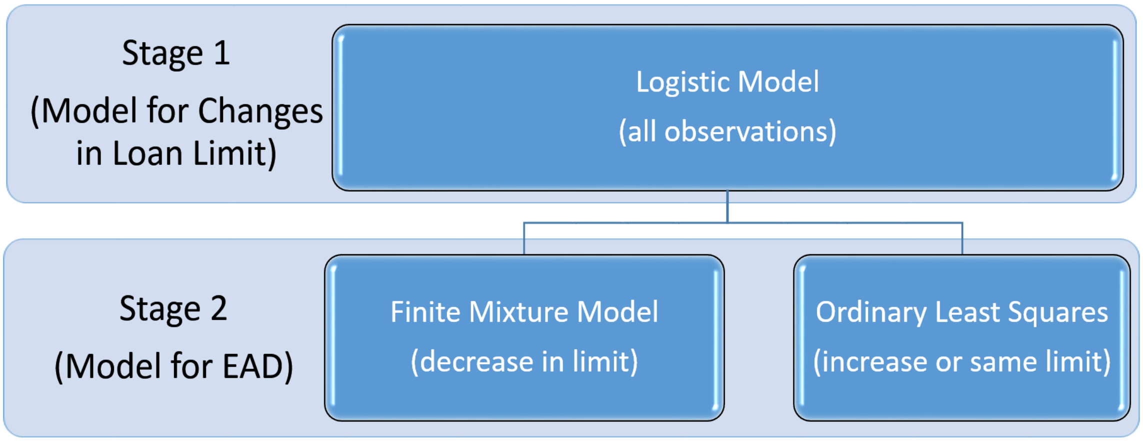


Figure 4.2: Schematic of Our Two-Stage Model

This model construct is similar to Leow and Crook (2013), however a key difference for our design is that it estimates separate models to explain the dynamics of the limits and balances. That is, the first stage focuses specifically on changes in the limit and the second stage focusses on conditional distribution of the balance given the change in limit. The Leow and Crook (2013) model estimates the probability of the balance being above the limit at default, and then estimates one model each for the balance and limit at default (see equation 3.4).

Because EAD is highly skewed (see table 4.2) we transform it using a logarithm of base 10. Let:

$$Y_i = log_{10}(B_{td,i}) \qquad i = 1, ..., 2,144 \tag{4.1}$$

We define the random that describes changes in the limit from time "t" at observation to time "t_d" at default, where for our data (t_d - t)=12 months. Let:

$$R_i = \begin{cases} 0 & \text{, if } \{ \ L_{td,i} \ / \ L_{t,i} < 1 \ \} \text{ , with probability } 1{-}p_i \\ 1 & \text{, if } \{ \ L_{td,i} \ / \ L_{t,i} \geq 1 \ \} \text{ , with probability } p_i \end{cases} \qquad i = 1, ..., 2,145 \tag{4.2}$$

We investigate a model of the joint behaviour of R_i (for changes in limit) and Y_i (for values of $log_{10}(EAD)$). Assume that there is a matrix $\mathbf{X}$ of covariates that are observed at time "t" along with these random variables that allows us to partition the joint density function conditional on $\mathbf{X}$ as follows:

$$f(y_i, r_i|\mathbf{X}) = f(y_i|r_i, \mathbf{X}).p(r_i|\mathbf{X}) \qquad i = 1, ..., 2,144 \tag{4.3}$$

We can formulate the conditional marginal distribution of $Y_i|\mathbf{X}$ by summing over the values of R_i:

$$f(y_i|\mathbf{X}) = \sum_{j=0}^{1} f(y_i|r_i = j, \mathbf{X}).p(r_i = j|\mathbf{X}) \qquad i = 1, ..., 2,144 \tag{4.4}$$

We can now formulate the conditional expectation of $Y_i|\mathbf{X}$ by integrating over the domain of Y_i and using equation 4.4:

$$\begin{aligned}
E[Y|\mathbf{X}] &= \int_0^\infty y.f(y_i|\mathbf{X}).dy \\
&= \int_0^\infty y.\sum_{r=0}^{1} f(y_i|r_i, \mathbf{X}).p(r_i|\mathbf{X}).dy \\
&= \int_0^\infty y.f(y_i|r_i = 0, \mathbf{X}).p(r_i = 0|\mathbf{X}).dy + \int_0^\infty y.f(y_i|r_i = 1, \mathbf{X}).p(r_i = 1|\mathbf{X}).dy \\
&= E[Y_i|R_i = 0, \mathbf{X}].P[R_i = 0|\mathbf{X}] + E[Y_i|R_i = 1, \mathbf{X}].P[R_i = 1|\mathbf{X}] \\
&\qquad i = 1, ..., 2,144
\end{aligned} \tag{4.5}$$

For the **stage one** model, we estimate $P[R_i = 1|\mathbf{X}]$ using a logistic regression (Nelder and Wedderburn, 1972):

$$R_i \sim Bernouli(p_i) \tag{4.6}$$

With:

$$G(p_i) = \underline{w}_i^T \underline{\alpha} \tag{4.7}$$

Thus:

$$E[R_i = 1|\underline{w}_i, \underline{\alpha}] = p_i = G^{-1}(\underline{w}_i^T \underline{\alpha}) \tag{4.8}$$

Where:

- G(.)=logit link function
- $\underline{w}_i$=tuple of explanatory covariates
- $\underline{\alpha}$=sensitivities to covariates

For the **stage two** model, we estimate $E[Y_i|R_i = 0, \mathbf{X}]$ and $E[Y_i|R_i = 1, \mathbf{X}]$. In the next section (see 4.2.2), we demonstrate that an ordinary lease squares model (OLS) and finite mixture model (FMM) respectively fit well these conditional distributions of $log_{10}(EAD)$. Let:

$$f_j(y_i|\underline{x}_{ij}, \underline{\theta}_{ij}) \qquad \text{be the pdf of the random variable } \{Y_i|R_i = j, \underline{x}_{ij}\} \tag{4.9}$$

Thus:

$$E[Y_{i,j}|\underline{x}_{ij}, \underline{\theta}_{ij}] = \mu_{ij} \tag{4.10}$$

Where:

- $\underline{x}_{ij}$=tuple of explanatory covariates for component j=0,1
- $\underline{\theta}_{ij}$=sensitivities to covariates

Note that the matrix of covariates **X** can be defined to be the tuples of covariate from the **stage one** and **stage two** models. We can also define $\underline{\phi}$ as the sensitivities to the covariates. Let:

- $\mathbf{X} = \{\underline{w}_i, \underline{x}_{i0}, \underline{x}_{i1}\}$; and
- $\underline{\phi} = \{\underline{\alpha}, \underline{\theta}_{i0}, \underline{\theta}_{i1}\}$

Thus with the introduction of these sensitivities, we can re-write equation 4.4 as follows:

$$f(y_i|\underline{X}, \underline{\phi}) = \sum_{j=0}^{1} f_j(b_i|\underline{x}_{ij}, \underline{\theta}_{ij}) \times P[R_i = j|\underline{w}_i, \underline{\alpha}] \tag{4.11}$$

Finally, after estimating $\hat{\underline{\phi}}$, predicted values of the mean for $Y_i|\mathbf{X}$ are given by:

$$E[\hat{Y}_i|\mathbf{X}, \hat{\underline{\phi}}] = (1 - E[R_i = 1|\underline{w}_i, \hat{\underline{\alpha}}]) \times E[\hat{Y}_{i,0}|\underline{x}_{i0}, \hat{\underline{\theta}}_{i0}] + E[R_i = 1|\underline{w}_i, \hat{\underline{\alpha}}] \times E[\hat{Y}_{i,1}|\underline{x}_{i1}, \hat{\underline{\theta}}_{i1}] \tag{4.12}$$

4.2.2 Distribution of Response Variables

In this section we explore response distributions for the changes in limit R_i , and the conditional distribution of EAD given changes in limit $Y_i|R_i = j$. Figure 4.3 below displays the distribution of changes in limits in the modelling dataset, showing that 33% of facilities had their limits decreased between time "t" and time "td", while 67% had limits remain the same or increased over the same period.

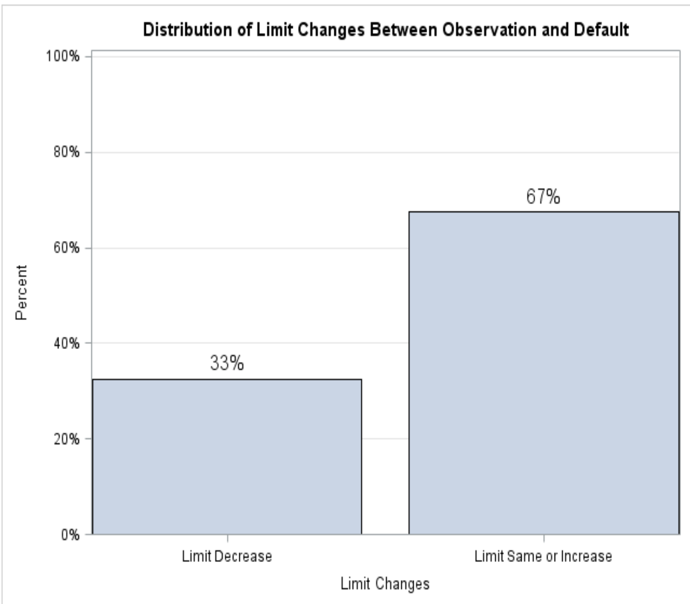
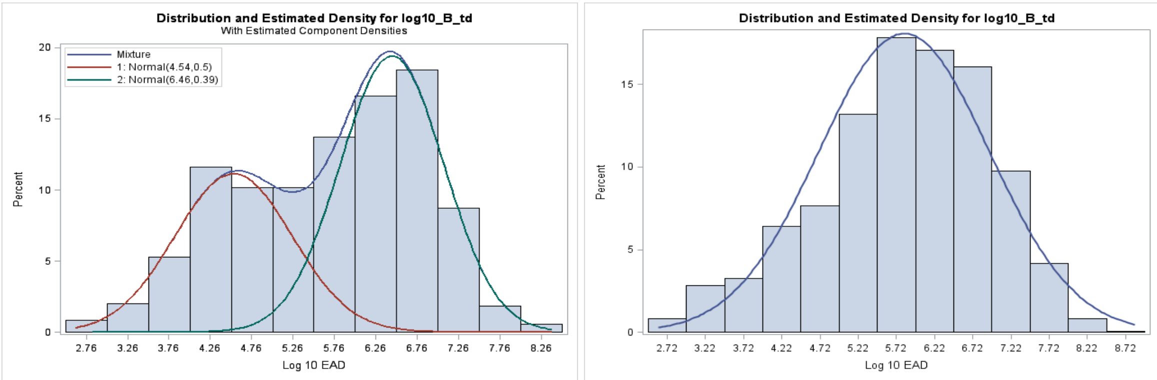


Figure 4.3: Distribution of Limit Changes Between Observation and Default

Figure 4.4(a) shows the distribution of $Y_i|R_i = 0$ for facilities that have a limit decrease between observation and default. This distribution is clearly bimodal, and we model this using a finite mixture model of two Gaussian distributions. Figure 4.4(b) shows the distribution of $Y_i|R_i = 1$ for facilities that have a limit increase between observation and default. This distribution is clearly uni-modal, and we model this using an ordinary least squares model.



(a) EAD Distribution for Decrease Limit **(b)** EAD Distribution for Same or Increase Limit

Figure 4.4: Distribution of Log 10 EAD Conditional on Changes in Limit

Figure 4.5 compares the empirical cumulative distributions of EAD when the limits decrease ($Y_i|R_i = 0$) to when they are the same/increase ($Y_i|R_i = 1$). The p-value for the Kolmogorov-Smirnov Two-Sample Test is shown in table 4.4 and rejects the null-hypothesis that the two empirical CDF's come from the same distribution.

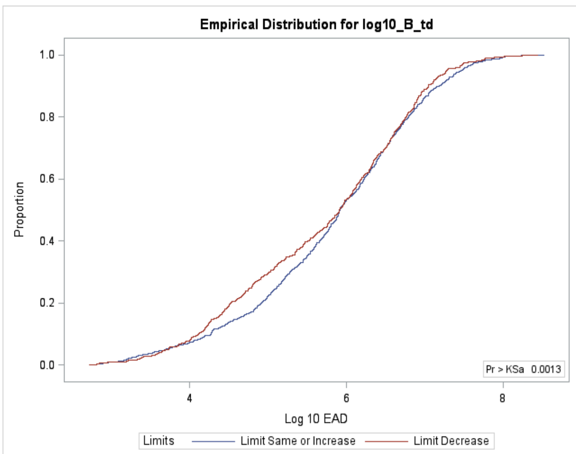


Figure 4.5: Comparison of Empirical Cumulative Distribution for EAD

Critical Value	Pr >Critical Value (upper tail)
1.92	0.0013

Table 4.4: Results of Kolmogorov-Smirnov Two-Sample Test

Continuing from equation 4.11, we now write down the pdf for $Y_i|R_i = j$, which is the distribution of $log_{10}(EAD)$ conditional on changes in limits. Equation 4.13 is the pdf of a two component Gaussian mixture model (McLachlan and Peel, 2004), and equation 4.14 is the pdf of a Gaussian distribution (Nelder and Wedderburn, 1972).

$$f_0(y_i|\underline{x}_{i0}, \underline{\theta}_{i0}) = \sum_{k=1}^{2} \pi_k \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_k} \exp\left(\frac{1}{2}\frac{(y_i - \mu_{k,i})^2}{\sigma_k^2}\right) \; , \; i = 1, ..., 699 \tag{4.13}$$

$$f_1(y_i|\underline{x}_{i1}, \underline{\theta}_{i1}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(\frac{1}{2}\frac{(y_i - \mu_i)^2}{\sigma^2}\right) \; , \; i = 1, ..., 1,445 \tag{4.14}$$

4.2.3 Univariate Analysis

The GCD database contains a wide variety of additional covariates beyond realised EAD and LGD outcomes. Some of these covariates are either not relevant for revolving facilities to large corporates or are not mandatory for ceding banks to supply. After removing variables that are either not relevant or not populated sufficiently, we are left with 14 candidate covariates. We also create an addition macroeconomic covariate based on World Bank's GDP Growth (WorldBank, 2015) by defining the year in which a facility defaulted as either: "downturn" when GDP growth is below 2%; or "expansion" when GDP growth is above 4%. We note that this definition is arbitrary, however this is simply a place-holder for the inclusion of a more elaborate macroeconomic model that would explain the sensitivity of EAD to the state of the macroeconomic cycle. Figure 4.6 demonstrates this by high lighting downturn as red and expansion as green while leaving "average" times uncoloured.

Figure 4.6: Growth Rate in Annual World Gross Domestic Product

The final shortlist of 15 covariates are displayed in table 4.5 below.

Entity (6)	Facility (8)	Macroeconomic (1)
Jurisdiction (strong/weak)	Limit	Economic State
Public/Private Company	Zero Balance	
Leveraged Finance Deal	Utilisation	
Lender Risk Rating	Syndication	
Operating Company	Guarantee/Collateral	
Number of Loans	Time to Maturity	
	Seniority	
	Loan Currency	

Table 4.5: List of Candidate Variables or Modelling

Appendix A details the univariate analysis for each of the 15 variables in table 4.5. For each variable we show both:

- the distribution of limit changes (R_i, for use in the **stage one** logistic regression model);
- the distribution of $log_{10}(EAD)$ $(Y_i|R_i = j)$ conditional on changes in the limit (for use in the **stage two** regressions)

To test the predictiveness of the covariates, we calculate a set of statistics and p-values from single variable regressions. For the **stage one** logistic regression, we calculate the following statistics which are routinely used to assess the univariate predictiveness of covariates in credit risk and show results in table B.1 in appendix B:

- Weight of Evidence (WoE) – a concept originally published in 1950 by the World War II codebreaker I.J. Good (Good (1950) and Good (1983)). Anderson (2007) explans that WoE converts the risk associated with a particular choice onto a linear scale that is easier for the human mind to assess. Higher values or WoE represent higher probability of an event occurring.
- Information Value (which is also known as the Kullback divergence measure), measures the difference between two distributions (Anderson, 2007). Siddiqi (2006) suggests covariates with values over 0.02 are likely to be predictive.
- Gini Co-efficient – a measure of separation, usually used to assess income disparties, but used in credit scoring to assess predictive power (Anderson, 2007). Higher values represent stronger predictiveness.
- Area Under the ROC Curve (AUC) – a measure related to the Gini Co-efficient which measures the area under the Receiver Operator Characteristic curve (Anderson, 2007). Higher values represent stronger predictiveness.

For the **stage two** models, we asses each covariate's predictiveness by estimating each individually in an OLS regressions. The overall significance of each covariate in the regressions, as per the p-value, is shown in appendix B. Table B.2 display results for individual regressions against $log_{10}(EAD)$ for limits remaining the same/increasing while and table B.3 display results for individual regressions against $log_{10}(EAD)$ for limits decreasing.

The results from univariate analysis show that most of the 15 candidate covariates are predictive, so we maintain them all as candidates in the next section for multivariate analysis.

4.2.4 Multivariate Model for Limit

We estimate the logistic regression for the **stage one** model as per equation 4.2 using PROC LOGISTIC in SAS/STAT 9.3. The target is the limit remaining the same or increasing between observation and default ($R_i = 1$) and we conduct step-wise variable selection with the threshold for entry to the model of 0.1 and a threshold to stay in the model of 0.2. The selected covariates, together with their Wald Chi-square statistics and p-values, are displayed in table 4.6 below. Table C.1 in appendix C details the parameter estimates or the logistic regression along with the significance of each effect.

Effect	DF	Wald Chi-Square	Pr >ChiSq
Jurisdiction	1	17.0542	<.0001
Leveraged Finance Deal	1	5.8627	0.0155
Lender Risk Rating	2	44.1402	<.0001
Operating Company	1	6.3277	0.0119
Number of Loans	2	12.8799	0.0016
Log 10 Limit	1	18.5775	<.0001
Zero balance	1	39.5674	<.0001
Syndication	1	3.7305	0.0534
Guarantee/Collateral	1	16.222	<.0001
Seniority	2	36.8339	<.0001
Economic State	2	11.7407	0.0028

Table 4.6: Covariates Obtained via Stepwise Selection for **Stage One** Logistic Regression

Figure 4.7 below shows the Receiver Operator Characteristic (ROC) curve and the area under the ROC curve of 0.7018. This suggests that the model has a high degree of predictive power in explaining movements in the facility limits.

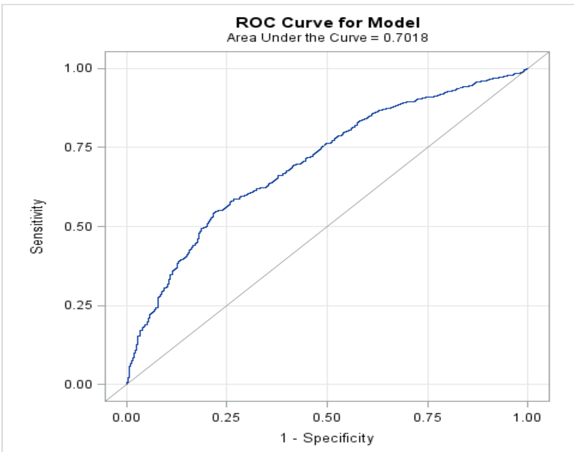


Figure 4.7: Receiver Operator Characteristic (ROC) Curve for **Stage One** Model

4.2.5 Multivariate Model for EAD Given Limit Same/Increase

There are two regressions for the **stage two** model, as outlined previously in equations 4.13 and 4.14. We use an OLS regression to explain EAD for facilities where the limit remains the same or increases. We undertake stepwise variable selection using PROC GLMSELECT from SAS/STAT 9.3 using a threshold for entry to the model of 0.1 and a threshold to stay in the model of 0.2. The selected effects are then used in PROC GENMOD from SAS/STAT 9.3 and the resulting Wald Chi-square statistics and p-values are displayed in table 4.7 below. Table C.2 in appendix C details the parameter estimates for the OLS regression along with the significance of each effect.

Efect	DF	Wald Chi-Square	Pr >ChiSq
Lender Risk Rating	2	8.19	0.0166
Log10 Limit	1	2959.49	<.0001
Zero Balance	1	15.87	<.0001
Syndication	1	8.15	0.0043
Log10 Months to maturity	1	4.71	0.0299

Table 4.7: Covariates Obtained via Stepwise Selection for **Stage Two** OLS Regression

4.2.6 Multivariate Model for EAD Given Limit Decrease

The second regression for the **stage two** model is a finite mixture model (FMM). We implement this using PROC FMM in SAS/STAT 9.3, but the procedure does not have any automatic variable selection. Thus we undertake variable selection manually beginning with the most significant variables identified from univariate analysis and applying judgement. For a mixture model it is not as straight forward to provide overall p-values for the selected covariates, so instead we provide a simple list of these covariates in table 4.8. Tables C.3, C.3 and C.5 in appendix C detail the parameter estimates and the significance of the effects in each of the two components and the probability model of the FMM.

Component	Parameter
1	Log10 Limit
1	Zero Balance
1	Lender Risk Rating
2	Log10 Limit
2	Leveraged Finance Deal
Probability	Loan Currency
Probability	Log10 Time to Maturity
Probability	Operating
Probability	Economic State

Table 4.8: Covariates Selected via Judgement for **Stage Two** FMM Regression

4.3 Results

4.3.1 Calculating the Model's Fitted Values

To assess the fit of the final model, we calculate predicted values ($E[\hat{Y}_i|\mathbf{X}, \underline{\hat{\phi}}]$) as per 4.12. This involves calculating for all $i = 1, ..., 2, 144$ the following three quantities:

- $E[R_i = 1|\underline{w}_i, \underline{\hat{\alpha}}]$ (from the **stage one** logistic regression);
- $E[\hat{Y}_i|\underline{x}_i, \underline{\hat{\theta}}_1]$ (from the **stage two** OLS model); and
- $E[\hat{Y}_i|\underline{x}_i, \underline{\hat{\theta}}_0]$ (from the **stage two** finite mixture model).

The parameter estimates $\underline{\hat{\alpha}}$, $\underline{\hat{\theta}}_{i1}$, and $\underline{\hat{\theta}}_{i0}$ are detailed in tables C.1, C.2, C.3, C.4 and C.5 in appendix C. Table 4.9 below details an example of calculating the overall fitted values.

Observation Number	A $E[R_i = 1 \mid \underline{w}_i, \underline{\hat{\alpha}}]$ Logistic	B $E[\hat{Y}_i \mid \underline{x}_i, \underline{\hat{\theta}}_1]$ OLS	C $E[\hat{Y}_i \mid \underline{x}_i, \underline{\hat{\theta}}_0]$ FMM	A×B + (1-A)×C $E[\hat{Y}_i \mid \mathbf{X}, \underline{\hat{\phi}}]$ Overall
1	0.88619	6.23810	6.05794	6.21760
2	0.67780	6.42065	6.09012	6.31416
3	0.71576	6.28705	5.84054	6.16013
4	0.80943	7.19365	7.03536	7.16348
5	0.64664	7.03440	6.78268	6.94545
...	...	...	...	...
2,144	0.45082	6.58330	6.35719	6.45913

Table 4.9: Calculation of Fitted Values

4.3.2 Model Diagnostics

In this section we assess the quality of the model's predictions of $log_{10}(EAD)$ by comparing observed values to predicted values.

Figure 4.8 compares a histogram of observed (top) and predicted (lower) $log_{10}(EAD)$ with overlaid kernel density estimate. Figure 4.9 compares the kernel density estimates from the two panels in figure 4.8 overlaid together on the same axis. These two graphs show there is a reasonably high level of predictive power by the model, as displayed by the similarity in histograms and kernel density estimates.

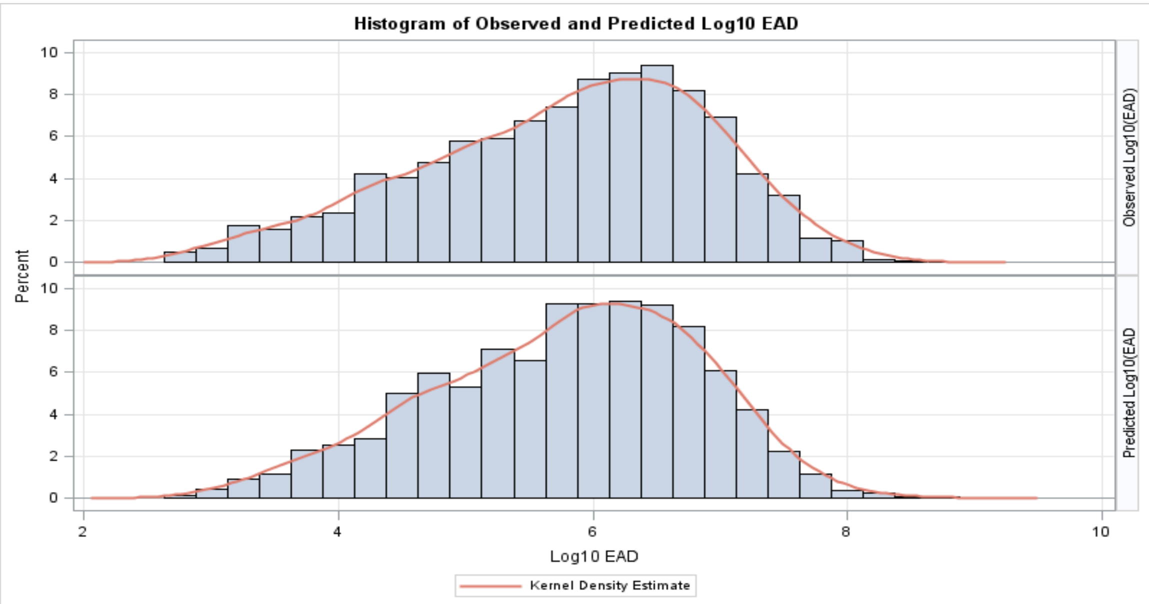


Figure 4.8: Histogram of Observed (Top) and Predicted (Lower) Log10 EAD

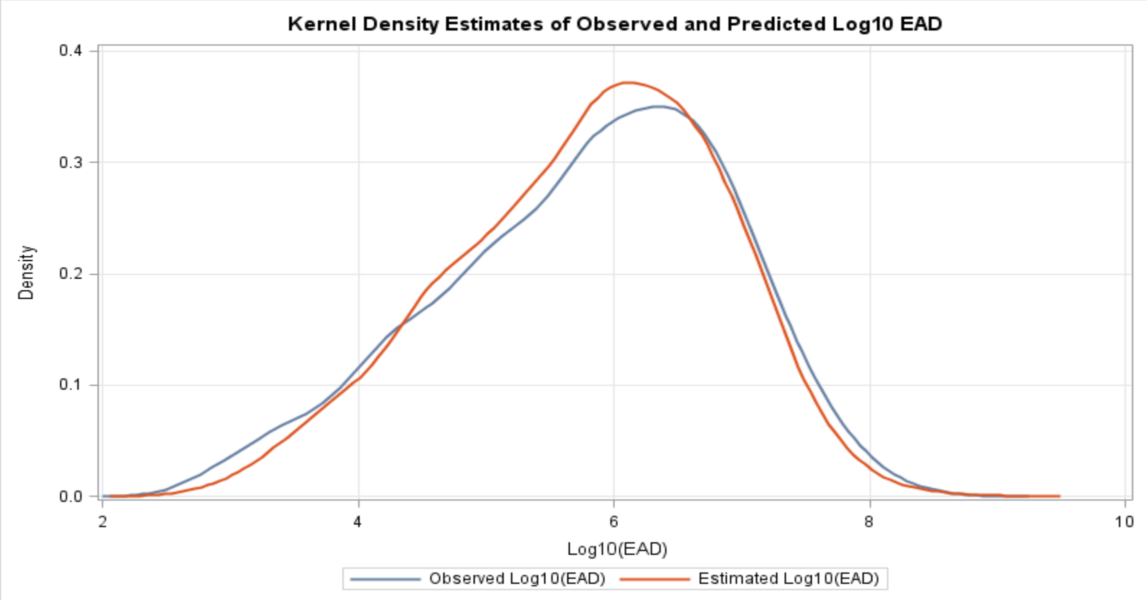


Figure 4.9: Kernel Density Estimates of Observed and Predicted Log10 EAD

Figure 4.10 shows a scatter plot of predicted vs observed $log_{10}(EAD)$ and figure 4.11 compares the empirical cumulative density functions for predicted vs observed $log_{10}(EAD)$. Table 4.10 shows p-value for the Kolmogorov-Smirnov Two-Sample Test that compares whether the predicted and observed empirical CDF's come from the same distribution. The test fails to reject this null-hypothesis, and we conclude that the observed and predicted values come from the same distribution.

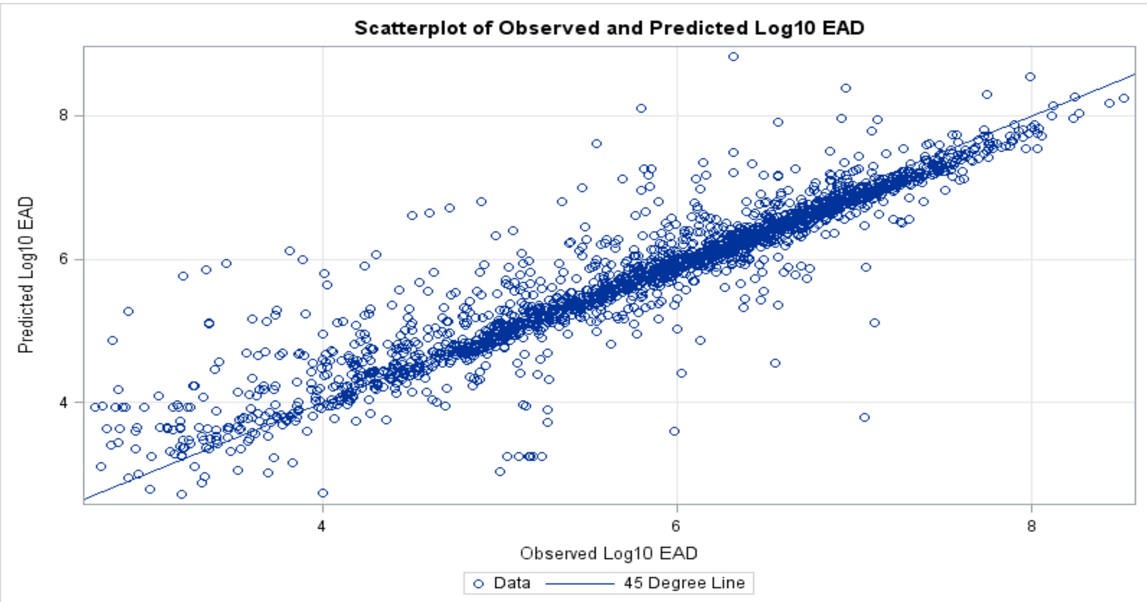


Figure 4.10: Scatterplot of Observed and Predicted Log10 EAD

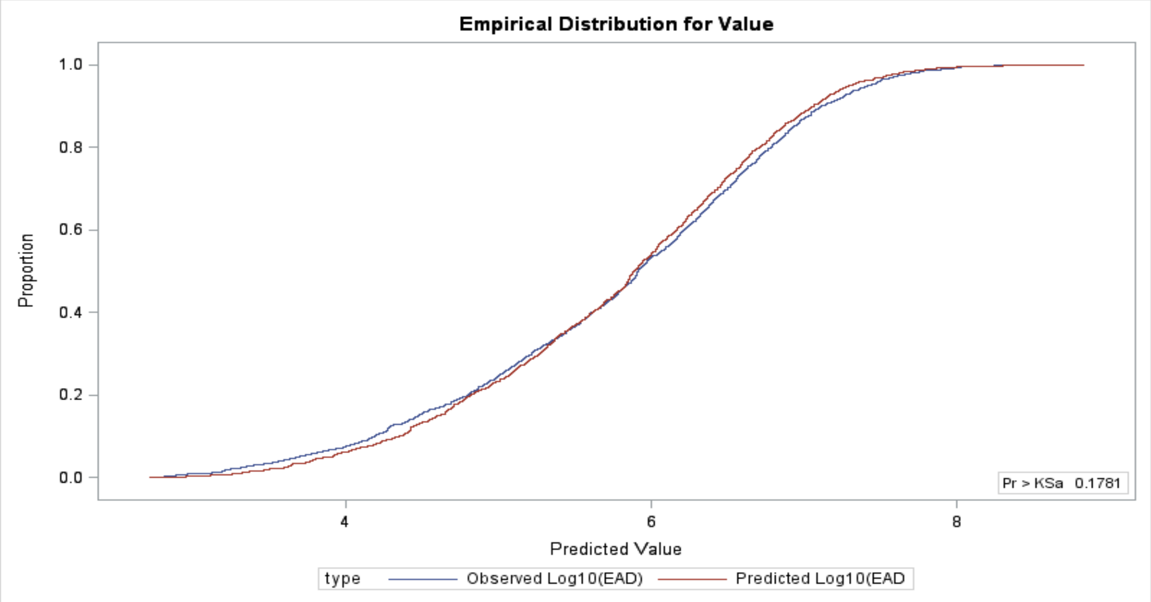


Figure 4.11: Cumulative Distribution Functions of Observed and Predicted Log10 EAD

Critical Value	Pr >Critical Value (upper tail)
1.1	0.1781

Table 4.10: Results of Kolmogorov-Smirnov Two-Sample Test for Observed vs Predicted

The below two figures analyse the residuals between the observed and predicted values. Figure 4.12 shows a histogram of residuals and there appears to be on average a slight over-estimation bias of our model. These features are also visible in the box plot of the residuals in 4.13.

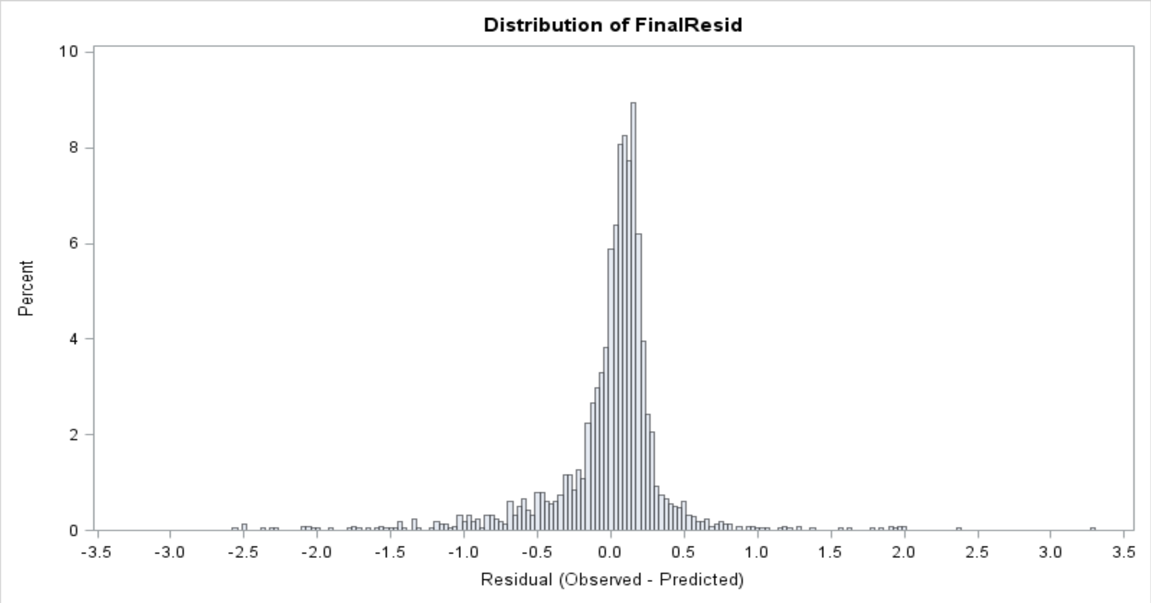


Figure 4.12: Histogram of Residuals

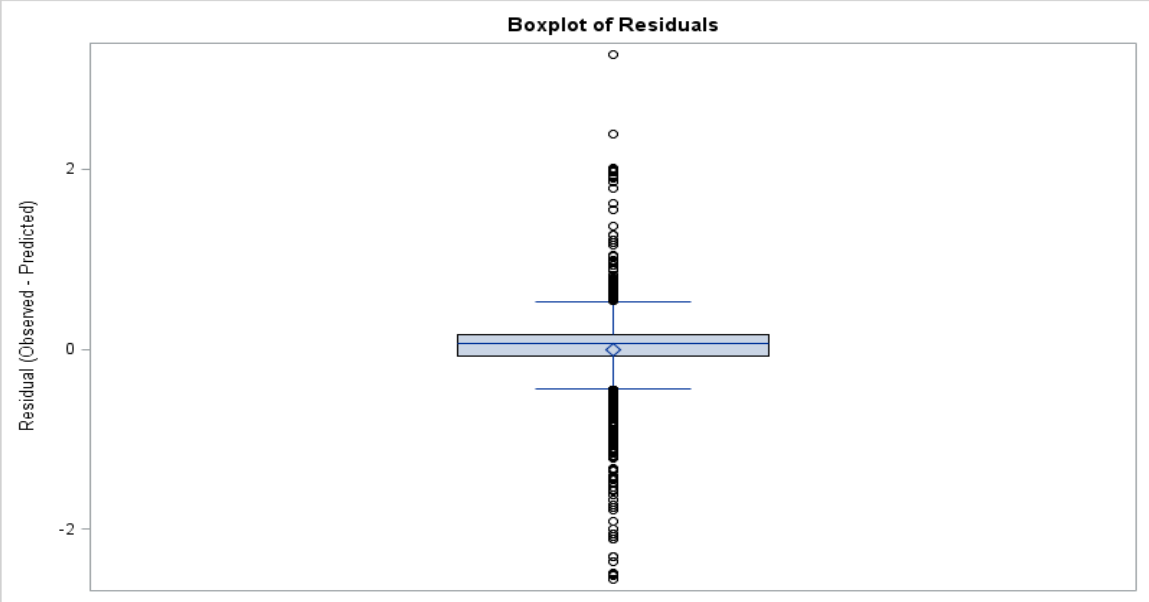


Figure 4.13: Boxplot of Residuals

In order to validate our model, we undertake a 1,000 sample cross validation. This involves, for each of the 1,000 random samples, re-estimating the final model on a random 70% of the data and validating on the remaining 30%. We then calculate the mean-square error from each of the $j = 1$ to $1,000$ samples of 30% validation sets as follows:

$$MSE_j = \sum_{i=1}^{N_j} (y_i - \overline{y})^2 \qquad N_j = size\ of\ j^{th}\ CV\ sample \tag{4.15}$$

Figure 4.14 below shows the histogram of the 1,000 samples of 30% validation sets. The green vertical line represents the average of the 1,000 MSE and the vertical red line represents the MSE from the model development. The closeness of the green and red vertical lines, together with the narrow dispersion of cross-validated MSE suggest that the mode has not been overfit.

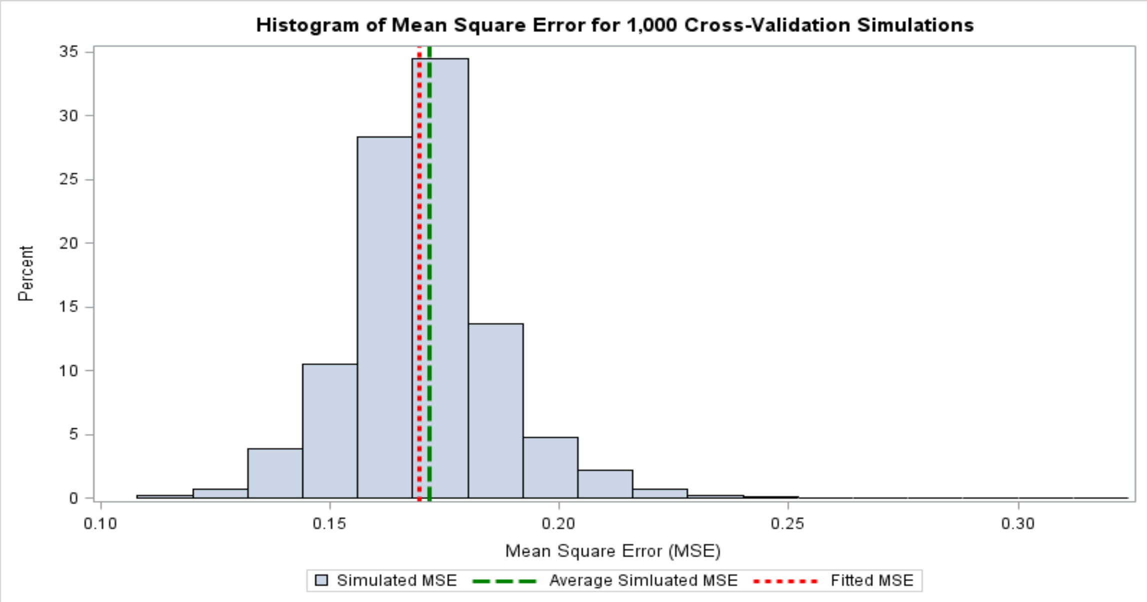


Figure 4.14: Histogram of Mean Square Error for 10,000 Cross Validation Simulations

4.3.3 Results and Interpretation

The analysis above gives a high degree of confidence that our model fits our observed data well. This is evidenced by the graphical and statistical tests that show that predicted and observed values are close. The cross-validation analysis also shows evidence that the model has not been overfit.

Our model suggests that the major drivers of EAD include:

- limit;
- utilisation;
- risk rating; and
- time to maturity.

While we also find evidence other significant variables, the list above represent the primary drivers. This list also generally agrees with other authors work that we have reviewed. We also find evidence that banks manage limits in the lead up to default, and that these changes in limits have substantial effects on the outcomes of EAD.

All the parameter estimates for our model are tabulated in appendix C, however tables 4.11, 4.12 and 4.13 detail the effects our model's covariates and their effect of an increase in limits and an increase in EAD respectively.

The Facility has a Greater Chance of a Limit Increase if it has the Following Features
Weak jurisdictions
Leveraged deals
Non-rated borrowers
Holding companies
More loans
Lower limit
Lower utilisation
Syndicated Deals
No Guarantee or collateral
Senior debt
Expansion (and to a lesser extent Recession)

Table 4.11: Drivers of an Increase in the Limit

Conditional on a Limit Increase the Facility Will Have a Higher EAD if it has the Following Features
Non-rated borrowers
Higher limit
Higher utilisation
Non-syndicated deals
Longer time to maturity

Table 4.12: Drivers of EAD, Conditional on a Limit Increase

Conditional on a Limit Decrease the Facility Will Have a Higher EAD if it has the Following Features
Higher limit
Higher utilisation
Non-rated borrowers
Syndicated deals
Currency other than EUR or USD
Longer time to maturity
Holding company
Expansion (and to a lesser extent Recession)

Table 4.13: Drivers of EAD, Conditional on a Limit Decrease

4.4 Conclusion and Discussion

4.4.1 Knowledge Discovered

Suitable risk management is vital for the survival and continued solvency of any business, and banks are no different. The events during the "Global Financial Crisis" (GFC) of 2007 and 2008 highlights some of the potential impacts when banks don't manage their risks appropriately and showed that the interdependencies inherent in the financial system meant the effects spread quickly to other industries and can affect entire economies.

With the advent of "Basel Accords" from the late 1980's onwards and its focus on estimating risk-based capital requirements, the quantification of risks that a bank faces (and in particular credit risk associated with the granting of loans which is typically the largest of its risks) requires the application of advanced analytics and statistical methods to help determine risk components. For lending to large corporates, where an individual bank's internal empirical data may be too thin to reliably estimate these, internationally active banks have formed consortia to pool data (such as the GCD) to assist informing these estimates.

This project has trained a statistical model to estimate the Exposure at Default (EAD) for large corporate counterparties to banks who are granted revolving facilities using one member bank's view of the GCD data. Apostolik et al. (2009). defines EAD as *"[t]he potential loss a bank would suffer if a borrower fails to meets its obligations"*. EAD is a key input parameter to not only estimation of regulatory credit capital that a bank's regulator stipulates it must hold in recognition for the risks it takes on when granting loans but also for other internal risk management purpose such as: economic capital; pricing; risk-adjusted return on capital (RAROC) calculations of profitability; stress testing; bad debt forecasting; loan loss provisioning and limit management.

Both academic and practitioner research in the area of EAD estimation is becoming a more attractive research topic, but there has been less focus on EAD to date than for the more readily estimable risk components Probability of Default (PD) and Loss Given Default (LGD). While the majority of existing authors estimate EAD indirectly via the Credit Conversion Factor (CCF) that is popularised in the Base Accord's standardised approach to credit risk capital, this may change in times to come. For example, several authors (such as Taplin et al. (2007)) point to the conceptual and numerical difficulties estimating CCF, and further the United Kingdom regulator the Financial Conduct Authority (2014) is now willing to consider own estimation of EAD rather than own estimation of CCF for advanced internal ratings based accreditation (A-IRB) approach to credit risk capital.

Our model has adopted such an approach by directly estimating EAD conditional on changes in limit, and despite this, our results largely agree with respect to key findings and main drivers from previous authors whose models predict alternate responses (mainly CCF). We find like other authors that that key drivers of EAD include: limit; balance; utilisation; risk rating; and time to maturity. Initial analysis (not contained in this report) did show facility type to be a key driver, and because of this our analysis chose to focus solely on revolving credit lines. We also find evidence of the "race to default" identified by Qi (2009) and Mantel (2012) and what Sufi (2009) identifies as a moral hazard problem. This describes the tendency for financially distressed customers to draw down remaining available funds and for banks to respond by managing limits.

Beyond the above concordance with existing research, we add to the literature in several key ways. First, we develop a model that directly estimates EAD conditional on changes in limit rather than like many of the existing papers that model alternate measures (such as CCF). Second, our model explicitly considers both the balance and (changes in) the limit as random variables and we adopt our methodology appropriately by constructing a two stage model – the first stage estimates the probability that limits decrease while the second stage estimates EAD conditional on changing limits. Third, we leverage the GCD database to estimate our model, and to the best of our knowledge this is only the second masters project or thesis regarding EAD estimation undertaken using this data. Fourth, our model shows good predictive power between model estimates and observed EAD. Finally, similar to Mantel (2012) and Jacobs (2010) we identify some evidence of relation between EAD and the state of the economy.

4.4.2 Limitations and Further Research

There are several avenues how our work could be extended. The first of which would be to extend the modelling to other facility types, such as term loans. Whether certain modelling aspects and parameterisations would carry over directly would need investigation but evidence from other researchers suggest that facility type is a key driver of EAD. A second enhancement would be the modelling of limits as a continuous response, rather than the dichotomous response adopted for our model. The third extension would be to conduct model back testing using both an in-time and out-of-time hold-out sample. Our choice not to undertake such testing is outlined in our paper and relates to the thinness of data for some covariates. We do however go on to propose that the frequent updating of GCD data via new defaults from existing members as well as the joining of new members to the consortium could provide the necessary holdout data to validate this model. The fourth and final extension would be to compare our model to other model constructs typically seen in practice. This would likely include comparison to models that indirectly estimate EAD via CCF.

Appendix A

This appendix details the univariate analysis for the 15 candidate covariates, with one paged focussing on each variable in turn. The first graph and table on each page detail the univariate predictive power of the random variable defining changes in the limit (R_i), while the remaining graphs show the conditional distribution of $log_{10}(EAD)$ conditional in changes in the limit (the random variable $Y_i|R_i = j$). Tables B.1, B.2 and B.3 collate our findings for each variable.

Jurisdiction

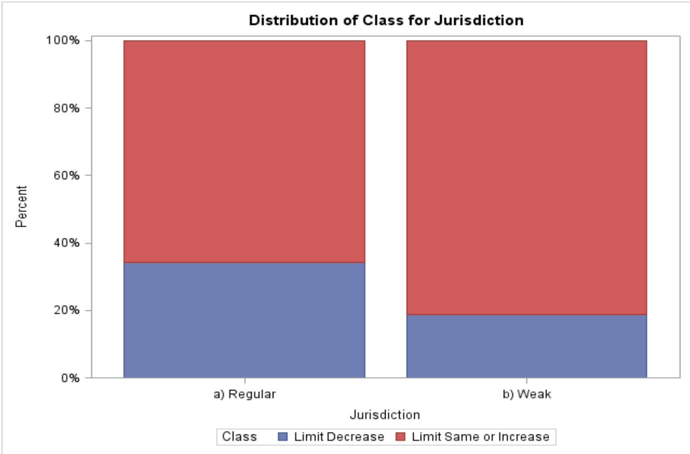


Figure A.1: Distribution of Limit Changes for Jurisdiction

Jurisdiction	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
Regular	1,906	1,252	654	89%	-0.077			
Weak	238	193	45	11%	0.730			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.056**	**0.069**	**0.535**

Table A.1: Summary Table for Jurisdiction

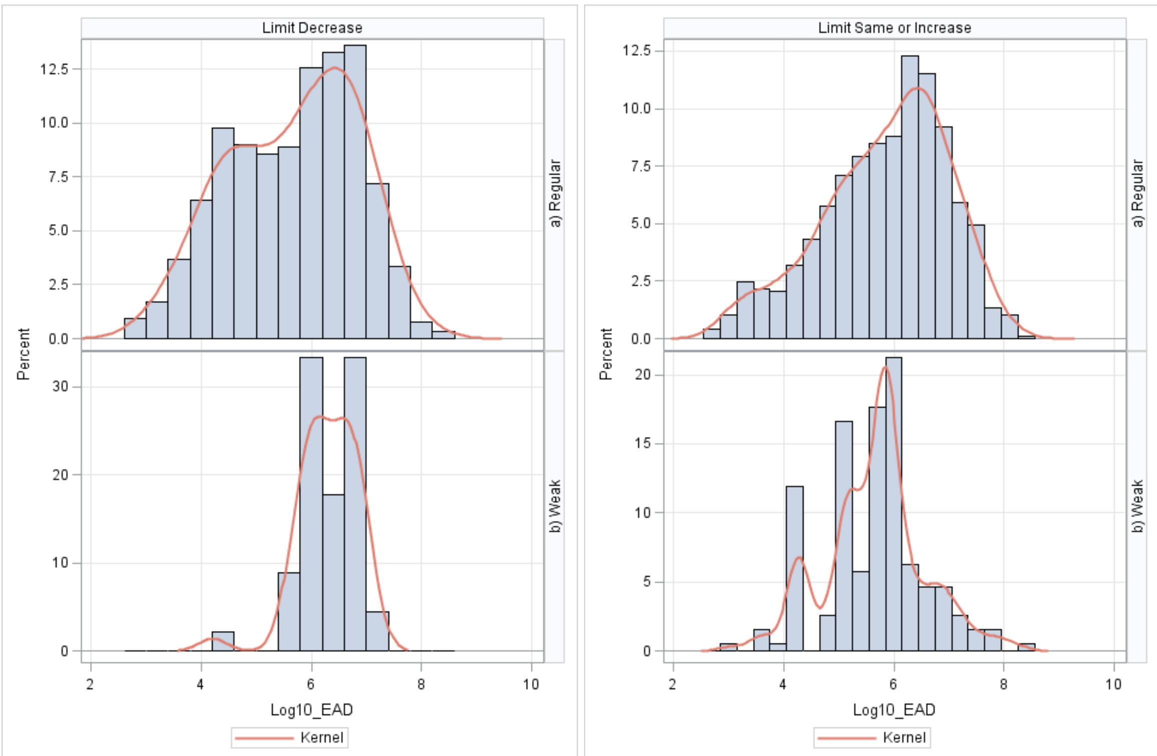


Figure A.2: Distribution of Log 10 EAD Conditional on Changes in Limit for Jurisdiction

Public or Private Indicator

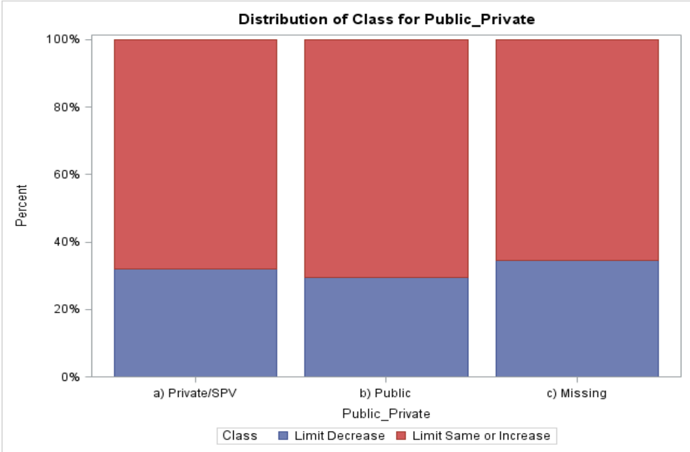


Figure A.3: Distribution of Limit Changes for Public/Private Indicator

Public/Private	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
Private/SPV	1,248	847	401	58%	0.022			
Public	241	170	71	11%	0.147			
Missing	655	428	227	31%	-0.092			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.005**	**0.037**	**0.518**

Table A.2: Summary Table for Public or Private Indicator

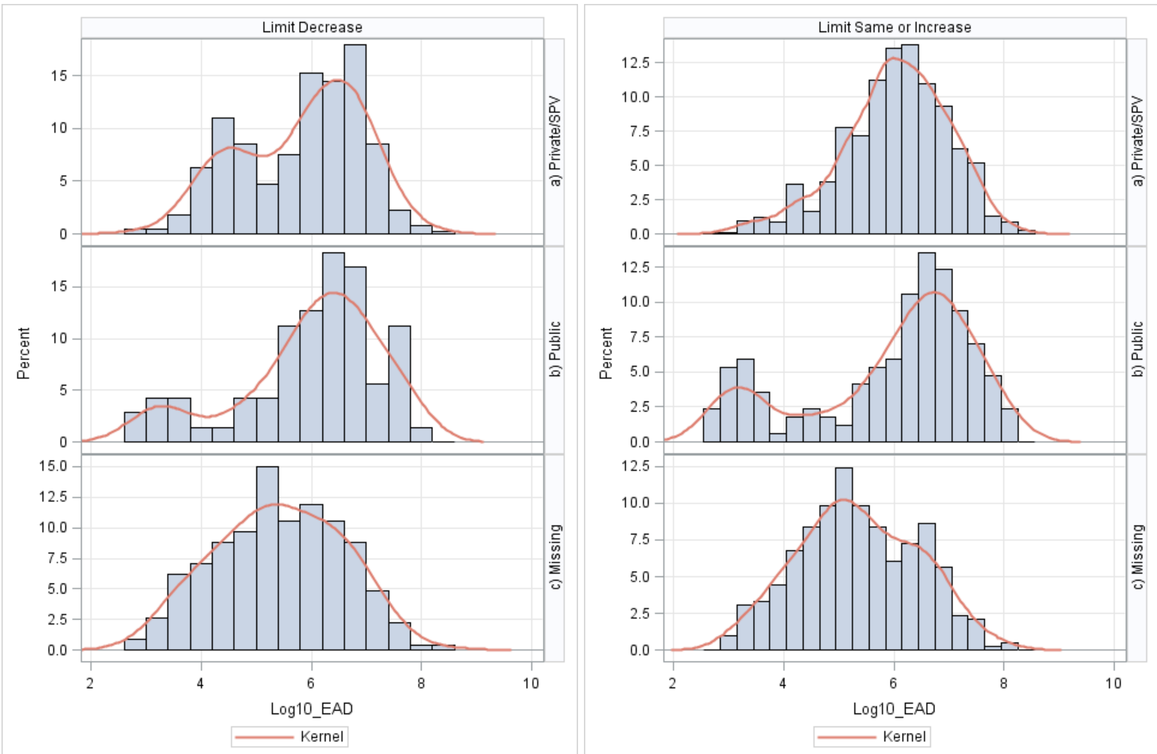


Figure A.4: Distribution of Log 10 EAD Conditional on Changes in Limit for Public/Private

Leveraged Deal Indicator

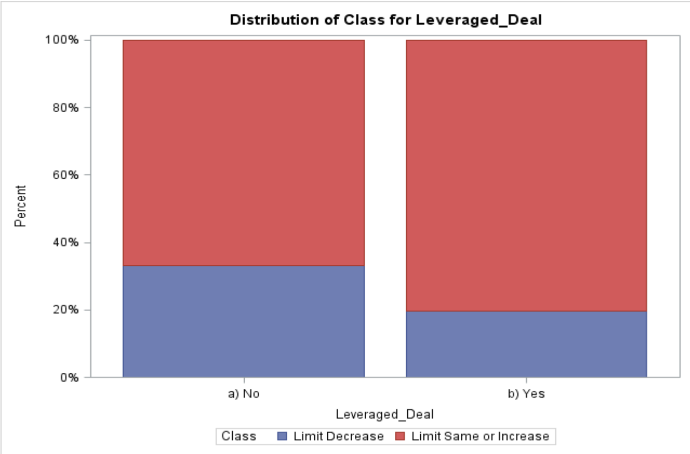


Figure A.5: Distribution of Limit Changes for Leveraged Deal Indicator

Leveraged Deal	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
No	2,048	1,368	680	96%	-0.027			
Yes	96	77	19	4%	0.673			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.018**	**0.026**	**0.513**

Table A.3: Summary Table for Leveraged Deal Indicator

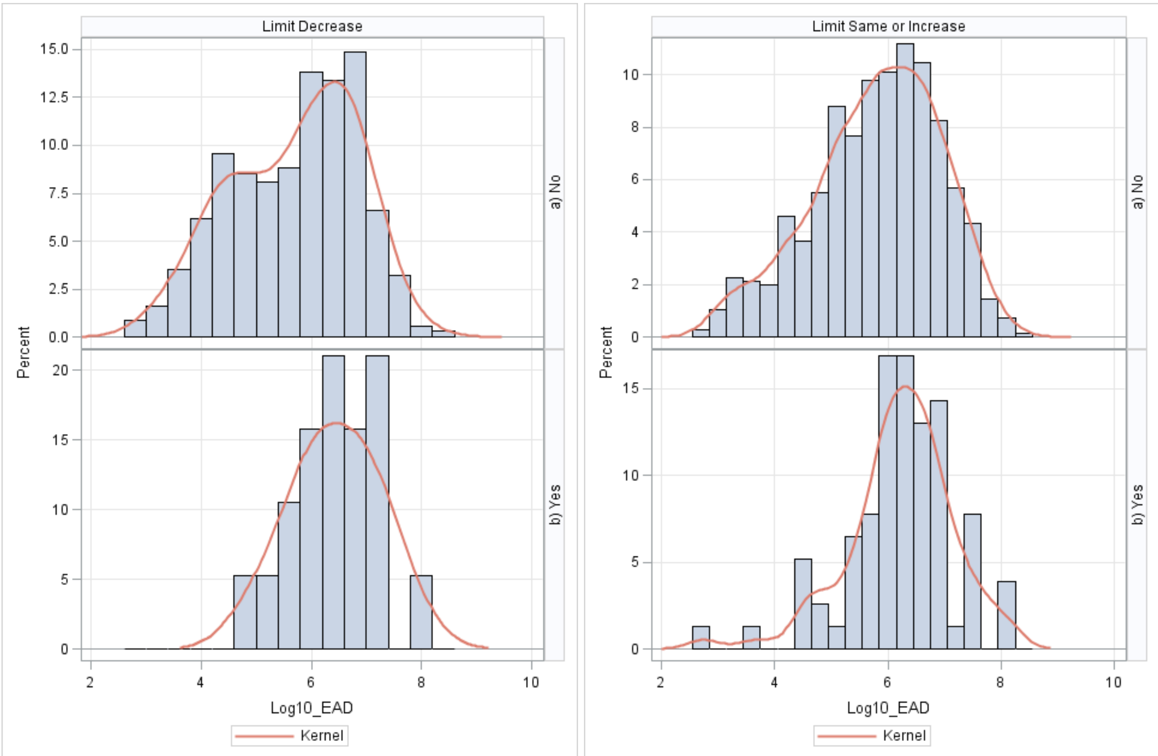


Figure A.6: Distribution of Log 10 EAD Conditional on Changes in Limit for Leveraged Deals

Lender Risk Rating

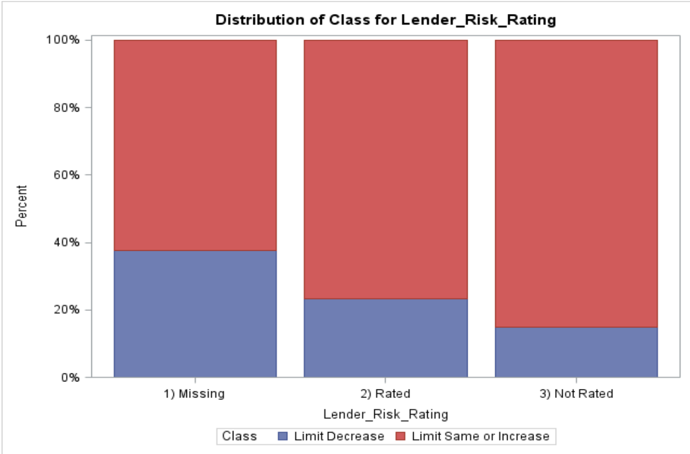


Figure A.7: Distribution of Limit Changes for Lender Risk Rating

Risk Rating	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
Missing	1,524	950	574	71%	-0.222			
Rated	392	301	91	18%	0.470			
Not Rated	228	194	34	11%	1.015			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.160**	**0.171**	**0.586**

Table A.4: Summary Table for Lender Risk Rating

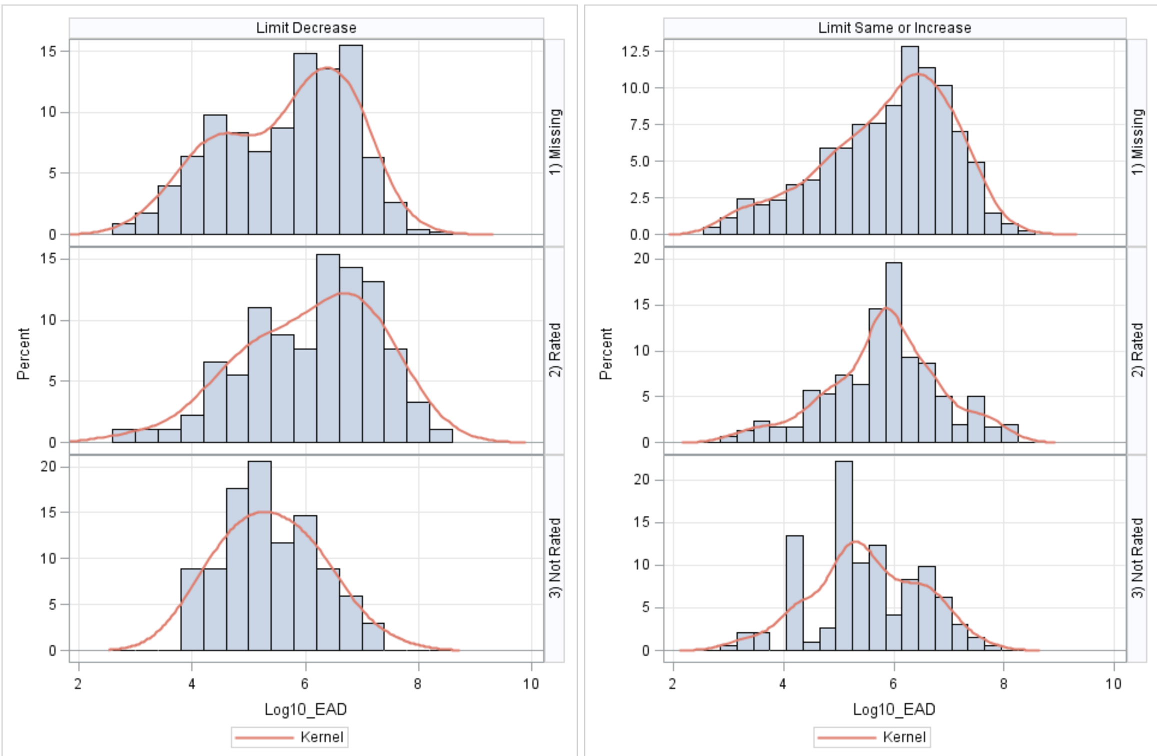


Figure A.8: Distribution of Log 10 EAD Conditional on Changes in Limit for Risk Rating

Operating or Holding Company Indicator

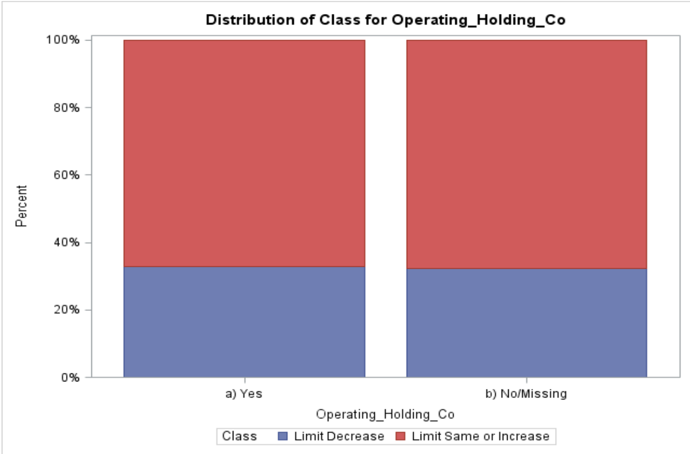


Figure A.9: Distribution of Limit Changes for Operating/Holding Company Indicator

Op/Hold Co	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
Yes	1,163	781	382	54%	-0.011			
No/Missing	981	664	317	46%	0.013			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.000**	**0.006**	**0.503**

Table A.5: Summary Table for Operating or Holding Company Indicator

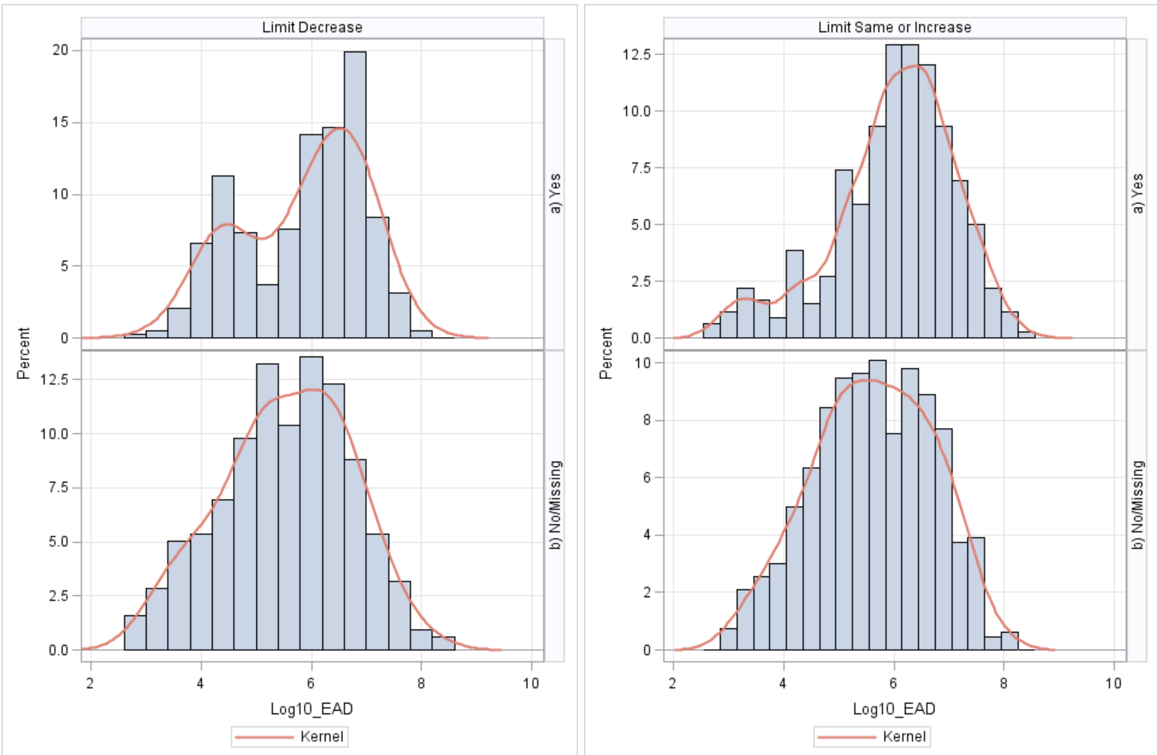


Figure A.10: Distribution of Log 10 EAD Conditional on Changes in Limit for Op/Hold Co

Number of Loans

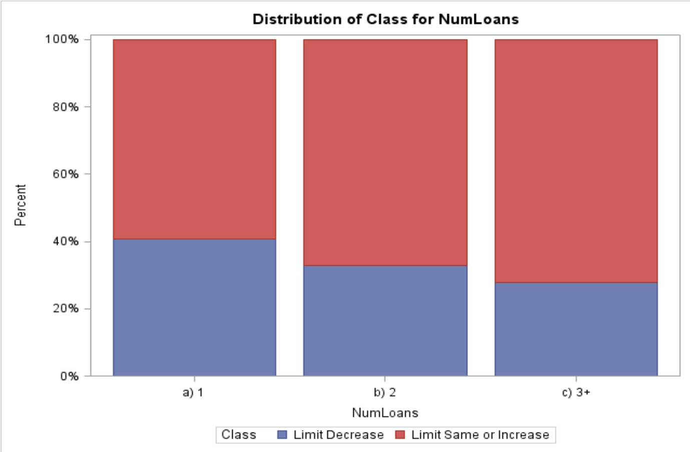


Figure A.11: Distribution of Limit Changes for Number of Loans

Number Loans	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
1	619	367	252	29%	-0.350			
2	443	297	146	21%	-0.016			
3	1,082	781	301	50%	0.227			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.062**	**0.131**	**0.565**

Table A.6: Summary Table for Number of Loans

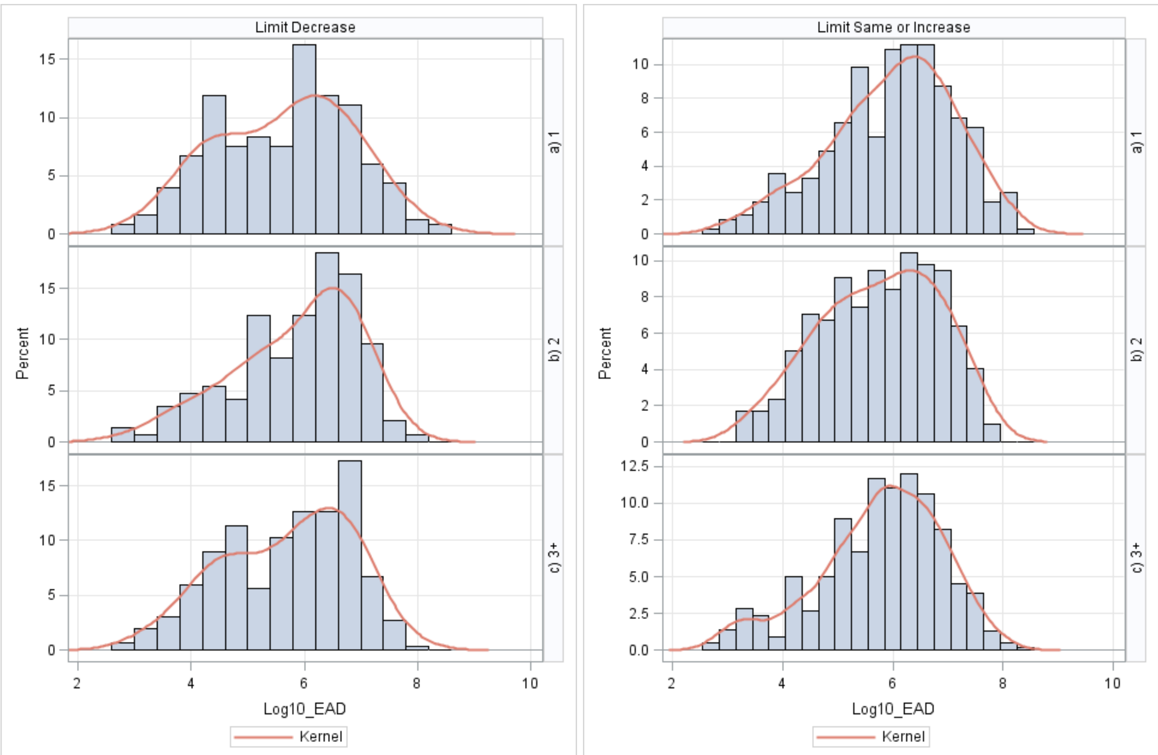


Figure A.12: Distribution of Log 10 EAD Conditional on Changes in Limit for Number of Loans

Log 10 Limit

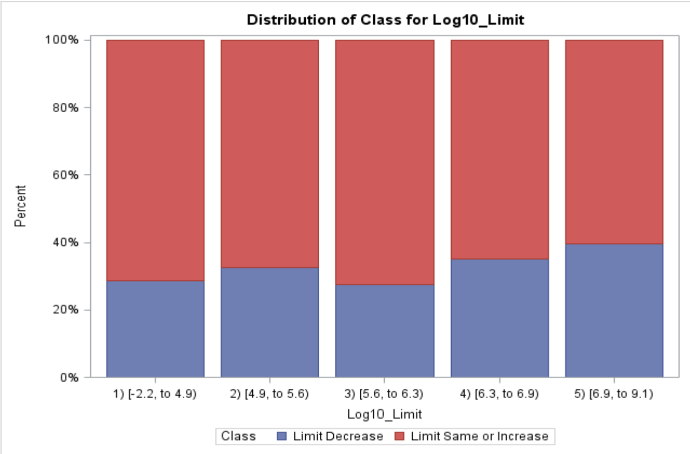


Figure A.13: Distribution of Limit Changes for Log 10 Limit

Log 10 Limit	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
[-2.2, to 4.9)	410	292	118	19%	0.180			
[4.9, to 5.6)	378	255	123	18%	0.003			
[5.6, to 6.3)	489	354	135	23%	0.238			
[6.3, to 6.9)	459	298	161	21%	-0.111			
[6.9, to 9.1)	408	246	162	19%	-0.308			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.040**	**0.111**	**0.556**

Table A.7: Summary Table for Log 10 Limit

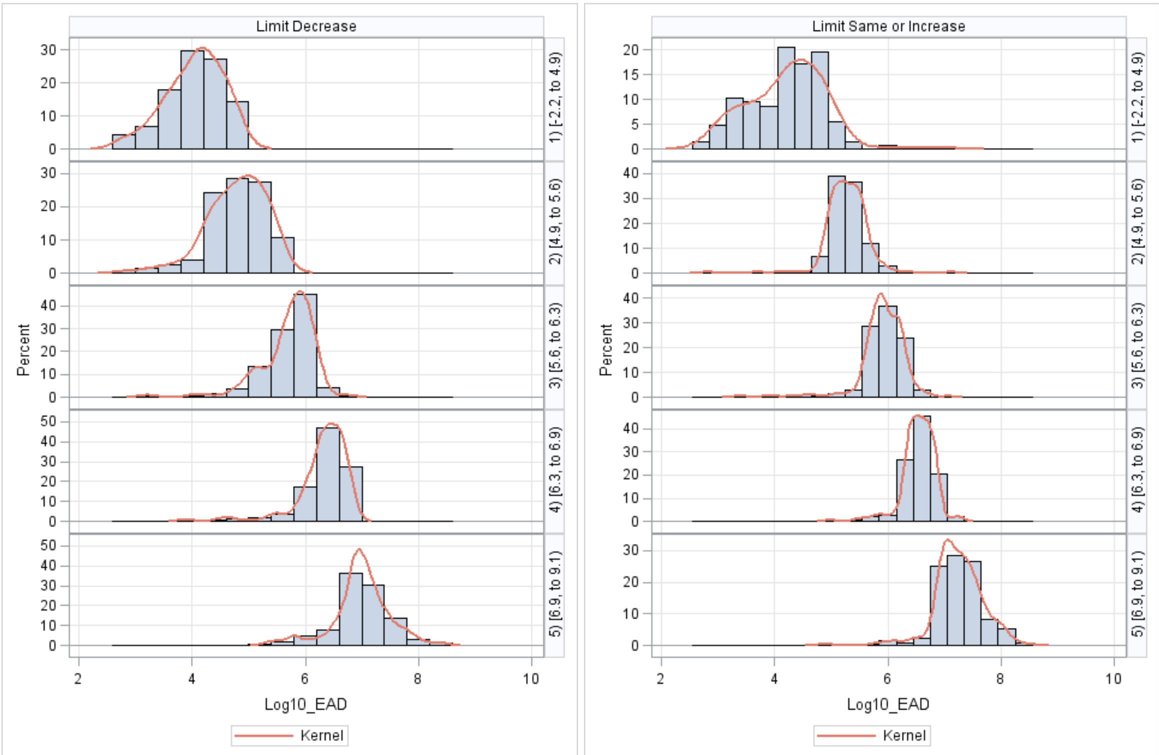


Figure A.14: Distribution of Log 10 EAD Conditional on Changes in Limit for Log10 Limit

Zero Balance Indicator

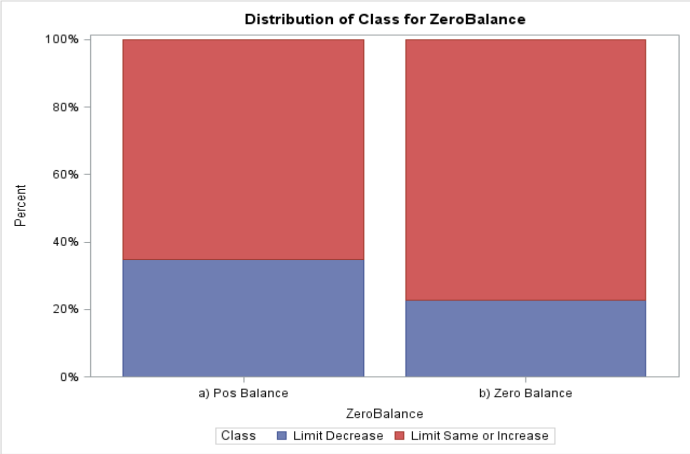


Figure A.15: Distribution of Limit Changes for Zero Balance Indicator

Zero Balance	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
Pos Balance	1,745	1,137	608	81%	-0.100			
Zero Balance	399	308	91	19%	0.493			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.049**	**0.083**	**0.541**

Table A.8: Summary Table for Zero Balance Indicator

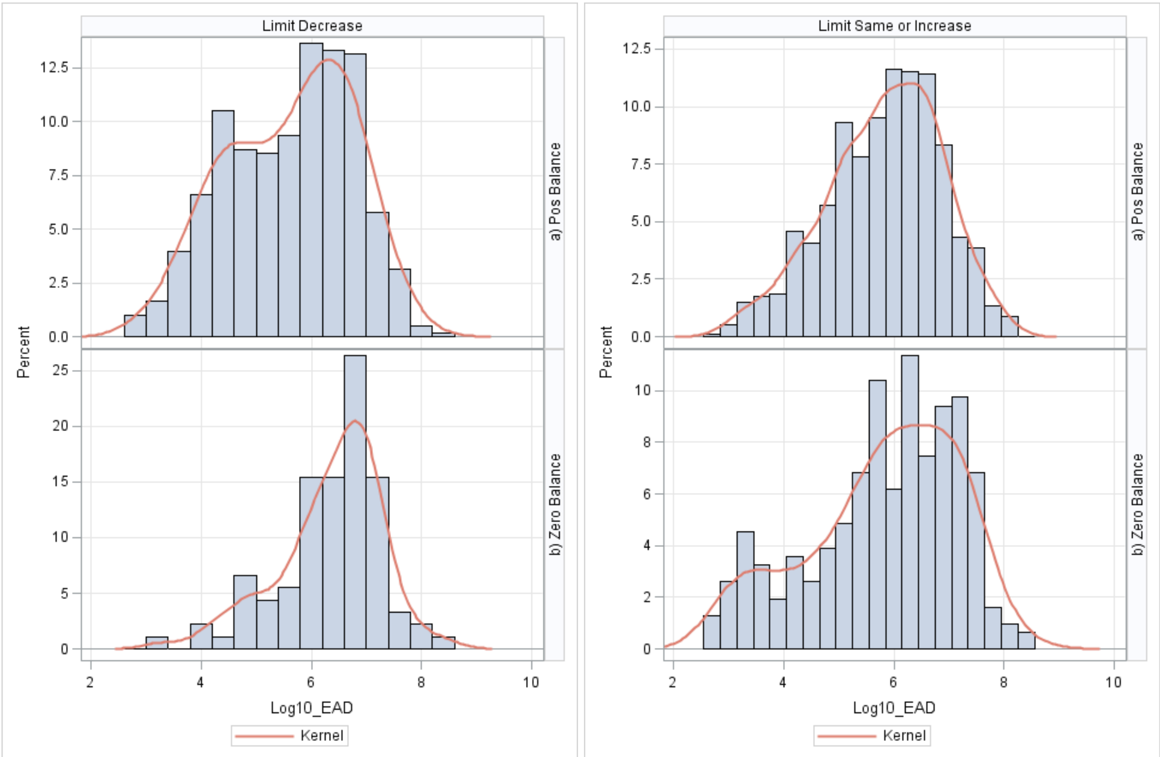


Figure A.16: Distribution of Log 10 EAD Conditional on Changes in Limit for Zero Balance

Utilisation

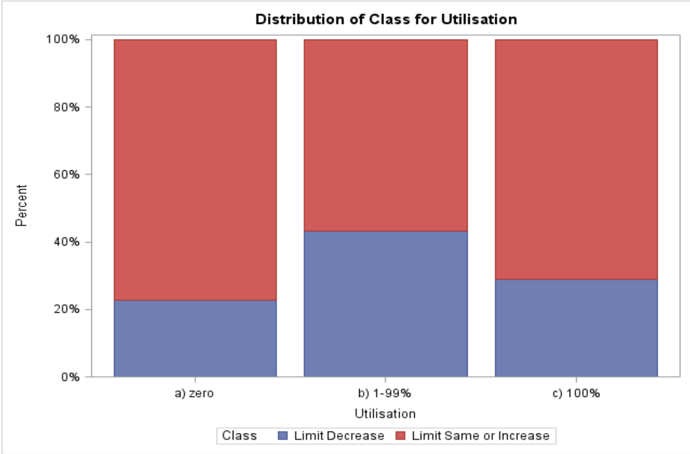


Figure A.17: Distribution of Limit Changes for Utilisation

Utilisation	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
Zero	399	308	91	19%	0.493			
1-99%	712	403	309	33%	-0.461			
100%	1,033	734	299	48%	0.172			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.130**	**0.188**	**0.594**

Table A.9: Summary Table for Utilisation

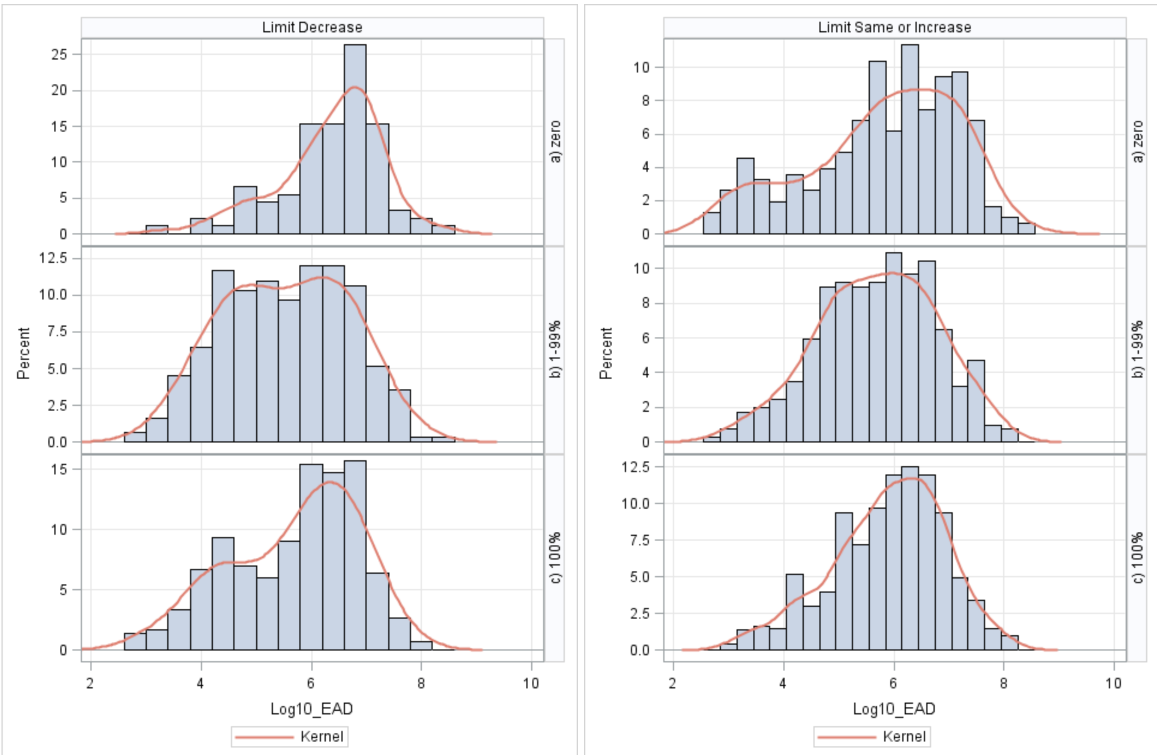


Figure A.18: Distribution of Log 10 EAD Conditional on Changes in Limit for Utilisation

Syndicated Deal Indicator

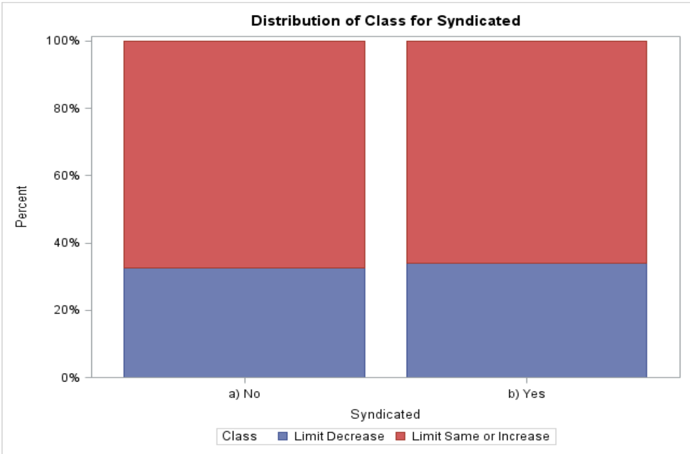


Figure A.19: Distribution of Limit Changes for Syndicated Deal Indicator

Syndicated	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
No	2,018	1,362	656	94%	0.004			
Yes	126	83	43	6%	-0.069			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.000**	**0.004**	**0.502**

Table A.10: Summary Table for Syndicated Deal Indicator

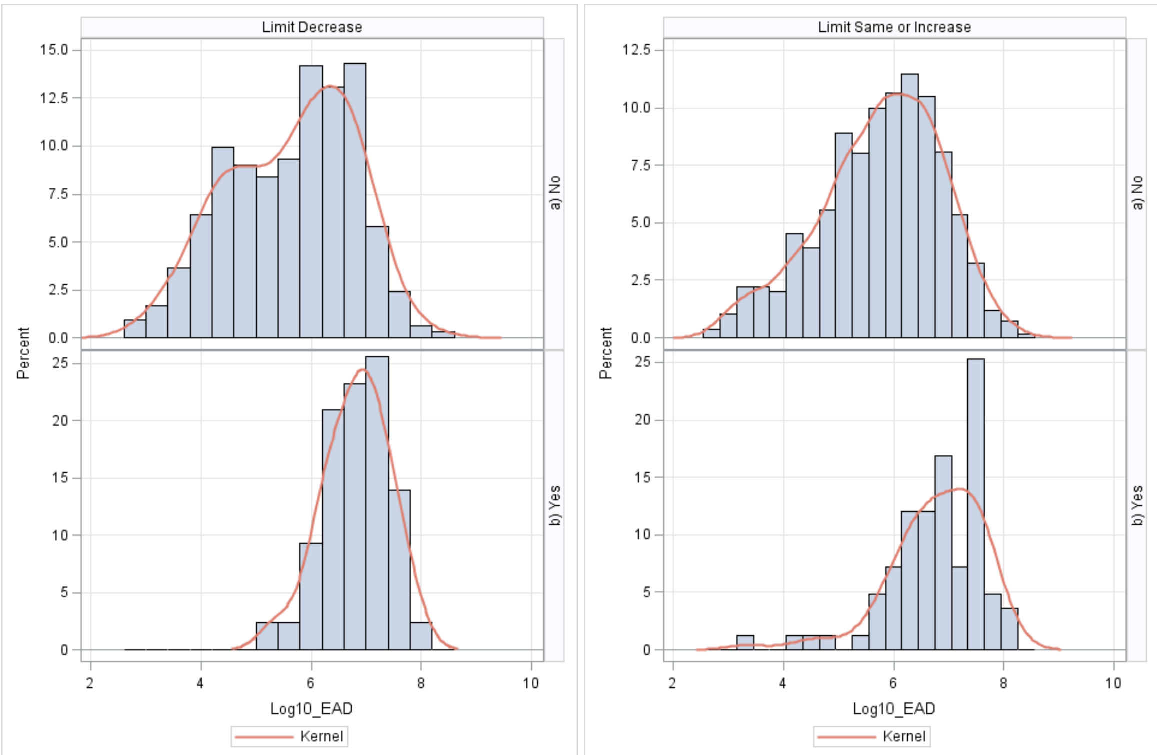


Figure A.20: Distribution of Log 10 EAD Conditional on Changes in Limit for Syndicated Indicator

Guarantee / Collateral Indicator

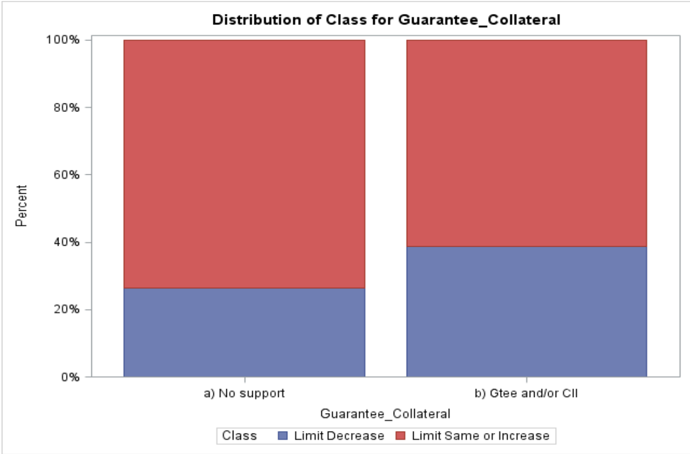


Figure A.21: Distribution of Limit Changes for Guarantee/Collateral Indicator

Support	N	Decrease	Same/Increase	Percent	WoE	IV	Gini	AUC
No support	1,060	781	279	49%	0.303			
Gtee and/or Cll	1,084	664	420	51%	-0.268			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.081**	**0.141**	**0.571**

Table A.11: Summary Table for Guarantee/Collateral Indicator

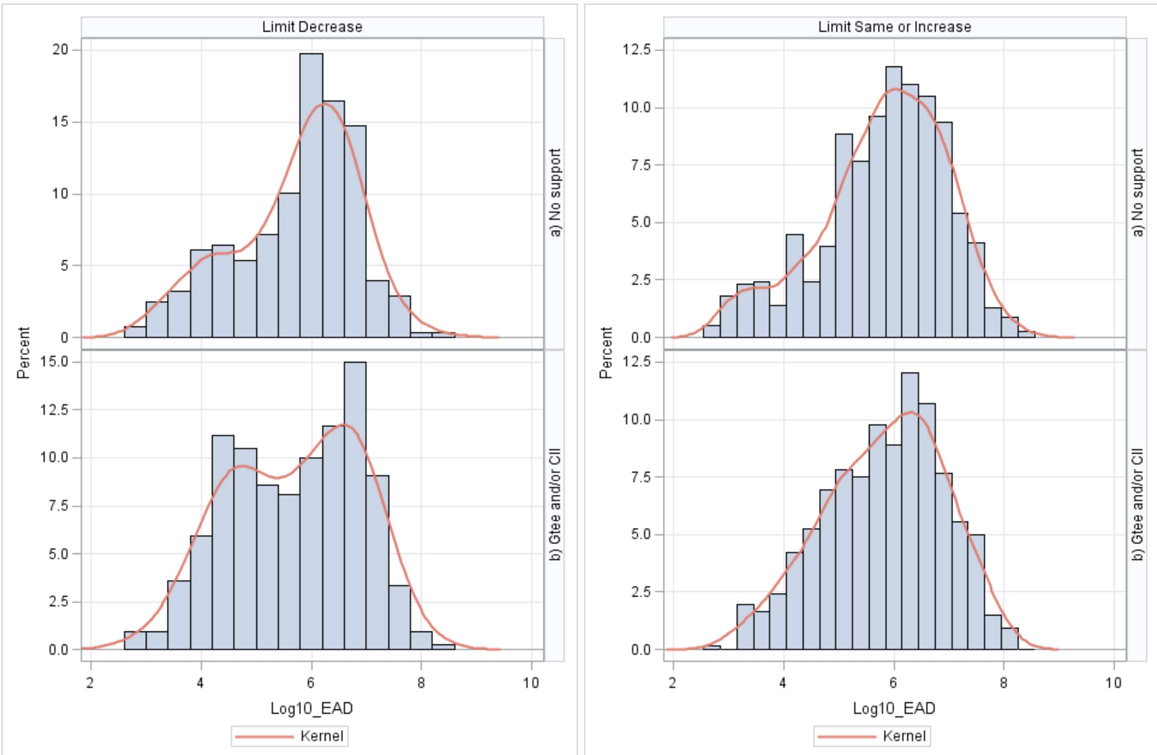


Figure A.22: Distribution of Log 10 EAD Conditional on Changes in Limit for Gtee/Collateral

Months to Maturity

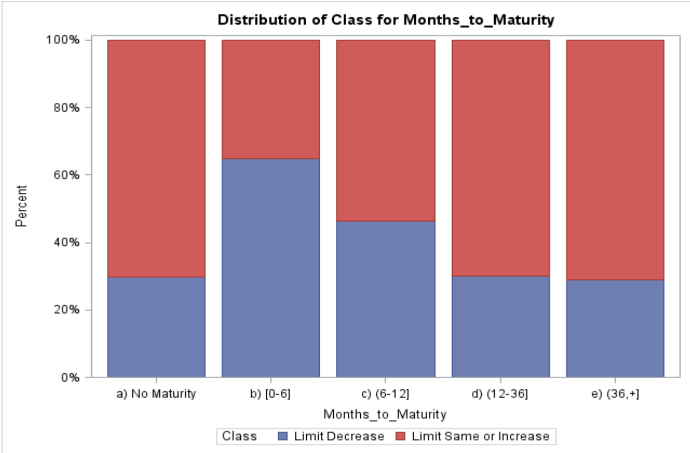


Figure A.23: Distribution of Limit Changes for Months to Maturity

Maturity	N	Decrease	Same/Increase	Percent	WoE	InfoValue	Gini	AUC
No Maturity	733	514	219	34%	0.127			
[0-6]	125	44	81	6%	-1.336			
(6-12]	121	65	56	6%	-0.577			
(12-36]	573	401	172	27%	0.120			
(36,+]	592	421	171	28%	0.175			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.152**	**0.132**	**0.566**

Table A.12: Summary Table for Months to Maturity

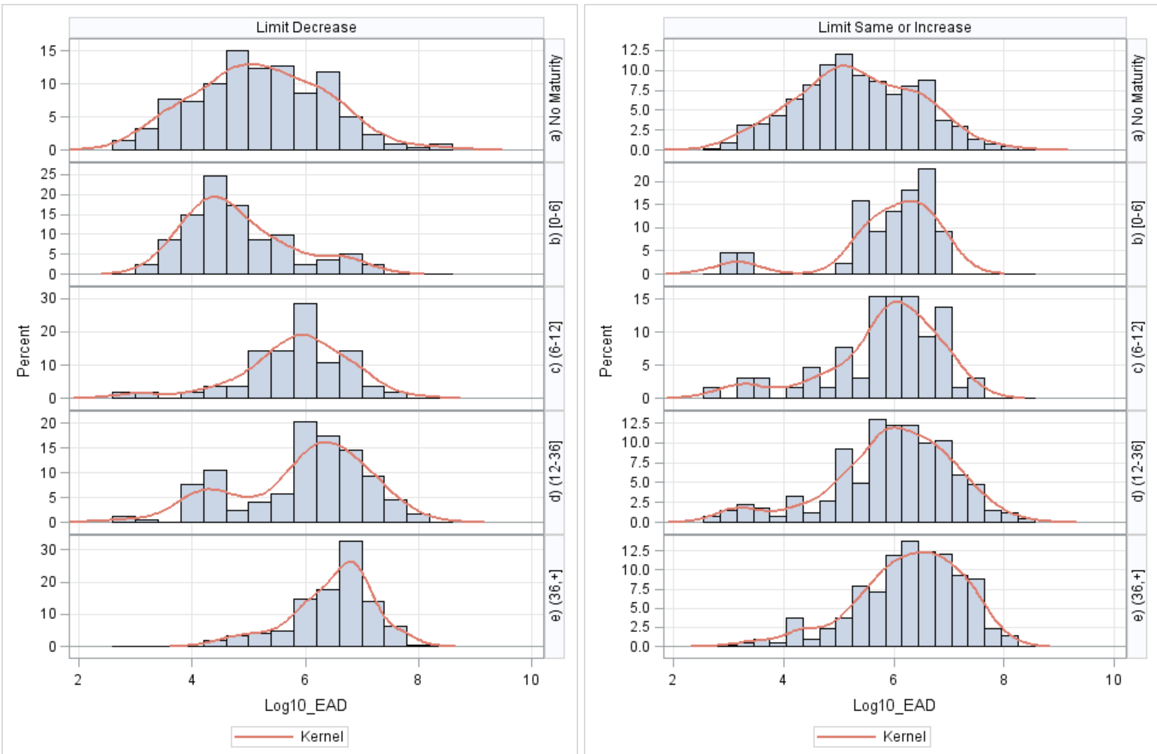


Figure A.24: Distribution of Log 10 EAD Conditional on Changes in Limit for Maturity

Seniority

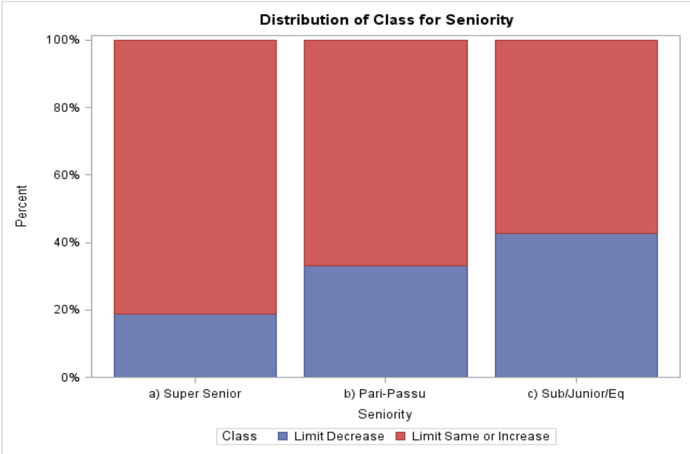


Figure A.25: Distribution of Limit Changes for Seniority

Seniority	N	Decrease	Same/Increase	Percent	WoE	InfoValue	Gini	AUC
Super Senior	208	169	39	10%	0.740			
Pari-Passu	1,765	1,178	587	82%	-0.030			
Sub/Junior/Eq	171	98	73	8%	-0.432			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.062**	**0.089**	**0.545**

Table A.13: Summary Table for Seniority

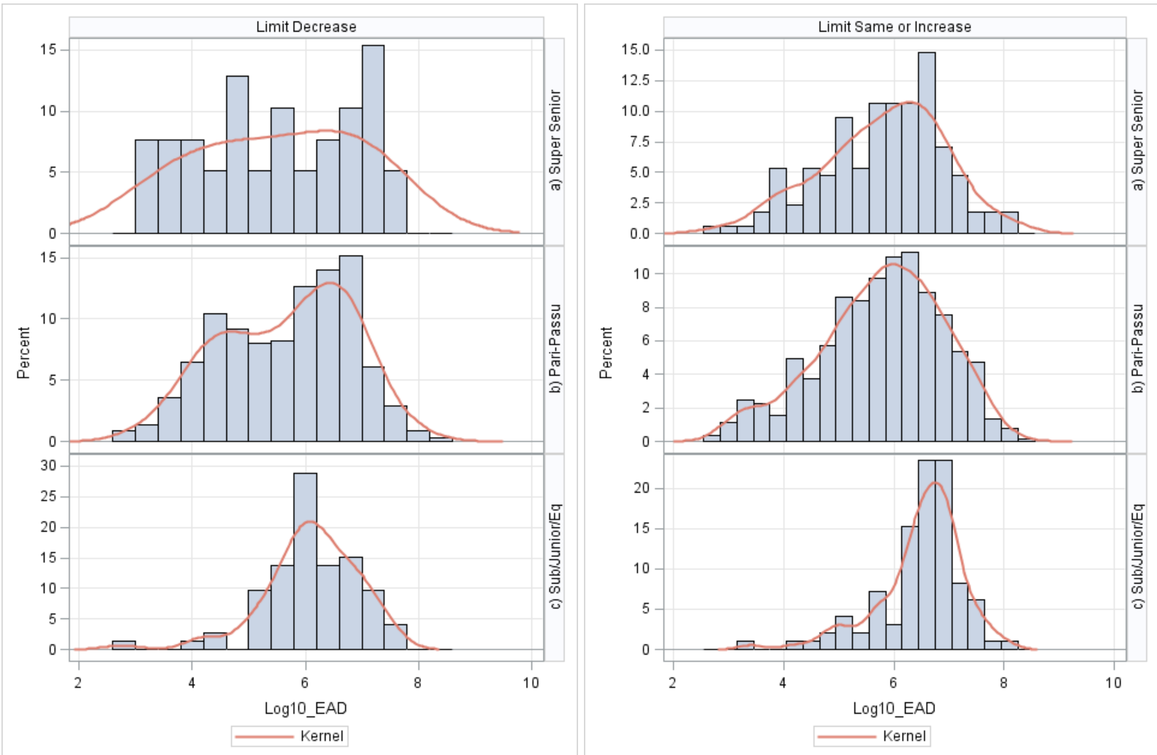


Figure A.26: Distribution of Log 10 EAD Conditional on Changes in Limit for Seniority

Jurisdiction

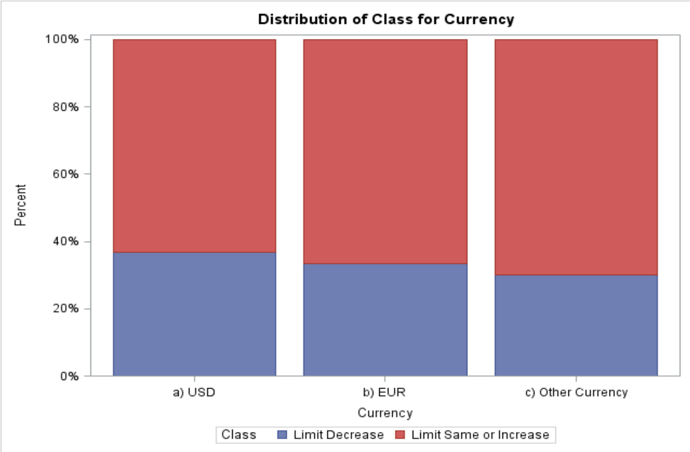


Figure A.27: Distribution of Limit Changes for Loan Currency

Currency	N	Decrease	Same/Increase	Percent	WoE	InfoValue	Gini	AUC
USD	374	236	138	17%	-0.190			
EUR	817	543	274	38%	-0.042			
Other	953	666	287	44%	0.116			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.013**	**0.060**	**0.530**

Table A.14: Summary Table for Loan Currency

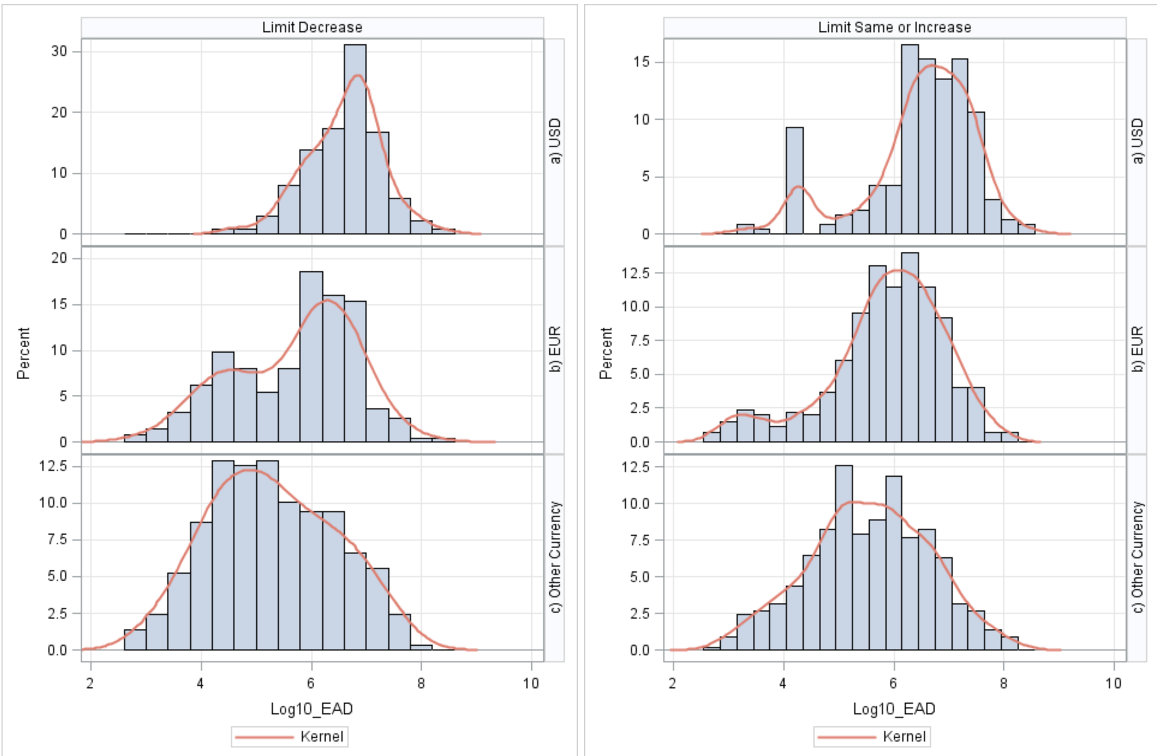


Figure A.28: Distribution of Log 10 EAD Conditional on Changes in Limit for Loan Currency

Economic State

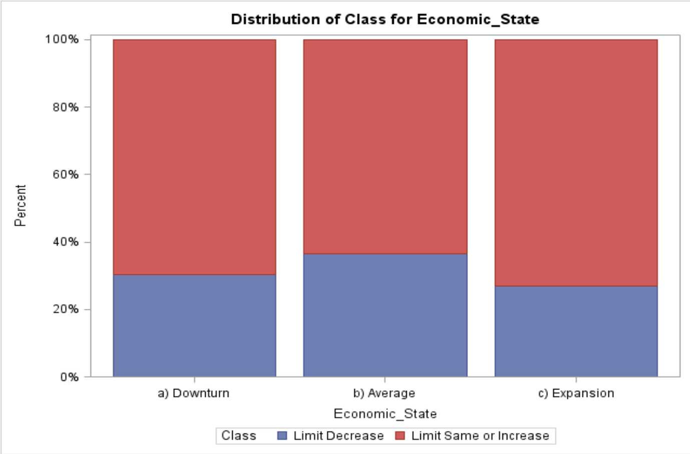


Figure A.29: Distribution of Limit Changes for Economic State

State	N	Decrease	Same/Increase	Percent	WoE	InfoValue	Gini	AUC
Downturn	1,007	702	305	47%	0.107			
Average	919	584	335	43%	-0.170			
Expansion	218	159	59	10%	0.265			
TOTAL	**2,144**	**1,445**	**699**	**100%**		**0.025**	**0.082**	**0.541**

Table A.15: Summary Table for Economic State

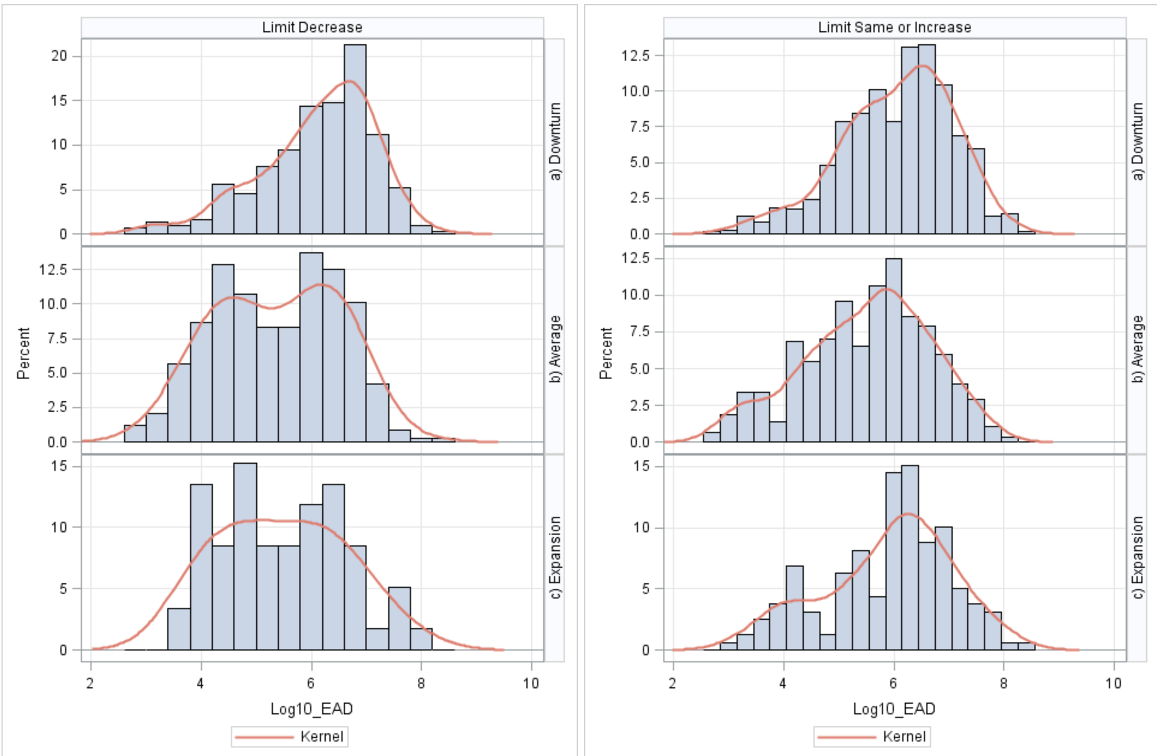


Figure A.30: Distribution of Log 10 EAD Conditional on Changes in Limit for Economic State

Appendix B

Covariate	Effect on Keeping Limit the Same or Increasing	IV	Gini	AUC
Jurisdiction	Weak jurisdictions are more likely to be granted a limit increase	0.056	0.069	0.535
Public/Private Company	No obvious risk signal	0.005	0.037	0.518
Leveraged Finance Deal	Leveraged deals more likely to be granted a limit increase	0.018	0.026	0.513
Lender Risk Rating	Facilities to unrated companies are more likely to be granted a limit increase	0.160	0.171	0.586
Operating Company	No obvious risk signal	0.000	0.006	0.503
Number of Loans	Counterparties with more facilites are more likely to be granted a limit increase	0.063	0.136	0.568
Limit	Facilities with a lower limit are more likely to be granted a limit increase	0.04	0.111	0.556
Zero Balance	Completely undrawn facilities are more likely to be granted a limit increase	0.049	0.083	0.541
Utilisation	Facilities either fully undrawn or fully undrawn are more likely to be granted a limit increase	0.13	0.188	0.594
Syndication	No obvious risk signal	0.000	0.004	0.502
Guarantee/Collateral	Non-supported facilities are more likely to be granted a limit increase	0.081	0.141	0.571
Time to Maturity	Facilities with longer maturity are more likely to be granted a limit increase	0.152	0.132	0.566
Seniority	Senior facilities are more likely to be granted a limit increase	0.062	0.089	0.545
Loan Currency	Other currencies are more likely to be granted a limit increase	0.013	0.060	0.530
Economic State	Downturn or Expansions are more likely to be granted a limit increase	0.025	0.082	0.541

Table B.1: Summary of Univariate Analysis for Changes in Limit

Covariate	Effect on EAD, Given Limit the Same/Increase	p-Value
Jurisdiction	Strong jurisdictions have higher EAD	0.0204
Public/Private Company	No obvious risk signal	$<.0001$
Leveraged Finance Deal	No obvious risk signal	0.0013
Lender Risk Rating	Unrated borrowers have lower EAD	$<.0001$
Operating/Holding Company	No obvious risk signal	$<.0001$
Number of Loans	Reversal in graphs and single regression	0.0018
Limit	Higher limit have higher EAD	$<.0001$
Zero Balance	No obvious risk signal	0.8981
Utilisation	Reversal in graphs and single regression	0.0168
Syndication	Syndicated deals have higher EAD	$<.0001$
Guarantee/Collateral	No obvious risk signal	0.8316
Time to Maturity	Loans with longer maturity have higher EAD	$<.0001$
Seniority	Reversal in graphs and single regression	$<.0001$
Loan Currency	USD and EUR have higher EAD	$<.0001$
Economic State	Downturn have higher EAD	$<.0001$

Table B.2: Summary of Univariate Analysis for EAD Given Limits Same or Increase

Covariate	Effect on EAD, Given Limit Decrease	P-Value
Jurisdiction	Strong jurisdictions have higher EAD	0.0003
Public/Private Company	No obvious risk signal	$<.0001$
Leveraged Finance Deal	Leverage deals have higer EAD	0.0051
Lender Risk Rating	No obvious risk signal	0.0006
Operating/Holding Company	Operating companies have higher EAD	0.0004
Number of Loans	Reversal in graphs and single regression	$<.0001$
Limit	Higher limit have higher EAD	$<.0001$
Zero Balance	Accounts with zero balance have higher EAD	$<.0001$
Utilisation	Accounts with lower utilisation have higher EAD	$<.0001$
Syndication	Syndicated deals have higher EAD	$<.0001$
Guarantee/Collateral	No obvious risk signal	0.6255
Time to Maturity	Loans with longer maturity have higher EAD	$<.0001$
Seniority	Junior loans have higher EAD	0.0052
Loan Currency	USD and EUR have higher EAD	$<.0001$
Economic State	Downturn have higher EAD	$<.0001$

Table B.3: Summary of Univariate Analysis for EAD Given Limits Decrease

Appendix C

Logistic Regression

Table C.1 displays the parameter estimates from the **stage one** logistic regression model.

Parameter	Level	DF	Estimate	Std Err	Wald	Pr >ChiSq
Intercept		1	1.7976	0.3531	25.9112	<.0001
Jurisdiction	Weak	1	0.763	0.1848	17.0542	<.0001
Leveraged Finance Deal	Yes	1	0.6703	0.2768	5.8627	0.0155
Lender Risk Rating	Rated	1	0.6855	0.144	22.6482	<.0001
Lender Risk Rating	Not Rated	1	1.1029	0.2093	27.767	<.0001
Operating Company	No/Missing	1	0.2766	0.11	6.3277	0.0119
Number of Loans	1	1	-0.3817	0.115	11.0184	0.0009
Number of Loans	2	1	-0.00394	0.1319	0.0009	0.9761
Log 10 Limit		1	-0.2229	0.0517	18.5775	<.0001
Zero balance	Zero Balance	1	0.9063	0.1441	39.5674	<.0001
Syndication	Yes	1	0.4184	0.2166	3.7305	0.0534
Guarantee/Collateral	Yes	1	-0.4159	0.1033	16.222	<.0001
Seniority	Super Senior	1	1.2112	0.1997	36.77	<.0001
Seniority	Sub/Jnr/Eq	1	0.0167	0.184	0.0082	0.9278
Economic State	Average	1	-0.3178	0.1081	8.6387	0.0033
Economic State	Expansion	1	0.141	0.1778	0.6293	0.4276

Table C.1: Parameter Estimates for **Stage One** Logistic Regression Model

Ordinary Least Squares

Table C.2 displays the parameter estimates from the **stage two** OLS regression model.

Parameter	Level	DF	Estimate	Std Err	Wald	Pr >ChiSq
Intercept		1	0.2601	0.0554	22.04	<.0001
Lender Risk Rating	Rated	1	-0.0101	0.025	0.16	0.6869
Lender Risk Rating	Not Rated	1	0.0829	0.0316	6.91	0.0086
Log10 Limit		1	0.9506	0.0096	9758.21	<.0001
Zero Balance	Zero Balance	1	-0.1037	0.026	15.95	<.0001
Syndication	Yes	1	-0.1233	0.0431	8.18	0.0042
Log10 Months to Maturity		1	0.0312	0.0144	4.72	0.0298

Table C.2: Parameter Estimates for **Stage Two** OLS Regression Model

Finite Mixture Model

Table C.3, C.4 and C.5 displays the parameter estimates from the **stage two** FMM regression model.

Component	Parameter	Level	Estimate	Std Err	z Value	Pr >|z|
1	Intercept		-0.06629	0.04311	-1.54	0.1241
1	Log10 Limit		1.0065	0.005389	186.79	<.0001
1	Zero Balance	Yes	-0.02976	0.01531	-1.94	0.0519
1	Zero Balance	No	0	.	.	.
1	Lender Risk Rating	Missing	-0.0449	0.02253	-1.99	0.0463
1	Lender Risk Rating	Rated	-0.03853	0.0252	-1.53	0.1262
1	Lender Risk Rating	Not Rated	0	.	.	.
1	Variance		0.007296	0.00129		

Table C.3: Parameter Estimates for First Normal Component **Stage Two** FMM Regression

Component	Parameter	Level	Estimate	Std Err	z Value	Pr >|z|
2	Intercept		0.07318	0.2546	0.29	0.7737
2	Log10 Limit		0.9262	0.03026	30.61	<.0001
2	Syndication	No	-0.2972	0.1521	-1.95	0.0507
2	Syndication	Yes	0	.	.	.
2	Variance		0.2625	0.02171		

Table C.4: Parameter Estimates for Second Normal Component **Stage Two** FMM Regression

Component	Parameter	Level	Estimate	Std Err	z Value	Pr >|z|
Probability	Intercept		0.6042	0.395	1.53	0.1261
Probability	Loan Currency	USD	-0.4075	0.3211	-1.27	0.2044
Probability	Loan Currency	EUR	-0.5418	0.2543	-2.13	0.0331
Probability	Loan Currency	Other	0	.	.	.
Probability	Log10 Time to Maturity		0.9217	0.1679	5.49	<.0001
Probability	Operating Company	Yes	-0.7968	0.2258	-3.53	0.0004
Probability	Operating Company	No	0	.	.	.
Probability	Economic State	Downturn	-0.2241	0.3939	-0.57	0.5693
Probability	Economic State	Average	-0.7277	0.385	-1.89	0.0587
Probability	Economic State	Expansion	0	.	.	.

Table C.5: Parameter Estimates for Probability Component **Stage Two** FMM Regression

Appendix D

The following SAS code, written in v9.3 (TS1M1), will refit the final model.

```
***================================================================;
***
*** Name:      04 Balance Models.sas
***
*** Date:      24/09/2015
***
*** Author:    MT
***
*** Purpose:   Fit the final models;
***
*** Step 1: Locations;
*** Step 2: Separate the modelling data into high and low;
*** Step 3: Estimate the OLS model;
*** Step 4: Fit FMM to decrease in limits;
*** Step 5: Chosen dummy logistic regression from stepwise selection;
*** Step 6: Score the final predicted model;
***
***================================================================;


***================================================================;
*** Step 1: Locations;
LIBNAME output "c:\output";

***Output destination;
ODS LISTING CLOSE;
ODS HTML;
***================================================================;


***================================================================;
*** Step 2: Separate the modelling data into high and low;
DATA high low;
   SET output.loanTable2;
   IF Flag_L_td_div_L_t3='b) (L_td / L_t)>=1' THEN OUTPUT high;
   IF Flag_L_td_div_L_t3='a) (L_td / L_t)< 1' THEN OUTPUT low;
RUN;
***================================================================;


***================================================================;
*** Step 3: Estimate the OLS model;
PROC GENMOD DATA=high NAMELEN=32 PLOTS=NONE;

   /*Classing for categorical variables*/
   CLASS
   Lender_Borrower_Risk_Rating4(REF='1) Missing')
   Syndicated_Indicator2(REF='a) No')
   Flag_B_t(REF='a) Pos Balance')
   / PARAM=REF;

   /*Linear predictor*/
   MODEL log10_B_td =
   log10_L_t
   log10_MonthsToMaturity7
   Syndicated_Indicator2
   Lender_Borrower_Risk_Rating4
```

```
    Flag_B_t
    / DIST=NORMAL LINK=IDENTITY TYPE3;

    /*Model output for scoring*/
    STORE ModelScoring;

RUN;
***=================================================================;


***=================================================================;
*** Step 4: Fit FMM to decrease in limits;
PROC FMM DATA=low NOITPRINT NAMELEN=32 PARMSTYLE=LABEL;

    /*Classing for categorical variables*/
    CLASS
    Flag_B_t
    Lender_Borrower_Risk_Rating4
    Syndicated_Indicator2
    Loan_Currency2
    Operating_Company_Indicator3
    EconomicState;

    /*First normal mixing component linear predictor*/
    MODEL log10_B_td=
    log10_L_t
    Flag_B_t
    Lender_Borrower_Risk_Rating4
    /DIST=NORMAL;

    /*Second normal mixing component linear predictor*/
    MODEL log10_B_td=log10_L_t
    Syndicated_Indicator2
    /DIST=NORMAL ;

    /*Probability model*/
    PROBMODEL
    Loan_Currency2
    log10_MonthsToMaturity7
    Operating_Company_Indicator3
    EconomicState;

    /*Model output for scoring*/
    ODS OUTPUT ParameterEstimates=ParameterEstimates MixingProbs=MixingProbs;

RUN;

*** Prepare for scoring;
DATA ParameterEstimates2;

    /*Allocate the length of some variables*/
    ATTRIB Parameter Effect LENGTH=$1000;

    /*Combine the parameter */
    SET ParameterEstimates
    (KEEP=ModelNo Parameter Effect Estimat
    WHERE=(Parameter~='Variance') IN=a)
```

```
      MixingProbs       (KEEP=        Parameter Effect Estimate
      WHERE=(Parameter~='Variance') IN=b);

      /*Assign probability model as model 3*/
      IF b THEN ModelNo=3;

      /*tailor the Estimate and Parameter variables*/
      Estimate2=SUM(INPUT(Estimate,8.6),0);
      Parameter2=Parameter;
      Parameter2=LEFT(TRANWRD(Parameter,COMPRESS(Effect),''));


      /*Combine Effect and Estimate to create linear predictors*/
      ATTRIB linpred LENGTH=$1000;
      IF Effect='Intercept' THEN linpred=Estimate2;
      ELSE IF Effect
      IN('log10_L_t','log10_MonthsToMaturity7','log10_L_t_sq') THEN
      linpred=COMPRESS(Effect || '*' || Estimate2);
      ELSE linpred='(' || COMPRESS(Effect) ||   '="' || TRIM(Parameter2) || '") * '
      || COMPRESS(Estimate2);


      /*Finalise labels and drop unnecessary variables*/
      ATTRIB _ALL_ LABEL='';
      DROP Parameter Estimate;

   RUN;

   /*Create the 3 linear predicors*/
   DATA ParameterEstimates3;

      /*Allocat the length of linear predictors*/
      ATTRIB Run_linpred LENGTH=$1000;
      RETAIN Run_linpred;

      SET ParameterEstimates2;

      /*Create linear predictor for each of the 3 models*/
      BY ModelNo;
      IF FIRST.ModelNo THEN Run_linpred=linpred;
      ELSE Run_linpred=TRIM(LEFT(Run_linpred)) || '+' || linpred;
      IF LAST.ModelNo THEN CALL SYMPUT(COMPRESS('linpred'||ModelNo),Run_linpred);
   RUN;
   ***==================================================================;


   ***==================================================================;
   *** Step 5: Chosen dummy logistic regression from stepwise selection;
   PROC LOGISTIC DATA=output.loantable2 PLOT=ROC NAMELEN=32;

      /*Classing for categorical variables*/
      CLASS
      Fitch_RR_Group2(REF='a) Regular')
      Leveraged_Finance_Indicator3(REF='a) No')
      Lender_Borrower_Risk_Rating4(REF='1) Missing')
      Operating_Company_Indicator3(REF='a) Yes')
      Flag_NumLoans3(REF='c) 3+')
```

```
    Flag_B_t(REF='a) Pos Balance')
    Syndicated_Indicator2(REF='a) No')
    Guarantee_or_Collateral2(REF='a) No support')
    Seniority_Code3(REF='b) Pari-Passu')
    EconomicState(REF='a) Downturn')
    /PARAM=REF;

    /*Linear predictor*/
    MODEL Num_L_td_div_L_t3(EVENT='1') =
    log10_L_t
    Fitch_RR_Group2
    Leveraged_Finance_Indicator3
    Lender_Borrower_Risk_Rating4
    Operating_Company_Indicator3
    Flag_NumLoans3
    Flag_B_t
    Syndicated_Indicator2
    Guarantee_or_Collateral2
    Seniority_Code3
    EconomicState

    /BINWIDTH=0;

    /*Model output for scoring*/
    OUTPUT OUT=scoredModel PREDICTED=Phat;
RUN;
***==================================================================;

***==================================================================;
*** Step 6: Score the final predeicted model;

*** Score the high model;
PROC PLM source=sasuser.model2;
    SCORE DATA=scoredModel OUT=scoredModel2;
RUN;

*** Score the low model;
DATA scoredModel3;
    SET scoredModel2;

    /*Linear predictors for the FMM*/
    log10_L_t_sq=log10_L_t*log10_L_t;
    linpred1=&linPred1;
    linpred2=&linPred2;
    linpred3=&linPred3;
    pi=EXP(linpred3)/(1+EXP(linpred3));
    PredLow=pi*linpred1 + (1-pi)*linpred2;

    *** Predicted and residuals;
    ATTRIB FinalPred  LABEL='Predicted Value';
    ATTRIB FinalResid LABEL='Residual (Observed - Predicted)';
    FinalPred=PredLow * (1-Phat) + Predicted*Phat;
    FinalResid=log10_B_td-FinalPred;
RUN;
***==================================================================;
```

Appendix E

The following R code will refit the final model.

<u>Details of R Version:</u>
R version 3.2.2 (2015-08-14) – "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

<u>Details of R Packages:</u>
base (version 3.2.2)
dplyr (version 0.4.3)
flexmix (version 2.3-13)
sas7bdat (version 0.5)

```
#==============================================================
#
# Name:   04 Balance Models.R
#
# Date:   24/09/2015
#
# Author: MT
#
# Purpose:Fit the final models
#
# Step 1: Packages and Locations
# Step 2: Separate the modelling data into high and low
# Step 3: Estimate the OLS model
# Step 4: Fit FMM to decrease in limits
# Step 5: Chosen dummy logistic regression from stepwise selection
# Step 6: Score the final predicted model
#
#==============================================================


#==============================================================
# Step 1: Packages and Locations

# Packages
library(flexmix)
library(sas7bdat)
library(dplyr)

# Location
outputLocation='C:/Users/Mark/Google Drive/Uni/Sem6/STAT825/01 SAS Code/02 Output'
#==============================================================


#==============================================================
# Step 2: Separate the modelling data into high and low

# Read the data
loantable2=read.sas7bdat(paste(outputLocation,'/loanTable2.sas7bdat',sep=''))

# Variables to keep
AllVars=c(
 'Num_L_td_div_L_t3',
 'DA_LOAN_ID',
 'Fitch_RR_Group2',
 'Leveraged_Finance_Indicator3',
 'Operating_Company_Indicator3',
 'Syndicated_Indicator2',
 'Guarantee_or_Collateral2',
 'log10_MonthsToMaturity7',
 'Seniority_Code3',
 'Loan_Currency2',
```

```
  'Lender_Borrower_Risk_Rating4',
  'Debt_Senior_Percentage3',
  'log10_L_t',
  'Flag_B_t',
  'EconomicState')

# High
high=subset(loantable2,
        Flag_L_td_div_L_t3=='b) (L_td / L_t)>=1',
        select=c('log10_B_td',AllVars))

# Low
low=subset(loantable2,
        Flag_L_td_div_L_t3=='a) (L_td / L_t)< 1')
#==============================================================


#==============================================================
# Step 3: Estimate the OLS model

# OLS Model variables
OlsVars=c('log10_L_t',
       'log10_MonthsToMaturity7',
       'Syndicated_Indicator2',
       'Lender_Borrower_Risk_Rating4',
       'Flag_B_t')

# OLS model
formula=paste('log10_B_td',paste(OlsVars,collapse=' + '),sep=' ~ ')
fit=lm(data=high, formula=formula)
coefsTable=data.frame(summary(fit)$coefficients)
View(coefsTable)
View(anova(fit))
View(summary(fit)$coefficients)
logLik(fit)
extractAIC(fit, k=2)
#==============================================================


#==============================================================
# Step 4: Fit FMM to decrease in limits

# Components
Model.sel = FLXMRglmfix(nested = list(formula = c( ~ log10_L_t+
                                      Flag_B_t+
                                      Lender_Borrower_Risk_Rating4,
                                     ~ log10_L_t+
                                      Syndicated_Indicator2),
                           k =c(1,1)))
# Probability model
Conc.sel  = FLXPmultinom(~ Loan_Currency2+
```

```
                        log10_MonthsToMaturity7+
                        Operating_Company_Indicator3+
                        EconomicState)

# Fit the finite mixture model
fmm=flexmix(log10_B_td~1,
          data=low,
          model=Model.sel,
          concomitant=Conc.sel)
rfmm <-refit(fmm)

# View the results
summary(rfmm)
rfmm@concomitant
AIC(fmm)
BIC(fmm)
fmm@logLik*-2
#==============================================================


#==============================================================
# Step 5: Chosen dummy logistic regression from stepwise selection

# Variables for logistic regression
LogisticVars=c('log10_L_t',
            'Fitch_RR_Group2',
            'Leveraged_Finance_Indicator3',
            'Lender_Borrower_Risk_Rating4',
            'Operating_Company_Indicator3',
            'Flag_NumLoans3',
            'Flag_B_t',
            'Syndicated_Indicator2',
            'Guarantee_or_Collateral2',
            'Seniority_Code3',
            'EconomicState')

# Logistic regression formula
logisticFormula=paste('Num_L_td_div_L_t3',paste(LogisticVars,collapse=' + '),sep=' ~ ')

#Logistic regression
logistic=glm(formula=logisticFormula,
           family=binomial(link='logit'),
           data=loantable2)
#==============================================================


#==============================================================
# Step 6: Score the final predicted model

# Score the Ols
ols.linpred=data.frame(ols.linpred=predict(fit,loantable2))
```

```
View(ols.linpred)

# Score the fmm
fmm.scored.linpred=predict(fmm,loantable2)
fmm.scored.prior=data.frame(prior(fmm,loantable2))
rownames(fmm.scored.prior)=NULL

# Score the logistic regression
logistic.logit=data.frame(logistic.logit=predict(logistic,loantable2))
logistic.prob=logistic.logit %>%
  mutate(logistic.prob=exp(logistic.logit)/(1+exp(logistic.logit)))

# Collate the final scored dataset
scored=data.frame(loantable2,
         ols.linpred,
         fmm.scored.linpred,
         fmm.scored.prior,
         logistic.prob)

# Calculate the final fitted values
scored2=scored %>%
  mutate(fmm.scored.linpred.overall=Comp.1*X1 + Comp.2*X2) %>%
  mutate(predicted= ols.linpred*logistic.prob +
           fmm.scored.linpred.overall*(1-logistic.prob))
#================================================================
```

Bibliography

Agarwal, S., Ambrose, B., and Liu, C. (2006). Credit lines and credit utilization. *Journal of Money, Credit and Banking*, 1(28):1–22. An optional note.

Anderson, R. (2007). *The Credit Scoring Toolkit*. Oxford.

Apostolik, R., Donohue, and C. Went, P. (2009). *Foundations of Banking Risk, An Overview of Banking, Banking Risks and Risk-Based Banking Regulation*. GARP.

APRA (2015). Aps 113: Capital adequacy: Internal ratings-based approach to credit risk (http://www.apra.gov.au/adi/prudentialframework/documents/basel-iii-prudential-standard-aps-113-(january-2013).pdf). [Online; accessed 20/08/2015].

Araten, M. and Jacobs, M. (2001). Loan equivalents for revolving credits and advised lines. *The RMA Journal*, 4(2):34–39. An optional note.

Bank of International Settlements (2005). An explanatory note on the basel ii irb risk weight functions. *Bank for International Settlements*.

Bank of International Settlements (2014). History of the basel committee (http://www.bis.org/bcbs/history.htm). [Online; accessed 20/09/2015].

BIS (2006). International convergence of capital measurement and capital standards a revised framework comprehensive version. *Bank for International Settlements*.

Brown, I., editor (2011). *Regression Model Development for Credit Card Exposure At Default (EAD) using SAS/STAT and SAS Enterprise Miner 5.3, Proceedings of the 2011 SAS Global Forum*.

Chapman, R. (2006). *Simple Tools and Techniques for Enterprise Risk Management*. Wiley Finance.

Committee of European Banking Supervisors (2006). Guidelines on the implementation, validation and assessment of advanced measurement (ama) and internal ratings based (irb) approaches (https://www.eba.europa.eu/cebs-archive/publications/consultations/2006/cp10-revised). [Online; accessed 26/08/2015].

Crouchy, M., Galai, D., and Mark, R. (2006). *The Essentials of Risk Management*. McGraw Hill.

Financial Conduct Authority (2007). Comments on ead models (www.fsa.gov.uk/pubs/international/ead_models.pdf). [Online; accessed 20/08/2015].

Financial Conduct Authority (2014). Ifpru 4.8 internal ratings based approach: Own estimates of exposure at default (ead) (https://fshandbook.info/fs/html/handbook/ifpru/4/8). [Online; accessed 26/08/2015].

Global Credit Data (2015). Global credit data analysis homepage (http://www.globalcreditdata.org/home.html). [Online; accessed 20/08/2015].

Good, I. J. (1950). *Probability and the Weighting of Evidence*. Charles Griffith, London.

Good, I. J. (1983). *Weight of Evidence: A Brief Survey. In Bayesian Statistics 2, J. M Bernardo, M. H. DeGrot, D . V. Lindley and A . F. M. Smith Eds.* North-Holland Amsterdam.

Heitfield, E. and Barger, N. (2003). Treatment of double-default and double-recovery effects for hedged exposures under pillar i of the proposed new basel capital accord. *Federal Reserve White Paper*.

Investopedia (2015). Definition of a line of credit - loc (http://www.investopedia.com/terms/l/lineofcredit.asp). [Online; accessed 24/08/2015].

Jacobs, M. (2010). An empirical study of exposure at default. *Journal of Advanced Studies in Finance*, 1:31–59. An optional note.

Jacobs, M. and Bag, P. (2011). What do we know about exposure at default on contingent credit lines? - a survey of the literature, empirical analysis and models. *The Journal of Advanced Studies in Finance*, 2(3):26–46. An optional note.

Jacobs, M. and Bag, P. (2012). Parsimonious exposure-at-default modeling for unfunded loan commitments. *The Journal of Risk Finance*, 13(1):77–94. An optional note.

Jimènez, G., Lopez, J., and Saurina, S. (2009). Ead calibration for corporate credit lines. *Federal Reserve Bank Of San Francisco Working Paper Series*.

Jorion, P. (2001). *Value at Risk: The New Benchmark for Managing Financial Risk*. New York: McGrath Hill.

Lam, J. (2003). *Enterprise Risk management: From Incentives to Controls*. Wiley Finance.

Leow, M. and Crook, J., editors (2013). *A Two Stage Mixture Model for Predicting EAD*.

Loukoianova, E., Neftci, S., and Sharma, S. (2013). Pricing and hedging of contingent credit lines. *IMF Working Paper*, 2(3). An optional note.

Mantel, K. (2012). *"Race to Default Determination of Exposure at Default", Proceedings of the 2012 PECDC Conference*.

McLachlan, G. and Peel, D. (2004). *Finite Mixture Models*. Wiley Series in Probability and Statistics. Wiley.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*.

Moral, G. (2006). *Chapter 10 in EAD Estimates for facilities with explicit limits, in "The Basel II Risk Parameters: Estimation, Validation, and Stress Testing"*. Springer.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc.*, (135):370–384.

Oxford English Dictionary (2015). Oxford english dictionary (http://www.oed.com/). [Online; accessed 19/09/2015].

Peirson, G., Brown, R., Easton, S., and Howard, P. (2002). *Business Finance, Eight Edition.* Mc Graw Hill.

Qi, M. (2009). Exposure at default of unsecured credit cards. *Working paper: U.S. Office of the Comptroller of the Currency*, 2(3). An optional note.

Rutkowski, M. and Tarca, S. (2014). Regulatory capital modelling for credit risk. *Working Paper*.

Siddiqi, N. (2006). *Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring.* John Wiley and Sons Inc.

Sufi, A. (2009). Bank lines of credit in corporate finance: An empirical analysis. *Review of Financial Studies*, 3(22):1057–1088. An optional note.

Taplin, R., To, M., and Hee, J. (2007). Modeling exposure at default, credit conversion factors and the basel ii accord. *Journal of Credit Risk*, 3(2):75–84.

Tong, E., Mues, C., Brown, I., and Thomas, L., editors (2015). *Exposure at Default models with and without the Credit Conversion Factor.*

Vasicek, O. (2002). Loan portfolio value. *Risk.net*, (2):160–162.

Witzany, J. (2011). European financial and accounting journal. *Working paper: U.S. Office of the Comptroller of the Currency*, 6(4):20–48. An optional note.

WorldBank (2015). World bank gdp (http://data.worldbank.org/indicator/ny.gdp.mktp.kd.zg/countries/1w?display=graph). [Online; accessed 24/08/2015].

Yeh, A., Twaddle, J., and Frith, M. (2005). Basel ii: A new capital framework. *Reserve Bank Bulletin*, 68(2):4–15.