Valuation of initial margin and model risk

MB Seitshiro

orcid.org / 0000-0001-9557-3714

Thesis accepted for the degree *Doctor of Philosophy in Business Mathematics and Informatics* at the North-West University

Promoter: Prof HP Mashele

Graduation December 2020

20146272

Declaration

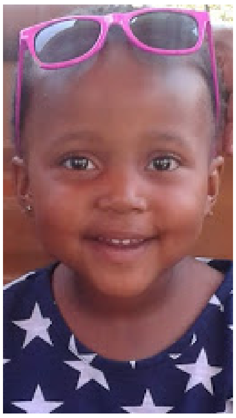
I hereby declear that the contents of this thesis is a record of my original work and have not been submitted in whole or in part for consideration of any other degree or qualification in the North-West University or any other university, except where citations are made to the work of others and acknowledgements indicates otherwise.

16 March 2020

Modisane Bennett Seitshiro | Date

Dedication

...Modimo Oreokametse...
...Omnipresence is God...



To "**Ore**" - life was short, but memories are everlasting. " **Montsi**"

I dedicate this thesis to my Family.
To my wife Tebogo, my children Mokgele, Botshelo, Kabelo for their endless love, support and motivation.

"Great is our Lord and mighty in power, his understanding has no limit."

Psalm 147:5.

Acknowledgements

I sincerely express my profound appreciation to Professor Hopolang Phillip Mashele as my supervisor for his valuable guidance, support and suggestions thoughout this research work. I thank him for his commitment and vision for future research. Because of his constant inspiration the four articles, which form part of this thesis, were submitted in accredited jounals for publication considerations. Prof, it is an honour working with you.

My deepest gratitude goes to my family. To my lovely loving wife Tebogo for her constant support and prayers, thank you - I love you. My inspirations to continue learning come from my children Mokgele, Botshelo and Kabelo for their inquisitive minds, wondering, and ready to poke around and figure somethings out. To my late father for providing unstinting support in acquisition of our knowledge and for always been spiritually present. To my daughter in heaven Oreokametse, *who has motivated me to always persevere in enhancing knowledge.* My acknowledgement also goes to my loving mother Maele Maboditsana for her spiritual touch in my way of life. To my Brother Dikgang Nelson and Sister Phindile Ruth for their moral support and everyday optimism of intellectual growth. To my nieces Maele and Tshiamo and nephew Sanele for being eager to learn more. My sincere gratitude goes to my parents-in-law Koko Sarah and Rremogolo David Molope for their endless inspirations.

I am highly grateful to the following people and institutions:

- To the Centre for Business Mathematics and Informatics (BMI) for giving me the opportunity to do the research with them.
- Warm appreciation goes to the director of the centre Prof. Riaan de Jongh for his support.
- To the dean's office in Faculty of Natural and Agricultural Sciences, especially

Prof. Eno Ebenso for his encouragement to pursue the Phd studies, and his continued support thereafter.

- to Banking Sector Education and Training Authority (Bankseta) for their financial support.
- to University Capacity Development Grant (UCDP) for assisting with conference and workshop attendance. Also, for assistance in staff buy-out, it was worth having it indeed.
- to all my colleagues in the School of Mathematical and Statistical Sciences for their morale support and words of encouragement.

Above all, I would like to thank my Almighty God for giving me the strength, good health, wisdom and ability to embark on this voyage of research work, to persist and complete it. He deserve to be praised as his blessings are miraculous.

Authorship

The entire work presented in this thesis is my own. I personally was engaged and worked in all the creation of this thesis which include the implementation of all computer codes, data analysis and entire chapters as a corresponding and lead author. My supervisor Professor Hopolang Phillip Mashele contributed enormously with the comments as the co-author of the manuscripts that came out of the chapters.

The content of Chapter 5 titled "Inappropriate parameter estimator" is written as an Econometrics research article and it was published in the Journal of Cogent Economics & Finance (Seitshiro & Mashele 2020). The contents of Chapter 2, Chapter 3 and Chapter 4 were submitted in peer-review journals for consideration.

Executive Summary

The research work of the thesis focuses on two themes: the valuation of initial margin and model risk quantification. The first theme addresses matters arising from the valuation of initial margin for over-the-counter derivatives in the real market with outstanding gross notional amount smaller than two billion, but acknowledging the present work for high outstanding gross notional amount in developed financial institutions. The initial margin requirement for uncleared derivatives is well in place for developed countries and for the high risk financial institutions, but the spill-over of the financial crises from developing institutions, the gradual phasing in of initial margin and the impact thereof are yet to be known. Hence, the major interest is drawn to this work.

To mitigate risk due to unforeseen financial markets turmoil, we propose a bootstrap initial margin valuation process that can be applicable during normal and stressed financial markets. The proposed parametric bootstrap method is in favour of the bootstrap initial margin (BIM) amounts for the simulated and real datasets. These BIM amounts are reasonably exceeding the traditional initial margin amounts whenever the significance level increases. The proposed valuation of initial margin reduces spill-over effects by ensuring that collateral, such as initial margin, is available to offset losses caused by the default of an over-the-counter derivatives.

The second theme of the thesis addresses three components of model risk quantification: model risk due to inappropriate statistical distribution; model misspecification; and inappropriate parameter estimation methods. Inappropriate statistical distribution was assessed using four bootstrap confidence interval techniques. The modified hybrid percentile bootstrap method was a superior technique because it reveals that with the same sample size and very small simulation iterations the other confidence methods produces similar goodness-of-fit results but completely different and insignificant performance measures. By way of illustrating the model misspecification for

some credit risk models, we carry out quantitative analysis of two specific statistical predictive response models using simulated balanced dataset and Taiwan credit card default dataset. The maximum likelihood estimation technique is employed for parameter estimation and inference, precisely the goodness of fit and model performance assessments. The binary logistic regression technique for the balanced datasets reveals prominent goodness of fit and performance measures as opposed to the complementary log-log technique. To deal with model risk due to parameter estimation methods, several statistical and mathematical numerical methods for determining the parameter values are utilized for predicting probability of default through binary logistic regression model and determining optimum parameters that minimize the objective model's cost function. The Mini-Batch Gradient Descent method is revealed to be the better parameter estimator among the chosen parameter estimation methods.

The banking industry utilises models on a daily basis. This research will assist banks to manage model risk better, particularly the selection of an appropriate statistical distribution, the identification of model misspecification and the quantification of an appropriate parameter estimation method. Researchers and practitioners will be able to compare the results of model risk techniques and choose the optimum method for their current market conditions. However, this practice needs to be validated and exercised regularly as the financial markets evolves rapidly.

Keywords: Binary logistic regression; bootstrap; complementary log-log; credit risk; inappropriate statistical distribution; initial margin; model misspecification; parameter estimation; probability of default.

Abbreviations

AIC	Akaike information criterion
ANN	Artificial neural network
ARCH	Autoregressive conditional heteroscedasticity
BCBS	Basel committee on banking supervision
BCP	Bias-corrected percentile
BCV	Bootstrap coefficient of variation
BFGS	Broyden, Fetcher, Goldfarb and Shanno
BGD	Batch gradient descent
BIC	Bayesian information criterion
BIM	Bootstrap initial margins
BLRM	Binary logistic regression model
BMCS	Bootstrap Monte-Carlo simulation
BP	Basic percentile
CCP	Central counterparties
CDF	Cumulative distribution function
CG	Conjugate gradient
Cloglog	Complementary log-log
Dev	Deviance
DMR	Distribution model risk
EAD	Exposure at default
EDF	Empirical distribution function
EM	Expectation maximization
ES	Expected shortfall
ESRM	Exponential spectral risk measures
FTSE100	Financial Times Stock Exchange with share index of the 100 companies listed on the London Stock Exchange
FVA	Funding valuation adjustment

GEV	Generalised extreme value
GoF	Goodness of fit
IM	Initial margin
IMR	Initial margin requirement
IOSCO	International organization of securities commissions
IRLS	Iteratively reweighted least squares
ISDA	International swaps and derivatives association
LDA	Linear discriminant analysis
LGD	Loss given default
LLR	Logarithm likelihood ratio
LM-BFGS	Limited-Memory Broyden, Fetcher, Goldfarb and Shanno
LR	Likelihood ratio
LRM	Logistic regression model
LTCM	Long-term capital management
MBGD	Mini-batch gradient descent
MC	Monte-Carlo simulation
MHP	Modified hybrid percentile
MLE	Maximum likelihood estimation
MR	Model risk
MVA	Margin valuation adjustment
NM	Nelder-Mead
NR	Newton-Raphson
OLS	Ordinary least squares

OTCD	Over-the-counter derivative
pc	Percentage change
PD	Probability of default
PDF	Probability density function
PFE	Potential future exposure
PRP	Polak and Ribiere parameter
PW	Powell's method
ROC	Receiver operating characteristic
RV	Random variable
SGD	Stochastic gradient descent
SHP	Standard hybrid percentile
SIMM	Standard initial margin model
TN	Truncated Newton
VaR	Value-at-Risk

Notations

S_t	the price of the over-the-counter derivatives at time t
R_t	the return of an over-the-counter derivative at time t
α	significance level
$\Phi(.)$	standard normal distribution
θ & γ	parameters of interest
$\hat{\theta}$ & $\hat{\gamma}$	statistic of interest
$\hat{\theta}^*$	statistic of interest from bootstrap sample
$\hat{\theta}^*(.)$	the mean of the bootstrap replicates
$\sigma_B^*(.)$	bootstrap standard error
B	bootstrap sample size
n	sample size
$I(.)$	indicator function
σ_t	standard deviation of the OTCD returns at time t
μ_t	mean of the OTCD returns at time t
z_t	standard normal variate
$\pi(.)$	probability of default
ρ	correlation coefficient
$\hat{\mathcal{U}}$	upper confidence level calculated from a bootstrap method
$\hat{\mathcal{V}}$	estimated credit value-at-risk
$\mathcal{L}$	likelihood function of the objective model
$\mathcal{A}$	accuracy of the objective predictive model
$\mathcal{P}$	precision of the model
$\mathcal{R}$	sensitivity or recall of the model
$\mathcal{S}$	specificity of the model
$\mathcal{F}_1$	harmonic mean of precision and recall for the model

Contents

List of Figures

List of Tables

1. Introduction

In this chapter the following concepts are introduced: Section 1.1 gives an overview of valuation of Initial Margin (IM) and the motivation for the study around IM. In Section 1.2, an overview and motivation for the quantification of model risk is given. In Section 1.3, the aims and objectives of the thesis are stated. Section 1.4 briefly provides details of the software, hardware and methods utilised for numerical calculations throughout the thesis. Lastly, Section 1.5 shows an outline of the thesis chapters to follow.

1.1 Initial Margin

The global economic and financial turmoil which started around 2007 to 2008 showed significant weaknesses in financial institution participants, particularly banks, to curb over-the-counter (OTC) derivatives. Over-the-counter derivatives (OTCDs) were thought to achieve high nominal returns without any significant increase of risk. As it later became evident, the risks inherent in these new products were not fully understood by banks themselves or by the regulators and supervisors (Norgren 2010). Therefore, the Committee on Payment and Settlement Systems (CPSS) and the Committee on the Global Financial System (CGFS) were consulted around 2012 by the Board of the International Organisation of Security Commissions, Basel Committee on Banking Supervision (BCBS) and International Organisation of Securities Commissions (IOSCO) through the second consultative document on margin requirements for derivatives that are not cleared through a central counterparty (CCP). Initial proposal was released and comments were received, suggesting a quantitative impact study to assess potential liquidity and mandatory margin requirements. This resulted in development of objectives, elements and principles of a margining framework for non-centrally cleared derivatives (BCBS 2015). This framework was later enhanced and required that participants of the global OTCD be subjected to bilateral margin rules (BCBS-IOSCO 2015) which govern the process of posting margins. These rules started phasing in from September 2016 with the main cost relating to both central clearing houses and the bilateral margin rules being the up-front posting of this initial margin.

The Initial margin (IM) is a certain percentage of the financial product price at the beginning of the transaction paid by both the market participants to each other or even to the CCP, as cash or equity to ensure that contractual terms are met. According to Kim & Oppenheimer (2002), margin requirement were meant to reallocate credit for more productive uses, to protect investors from having too much debt and also to reduce stock price variations. However, some of the OTC market participants had the luxury of not posting any initial margin prior to the 2008 financial crises. This means that OTC market participants who were believed to be creditworthy were not asked to post IM, but those the bank considered to be risky had to. In the past, even currently the IM is viewed as the good faith deposit required by the exchange house in order for the counterparty to transact a particular derivative, in a way of protecting the exchange house and the other counterparty in the event of default or client's failure to cover losses. The precise margin required varied from one exchange house traded product to another, and between exchanges houses. The difference lies in the volatility of the underlying instruments, i.e., the greater the potential movement in the underlying asset the greater the potential for a loss may be on the position. Hence, the financial regulator, BCBS-IOSCO, initiated the initial margin requirement for all OTC derivatives not cleared through the CCP.

According to McBride (2010), CCP-IMs purpose is to reflect the possibility that a defaulting buyer may not post margin necessary to cover the fluctuations in the market value of the buyer's position while the CCP/seller liquidates such position. In such a situation, the losses incurred by the CCP/seller during the liquidation will be absorbed by the buyers initial margin. However, Gregory (2016) cautions that posting isolated initial margin creates a wealth transfer between derivatives seller and other sellers since the derivatives creditors receive a higher recovery in the event of default at the expense of the other creditors. Furthermore, IM is another way of protection against the risk of large adverse price movements that might occur over a long period of time (Longin 2000*a*). The main cost connected to both CCP and the bilateral margin rules globally is the up-front posting of IM and the change according to the portfolio in question and market conditions. Green & Kenyon (2015) computed initial margin using the fixed determined historical scenarios to simulated VaR. As a result this method became questionable because the historical prices changes rapidly and so are the stress scenarios. A mathematical initial margin pricing measurement as a quantile of a normal variate was considered by (Brigo & Pallavicini 2014). The variance thereof were calculated from the conditional close-out amount in the margin period of risk. Whereas ISDA standard initial margin model (SIMM) has already been standardised in such a manner that, it uses risk factors and sensitivities from the given asset classes (Albanese et al. 2016). All these financial models for determining IM were built with the focus of reducing systemic risks in derivatives markets. Also to shift the clearing and trading of OTCD instruments to CCP and organised exchange houses for better management and oversight by the regulatory bodies, as requested by the policy developers (FSB 2015). A couple of recent International Monetary Fund (IMF) papers on counterparty risk relating to OTCDs, find that a large part of the counterparty risk in OTCDs market is under collateralized by up to $2 trillion in relation to the risk from the global system (Singh 2010).

The latest trend of the exchange traded derivatives outstanding notional amounts are shown in Figure 1.1. Here, the OTCDs outstanding positions dataset of the traders/dealers has been captured, mainly from banks around the world. The dataset captured from the BIS-Statistics are the variable such as outstanding notional value, market value and credit exposure of OTC foreign exchange, interest rate, equity, commodity, and credit derivatives. All these variables are shown in Appendix A - Figure A.1 to A.4. Figure 1.1 shows the semi-annual average notional amount outstanding of the global OTCD market in US dollars. According to Figure A.1

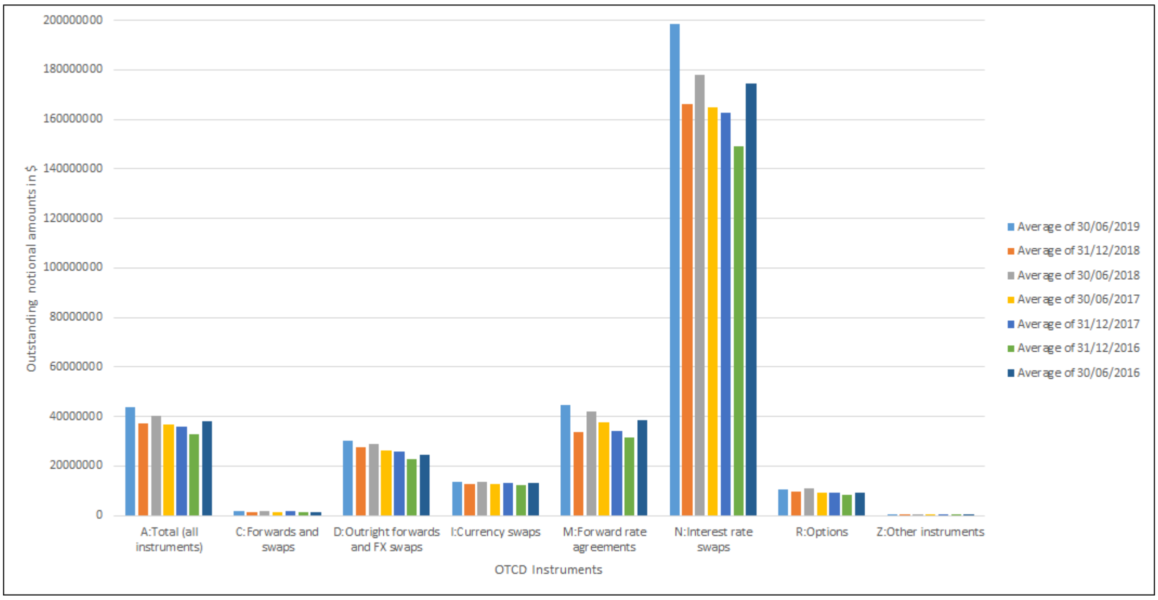


Figure 1.1: Trends of the world-wide OTCD outstanding notional amounts

in the appendix, OTC Interest Rate derivatives markets has been dominating the OTC market with very high outstanding notional amounts recorded. Interest rate derivatives markets have undergone significant structural shifts between April 2013 and April 2016, with the turnover measured in notional amounts nearly doubled for US dollar denominated interest rate derivatives (Wooldridge 2016). Globally, average daily turnover in OTC interest rate derivatives markets increased by 16%, to $2.7 trillion, between the preceding Triennial Survey in April 2013 and now 2020.

Initial margin protects the transacting parties from the potential future exposure that could arise from future changes in the mark to market value of the contract during the time it takes to close out and replace the position in the event that one or more counterparties default. The amount of initial margin reflects the size of the potential future exposure. It depends on a variety of factors, including how often the contract is revalued and variation or maintenance margin exchanged, the volatility of the underlying instrument, and the expected duration of the contract closeout and replacement period, and can change over time, particularly where it is calculated on a portfolio basis and transactions are added to or removed from the portfolio on a continuous basis (BCBS 2015).

When computing the initial margin requirement for particular futures and option contracts on the exchange, it is assumed that the returns thereof follow a normal distribution with the parameters calibrated, then IMR is computed as follows:

$$IMR = v \times \left[\exp\left(\Phi^{-1}\left(\frac{1}{2}(1+\alpha) \right) s \right) - 1 \right] \times \sqrt{n}$$

where

- v is the daily profit or loss from the contract,
- Φ^{-1} is the inverse cummulative normal distribution function,
- s is the standard deviation of the fitted log-returns,
- α is the significance level, and
- n is the holding period in business days.

The daily simulated 1250 returns will be used when contracting the normal distribution of a contract or portfolio levels. The days of returns is an own choice which will yield a reasonable normal distribution.The current South African exchanges uses a risk parameter of 3.5 standard deviation which corresponds to a confidence level of 99.95%. The chance that larger moves will occur in practice (translating to margins being insufficient to cover losses) is 1 in 1250 in any day and 1 in 9 over an entire year. The same method is being applied by the financial institutions in the OTC derivative markets to determine the fit initial margin for their contracts. It is noted by most of the South African exchange traded companies, academics and regulatory bodies that the parametric VaR methodologies are now scaled down as compared to historical

VaR methodologies for the calculations of initial margin requirement. The confidence levels in the use for latter institutions are 99.95% and 99.97%. The rolling 1250 day look-back period should be enhanced with 250 day turmoil or stressed look back period that took place during June 2008 and June 2009. Volatility scaling should be used to scale all returns (stressed and rolling) in the look-back period. In particular, each return should be multiplied by the ratio of the current 90 day volatility to the 90 day volatility that prevailed at the time when the return was observed. However, in order to prevent the extent to which a low volatility environment can decrease IMRs, a floor should be introduced whereby none of the returns in the look-back period can be scaled down by more than 30%.

In this research work, the parametric bootstrap methods developed by Efron (1979), will be utilized for the valuation of OTCD initial margin in the financial market, considering the low outstanding notional amounts. That is, an average monthly aggregate outstanding gross notional amount of OTCD instruments not exceeding a certain threshold amount, such as R20 billion (approximately 1.23 Euro billion). Also, the new requirements for initial margin between parties only applied to new financial contracts entered into after 1 September 2016 as given in Appendix A - Figure A.5. The low threshold should be considered because even though the small OTCD markets are not likely to cause systemic risk but there may be substantial growth in the market for the risk to be encountered. Nevertheless, the small OTCD market participants should also be used to the requirement of the initial margin, while the markets might not be sufficiently large to withstand the implementation of dedicated CCPs or financial systems infrastructure to securely exchange bilateral margin. Here the OTCD prices will be assumed to have a Gaussian probability distribution with the mean and standard deviation parameters. The bootstrap Value-at-Risk (BVaR) model will be applied as a risk measure that generates bootstrap initial margins (BIM) sufficient enough for the small OTCD markets.

1.2 Model Risk

Model Risk (MR) occurs because of inappropriateness of modelling (Derman, 1996). It implies that a model will not be fit enough to solve the problem at hand, such as predicting financial results with high accuracy. Allen (2012) regards MR as the risk that theoretical models in pricing, trading, hedging, and estimating risk will turn out to produce misleading results. Therefore model risk management in risk space is of paramount importance for the steadiness of the global financial system. This steadiness is related to the total exposure from financial institutions relating to credit market and the amount of capital that is available as a fallback against turmoil market or default events. As a result financial institutions and regulatory authorities make use of mathematical models to express the needed capital buffers.

Model risk has a history of literature, perhaps as model error in mathematical and statistical sciences or as model uncertainty in financial and economic management sciences and so on. Model risk research has been enhanced by the work of Wu & Olson (2010), Berkowitz et al. (2011), Tunaru et al. (2015), Danielsson et al. (2016) and Morini (2011) from the academics, together with the work done by Office of the Comptroller of the Currency (2011), Krishnamurthy (2013), Derman (1996) and Basel (2004) from the regulators, the list of these research work is not limited to the latter only. However, they provide substantial information for which financial models need improvement against those that are doing exceptionally well in the financial markets. Furthermore, they give guidelines in relation to the development of a financial model, its validation, implementation and how it should be used and governed through the relevant risk watch departments (such as, Market, Credit, Operational risk departments) and also consolidated and documented for reporting purposes to the likes of national regulatory bodies.

There are various sources of financial model risk (FMR). FMR may be coming from model development, statistical distribution risk, model specification, model imple-

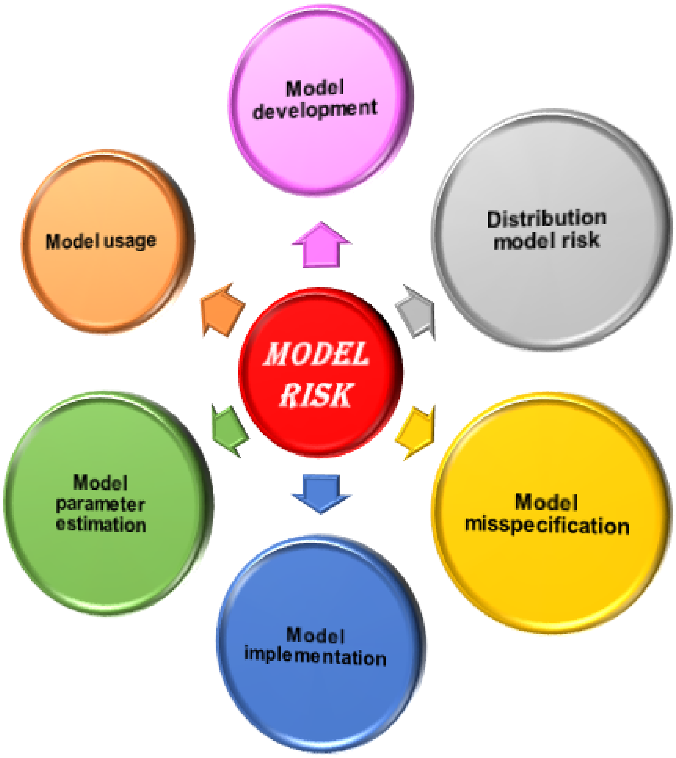


Figure 1.2: Model risk sources

mentation, model parameter estimation, or model usage. These are some of the sources highlighted by Figure 1.2, however the list may be expanded further. Model development have several steps that one can consider, for an example understand the objective and goals of building a model, data collection, data preparation, transforming and examination of variables to include in the model. Whenever anything goes wrong on the given steps, then the results will be referred to as FMR. According to Mileris & Boguslauskas (2011), models can be effectively created to measure credit risk by giving clear steps of the credit risk estimation model development. This will also reduce FMR in details. According to Breuer & Csiszár (2016), the distribution model risk (DMR) of a portfolio is defined by the highest estimated loss over a set of plausible distributions in terms of some deviation from an estimated distribution.

This may also be referred to as model risk due to wrong assumption of the statistical distributions. Model misspecification is the model specification error, which implies that the most valuable variable as an input to the model may be missing or key assumptions about the model may be incorrect. Thus, model results that are incorrect or misleading may be avoided through the processes of specifying and estimating a good fit model. Model implementation imply that the model is correct, however the implementation of this model has been done incorrectly. For an example through the mistake in programming mathematical equations or deliberately making mistake for personal gains (such as doing fraud). There are various techniques for financial model parameter estimation that are utilised by practitioners and academics. Any given technique that can be applied to a dataset to construct an estimate of the parameter for the financial model is known as an estimator. However, the estimated parameters for the financial model may not be the true representative of the parameter, which will affect the future outcomes of the model. Thus, an uncertainty of estimating the correct parameter value given a model structure is known as parameter estimation risk which is found under model risk (Tunaru et al. 2015, Glasserman & Xu 2014). The main reason for financial modellers to have sufficient documentation and training manuals for their models is to avoid model misuse. Model misuse, comprise of a quantitative analyst applying models not within the scope of uses for which they were developed.

Model risk is very important in this research work because it might reduce or mitigate systemic risk caused by models. Financial model's results have serious implications on the economy, therefore it is of paramount importance to realize the decision making done based on the model.

1.3 Research Aims and Objectives

At the time of writing this thesis, the initial margins for the OTC which are not cleared through the exchange houses or through the CCPs were burning subject and it still remains so to date. Financial model risk should continuously be assessed, especially by regulators, practitioners and academics. The above discussed sub-titles remain argumentative issues, hence the major interest to this research work.

Aim of the Study

The main aim of the thesis is to value initial margin for counterparties participating in the over-the-counter derivative (OTCD) market and quantify model risk (MR) for financial models, especially those models found in financial credit departments. The research work uses mathematical and statistical models to measure financial risk. The parametric approach, simulation and optimization techniques are utilised to estimate the ideal distribution for risk measurements, model misspecification and model parameters such that the end results are satisfactory for financial models.

Objectives of the Study

The following are secondary objectives (main contributions) of the thesis:

1. to provide an initial margin valuation methodology that can be used in large and small OTCD markets, which are found in emerging and developed financial markets;

2. to calculate the confidence interval for a risk measure in credit risk;

3. to investigate different model risk measures existing in credit risk management;

4. to develop a protocol that can be followed for identifying and quantify the model misspecification;

5. to analyse asymmetry and symmetrical distributions when there exist model misspecification; and

6. to compare the parameter estimation techniques for financial risk models in order to attain a good fit model.

1.4 Methods of investigation

From a mathematical and statistical framework, resampling techniques, parametric and non-parametric approaches will be used for simulating the dataset where needed and building the relevant financial models. Secondary credit history data from global and South African data portals will be extracted to use as input for testing models and in the numerical studies. Computational tools that will be used for numerical studies or quantitative analysis are the VBA-Excel 2013 supported by Löeffler & Posch (2011) and the IPython notebook programming language supported by (Pérez & Granger 2007).

1.5 Outline of the thesis

This thesis is presented in six chapters that include a published article that came out of this work. Each chapter is written as a peer-reviewed article and can be read independently of the entire thesis. The rest of the thesis is organized as follows:

- **Chapter 2**: Valuation of initial margin. In this chapter we propose a dynamic and generic model of initial margin for over-the-counter derivative instruments through the resampling method. This method is suggested following regulatory body's requirements, the Basel Committee on Banking Supervision (BCBS) and the International Organization of Securities Commissions (IOSCO) requested that financial institutions and their counterparty involved in trading over-the-counter derivatives should adhere to initial margin and variation margins at all times.

- **Chapter 3**: Model risk due to inappropriate statistical distribution. In this chapter we propose the bootstrap upper bound confidence level for the credit value-at-risk derived from a credit risk model. We show that the distribution model risk is of paramount importance for consideration when the data is from a symmetrical and asymmetrical distribution. For asymmetrical distribution other forms of confidence levels should be further investigated.

- **Chapter 4**: Model misspecification. We choose the credit risk models as an example that exhibits model misspecified and the other without model misspecification in a financial institution. The goodness of fit and model performance measurements are assessed.

- **Chapter 5**: Inappropriate parameter estimation. Inappropriate parameter estimators can be dealt with through considering several statistical and/or numerical methods. In this study we propose other several ways using the credit risk model as an illustration.

- **Chapter 6**: The chapter gives a summary of all chapters combined and recommendation about the future research discovered within this research thesis.

References are provided at the end of the thesis. The citation utilised in this thesis are listed according to the requirements mentioned by the North-West University manual for post-graduate studies.

2. Valuation of initial margin

In this chapter, the research work proposes parametric bootstrap method for valuation of over-the-counter derivative (OTCD) initial margin in financial markets with low outstanding notional amounts. That is, an aggregate outstanding gross notional amount of OTC derivative instruments not exceeding R20 billion. The OTCD market is assumed to have a Gaussian probability distribution with the mean and standard deviation as parameters. The bootstrap Value at Risk (BVaR) model is applied as a risk measure that generates bootstrap initial margins (BIM).The proposed parametric bootstrap method is in favour of the BIM amounts for the simulated and real datasets. These BIM amounts are reasonably exceeding the IM amounts whenever the significance level increases. This research work assumed that the OTCD returns only come from a normal probability distribution. The OTCD initial margin requirement in respect to transactions done by counterparties may affect the entire financial market participants under uncleared OTCD, while reducing systemic risk. Thus, reducing spillover effects by ensuring that collateral (IM) is available to offset losses caused by the default of a OTCDs counterparty. This research contributes to the literature by presenting a valuation of initial margin for the financial market with low outstanding notional amounts by using the parametric bootstrap method.

Keywords: Bootstrap, Gaussian probability distribution, Initial Margin, Over the counter derivatives, Value at Risk.

2.1 Introduction and background

The economic and financial crisis shocks that began somewhere between 2007 and 2009 revealed significant weaknesses in the resilience of financial institutions. It is in this regard that the Group of Twenty (G20) initiated a reform programme and the expansion of Central Counterparties (CCP's) scope during the year 2009 with the aim of reducing the systemic risk, especially for over-the-counter derivatives (OTCD). The business activities of the large OTCD market and volatility for the market value of outstanding OTCD exposures was significantly higher than bank assets and economic output (Lin & Surti 2015). The cost of hedging or replacing a contract at the time of default (i.e., Credit Exposure) is the potential future exposure (PFE) and is covered by the initial margin (IM). This cost of IM is known as the margin valuation adjustment (MVA), and it has therefore been considered into instrument model pricing by banks (Green & Kenyon 2015). The debate around the cost of funding for OTCD, that is MVA and similar funding cost like Funding Valuation Adjustment (FVA) illustrated by Lou (2016), is in progress as many researchers and regulatory bodies including Basel Committee on Banking Supervision (BCBS) and the International Organization of Securities Commissions (IOSCO), emphasize that costs should be mathematically captured through model pricing and valuation (BCBS 2015).

According to McBride (2010), IM determined through Central Counterparties (CCPs) has the main purpose of reflecting the possibility that a defaulting buyer may not post margin necessary to cover the fluctuations in the market value of the buyer positions while the CCP or seller liquidates such positions. Thus, IM is another way of protection against the risk of large adverse price movements that might occur over a long period of time (Longin 2000*b*). The amount of OTCD IM reflects the size of the PFE. This amount depends on a variety of factors, including how often the contract is revalued, variation margins exchanged, the volatility of the underlying OTCD instrument and the expected duration of the contract closeout. Particularly when it is calculated on a portfolio basis and transactions are added to or removed from the portfolio on a continuous basis (BCBS 2015). The IM requirements might be set and determined through the use of the parametric technique such as the Gaussian probability distribution applied by Duffie & Pan (1997), student t-distribution used by Del Brio et al. (2014) and delta-approximated Value at Risk used by Lou (2016), among many others. Alternatively nonparametric methods may be employed as well, this includes but not limited to the techniques named, kernel density estimation, bootstrap method founded by Efron & Tibshirani (1986) and later improved by Swanepoel & De Beer (1993), financial risk measures known as Value at Risk (VaR_α), Expected Shortfall (ES) and Exponential Spectral Risk Measures (ESRM) (Cotter & Dowd 2006). Some of the margins models come from the risk models (Murphy et al. 2014). These latter technique measurements have one common assumption, which suggest that the investors' defaults happen at extreme returns. The consideration of heavy tail distribution free often leads to better description, accuracy and efficient parameter estimation results.

In this chapter we propose the dynamic and generic model of initial margin for OTCD instruments through the resampling method called bootstrap, as opposed to the traditional full valuation VaR_α methods shown by Ball & Fang (2006), Albanese et al. (2016) and Berkowitz et al. (2011), with an assumption that the data come from normal distribution. It is currently well known that the normality assumption is unrealistic in situations where there exist smaller historical sample sizes and more difficult when the quantile level is high. Although consistent estimates of parameters can be obtained through other candidate models (Bignozzi & Tsanakas 2016). We therefore show that the bootstrap method for valuing the OTCD initial margin (OTCD-IM) can be considered under realistic business day's history, stressed business days and realised volatility. Using nonparametric statistics, Alemany et al. (2013) showed that asymptotic properties hold whenever we estimate extreme quantiles or parameters using large sample sizes. Research showed that determining initial margin using the VaR_α historical scenarios, considering the high shocks in historical prices and scenarios that changes rapidly leads to funding cost of the IM (Green & Kenyon 2015). IM corresponding to a given level of risk tolerance increase significantly when it moves away from point-in-time towards a stress-period calibration of the volatility of adjusted returns on OTCD contracts. Lin & Surti (2015) found it to be an issue for capital requirement to the methodology of calibrating key risk parameters. Other researchers such as Brigo & Pallavicini (2014), included a mathematical initial margin pricing measure to be a quantile of a normal variate with their variance calculated from the conditional close-out amount in the margin period of risk. Whereas International Swaps and Derivatives Association (ISDA) standard initial margin model (SIMM) has already been standardised in such a manner that it uses risk factors and sensitivities from the given asset classes (Albanese et al. 2016). In this work we also analyse the precisions of IM estimations provided by bootstrap methods.

The sections of the chapter are structured as follows. In Section 2.2 we present the review of the traditional full valuation for the IM through the use of the historical VaR_α which assumes that the returns come from independent and identical standard normal distribution ($R_t \sim N(0;1)$), based on the OTCD trade by trade level. An example of the current valuation and the disadvantage of using the VaR_α are also discussed. The implementation and the evaluation of the proposed standard bootstrap method for valuing the OTCD Initial Margin is described in detail though Section 2.3. In Section 2.4, we illustrate the simulation study by showing the setup of the study, the results of the simulation through graphs and tables, and conclude with the simulation discussions. Section 2.5 illustrates the bootstrap using the application to real dataset. Finally, in Section 2.6 we conclude the study and make recommendation.

2.2 The Initial Margin through VaR

We assume that the financial asset processes may be stochastic in nature (i.e. OTCD), which imply they have a Random walk or Brownian motion or Wiener process, S_t. Since in practice there are huge amounts of unknown movements of asset prices which influence the phenomena of interest, we therefore consider the central limit theorem technique. Thus, we assume that the unknown influences build up to become normally distributed. This process is known as the Gaussian distribution.

Let $S_t > 0$ represent the OTCD prices with returns given as

$$R_t = \ln S_t - \ln S_{t-1}, \text{ for } t = 0 \text{ to } T. \tag{2.2.1}$$

If $R_1, R_2, \ldots, R_n$ are observed returns following a Gaussian probability distribution G_R with the mean μ and variance σ^2, i.e. $R_t \sim iid\ N(\mu; \sigma^2)$, then we define the VaR_α for OTCD returns.

2.2.1 Definition. Value at risk measurement (VaR_α). Given $\alpha \in [0,1]$ and $G_R(r) = P(R_t \leq r)$ for $r \in \Re$, the Value-at-Risk at level α of the OTCD returns with distribution G_R is the smallest real value of the return percentile given by

$$VaR_\alpha(R_t) = G_R^{-1}(\alpha) = inf\{r \in \Re : G_R(r) > \alpha\}. \tag{2.2.2}$$

See Artzner et al. (1999), for details of the quantile properties and Appendix A.2 for the basic statistical method of the VaR_α. The VaR_α is a statistical technique used to approximate the probability of loss for the asset or portfolio of assets, based on the analysis of historical price trends and volatilities described by Khindanova et al. (2001), in our case OTCD returns. For the case of OTCDs, an initial margin through the VaR_α will be considered as the margin intended to cover the largest loss (in %) that may be encountered by an investor over a single day with a $(1-\alpha)100\%$ confidence level. The initial margin amount is collected on an upfront basis, at the time of trade. Most CCPs or stock exchange of derivatives determines the losses that might arise in a given day with a confidence level of $(1-\alpha)100\%$ which imply that only $\alpha\%$ of daily losses is far away than 3.5 standard deviations from the mean. Thus saying that the initial margin amount will cover $(1-\alpha)100\%$ of all possible daily changes in the asset market of interest. Therefore, IM amount is estimated with the following expression:

$$IM = \widehat{VaR_\alpha} \times OTCD\ Notional\ Amount. \tag{2.2.3}$$

Given that the VaR_α for the returns is

$$VaR_\alpha = E[R_t] + \Phi^{-1}(\alpha)\sigma_t[R_t].$$

Therefore, the estimated VaR_α is given as

$$\widehat{VaR_\alpha} = \bar{R} + \Phi^{-1}(\alpha)s,$$

where $\bar{R}$ is the mean, s is the standard deviation for the historical OTCD returns and Φ^{-1} is the inverse cumulative standard normal distribution at a given significance level α.

2.2.2 Evaluation of IM

The evaluation of measures such as IM and VaR_α triggers questions about the biasness, accuracy and efficiency of the estimator. We use the Rao–Blackwell theorem to evaluate the estimator $\hat{\theta}$ of interest (Wackerly et al. 2014).

Since the unknown parameter of interest $\theta(G) = VaR_\alpha$ is from the unknown probability distribution G and can be estimated using an empirical distribution of returns $\hat{G}$. Then, biasedness of $\hat{\theta}$ for estimating θ is defined as:

$$Bias(\hat{\theta}) = E_G(\hat{\theta}) - \theta(G). \tag{2.2.4}$$

Contrary, an estimator is unbiased if the expected value of the estimator is the same or almost the same as the actual value of the population parameter, i.e. $MSE(\hat{\theta}) \approx Var(\hat{\theta})$.

A general statistical measure for the size of the error that is often used is the mean squared error (MSE), which is defined as

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Bias(\hat{\theta})^2 + Var(\hat{\theta}). \tag{2.2.5}$$

Other measures of statistical errors for evaluation can be seen from (Efron & Tibshirani 1986). The bootstrap techniques will be employed to evaluate the proposed Initial Margin amount.

2.2.3 Practical IM example

Let an Initial Margin (IM) amount which is collected on either short or long positions kept constant at a significance level of 1%, such that there exist the same probability of IM being exceeded.

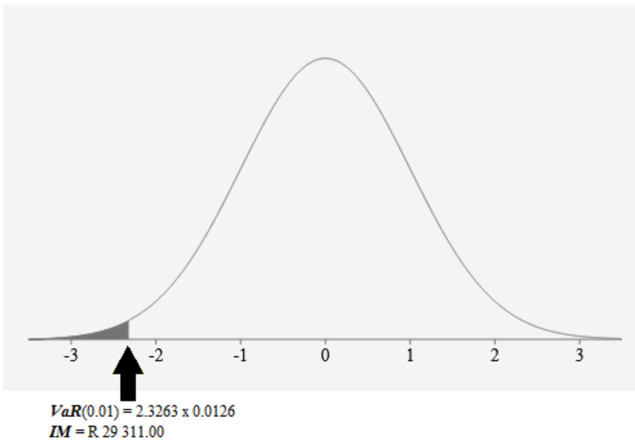


Figure 2.1: Initial Margin and its VaR_{α} from Normal distribution of the returns.

Consider the distribution of the returns as shown in Figure 2.1 above, which imply that the IM amount will be kept constant at $2.3263 \times \sigma$, where σ is the daily standard deviation of the returns and 2.3263 is the inverse standard normal score (Φ^{-1}) at 1%

level of significance. OTCD poses an annualised standard deviation of 0.2, then the daily standard deviation is $s = 0.2 \times \sqrt{\frac{1}{250}} = 0.0126$ and the average returns is zero since they follow a standard normal distribution. Therefore, the VaR_α percentage is kept constant for the OTCD of interest at

$$\begin{aligned}\widehat{VaR_\alpha} &= \bar{R} + \Phi^{-1}(\alpha)s \\ &= 2.3263 \times 0.0126 \\ &= 2.9311\end{aligned}$$

Setting the notional amount to R1 million, the Initial Margin requirement amount is calculated using equation (2.2.3) above, given as follows

$$IM = 2.9311\% \times 1000000.$$

This amount imply that the trader or counterparty engaging in this transaction has to deposit R 29 311.00 in the margin account at the start of the transaction. The above example shows the calculation of the Initial Margin for an asset through the use of the traditional full valuation VaR_α using an assumed normal distribution of the asset returns.

2.2.4 Disadvantage of the IM using VaR_α

Non-centrally cleared derivatives contracts should be subject to higher capital requirements. IM is one such part that will pump up the capital requirement for OTCD. The use of IM models is intended to produce appropriately risk-sensitive assessments of potential future exposure (PFE) so as to promote robust margin requirements (BCBS 2015).

The distribution of the returns for different types of instruments in the financial markets are characterized as fat-tailed or they have time varying volatility and skewed to leptokurtic nature from the empirical research. The initial margin models, such as constant volatility, exponentially weighted moving average, historical simulation VaR_α and Hull and White approaches, which passed a standard risk-sensitivity test can vary quite widely in their degree of procyclicality (Murphy et al. 2014). The fractional-stable Gaussian ARCH models was identified and considered as the solution for drawback of heavy tailed, skewed and leptokurtic nature of returns. The conditional heteroskedastic models based on the stable hypothesis can be applied to describe both thick tails and time-varying volatility (Khindanova et al. 2001). However, this research work considers that the historical returns of the OTCD instruments that are subject to IM are normally distributed with a mean and variance estimated from the returns.

We note that no information about the distribution of financial instrument returns may be at our disposal, it is also less likely to use the asymptotic properties of estimators of the parameters which are being estimated. In this case through diagnostics and statistical analyses, we are of the view that the bootstrap estimation methods may turn out to be more effective compared to the classic parametric method shown by Bank of England and some researchers (Murphy et al. 2014). The majority of research showed the bootstrap estimation methods to be more effective than the classical non-parametric method for the confidence interval obtained having much smaller span while having similar estimation likelihood (Pekasiewicz 2016). The impact of margin changes in OTC equity options and commodity futures markets, respectively, were examined (Hedegaard 2011). The finding retrieved were significantly high and there was an increase in margin requirements as the markets become more volatile.

The studies showed the lack of increasing estimation accuracy as a result of parametric and nonparametric method application to financial instruments causing an increase in the estimation likelihood, especially for random variables with very heavy tail distributions (Pekasiewicz 2016). The bootstrap method is confirmed to work in agreement to the latter by increasing the estimation likelihood.

2.3 The Initial Margin through bootstrapped VaR

The proposed bootstrap method is a resampling method that is used to approximate the true distribution of the statistics of interest, by drawing samples from the existing sample data where the statistic was measured. Efron & Tibshirani (1994) defined the method as "A computer-based method for assigning measures of accuracy to statistical estimates". Our estimate of interest is the Initial Margin computed using the bootstrapped VaR_α as described below.

We consider $\mathbf{R_n} = (R_1, R_2, \ldots, R_n)$ generated using equation (2.2.1) to be the OTCD returns data drawn from an unknown population distribution G. Suppose that the parameter of interest is denoted as $\theta = T(\mathbf{R_n}, G)$, which is simply the function, T, of the unknown distribution function G. The move from the *Real World* to the *bootstrap world* as shown in Figure 2.2 below asserts that the bootstrap estimator of the parameter θ is given as $\hat{\theta} = T_n(\mathbf{R_n}, \hat{G})$. More details of the procedure for bootstrap method can be found in (Seitshiro 2006). The bootstrap sample which are drawn with replacement from the original sample data, $\mathbf{R_n}$, is denoted by $\mathbf{R_n^*} = (R_1^*, R_2^*, \ldots, R_n^*)$. The term $\hat{G}$ is regarded as the empirical distribution function (EDF) defined as:

$$\hat{G} = \frac{1}{n}\sum_{i=1}^{n} I(R_i \leq r), \tag{2.3.1}$$

where $I(.)$ is the indicator function given by

$$I(.) = \begin{cases} 0, & \text{if } (.) \text{ is False} \\ 1, & \text{if } (.) \text{ is True} \end{cases}.$$

We chose the EDF as a primary approximation distribution of G mainly because it has other many different desirable properties as an estimator for G. This imply that $\hat{G}$ converges uniformly, with probability 1, to G as the sample size n becomes large. Furthermore, independently drawing samples from the EDF reduces to drawing samples with replacement from the original sample (Efron & Tibshirani 1994).

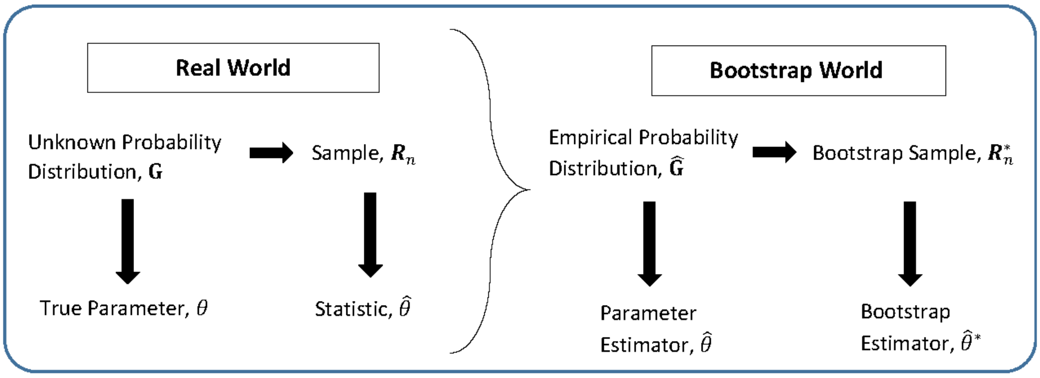


Figure 2.2: Schematic diagram showing the summary of the bootstrap principle.

We make use of the following Monte-Carlo simulation algorithm which relies on the EDF to generate the bootstrap probability distributions and propose the new method called bootstrap Initial Margin amount (BIM):

1. We start by generating the sample $\mathbf{R}_n = (R_1, R_2, \dots, R_n)$, with $1/n$ as the probability of an observation being selected. The sample can be 1 day return or n day returns, generated by taking the difference between logarithm of today's asset price and logarithm of yesterday's (or n-prior days) asset price.

2. Generate the first random bootstrap sample of n independent observations $\mathbf{R}_n^*(1) = (R_{11}, R_{21}, \dots, R_{n1})$ from fixed EDF, $\hat{G}$. Thus, sampling with replacement from $\mathbf{R}_n$.

3. Calculate the statistic of interest $\hat{\theta}^*(1)$ from the first bootstrap sample generated in step 2.

4. Independently repeat step number 2 and 3 until the following B number of samples $\mathbf{R}_n^*(B)$ are drawn and corresponding bootstrap statistics $\hat{\theta}^*(1), \hat{\theta}^*(2), \ldots, \hat{\theta}^*(B)$ are calculated respectively

Bootstrap Sample	**Replications**
$\mathbf{R}_n^*(1) = (R_{11}, R_{21}, \ldots, R_{n1})$	$\hat{\theta}^*(1)$
$\mathbf{R}_n^*(2) = (R_{12}, R_{22}, \ldots, R_{n2})$	$\hat{\theta}^*(2)$
$\vdots$	$\vdots$
$\mathbf{R}_n^*(B) = (R_{1B}, R_{2B}, \ldots, R_{nB})$	$\hat{\theta}^*(B)$

5. Sort the bootstrap replicates from step 4, such that the order statistics is given by

$$\hat{\theta}^*(1) \leq \hat{\theta}^*(2) \leq \ldots \leq \hat{\theta}^*(B).$$

6. Estimate the bootstrap value at risk ($BVaR_\alpha$) as follows

$$BVaR_\alpha = \hat{\theta}^*(\beta), \tag{2.3.2}$$

where $\beta = (B+1)\alpha$.

7. The BIM is then given by

$$BIM_\alpha = BVaR_\alpha \times V, \tag{2.3.3}$$

where V is the OTCD value.

8. The bootstrap estimate of the standard error is calculated as

$$\sigma_B^*(\hat{\theta}^*) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}[\hat{\theta}^*(b) - \hat{\theta}^*(.)]^2}, \tag{2.3.4}$$

where

$$\hat{\theta}^*(.) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*(b).$$

9. The bootstrap coefficient of variation is given as:

$$BCV = \frac{\sigma_B^*(\hat{\theta}^*)}{\hat{\theta}^*(.)}. \tag{2.3.5}$$

The above Monte Carlo algorithm used to compute the bootstrap replications give the approximation of the sampling distribution of the estimator $(\hat{\theta})$ and not the exact estimator or the parameter (Seitshiro 2006). As the bootstrap sample (B) goes to infinity the bootstrap approximated distribution of the bootstrap replication should lead to better precision and more accurate sample statistic, i.e.,

$$\lim_{n\to\infty}\hat{\theta}^* \approx \hat{\theta}. \tag{2.3.6}$$

The number of resamples is supposed to be as many as possible and is mainly limited by available computing power and time.

2.4 Simulation Study

Using the methodology described in Section 2.3 above, a simulation study was conducted to produce the proposed initial margin for OTCD. The following chronological sub-sections are simulation setup, results and discussion of the results. All the numerical computation given in the results were implemented through Visual Basic for Applications in Microsoft Excel 2013.

2.4.1 Setup

The bootstrap Monte Carlo Simulation (BMCS) method is a nonparametric technique that will be used to generate the returns of the OTCD Müller et al. (2017), drawn with replacement from an assumed normal distribution. It is different from traditional parametric approaches, because it employs a large number of repetitive computations to estimate the shape of a statistic's sampling distribution, as opposed to strong distributional assumption.

For the BMCS, we select sample size $n = 252$ as a representation of business daily of OTCD returns for the year generated using the standard normal distribution with the mean zero and variance of one. From this sample, the normal VaR_α and its corresponding IM are computed using different significance levels though equations (2.2.3) and (2.2.4) respectively. Whenever the number of bootstrap replications and the number of Monte Carlo replications are high enough (i.e., $B \geq 200, MC \geq 1000$), then the method provides acceptable results as an approximate method (Müller et al. 2017). Therefore, for this study we set the bootstrap replicate samples of OTCD return to $B = \{250, 750, 1250\}$ and the varied significant levels for the respective replicated OTCD return samples for the Bootstrap VaR_α in equation (2.3.2) and bootstrap IM in equation (2.3.3) to $\alpha = \{0.001, 0.005, 0.010, 0.025, 0.050, 0.100\}$. The α–percentile is used to identify the efficient worst outcome known as the bootstrap Value at Risk ($BVaR_\alpha$) and we apply equation (2.3.3) for calculating the measure of interest for OTCD called Bootstrapped Initial Margin (BIM). Furthermore, equations (2.3.4) and (2.3.5) are applied to evaluate the precision and accuracy of the method. The time for every simulation iteration is also recorded. We assume a principal notional amount of R 1 million to a simple asset class containing a single risk factor to assert the input shocks are sufficient for the IM calculations.

2.4.2 Results

Table 2.1 and Figure 2.3 reveal the behaviour of the value at risk and the respective initial margin amounts computed using the classical full valuation and the proposed BIM for a single position of OTCD. The precision and accuracy of the proposed bootstrap method is evaluated through the bootstrap standard error and coefficient of variation.

Table 2.1: Simulation summary results: BIM for OTCD.

				n = 252			
α	B	$VaR_\alpha(\hat{\theta})$	$BVaR_\alpha(\hat{\theta}^*)$	$BSE(\sigma_B^*(\hat{\theta}^*))$	IM	BIM	BCV
0.001	250	-3.48558	-2.7670	0.1237	**34 855.79**	27 670.04	4.59E-04
	750	-3.48558	-2.7670	0.1150	**34 855.79**	27 670.04	4.26E-04
	1250	-3.48558	-2.7670	0.1270	**34 855.79**	27 670.04	4.72E-04
0.005	250	-3.36459	-2.7670	0.1374	**33 645.87**	27 670.04	5.11E-04
	750	-3.36459	-2.7670	0.1189	**33 645.87**	27 670.04	4.16E-04
	1250	-3.36459	-2.7670	0.1159	**33 645.87**	27 670.04	4.42E-04
0.010	250	-2.66412	-2.7670	0.2147	26 641.16	**27 670.04**	8.69E-04
	750	-2.66412	-2.7670	0.2106	26 641.16	**27 670.04**	8.55E-04
	1250	-2.66412	-2.7670	0.2131	26 641.16	**27 670.04**	8.57E-04
0.025	250	-1.93626	-2.6548	0.2248	19 362.59	**26 548.03**	1.05E-03
	750	-1.93626	-2.4935	0.2263	19 362.59	24 935.00	1.05E-03
	1250	-1.93626	-2.4935	0.2298	19 362.59	24 935.00	1.06E-03
0.050	250	-1.75225	-1.9257	0.1806	17 522.54	19 256.66	1.13E-03
	750	-1.75225	-2.0033	0.2006	17 522.54	**20 032.51**	1.14E-03
	1250	-1.75225	-2.0033	0.2005	17 522.54	**20 032.51**	1.22E-03
0.100	250	-1.40207	-1.3245	0.0904	**14 020.67**	13 244.73	7.37E-04
	750	-1.40207	-1.3979	0.0877	**14 020.67**	13 979.45	7.60E-04
	1250	-1.40207	-1.3245	0.0932	**14 020.67**	13 244.73	7.09E-04

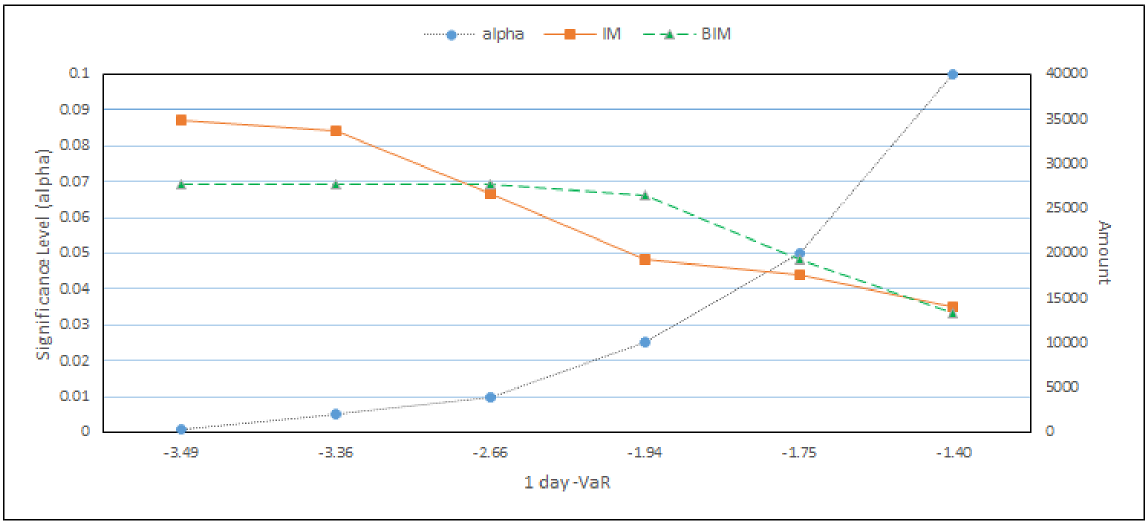


Figure 2.3: Simulation summary results: Value at Risk against Significance level, Initial Margin and BIM for OTCD.

2.4.3 Discussion

Every VaR_α measure generated with Financial Market inputs makes assumptions about some return distributions, which whenever violated, result will be inappropriate estimates of the VaR_α and eventually leads to inappropriate Initial Margin amount for the OTCD instrument of interest. Many financial market reports and research have revealed substantial evidence that asset returns are non-normally distributed and outliers are more common, but they are much larger than expected, given the distribution of different types of financial assets. An innovative distribution with heavier tails than the normal, as well as possible asymmetry, often provides a better description and therefore would lead to more efficient estimation results for financial market instruments (De Jongh & Venter 2015).

Table 2.1 reveals that as the significance level increases, then the IM calculated through VaR_α decreases. So, when the significance level increase the absolute bootstrap VaR_α decreases and the resulting bootstrap IM as well decreases, as expected. According to the BCBS (2015), the Non-centrally cleared derivatives contracts should be subject to higher capital requirements, starting with the IM and all other capital amounts. The maximum IM to be selected is given by the highest figures among IM and BIM. These amounts are bolded in Table 2.1. The extent of variability in relation to the mean of each bootstrap sample is shown by bootstrap Coefficient of Variation (BCV). The higher the BCV the higher the variability in relation to the mean. Table 2.1 reveals a sticking point that the significance level of $\alpha = \{0.01, 0.025, 0.05, 0.1\}$ have high BCV. Bootstrap method work precisely with high significance level and more exceptions. The larger the bootstrap sample one chooses the longer the time in seconds the simulation code in excel will have to spend for the results to be produced.

Figure 2.3 findings show that as the significance level increases the VaR_α increases. It reveals more enlightening information about the Initial Margin calculated through the normal traditional VaR_α, because as the VaR_α increase then the IM decreases. The normal IM is high than the BIM at significant level of $\alpha = \{0.001, 0.005, 0.1\}$ and thereafter, the BIM take precedence over the normal IM at significance level of $\alpha = \{0.01, 0.025, 0.05\}$.

2.5 Application to real data

The purpose of this section is to give an insight example of bootstrap IM valuation using the OTCD called variance swap on an index of stocks. They are instruments which offer shareholders direct exposure to the standard deviation of an underlying asset. The OTC variance swap instruments took place as a product in the aftershock of the Long Term Capital Management (LTCM) meltdown in late 1998 during the Asian crises. More information about the lessons concerning OTCD can be found from an article of Shirreff (1999). In this research we consider variance swap on FTSE100, whereby if the index amount is positive the variance seller will pay the variance buyer the index amount, otherwise the variance buyer will pay the variance seller an amount equal to the absolute value of the index amount. The payoff from a variance swap at time T to the payer of the fixed variance rate is $N(\bar{V} - V_K)$, where N is the notional principal amount, V_K is the fixed variance rate and $\bar{V}$ is the annualised realised variance given as

$$\bar{V} = \frac{252}{T} \sum_{t=1}^{T} R_t^2,$$

where R_t is the natural log returns defined in equation (2.2.1). The square root of the variance is the volatility. Note, unlike the strike of an option, the strike of the variance swap is known as the fixed variance rate which imply the level of variance/volatility bought or sold. The buyer of a variance swap (i.e. going long position on volatility) will be in profit whenever the realised volatility exceeds the level set by the fixed variance rate otherwise the buyer will be in loss. On the contrary, the seller of the variance swap is said to be going short on volatility and will therefore profit whenever the level of the variance sold exceed the realised variance rate.

The time series of FTSE100 realised variance is obtained from Oxford-Man Institute of Quantitative Finance Realized Library (Heber et al. 2018). The time series captures a window period from 19 July 2017 to 17 July 2018.

Table 2.2 and Figure 2.4 of the FTSE 100 on variance swap conforms to the findings of the simulation found in Table 2.1 and Figure 2.3. The findings reveal that as the significance level increases, the VaR_α increases. The IM measure is the same as the BIM at significant level of $\alpha = \{0.001, 0.005\}$ and thereafter, the BIM take precedence over the normal IM. The intriguing part is the convergence closeness of the value at risk through the use of the parametric method and the bootstrap method at significance level of $\alpha = \{0.05, 0.1\}$. Thus, the BIM seems better suited to enable financial institutions and non-financial counterparties to better manage high margin requirements in this new market environment.

Table 2.2: Variance swaps on FTSE 100 summary results: Initial Margin through Normal VaR_α and Bootstrap VaR_α.

				n = 252			
α	B	$VaR_\alpha(\hat{\theta})$	$BVaR_\alpha(\hat{\theta}^*)$	$BSE(\sigma_B^*(\hat{\theta}^*))$	IM	BIM	BCV
0.001	250	-2.44	-2.44	0.2695	**24 400.00**	**24 400.00**	0.0012
	750	-2.44	-2.44	0.2471	**24 400.00**	**24 400.00**	0.0011
	1250	-2.44	-2.44	0.2710	**24 400.00**	**24 400.00**	0.0012
0.005	250	-2.44	-2.44	0.2717	**24 400.00**	**24 400.00**	0.0012
	750	-2.44	-2.44	0.2732	**24 400.00**	**24 400.00**	0.0012
	1250	-2.44	-2.44	0.2677	**24 400.00**	**24 400.00**	0.0012
0.010	250	-1.67	-2.44	0.3597	16 000.00	**24 400.00**	0.0019
	750	-1.67	-2.44	0.3559	16 000.00	**24 400.00**	0.0019
	1250	-1.67	-2.44	0.3522	16 000.00	**24 400.00**	0.0019
0.025	250	-1.52	-2.28	0.1909	15 200.00	**22 800.00**	0.0013
	750	-1.67	-1.67	0.1717	15 200.00	16 700.00	0.0012
	1250	-1.67	-1.67	0.1669	15 200.00	16 700.00	0.0011
0.050	250	-1.24	-1.33	0.0936	12 400.00	**13 300.00**	0.0008
	750	-1.24	-1.33	0.0922	12 400.00	**13 300.00**	0.0008
	1250	-1.24	-1.33	0.0901	12 400.00	**13 300.00**	0.0007
0.100	250	-0.90	-1.09	0.1162	9 000.00	**10 900.00**	0.0013
	750	-0.90	-1.09	0.1222	9 000.00	**10 900.00**	0.0013
	1250	-0.90	-1.07	0.1167	9 000.00	9 030.00	0.0013

2.6 Conclusion

Centrally cleared derivatives contracts such as fixed income instruments traded through the Johannesburg Stock Exchange, applied a VaR_α approach with a 99.7% lower percentile to calculate the IM amount (Garcia Trillos et al. 2016). We compute the IM on the basis of OTCD processes whereby the emphasis is on the BIM computation given the probabilistic nature of general financial current risk exposures and we do not settle a lot on the financial future risk exposure or future initial margin. The proposed BIM is applicable during normal and stressed financial markets. The process makes no assumption about the distribution of the changes for the financial distri-

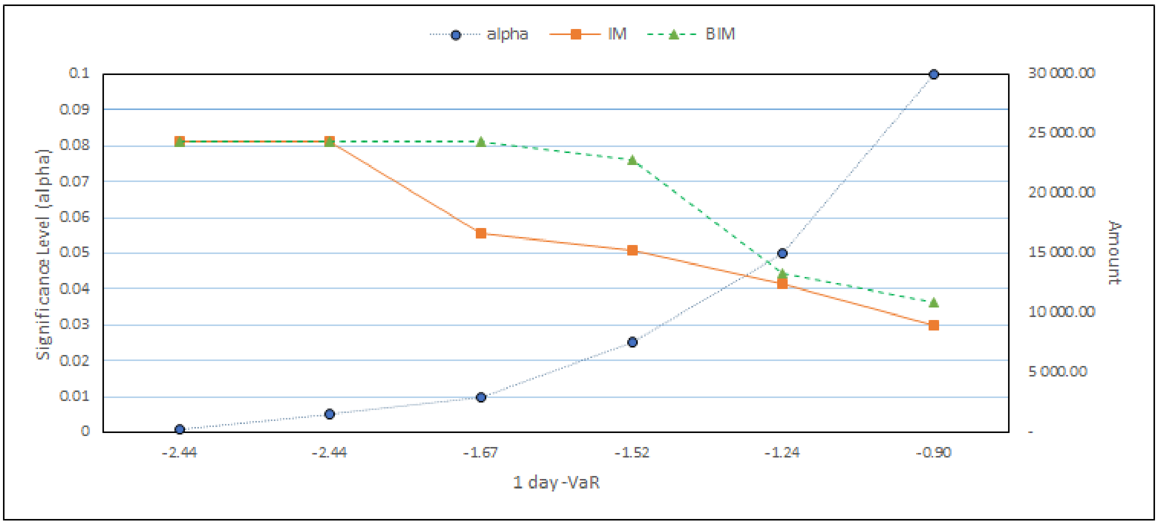


Figure 2.4: Variance Swap on FTSE 100 summary results: Value at Risk against Significance level, Initial Margin and BIM amounts.

bution, but we acknowledge that there might be heavy tail distributions of returns. Theoretically, the financial return distribution is assumed to be the Gaussian normal distribution in this chapter and it has shown to be a statistical model that entails an asymptotic distribution properties.

Our proposed method of computing the initial margin amount through the bootstrapped value at risk have some advantages over traditional methods. The method gives an acceptable degree of percentage change over initial margin amount, which will not leave the posted party over exposed in the event of counterparty default. The BIM method is free of the model risk, as such the initial margin required will go up under the times of market stress and become reasonably low under calm market conditions. The method should be applied with careful consideration that even very large unexpected market shocks will not be neglected, this will raise the initial margin amount to a high amount. The bootstrap method applied to FTSE100 on swaps time series data conforms very well with the simulated data. Thus, the proposed methodology will enable counterparties to meet regulatory and risk management margin requirements under small OTCD outstanding notionals.

For further work on the analysis of the bootstrap method, we think that the method should be applied on the real data set that encompasses the different financial classes such as foreign exchange derivative, interest rate derivative and so on.

3. Model risk due to inappropriate statistical distribution

Despite the weaknesses, Value-at-Risk (VaR) is still the risk measure that practitioners and regulatory bodies customarily utilize for calculation of risk capital in the financial institutions such as banks. This chapter is concerned with the forecasting of credit Value-at-Risk and assessment of distribution model risk (DMR). Therefore, the DMR for credit risk model is assessed by estimating upper confidence level using the bootstrap method. Random variables coming from different parametric distributions are assumed as proxies to the systematic common factor for Vasicek single factor model. The Bootstrap Monte-Carlo simulations are performed for assessing the accuracy of the credit VaR using the proposed bootstrap upper confidence level and other methods for bootstrap confidence levels. The methods include the basic percentile, standard hybrid percentile and bias-corrected percentile approaches. Furthermore, a new method called modified hybrid percentile (MHP) is proposed for forecasting the credit VaR using the bootstrap upper confidence levels. The result show that the proposed MHP method is highly in favour of both asymmetrical and symmetrical distributions.

Keywords: Confidence level, Credit Risk, Credit Value-at-Risk, Distribution Model Risk, Symmetrical and Asymmetrical distributions, Parametric Bootstrap.

3.1 Introduction and background

Model risk management in the credit risk departments is of paramount importance for the steadiness of the global financial system. This steadiness is the total exposure from financial institutions point of view that relates to the credit market and the amount of capital available as a remedial measure against turmoil or default events. The financial institutions and regulatory authorities make use of financial, mathematical and statistical models to express the needed capital funding barriers. While models ought to be simplified descriptions of systems or processes to assist in calculations and prediction, they are also exposed to numerous risk factors. According to Fontana et al. (2019), model risk in credit risk relate to adopting the incorrect model for the default events and calibrating or/and estimating a given model in an incorrect way. Model risk is the uncertainty for a model to perform a given task and yielding accurate and acceptable results. Model risk is also the risk of inadequacies in assumptions made about the formulation of the model, together with the choice of different suspicions on the probability distributions of the measure of interest.

In this research work, the distribution model risk (DMR) is utilised to estimate the bootstrap upper confidence level that can be used to set the credit value-at-risk (VaR) of a credit loss portfolio. According to Breuer & Csiszár (2016), the DMR of a portfolio is defined by the highest estimated loss over a set of plausible distributions in terms of some deviation from an estimated distribution. The credit-VaR is the worst expected loss amount as a result of counterparty default per given holding period with a given confidence level. In the credit market, value changes to financial institutions are relatively small when there are minor upgrade or downgrades, however they can be substantial when there are defaults. Thus, the probability of large credit losses produces asymmetrical distributions with fat-tails that differ significantly from the usual normal distribution and similar distributions. We consider a Vasicek single factor model with losses only occurring when an obligor or a contructual debtor defaults from an obligation in a fixed time horizon and mimic the qualitative behaviour of empirical credit loss distributions with fat tail and high skewness. The asymptotic approximation of the model performs very well when large number of small exposures are encountered, however when the portfolio is dominated by a few debtors the analytic approximation of the Vasicek model can significantly under-estimate risks in the presence of exposure concentrations (Bluhm et al. 2016). A variety of other methods for estimating the portfolio credit risk have been looked at (Glasserman & Ruiz-Mata 2006, Huang et al. 2007). They provide very relevant and distinct comparison of methods for computing credit loss distributions, taking interaction into consideration. However the systemic factors which are the explaining factors of the asset return are assumed to be normally distributed.

The financial institutions whose failure produces spillovers and trigger high uncertainty in the financial system ought to be controlled and managed, either through internal risk management or through the regulatory bodies. Compared to other industrial sectors the systematic component of default is significant in the financial industry. In accordance with Memmel et al. (2015), systematic factors influence the magnitude of defaults events. These systematic factors influence the entire borrower's cohort and it plays an important role in the default risk of credit portfolios. Also, the turmoil of financial markets, especially those found in banks, tend to experience a systemic failure when a breakdown of the financial system is triggered by a strong systemic event, thus the collapse of an entire financial system. The outcome of this systemic failure turns out to be severe and negative impact to the worldwide economy. The adverse shock from the economy may cause significant losses in financial institutions and the insolvent bank may default on its interbank payment obligations to other banks due to the systemic nature of financial institutions, such as banks. This is what causes more banks to fail. In the lending practice high credit risk is associated with tied collateral required by banks (Ortiz-Gracia & Masdemont 2011). The recent financial crisis prompted the banking supervisors to designate the credit risk management guidelines. They are aimed to ensure the stability of a whole financial system. Therefore, understanding the measures of risk for the diverse parametric distributions help in decision making, whenever the stress period is experienced. DMR can be controlled when the measures of risk, such as credit-VaR, are accurately quantified and the assumed systemic factor from the credit risk model is well understood and interpreted.

The study assesses model risk when the estimation of credit-VaR using portfolio of loans in which the individual default probabilities are unknown. The credit-VaR of the random variable say, portfolio of losses, is the worst loss expected due to the other counterparty default over a given period with a given probability (Ortiz-Gracia & Masdemont 2011). The credit-VaR estimate is a random variable because the sample size is finite. The estimated credit-VaR in the case of this research is an asymptotically convergent estimator of the true measure of risk credit-VaR, i.e. depending on the particular sample. A different credit-VaR estimate will be observed every time the sample is run, fluctuating around the true credit-VaR (Siegl & West 2001). Credit risk management is concerned about the risk of loss arising from a debtor's inability to respect an obligation. We develop a resampling process known as the parametric Bootstrap Monte-Carlo algorithm (Efron & Tibshirani 1994, Siegl & West 2001). An extensive review around the methods is also presented by Efron (1979), Efron & Tibshirani (1986), Müller et al. (2017), Flowers-Cano et al. (2018) and Pekasiewicz (2016). The bootstrap replicates thereof are then used to estimate the upper confidence level that can be considered when determining the required capital for default events in homogeneous loss distribution. The Basel III and Solvency II guidelines for capital requirements suggest that financial institutions can use their internal models for determining the credit-VaR. However, model risk is a major obstacle when credit-VaR is estimated from the unknown or contaminate loss distribution, hence DMR is considered. We simulate credit losses distributions from different statistical parametric distributions to assess the model results. In some cases, the normality assumption is poor proxy of the actual and unobservable distributions. This usually leads to a higher credit-VaR when the upper level tail of the latter distribution is underrepresented (Batiz-Zuk et al. 2013). Therefore, the parametric bootstrap technique is utilized to create the acceptable sample size to be fed into the credit risk model of interest and analysed.

The contributions to the research work are two-fold. Firstly, we assume RVs from different distributions for the systematic factor found in the Vasicek single factor model for the credit default events and determine the credit-VaR. The DMR is observed in cases where there are significant difference between the bootstrapped upper confidence level and the usual credit-VaR with a certain significant level. Secondly, we develop a modified hybrid percentile (MHP) based on some known bootstrap percentiles described by (Efron & Tibshirani 1994). MHP is the bootstrap upper confidence level with a modified margin of error developed with the bootstrap sample size. The method is compared with other bootstrap methods for calculating the confidence level, namely the Basic Percentile (BP), Standard Hybrid Percentile (SHP) and the Bias-Corrected Percentile (BCP). The remaining sections of this chapter are structured as follows: In Section 3.2 we present the review of the credit risk modelling and the related factors together with their assumed statistical probability distributions. Section 3.3 provides a review of the Credit Value-at-Risk through Bootstrap method and their bootstrap confidence interval. The newly proposed MHP is reflected in this section. In Section 3.4, we present numerical results through graphs and tables, interpretation and discussions. Finally, in Section 3.5 we conclude our research study.

3.2 Credit Risk Modelling

In this section we review the loss distribution of a portfolio, considering the Vasicek one factor model for credit defaults and its assumed different distributions for systematic common factor. Credit-VaR is a measure of credit risk, which uses the statistical concepts to estimate the likelihood that a given portfolio's losses (L) will exceed a certain amount x given the period and the dataset, such as historical credit trends and volatilities. The time period is known as the holding period and the probability or likelihood is known as the confidence interval. credit-VaR is not an estimate of the worst possible loss, but the largest likely loss. For an example, reflect on a credit loss portfolio that consists of default sensitive instruments such as bonds with a rating of AAA. The credit-VaR, is the minimum loss of next year when the worst 0.05% event occurs. Alternatively, in imply that 99.95% of the time the loss will not be greater than credit-VaR. The credit-VaR is measured at the time span of one year and is different from the 10-day convention adopted by market risk VaR.

The loss variable of an instrument can be disintegrated into the debtor's probability of default (PD), the loss given default (LGD), and the exposure at default (EAD). The EAD is assumed to be a constant in this study. LGD is the part of EAD that yields negative impact when a default event is experienced. LGD is usually less than one because many default debtors are originally backed by securities. PD and LGD are positively correlated and this imply that PD and the recovery rate (RR) are negatively correlated. Thus, the RR is the extent to which notional on the debtor who has defaulted can be recovered. If the RR is $a\%$ then the LGD is $(1-a)\%$.

The main issue in computing credit-VaR for a variable of loss portfolio is that the joint PD for two debtors does not follow the independence law because institutions in the same sector tend to default together. The latter is known as concentration risk of institution assets, which is one of the most important factors contributing to systemic banking risk (Beck et al. 2018). Although some of the institutions wish to apply credit-VaR, their involvement in a credit loss portfolio could have turned out to be public recently and their equity return may not be available before the initial public offering. According to Tanner & Wong (1987), there are methods that can be employed to assign the contaminated and/or missing values and estimate the parameters of the predictive models. Statistical methods, such as the expectation maximization and maximum likelihood estimations can be employed to estimate the optimum parameters (Dempster et al. 1977). Model risk for parameter estimation and implementation of model with known parametric distribution is of paramount importance in this regard to be well assessed. Since the time span of credit-VaR is usually one year, it is not feasible to collect enough historic credit loss data for validation purpose but data resampling such as bootstrap methods may be feasible.

The Figure 3.1 reprinted from Bluhm et al. (2016) show all the risk quantities of the credit portfolio that can be identified by means of the loss distribution of the portfolio.

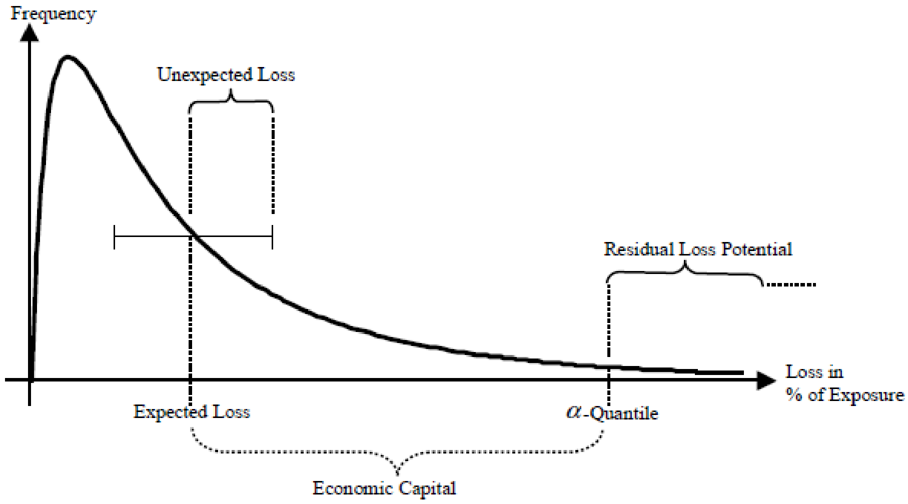


Figure 3.1: The portfolio loss distribution

According to Bluhm et al. (2016), Figure 3.1 shows that empirical statistical quantities can be utilized as a proxy for the respective "True" risk quantities, since distribution of the portfolio loss can be determined in an empirical way. In this research work, the Monte-Carlo simulation is considered to generate a loss distribution. We follow the steps suggested by Morris et al. (2019). The following subsection highlights the single factor model of credit risk.

3.2.1 Vasicek single factor model for credit defaults

Let a credit portfolio consisting of debtors that have an an obligation to honour the contract. The losses incurred due to the default events of debtors are expressed as

$$L_i = c_i Y_i, \qquad \text{for } i = 1, 2, \ldots, n \tag{3.2.1}$$

c_i represent the product of the EAD and LGD and the debtor is subject to default after a fixed time horizon (usually a year). The default can be modeled as a Bernoulli random variable Y_i such that

$$Y_i = \begin{cases} 1, & \text{with probability } \pi_i \\ 0, & \text{with probability } 1 - \pi_i \end{cases},$$

where π_i denotes the debtors probability of default. It follows that the portfolio loss is then given by

$$L = \sum_{i=1}^{n} L_i.$$

Basel II Accord for the evaluation of capital requirements in risk management considered the VaR as an appropriate risk measure to be reported. We consider α to be the confidence level, then the credit-VaR is the α-quantile of the loss distribution of L as shown from Figure 3.1. According to Huang et al. (2007), it is expressed as

$$VaR_\alpha = \inf\{x : P(L \leq x) \geq \alpha\} \tag{3.2.2}$$

Vesicek single factor model proposed and properties considered by Vasicek (2002) derives the loss distribution for a homogeneous portfolio of n loans of equal amount. The distribution of losses is derived with the assumption that all loans in the portfolio have the same expiry date, the same probability of default, and the same pairwise correlation of the debtor's assets. The modeling of the dependence structure among debtors in the portfolio is simplified by the introduction of a common factor that affects all debtors. It is assumed that the standardized asset log-return X_i of debtor i can be decomposed into a systematic common factor Z_i and an idiosyncratic factor ϵ_i independently from Z_i such that

$$X_i = \sqrt{\rho_i} Z_i + \sqrt{1 - \rho_i} \epsilon_i, \tag{3.2.3}$$

where $\epsilon_i \sim N(0,1)$ and Z_i is assumed to be coming from any distribution depending on the credit market standing, any two debtors asset values are correlated with a coefficient ρ_i. For simplicity, let us apply our findings to uniform portfolios by assuming that $\rho_i = \rho$ for all debtors i.

We further assume that Z_i is a predictor variable that is observable. Debtor i will default if its asset value X_i breach the default threshold π_i. Then, conditional on $Z_i = z_i$, the probability of default for the debtor is expressed as

$$\begin{aligned}
\pi_i(z) &= P(X_i \leq \pi_i | Z_i = z_i) \\
&= P(\sqrt{\rho} z_i + \sqrt{1-\rho}\epsilon_i \leq \pi_i) \\
&= P\left(\epsilon_i \leq \frac{\pi_i - \sqrt{\rho} z_i}{\sqrt{1-\rho}}\right)
\end{aligned}$$

Thus,

$$VaR_\alpha(L) = c_i G_L\left(\frac{\pi_i - \sqrt{\rho} z}{\sqrt{1-\rho}}\right) \tag{3.2.4}$$

where $G_L(.)$ is the assumed cumulative distribution function (cdf) representing possible losses L.

Therefore, with above equations (3.2.2) and (3.2.4), the credit-VaR at the α-percentile for an infinitely large portfolio losses without exposure concentration may be expressed as:

$$\mathcal{V}_\alpha(L) = c_i G\left(\frac{G^{-1}(\pi_i) - \sqrt{\rho} G^{-1}(\alpha)}{\sqrt{1-\rho}}\right) \tag{3.2.5}$$

where $G(.)$ is the assumed cumulative distribution function (cdf) representing possible losses L. According to Huang et al. (2007), the equation (3.2.5) has a weakness of underestimating risks for portfolios with few debtors or those dominated by a few large exposures.

3.2.2 Systematic factor distributions

This subsection gives a breif review of some parametric distributions. According to Shi et al. (2017), the usual heavy tailed distributions are Pareto-type such as Pareto, Student's t, Burr, log-gamma, Frechet, and inverse gamma distributions. The right tail of these latter distributions are heavier than that of any exponential distributions.

Let Z_{ik} be the random variable generated from the k^{th} pdf. We consider the random variable to be coming from several parametric distributions which are considered as proxies to the systematic common factor contributing to the losses for equation (3.2.3). All pdfs are given in the Table 3.1 below for the analysis.

In Table 3.1, the first column denotes the random variable (R.V.) as Z_{ik} for i^{th} debtor and the k^{th} pdf given in column two. The third column is the parametric formula of pdf and the fourth column gives the parameter conditions. The last column gives the references of the pdf. Furthermore, the cdfs of the distributions and their properties are explained by the references.

3.3 Credit Value-at-Risk Bootstrap

In this section, we give a review of the statistical resampling method known as the bootstrap technique and the upper bound of four chosen bootstrap confidence intervals. This upper bound is considered to be our statistic of interest, i.e., credit-VaR for the credit losses distribution.

Suppose a random sample drawn from an unknown credit losses distribution function G is represented by $\mathcal{X}_n = \{x_1, x_2, \dots, x_n\}$. The aim is to construct a $100(1-\delta)\%$ upper confidence level for the Credit Value at Risk denoted by $\mathcal{V} = \mathcal{V}(G)$, where $\mathcal{V}(.)$ is the known Credit Value at Risk functional form. The plug-in estimator for $\mathcal{V}$ is given by $\hat{\mathcal{V}}_n = \mathcal{V}(G_n)$, where $G_n(x) = \sum_{j=1}^{m} I(X_j \leq x)$, where $I(.)$ is the

Table 3.1: PDFs assumed for the systematic common factor

Z_{ik}	PDF	$g(Z_{ik})$	Parameter Conditions	Reference
Z_{i1}	Normal	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right]$	$z \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma > 0$	Lyon (2013)
Z_{i2}	Student-t	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$v \in \mathbb{N}$, $\Gamma(.)$ Gamma	Jorion (1996)
Z_{i3}	Cauchy	$\frac{1}{\pi\sigma\left[1+(\frac{z-\mu}{\sigma})^2\right]}$	$\mu \in \mathbb{R}$, $\sigma > 0$	Freue (2007)
Z_{i4}	Log-Normal	$\frac{1}{\sqrt{2\pi\sigma^2}z} \exp\left[-\frac{1}{2}\left(\frac{\ln(z)-\mu}{\sigma}\right)^2\right]$	$\mu \in \mathbb{R}$, $\sigma > 0$	Lyon (2013)
Z_{i5}	Exponential	$\frac{1}{\mu}\exp(-\frac{z}{\mu})$	$\frac{1}{\mu} > 0$	Rice (2006)
Z_{i6}	Gumbel	$\frac{e^{-(z-\mu)/\sigma}}{\sigma} e^{-e^{-(z-\mu)/\sigma}}$	$\mu > 0$, $\sigma > 0$	Szegö (2002)
Z_{i7}	Gamma	$z^{\mu-1}\frac{e^{-z/\sigma}}{\sigma^{\mu}\Gamma(\mu)}$	$\mu > 0$, $\sigma > 0$	Efron & Hastie (2016)
Z_{i8}	Beta	$\frac{1}{B(\mathfrak{a},\mathfrak{b})}z^{\mathfrak{a}-1}(1-z)^{\mathfrak{b}-1}$ where $B(\mathfrak{a},\mathfrak{b}) = \int_0^1 t^{\mathfrak{a}-1}(1-t)^{\mathfrak{b}-1}dt$.	$\mathfrak{a} > 0, \mathfrak{b} > 0$	Efron & Hastie (2016)
Z_{i9}	Weibull	$\frac{\mu}{\sigma}(\frac{z}{\sigma})^{\mu-1}e^{-(z/\sigma)^{\mu}}$	$\mu \in \mathbb{R}$, $\sigma > 0$	McNeil (1999)
Z_{i10}	Exponweib	$\xi\tau(1-\exp(-z^{\tau}))^{\xi-1}\exp(-z^{\tau})z^{\tau-1}$	$\xi > 0$, $\tau > 0$	Cox & Matheson (2014)
Z_{iS}	Sum	$\sum_{k=1}^{10} g(z_k)$	$-\infty \leq z \leq \infty$	
Z_{iP}	Product	$\prod_{k=1}^{10} g(z_k)$	$-\infty \leq z \leq \infty$	

indicator function. Furthermore, suppose the random sample from the known credit losses distribution function G_n is represented by $\mathcal{X}_n^* = \{x_1^*, x_2^*, \ldots, x_n^*\}$, such that the bootstrap estimator of $\mathcal{V}$ is denoted by $\hat{\mathcal{V}}_n^* = \mathcal{V}(G_n^*)$, where $G_n^*(x) = \sum_{j=1}^{m} I(X_j^* \leq x)$. Let the estimator of the asymptotic standard error $\sigma = \sqrt{n}\hat{\mathcal{V}}_n$ be $\hat{\sigma}_n$. We define the standardised statistic of $\hat{\mathcal{V}}_n$ as

$$T_n = (\hat{\mathcal{V}}_n - \mathcal{V})\sqrt{\frac{n}{\sigma^2}}. \tag{3.3.1}$$

When the statistic is standardised but the denominator is an estimated standard deviation rather than the true standard deviation then it is known as the studentised statistic expressed as

$$\mathcal{T}_n = (\hat{\mathcal{V}}_n - \mathcal{V})\sqrt{\frac{n}{\hat{\sigma}_n^2}}. \tag{3.3.2}$$

The theoretical δ-level percentiles of the statistics, T_n and $\mathcal{T}_n$, are respectively denoted by ω_δ and υ_δ as follows

$$P(T_n \leq \omega_\delta) = P(\mathcal{T}_n \leq \upsilon_\delta) = \delta. \tag{3.3.3}$$

Similarly, we consider the bootstrap version of the standardised and studentised statistics of the estimator from bootstrap $\hat{\mathcal{V}}_n^*$ to be respectively denoted by T_n^* and $\mathcal{T}_n^*$ and expressed as

$$T_n^* = (\hat{\mathcal{V}}_n^* - \hat{\mathcal{V}}_n)\sqrt{\frac{n}{\hat{\sigma}_n^2}} \qquad \text{and} \qquad \mathcal{T}_n^* = (\hat{\mathcal{V}}_n^* - \hat{\mathcal{V}}_n)\sqrt{\frac{n}{\sigma_n^{*2}}}, \tag{3.3.4}$$

where σ_n^{*2} is an estimator from the function of $\mathcal{X}_n^*$, represented the same as σ_n^2 is from the function of $\mathcal{X}_n$. The δ-percentiles of the bootstrap distributions of T_n^* and $\mathcal{T}_n^*$ are respectively denoted as $\hat{\omega}_\delta$ and $\hat{\upsilon}_\delta$, such that

$$P^*(T_n^* \leq \hat{\omega}_\delta) = P(\mathcal{T}_n^* \leq \hat{\upsilon}_\delta) = \delta. \tag{3.3.5}$$

Let

$$\hat{F}_n(z) = P^*(\hat{\mathcal{V}}_n^* \leq z) = P(\hat{\mathcal{V}}_n^* \leq z | G_n),$$

then the bootstrap standardised upper level percentile $\hat{\omega}_\delta$ can be expressed explicitly as

$$P^*\left((\hat{\mathcal{V}}_n^* - \hat{\mathcal{V}}_n)\sqrt{\frac{n}{\hat{\sigma}_n^2}} \leq \hat{\omega}_\delta\right) = \delta \implies P^*\left(\hat{\mathcal{V}}_n^* \leq \hat{\mathcal{V}}_n + \hat{\omega}_\delta\sqrt{\frac{\hat{\sigma}_n^2}{n}}\right) = \delta,$$

where P^* is the conditional law of $\mathcal{X}_n^*$ given $\mathcal{X}_n$. Furthermore,

$$\Longrightarrow \hat{F}_n\left(\hat{\mathcal{V}}_n + \hat{\omega}_\delta\sqrt{\frac{\hat{\sigma}_n^2}{n}}\right) = \delta$$

and

$$\Longrightarrow \hat{F}_n^{-1}(\delta) = \hat{\mathcal{V}}_n + \hat{\omega}_\delta\sqrt{\frac{\hat{\sigma}_n^2}{n}}$$

Therefore,

$$\hat{\omega}_\delta = \frac{\left(\hat{F}_n^{-1}(\delta) - \hat{\mathcal{V}}_n\right)}{\sqrt{\frac{\hat{\sigma}_n^2}{n}}} \tag{3.3.6}$$

3.3.1 Confidence Level Bootstrapping

There is an extensive research on numerous ways to calculate bootstrap confidence intervals (Davison & Hinkley 1997). The authors discuss the literature about introduction to bootstrap confidence intervals in detail.

The upper level of the confidence interval $\mathcal{U}$ will be approximated using Monte-Carlo (MC) process suggested by Efron (1979) and Efron & Tibshirani (1994). Among various bootstrap methods to construct the confidence intervals, we consider the Basic Percentile (BP), Standard Hybrid Percentile (SHP) and the Bias-corrected Percentile (BCP) (Samart et al. 2018, Flowers-Cano et al. 2018, Dey et al. 2018). Furthermore, we propose the Modified Hybrid Percentile as an alternative method. The procees for Bootstrap MC is as follows: (i) Draw B independent bootstrap random samples of size n from $\mathcal{X}_n$; (ii) For each bootstrap sample, calculate the bootstrap replicates $\hat{\mathcal{V}}_1^*, \hat{\mathcal{V}}_2^*, \ldots, \hat{\mathcal{V}}_B^*$ and the corresponding order statistics $\hat{\mathcal{V}}_{(1)}^* \leq \hat{\mathcal{V}}_{(2)}^* \leq \cdots \leq \hat{\mathcal{V}}_{(B)}^*$; and (iii) approximate $100(1-\delta)\%$ bootstrap upper confidence level $(\mathcal{U}(\delta))$ with the equation (3.3.6) for the credit value at risk parameter $\mathcal{V}_n$ by the following bootstrap confidence levels:

1. Basic percentile (BP)

$$\hat{\mathcal{U}}_{BP} = \left(-\infty; \hat{\mathcal{V}}_n + \hat{\omega}_{1-\delta}\sqrt{\frac{\hat{\sigma}_B^2}{n}}\right] \tag{3.3.7}$$

where $\hat{\sigma}_B^2 = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\mathcal{V}}_{(b)}^* - \hat{\mathcal{V}}_{(.)}^*\right)^2$ and $\hat{\mathcal{V}}_{(.)}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\mathcal{V}}_{(b)}^*$. Standard errors are crude but useful measures of statistical accuracy. According to Dey et al. (2018) the bootstrap estimate of standard error requires no theoretical calculations, and no matter how mathematically complicated the estimator may be.

2. Standard Hybrid Percentile (SHP) Research shows that the hybrid percentile confidence level have coverage errors close to the nominal value for most of the cases considered (Chuang & Lai 2000, Hall 1988). The SHP is given as

$$\hat{\mathcal{U}}_{HM} = \left(-\infty; 2 \times \hat{\mathcal{V}}_n - \hat{\mathcal{V}}_{(r)}^*\right], \tag{3.3.8}$$

where $r = B \times \delta$.

3. Bias-corrected percentile(BCP) Allowing for the skewness of the distribution of the statistic of interest, such as the credit value at risk $\hat{\mathcal{V}}_n$, BCP method is intended to produce a shorter confidence interval given as

$$\hat{\mathcal{U}}_{BCP} = \left(-\infty; \hat{F}_n^{-1}(\Phi(2\hat{z}_B + z_{1-\delta}))\right], \tag{3.3.9}$$

where $\hat{z}_B = \Phi^{-1}\left[\frac{1}{B}\sum_{b=1}^{B} I(\hat{\mathcal{V}}_{(b)}^* \leq \hat{\mathcal{V}}_n)\right]$.

4. Modified Hybrid Percentile (MHP) Here we consider including bootstrap sample size instead of the random variable sample size in the *margin of error* for the upper bound of the confidence interval given in equation (3.3.7). The margin of error is subtracted from the credit value at risk, $\hat{\mathcal{V}}_n$ to give the bootstrap MHP as

$$\hat{\mathcal{U}}_{HM} = \left(-\infty; \hat{\mathcal{V}}_n - \hat{\omega}_{\delta}\sqrt{\frac{\hat{\sigma}_B^2}{B}}\right]. \tag{3.3.10}$$

3.4 Numerical results and discussions

In this section we perform the numerical experiments and then display and discuss the results for the credit-VaR distributions. We generate the $B = 1000$ bootstrap samples of size $n = 252$ from each of the selected distributions given in Table 3.1. The RVs were generated using the standardized numpy and scipy libraries in Python programme. The numerical results of this study are reported in Tables 3.2, 3.3 and 3.4.

Table 3.2: Descriptive statistics of X_i.

				n = 252	$\rho = 0.25$				
RV	Min	Max	$\bar{X}_k$	s_k^2	γ_k	κ_k	$\hat{\mathcal{V}}_{n,99.00\%}$	$\hat{\mathcal{V}}_{n,99.90\%}$	$\hat{\mathcal{V}}_{n,99.99\%}$
X_1	-2.3819	2.7016	-0.0728	0.9047	0.0611	-0.0409	2.3686	2.6712	2.6986
X_2	-29.7905	162.0119	0.6402	123.0575	12.2130	176.6222	17.9094	127.1956	158.5303
X_3	-13.9657	8.2439	-0.1690	5.7704	-1.8205	7.8443	4.2398	7.4540	8.1649
X_4	-2.2427	8.1548	0.7927	1.8611	1.8550	5.9208	5.9197	7.6667	8.1059
X_5	-2.3367	3.0105	0.4946	0.8686	0.0516	-0.1822	2.6528	2.9274	3.0022
X_6	-2.4435	5.0608	0.2928	1.2677	0.6532	0.8315	3.2686	4.6157	5.0163
X_7	-2.4512	2.4509	-0.0288	0.6888	0.1398	0.1126	2.0933	2.4181	2.4476
X_8	-2.2798	2.7127	0.2122	0.6993	0.1124	-0.0070	2.2006	2.6688	2.7084
X_9	-2.3583	2.9707	0.4622	0.8648	0.1859	0.1606	2.7161	2.9550	2.9691
X_{10}	-2.4511	3.1713	0.0822	0.8170	0.2820	0.3151	2.4104	3.0020	3.1544
X_S	-27.6969	164.5634	2.9654	133.8045	10.8094	149.9406	19.9163	131.1717	161.2243
X_P	-2.4512	2.4509	-0.0288	0.6888	0.1398	0.1126	2.0933	2.4181	2.4476

Table 3.2 show Random Variable (RV) descriptive statistics for a sample size $n = 252$. Min, Max, $\bar{X}_k$, s_k^2, γ_k, κ_k, $\hat{\mathcal{V}}_{n,99.00\%}$, $\hat{\mathcal{V}}_{n,99.90\%}$ and $\hat{\mathcal{V}}_{n,99.99\%}$ are respectively denoting minimum, maximum, sample mean, variance, skewness, kurtosis, credit value at risk at confident level 99.00%, 99.90% and 99.99% of the RVs.

Table 3.2 reveals that the RVs that have heavy tails distributions are given by X_2, X_3 and X_4. These are the standardised asset log-return of the debtor coming from the systematic common factors generated respectively by the Student-t, Cauchy and log-normal distributions. When the sum of all the RV's is assumed then the heavy tail is embedded within the X_s distribution. However, if the product of all the RV's is assumed then the X_p converges to normal distribution. Credit-VaR is the upper confidence level at a certain percentile of the underlying loss distribution. For a credit loss distribution, the duration in terms of time span of concern is often over a number of years, however we consider one year. Therefore, the heavy tailed distribution highlighted give some guidance on the estimate of how large the potential loss could be given some confidence level. Here, we consider the credit-VaR at 99%, 99.9% and 99.99% levels respectively. Thus, an estimation of the credit loss level which will not be exceeded with certain percentile levels. Heavy tail distributions show to exhibit high credit losses. Formal tests (e.g., Shapiro-Wilk Test and Anderson-Darling goodness-of-fit test) and informal tests (e.g., assessing coefficients of skewness and variation, also descriptive plots) for distinguishing whether the data is normally distributed or diversify from normality may be performed (Braione & Scholtes 2016, Premaratne & Bera 2017). In this research work, the informal testing was performed for the RVs given in Table 3.2. The RV X_s clearly shows to diversify from normality assumption and RV X_p is normally distributed.

Table 3.3 shows the summary of the bootstrap credit-VaR replicates at the upper percentile confidence level of 1%, 0.1% and 0.01%. These upper confidence levels are shown with the black line in Figure B.10 to Figure B.13, only for the RVs given by X_s and X_p. From the distribution of the bootstrap credit-VaR replicates the upper confidence levels of 99.00%, 99.90%, and 99.99% are chosen, i.e. $\hat{\mathcal{V}}_n^*$. The respective distributions of the assumed RV's X_{ik} are illustrated graphically in Appendix B.4 from Figures B.1 to B.5. The results of the descriptive statistics in Table 3.2 for the RV's assumed, i.e., the sum X_s and product X_p, are used as systemic common factors to the Vasicek single factor model. Large skewness value ($\gamma_k > 1$) and large kurtosis values ($\kappa_k > 3$) are noted for X_s. This imply that the X_s distribution has heavy and long tails with higher and sharp peak than the normal distribution. The RV X_p distribution reveals to be close to normality since there is skewness value ($\gamma_k \approx 0$) and large kurtosis value ($\kappa_k \approx 3$). The RV X_s show to be asymmetric distribution and X_p show symmetrical distribution, we continue to interpret the bootstrap upper confidence levels given in Table 3.4 based on the two RV distributions as described in Section 3.3. The graphical representation of their distributions are illustrated in Figures B.4 and B.5 respectively.

Table 3.4 shows the RVs of sample size $n = 252$, bootstrap sample size $B = 1000$ and fixed correlation coefficient $\rho = 0.25$ are used to produce the bootstrap descriptive statistics for the credit-VaRs. δ denote the lower confidence level of thecredit-VaR boostrap samples. $\hat{\mathcal{V}}_n^*$ denote the bootstrap credit-VaR. The percentage change pc of the credit-VaR (i.e., $\hat{\mathcal{V}}_n$) from RV to the Bootstrap upper confidence level (i.e., $\hat{\mathcal{U}}_{BP}$, $\hat{\mathcal{U}}_{SHP}$, $\hat{\mathcal{U}}_{BCP}$ and $\hat{\mathcal{U}}_{MHP}$) are given in parenthesis.

Table 3.4 shows the results of four different bootstrap methods (i.e., BP, SHP, BCP and MHP) for determining the confidence levels with their respective percentage change, pc, values from the credit-VaR, $\hat{\mathcal{V}}_n$, to the Bootstrap credit-VaR, $\hat{\mathcal{V}}_n^*$. The percentage change given in parenthesis is calculated as

$$pc = \frac{\hat{\mathcal{U}}_{(.)} - \hat{\mathcal{V}}_n}{\hat{\mathcal{V}}_n} \times 100\%.$$

The negative pc value represent the under-estimation of the credit-VaR and the positive pc represent the over-estimation of credit-VaR. $\hat{\mathcal{U}}_{SHP}$ and $\hat{\mathcal{U}}_{MHP}$ values for X_s at all confidence levels are over-estimated. $\hat{\mathcal{U}}_{SHP}$ values for X_p are over-estimated, while $\hat{\mathcal{U}}_{MHP}$ values for X_p at all confidence levels are under-estimated. It make fundamental sense that the $\hat{\mathcal{U}}_{MHP}$ react like this, because when there is financial crises and stressed markets the defaults or losses tend to increase and create heavy skewed distributions. But, under calm and normal market conditions, the losses distributions are normally distributed. Hence the under-estimation is experienced. $\hat{\mathcal{U}}_{MHP}$ as a credit-VaR can be utilised by regulatory bodies or risk managers in the financial institution to determine the capital charge. Thus, the worst loss expected because of credit counterparty default over some given holding period with a given probability or confidence level. $\hat{\mathcal{U}}_{BP}$ under-estimate the credit-VaR for both the RV's X_s and X_p, which is not an ideal method especially of X_s that exhibits extreme values. $\hat{\mathcal{U}}_{SHP}$ and

Table 3.3: Bootstrap credit-VaR Descriptive Statistics for the given confidence levels.

					$n = 252$	$\rho = 0.25$	$B = 1000$			
RV	δ	Min	Max	$\bar{X}_k$	s_k^2	γ_k	κ_k	$\hat{\mathcal{V}}^*_{n,99.00\%}$	$\hat{\mathcal{V}}^*_{n,99.90\%}$	$\hat{\mathcal{V}}^*_{n,99.99\%}$
X_1	1%	-0.1987	1.1942	0.2141	0.0628	1.0393	0.8536	0.9605	1.1483	1.1896
	0.10%	-0.1962	2.4452	0.4764	0.2196	1.4428	2.6243	2.2157	2.3738	2.4380
	0.01%	-0.1960	2.6682	0.5220	0.2690	1.5336	2.9928	2.5408	2.6645	2.6678
X_2	1%	-0.5523	54.2777	3.2479	33.9650	5.2499	32.8047	36.1059	48.4261	53.6925
	0.10%	-0.5440	142.1234	5.0973	185.6278	7.9211	71.6644	73.6963	141.9421	142.1052
	0.01%	-0.5430	160.0230	5.4176	229.6153	8.1878	76.0801	79.7837	160.0049	160.0212
X_3	1%	-0.5374	2.8221	0.4521	0.2641	1.0423	1.2845	1.9933	2.5651	2.7964
	0.10%	-0.5339	7.4353	0.9452	0.8822	1.7877	5.7852	3.9246	7.0635	7.3981
	0.01%	-0.5334	8.1631	1.0324	1.0902	1.9430	6.9184	4.3384	8.1259	8.1593
X_4	1%	0.6531	3.4849	1.2072	0.1807	1.6990	3.6335	2.7709	3.2545	3.4619
	0.10%	0.6561	7.2899	1.5958	0.8818	2.4528	7.5181	5.3340	7.1984	7.2808
	0.01%	0.6565	8.0683	1.6661	1.0989	2.5377	8.0302	5.8444	8.0591	8.0674
X_5	1%	0.3463	2.0636	0.7823	0.0698	1.2958	1.8937	1.6156	1.9226	2.0495
	0.10%	0.3608	2.9178	1.0430	0.2335	1.2911	1.3466	2.5500	2.8072	2.9067
	0.01%	0.3612	3.0012	1.0893	0.2837	1.3098	1.3434	2.6706	2.9901	3.0001
X_6	1%	0.1589	2.3454	0.6278	0.1008	1.4647	2.7434	1.7343	2.1102	2.3219
	0.10%	0.1602	4.5843	0.9462	0.3717	1.6241	3.8207	2.8789	4.4610	4.5720
	0.01%	0.1605	5.0131	1.0053	0.4620	1.6926	4.1698	3.2444	5.0008	5.0119
X_7	1%	-0.1746	1.2675	0.2142	0.0507	1.2878	2.1853	0.9514	1.2495	1.2657
	0.10%	-0.1742	2.1786	0.4404	0.1705	1.3478	1.9108	1.8026	2.1591	2.1767
	0.01%	-0.1742	2.4217	0.4816	0.2095	1.4000	2.0716	1.9814	2.4132	2.4209
X_8	1%	0.0977	1.3380	0.4511	0.0462	1.2245	1.6737	1.1662	1.2814	1.3323
	0.10%	0.1135	2.4393	0.6524	0.1440	1.2861	2.0285	1.8899	2.4250	2.4379
	0.01%	0.1137	2.6854	0.6863	0.1737	1.3798	2.4864	2.1009	2.6840	2.6853
X_9	1%	0.3512	2.0277	0.7382	0.0608	1.3659	2.5094	1.4735	1.9697	2.0219
	0.10%	0.3541	2.8738	0.9828	0.2167	1.3157	1.5341	2.5108	2.7302	2.8595
	0.01%	0.3548	2.9610	1.0266	0.2669	1.3656	1.6608	2.7153	2.9414	2.9591
X_{10}	1%	-0.0233	1.4041	0.3570	0.0592	1.2362	1.6625	1.1469	1.3194	1.3956
	0.10%	-0.0099	2.8630	0.5975	0.2043	1.6128	3.5326	2.1801	2.8575	2.8625
	0.01%	-0.0098	3.1405	0.6408	0.2510	1.6861	3.9211	2.4350	3.1399	3.1404
X_S	1%	1.7336	73.4698	5.5095	24.9614	5.8943	51.8297	30.2334	47.3393	70.8568
	0.10%	1.7646	144.2689	6.9105	78.3937	8.2639	91.8926	43.2266	103.7985	140.2219
	0.01%	1.7662	162.5340	7.1412	91.6574	8.6837	102.1522	45.0345	109.9128	157.2719
X_P	1%	-0.1457	1.5748	0.2242	0.0514	1.3557	2.6764	0.9163	1.3114	1.5485
	0.10%	-0.1424	2.2718	0.4496	0.1848	1.5016	2.5913	1.9525	2.2317	2.2678
	0.01%	-0.1421	2.4330	0.4881	0.2234	1.5154	2.5291	2.1565	2.4151	2.4312

Table 3.4: Summary results: Bootstrap confidence levels.

		$n = 252$	$\rho = 0.25$	$B = 1000$		
RV	δ	$\hat{\mathcal{V}}_n$	$\hat{\mathcal{U}}_{BP}$	$\hat{\mathcal{U}}_{SHP}$	$\hat{\mathcal{U}}_{BCP}$	$\hat{\mathcal{U}}_{MHP}$
X_S	1%	19.9163	30.2334	36.4580	14.7371	73.4698
			(51.80%)	(83.06%)	(-26.00%)	(268.89%)
	0.10%	131.1717	103.7985	259.8582	144.9129	144.2689
			(-20.87%)	(98.11%)	(10.48%)	(9.98%)
	0.01%	161.2243	157.2719	319.9155	163.2083	162.5340
			(-2.45%)	(98.43%)	(1.23%)	(0.81%)
X_P	1%	2.0933	0.9163 (%)	3.8240 (%)	3.1825 (%)	1.5748 (%)
			(-56.23%)	(82.68%)	(52.03%)	(-24.77%)
	0.10%	2.4181	2.2317 (%)	3.8272 (%)	3.6996 (%)	2.2718 (%)
			(-7.71%)	(58.27%)	(53.00%)	(-6.05%)
	0.01%	2.4476	2.2678 (%)	3.8726 (%)	3.7474 (%)	2.4330 (%)
			(-7.35%)	(58.22%)	(53.11%)	(-0.60%)

$\hat{\mathcal{U}}_{BCP}$ methods over-estimates credit-VaR for all conditions of the confidence levels chosen and for both X_s and X_p. These two methods are not ideal because they will penalise the counterparty with high capital even under calm market conditions, i.e., when the RV is normally distributed. The result imply that the RV that is normally distributed will be treated like the RV that has fat-tails, skewed and peaked than the normal distribution. $\hat{\mathcal{U}}_{MHP}$ method is an ideal technique to be used for calculating the capital requiment in an event of obligor's default because when there are high extreme defaults in the credit market, it will over-estimate the credit-VaR and when there are normal losses then the credit-VaR is moderately under-estimated. DMR is minimised by using different bootstrap methods for calculating the ideal confidence levels.

3.5 Conclusion

The two-fold contributions to the chapter were achieved. Firstly, we have assessed different RVs produced from the parametric standard symmetric and asymmetric distributions as proxies to the systematic factor for the Vasicek single factor model. The

RVs were further assessed using the descriptive statistics for the purpose of classifying the symmetrical from asymmetrical distributions. Asymmetrical distribution were identified to have heavy or long tailed skewness and kurtosis, while symmetrical have normally distributed. Credit VaR was first determined using the ordered RVs at the upper confidence levels of 99%, 99.9% and 99.99%, then the Bootstrap Monte-Carlo technique was applied to retrieve 1000 replicates. Secondly, the bootstrap replicates distributions were used to calculate the bootstrap confidence levels illustrated in Section 3.3. DMR was highlighted where the bootstrap method underestimated or over-estimated the upper confidence level for asymmetric and symmetrical distributions at the same time. This situation was clearly illustrated by the SHP and BCP bootstrap methods for over-estimating and the BP bootstrap method for under-estimating credit VaR for all given types of distributions. MHP bootstrap method was a superior model in terms of distinguishing between the symmetric and asymmetric distribution and respectively under-estimating and over-estimating the credit-VaR. Furthermore, when the percentile confidence level get smaller, the credit VaR calculated from the RV gets closer to the credit-VaR determined through the MHP bootstrap method. The asymptotic properties of the MHP bootstrap method may be explored for further research studies.

4. Model misspecification

In this chapter we suggest a technique to quantify model risk, particularly model misspecification for binary response regression problems found in financial risk management, such as in credit risk modelling. We choose the probability of default model as one instance of many other credit risk models that may be misspecified in a financial institution. By way of illustrating the model misspecification for probability of default, we carry out quantification of two specific statistical predictive response techniques, namely the binary logistic regression and complementary log-log. The maximum likelihood estimation technique is employed for parameter estimation. The statistical inference, precisely the goodness of fit and model performance measurements, are assessed. Using the simulation dataset and Taiwan credit card default dataset, our finding reveal that with the same sample size and very small simulation iterations the two techniques produce similar goodness-of-fit results but completely different performance measures. However, when the iterations increase the binary logistic regression technique for balanced dataset reveal prominent goodness of fit and performance measures as opposed to the complementary log-log technique for both simulated and real datasets.

Keywords: Complementary log-log, binary logistic regression, model misspecification, probability of default.

4.1 Introduction and background

Lately, model risk (MR) has received considerable attention from both academic researchers and practitioners. McNeil et al. (2005) referred to MR as the risk taken by a financial institution that incurs losses due to risk management models being misspecified or due to some assumptions underlying these models not being met in practice. According to Barrieu & Scandolo (2015), the danger of working with a potentially *not well-suited model* is referred to as model misspecification which is one of the components of MR. Also, model misspecification can be defined as the observational indistinguishable predictive models that are having consequences (Kerkhof et al. 2010). MR occurs because of inappropriateness of modelling, which means that a model will not be capable to solve the problem at hand, such as giving high accuracy of model predictions. According to the Office of the Comptroller of the Currency (2011) and Derman (1996), incorrect model is the risk of using a model that is inappropriate under current dataset conditions. The likelihood of the validity for different models to choose from is an act of quantifying MR, i.e., a probability distribution on the set of models being known. Model uncertainty is when a probability distribution on the set of models is unknown. The penalties of using a misspecified model are of concern in the risk predictive models, where a misspecified model can lead to a highly erroneous prediction or low accuracy. In this chapter we assess MR triggered by model misspecification through quantifying probability of default (PD) modelled from two predictive techniques which are given by Binary Logistic Regression (Logit) and Complementary log-log (Cloglog). The PD model is used to model the binary response variable given the inputs known as the predictor variables. The two predictive techniques allow an easy check as to whether the dependence between the potential predictor variables and PDs are meaningful enough to make a sensible decision (Ohlson 1980). The sensible decision will be of predicting the classification of defaults and non-defaults correctly. Furthermore, we show that given the simulation dataset and Taiwan credit card default dataset which is used by other reseachers Yeh & Lien (2009), Islam et al. (2018), Yang et al. (2018), one of the two predictive techniques exhibits model misspecification characteristics due to its distribution which form part of the assumptions. The logit has symmetrical distribution and the Cloglog has an asymmetrical distribution. The violation of either assumption leads to model mispecification. Thus, having an inappropriate technique to be used for predictions. We further consider a model to be misspecified when the performance measures are unstable or there exist low prediction accuracy levels as the parameter of the predictive techniques are optimised.

Predictive models such as the Logit and Cloglog are the most common techniques used to estimate the PDs based on a historical data of credit defaults (Mullahy 1986). The Logit and Cloglog techniques have been used extensively in a wide variety of applications and in fields that are diverse, such as in quantitative risk management, finance, medicine, engineering, economics, psychology, education among many other fields. PDs may be estimated from the observable prices of credit default swaps, bonds, and options on common stocks. The simplest approach, taken by many financial institutions, is to use external ratings agencies such as Standard and Poors, Fitch or Moody's for estimating PDs. However, their approaches are both developed by financial institutions internally and supplied by third party organizations. Similar approaches can be taken to retail sector to extract default events, utilizing the term **credit-score** as an alternative expression for the PD which draws true attention of the lender. Desai et al. (1997) compared statistical and non-statistical techniques for PD and showed that statistical predictive models, more specifically the Logit technique, is the most appropriate approach to model PDs. Significant support for predictive models that are asymmetrical, such as Cloglog technique, were considered to produce similar results of predictive accuracy and power as the Logit technique (Calabrese 2014). However, we argue and show that given the Logit and Cloglog assumptions, the techniques could be misspecified and give less predictive accuracy or unstable predictive performance measures. The compatibility of the Cloglog technique was supported by researchers because of its natural distribution being asymmetrical (Kitali et al. 2017, Martin & Wu 2018). While, Heilbron (1994) suggest that a class with more frequencies than the other should be given high likelihood. Kitali et al. (2017), Martin & Wu (2018), Calabrese (2014), Yang et al. (2018) discuss other compatible models for different setup of binary and ordinal response variables and further suggests different perfomance assessments that can be performed. Hence, we quantify the predictive techniques through assessing their stability in terms of performance measurements and goodness of fit.

Academic researchers and practitioners across financial institutions have pursued to improve bankruptcy predictive models using various quantitative approaches. One of the first researchers, Ohlson (1980), applied Logit technique to default prediction whereby PDs for the potential borrower were determined. According to Begley et al. (1996) there was a need for alternative improvements in modelling of PDs. Several successive studies have performed similar tests using Logit and Cloglog (Memić 2015, McNeil & Wendin 2007, Müller 2012, Berg 2007, Calabrese et al. 2011). Logit technique is a sophisticated statistical method, and concerns have been expressed regarding its application and interpretation. The concerns have concentrated on assumptions related to the appropriate use, accurate interpretation, and complete reporting (Arjas & Haara 1987). Predictive models' assumptions in financial institutions are of paramount importance (Jefferys & Berger 1991). We acknowledge the improvement of predictive models done over the years by researchers, however in this chapter we draw our attention to model misspecification of predictive models. The model assumptions should be a guide of discovering an unsuitable model and less accuracy prediction model for the simulated and the credit card default datasets.

The outline of the chapter is given as follows: Section 4.2 describes the theoretical PDs models given by the Logit and Cloglog, considering a binary response variable with a predictor variable. Section 4.3 discusses a maximum likelihood estimation methodology for estimating the parameters of the two predictive techniques. Section 4.4 outlines the goodness-of-fit tests, model selection criteria and the performance measures. Section 4.5 outlines the simulation steps followed, experimental results and discussion of results. Finally, Section 4.6 concludes the chapter.

4.2 Statistical techniques for PD

In this section, we review the two statistical predictive techniques chosen to predict PDs and outline some of their properties. The PD is defined broadly by the Basel III as the probability of one or both of the events below occuring:

- The bank considers that the obligor is unlikely to pay its credit obligations in full, without the bank choosing to action such realizing security, if held.
- The obligor has past payments more than 90 days (3 months) on any credit obligation to the bank. For example, overdrafts will be considered as being past due once the obligor has breached an advised limit or been instructed of a limit smaller than current outstanding amounts.

Binary predictive model is a valuable mathematical and statistical scientific method used by different disciplines to model the probability of an event of interest occurring, such as PD. PD is close or equal to a categorical response of interest given one or more predictor variables. The mean response functions are sometimes referred to as the link functions or the cumulative probability distribution (cdf). To simplify, we only focus on the case of a dichotomous (or binary) response variable.

Consider the usual form of regression model suggested by Neter et al. (1996) given by

$$Y_i = F(\mathbf{X}_i, \boldsymbol{\gamma}) + \epsilon_i, \tag{4.2.1}$$

where $i = 1, 2, \ldots, n$ are cases or trials and n is the sample size of the cases. $Y_i \sim Bern(\pi_i)$ is the Bernoulli random response vector with only two values denoted as 0 and 1 with probabilities π_i and $1 - \pi_i$ respectively. ϵ_i is the error terms which depend on the Bernoulli distribution of the response variable, Y_i, and the mean response function $F(\mathbf{X}_i, \boldsymbol{\gamma})$ is the widely used nonlinear function with $\mathbf{X}_i$ as the design matrix of predictor variables and $\boldsymbol{\gamma}$ as the parameter vector. The probability that the response variable takes on the unit value is modeled as $P(Y_i = 1|X_i) = F(\mathbf{X}_i, \boldsymbol{\gamma}) = \pi_i$. The mean response function F(.) maps the single index into [0,1] and generally satisfies

$$F(-\infty) = 0, F(\infty) = 1, \frac{\partial F(.)}{\partial(.)} = f(.) > 0,$$

where $F(.)$ is the cumulative distribution function (cdf) and $f(.)$ is the probability distribution function (pdf). The considered two mean response functions, referred to as Logit and Cloglog techniques, for the PDs are discussed briefly and their graphical representations are compared for balanced simulated dataset in the following Subsections 4.2.1 and 4.2.2.

4.2.1 Logistic mean response function

The main purpose of Logit is to find the best fit and most parsimonious model to predict the response variable. The method is relatively robust, flexible and easily used, and it lends itself to a meaningful interpretation (Pohar et al. 2004). There are usually no assumptions made regarding the distribution of the predictor variables. However, in this chapter we highlight the assumption and prperties for the techniques. Logit technique can be applied to two or more categories of the response variables. Our focus is on the binary response variable with a predictor variable. Harrell & Lee (1985) found that, even when the assumptions are made, most should be satisfied and therefore it will be almost as efficient as other predictive techniques such as the Linear Discriminant Analysis (LDA). Some studies have found that, when using mixed binary, categorical, and continuous random variables (r.v's), the Logit technique gives slightly better results than linear or quadratic discriminant analysis (Fan & Wang 1999). Logit require larger sample size to achieve stable classification results. It is of paramount importance to take the properties of Logit into consideration when building a PD model. The properties are listed below without any order of importance.

Properties of the Logit

- **Logit** requires Y_i to be ordinal, such as binary variable with $1 =$ default and $0 =$ non-default cases. The relationship between the Y_i and X_i is clearly nonlinear.

Y_i in Logit is not measured on an interval or ratio measurement scale. According to Cox (2018) a binary Y_i sometimes arises by considering a more complex response variable, such that a component may be classed as default when a quantitative test observation falls outside specification limits or, more generally, when a set of test observations falls in an unacceptable region.

- For Logit, the class category level 1 ($Y_i = 1$) represent the desired response value, that is default events.
- Only the meaningful X_i's should be included in the predictive model of PD.
- The X_i's should be independent of each other if more than one are considered. (i.e., no multicollinearity).
- The X_i's are linearly related to the log-odds, given by $\frac{\pi_i}{1-\pi_i}$.
- X_i can take on all forms of distributions.
- Logit requires quite a large sample size. Therefore, the fewer the observations per X_i, the greater the likelihood for the estimates of the parameters to be unreliable. The standard errors of the estimates and confidence intervals will also be less accurate. A general guideline is that a minimum of 10 cases with the least frequent response for each X_i in the predictive model is needed. For example, if 2 X_i's are obtained and the predicted probability of the least frequent response is 0.10, then a minimum sample size of $200 \times (10 \times 2/0.10)$ cases will be needed.
- Finally, Logit is symmetrical around the value 0.5, thus the PD approaches 0 at the same rate as it approaches 1. Therefore, the model assigns the same importance to the characteristics of defaults as it does to non-default events.

The Logit mean response function is denoted by

$$F(\mathbf{X}_i, \boldsymbol{\gamma}) = P(Y_i = 1|\mathbf{X}_i) = \pi_i = \frac{1}{1 + e^{(-\mathbf{X}_i^T \boldsymbol{\gamma})}}. \tag{4.2.2}$$

Its intrinsically transformation linear form is shown as

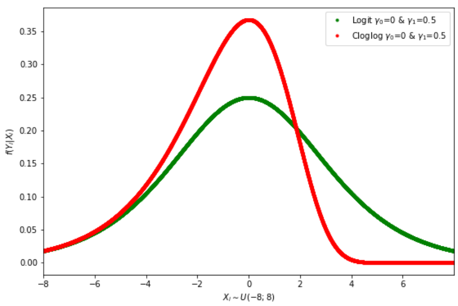
$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{X}_i^T\boldsymbol{\gamma}. \tag{4.2.3}$$

Thus, from equation (4.2.3), the complement of π_i given as the denominator of the odds ratio is expressed as

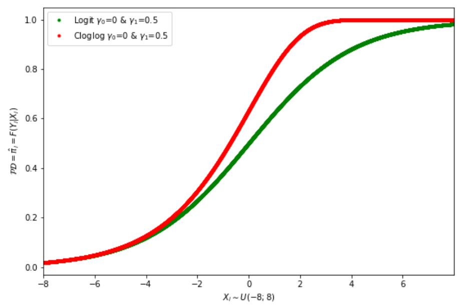
$$1-\pi_i = \frac{e^{-\mathbf{X}_i^T\boldsymbol{\gamma}}}{1+e^{-\mathbf{X}_i^T\boldsymbol{\gamma}}} = \frac{1}{1+e^{\mathbf{X}_i^T\boldsymbol{\gamma}}} \tag{4.2.4}$$

The Y_i distribution is defined by the true probability that $Y_i = 1$ and the model makes no assumption about the distribution of the $\mathbf{X}_i$ - design matrix.

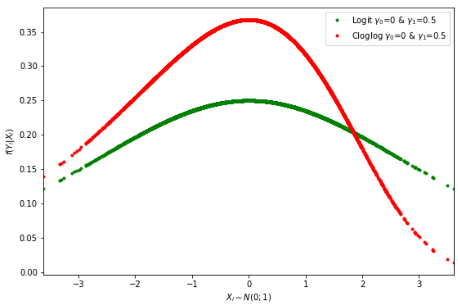
Figure 4.1(a) generated using equation (4.2.2) reveal the way Logit function behaves according to the predictor variable and the parameter. The Logit function acts similar to the standard normal distribution function which is symmetrical with a mean of zero. As the scaled predictor (x_i) moves though the real number axis, π_i increases monotonically between the limits of 0 and 1. The prediction will always fall between 0 and 1 due to the logit transformation given in equation (4.2.3). Whenever π_i is above the threshold 0.5 then it is inferred that an event of default will occur otherwise it is inferred that the event will not occur. Considering one predictor variable, the probability of a binary response variable has γ_0 denoting the location of the curve on the X-axis and γ_1 is the gradient/slope of the curve. Whenever $\gamma_1 = 0$ the curve will represent the threshold for balanced response classes (i.e., $\pi_i = 0.5$) and if $\gamma_1 = 1$ then the curve revert back to a linear regression model. As γ_1 increases then the s-shape curve become stronger.



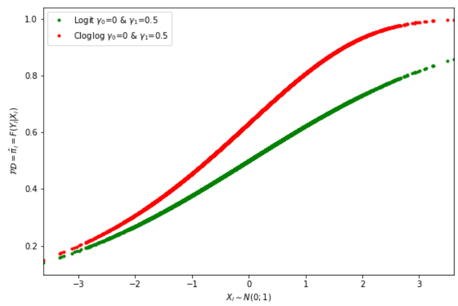
(a) PDFs using uniform random variable.



(b) CDF using uniform random variable .



(c) PDFs using normal random variable.



(d) CDF's using normal random variable .

Figure 4.1: Behavior of the Logit and Cloglog as the two statistical techniques for PD with parameters $\gamma_0 = 0$ and $\gamma_1 = 0.5$.

4.2.2 Complementary Log-Log mean response function

The Cloglog has the similar shape as the Logit except that it is skewed distributed in nature. It is sometimes referred to as the proportional hazards function since it is developed from a generalization of the proportional hazards model for survival datasets to handle grouped survival times (Agresti 2003). This mean response function assumes that the hazard rate increases exponentially with the increasing predictor values, which is the case for extreme value distribution known as the Gompertz. The density function of the Cloglog is sometimes referred to as the *extreme value* or the *Gumbel* probability distribution having zero mean and a unit variance. Most of the properties from the Logit in Subsection 4.2.1 above will also apply in the Cloglog function. The extended Cloglog properties are highlighted below.

Properties of the Cloglog

- **Binary complementary log-log regression** (Cloglog) requires the Y_i to be binary in nature that is 1 for default and 0 for non-default cases.
- Parameter estimates in a Cloglog can be interpreted as effects of the X_i which follows a linear model with reverse extreme value errors.
- To compare parameter estimates with Logit analysis we should divide them by $\sqrt{2}$, or standardize both Cloglog and Logit parameter estimates.
- Figure 4.1 compares the Cloglog function with the Logit after standardizing it to have mean zero and variance one. Although the Cloglog differs from the other technique, one would need extremely large sample sizes to be able to discriminate empirically between these techniques.
- The Cloglog intrinsically linear form has a direct interpretation in terms of PD model parameter estimates.

- Finally, the Cloglog distribution is asymmetric, with a long tail to the left. It has an arithmetic mean equal to Euler's constant of approximately 0.5772 and variance of $\pi^2 = 1.6449$. The median is $-\ln(\ln(2)) = 0.3665$ and the lower and upper quartiles are respectively given as -0.327 and 1.246. Thus, the model assigns more importance to non-default events than it does to default events.

The Cloglog is denoted as

$$F(\mathbf{X}_i, \boldsymbol{\gamma}) = P(Y_i = 1|\mathbf{X}_i) = \pi_i = 1 - e^{-e^{\mathbf{X}_i^T \boldsymbol{\gamma}}}. \tag{4.2.5}$$

Its intrinsically transformation linear form is shown as

$$\ln[-\ln(1 - \pi_i)] = \boldsymbol{\gamma}^T \mathbf{X}_i. \tag{4.2.6}$$

Thus, from equation (4.2.5), the complement of π_i can be expressed as

$$1 - \pi_i = -e^{-e^{\mathbf{X}_i^T \boldsymbol{\gamma}}} \tag{4.2.7}$$

There is one main disadvantage to encounter which is associated with the application of PD models for default prediction, and that is the number of observed default events could be very small. Hence, the use of the Logit techniques may not be appropriate because of its symmetrical nature. It implies that the predicted PD will be approaching zero at the same rate as it approaches one. In other words, the symmetrical models assign the same importance to the characteristics of non-defaults and defaults events, ignoring the scarcity of defaults event. The pdfs plots in Figure 4.1(c) generated using equation (4.2.5) reveals the behaviour of the Cloglog which is skewed to the left. It is asymmetrical, π_i for the cdf shows to approach 0 fairly slow but approaches 1 quite sharply. As the predictor values increases, the predicted mean response is monotone decreasing when $\gamma_1 > 0$ and monotone increasing when $\gamma_1 < 0$.

According to Calabrese & Osmetti (2013) it is not ideal, as the characteristics of defaults in this case are more informative than those of non-defaults. The disadvantage suggests using an asymmetric function, such as that proposed by Calabrese & Osmetti (2013) and Calabrese & Giudici (2015), where the main concentration of the prediction focusses on the tail of the distribution. Thereby implementing a function that lets the predicted PD to approach zero slower than it approaches one. To achieve this, researchers suggest using the inverse of the cumulative distribution function of a generalised extreme value (GEV) random variable as function in a generalised linear model for Bernoulli response variables. Penman & Johnson (2009) suggested that an evaluation of the performance of Cloglog technique is needed to further characterize the bias in the point estimator and to determine the power of goodness of fit tests. For now, it is still viewed as the subject of an ongoing study.

4.3 Maximum Likelihood Estimation

In this section we briefly review the parameter estimation method used in this study. A conventional way to estimate the parameters of a predictive function is through the use of the maximum likelihood estimation (MLE) method (Agresti 2015). Maximizing the log-likelihood function is the same as minimizing the cross-entropy, which will be used in the results. The predictive models are derived in such a way that for n observations the likelihood function $L(\boldsymbol{\gamma})$ or $ln[L(\boldsymbol{\gamma})]$ is maximized with respect to $\boldsymbol{\gamma}$. The process of estimating the parameters $\boldsymbol{\gamma}$ given paired training data set $(X_1, Y_1), \ldots, (X_n, Y_n)$ through maximum likelihood is referred to as learning the PD (Lindsten et al. 2018).

Let $\pi_i = P(Y_i = 1|\mathbf{X}_i)$ alternatively $P(Y_i = 0|\mathbf{X}_i) = 1 - \pi_i$ can be considered if the non-defaults events are of interest. The probability distribution function of the

Bernoulli response variable Y_i is then given by

$$P(Y_i|x_i) = \pi_i^{Y_i}(1-\pi_i)^{(1-Y_i)}. \tag{4.3.1}$$

Given that Y_i have mutually exclusive outcomes, i.e. $P(Y_i = 1|Y_i = 0) = 0$, the likelihood function is

$$L(\boldsymbol{\gamma}) = \prod_{i=1}^{n} \pi_i^{Y_i}(1-\pi_i)^{(1-Y_i)}. \tag{4.3.2}$$

The log-likelihood function which is practically referred to as the loss function is

$$\mathcal{L}(\boldsymbol{\gamma}) = ln[L(\boldsymbol{\gamma})] = \sum_{i=1}^{n}[Y_i \ln(\pi_i) + (1-Y_i)\ln(1-\pi_i)]. \tag{4.3.3}$$

The log-likelihood function in equation (4.3.3) can be expressed in terms of the Logit function given in equation (4.2.2) and Cloglog function given in equation (4.2.5) respectively as

$$\mathcal{L}_1(\boldsymbol{\gamma}) = \sum_{i=1}^{n}\left[Y_i(\mathbf{X}_i^T\boldsymbol{\gamma}) - \ln(1+e^{\mathbf{X}_i^T\boldsymbol{\gamma}})\right], \textit{and} \tag{4.3.4}$$

$$\mathcal{L}_2(\boldsymbol{\gamma}) = \sum_{i=1}^{n}\left[Y_i \ln\left(e^{e^{\mathbf{X}_i^T\boldsymbol{\gamma}}} - 1\right) - e^{\mathbf{X}_i^T\boldsymbol{\gamma}}\right]. \tag{4.3.5}$$

Further details of the work is found in Appendices B.1 and B.2. The MLE with respect to $\hat{\boldsymbol{\gamma}}$ are obtained through numerical optimization method, known as Newton-Raphson (NR) method. The ultimate goal is to compute the MLE of the parameters $\boldsymbol{\gamma}$ such that

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \mathcal{L}(\boldsymbol{\gamma}) \tag{4.3.6}$$

Hastie et al. (2001) show that the standard choice is to use the NR algorithm because it is equivalent to the iteratively reweighted least squares (IRLS) algorithm. The NR method iteratively solves nonlinear equations and approximate the parameter estimates through the expression

$$\hat{\boldsymbol{\gamma}}_{k+1} = \hat{\boldsymbol{\gamma}}_k + [-\mathcal{L}''(\boldsymbol{\gamma}_k)]^{-1}\mathcal{L}'(\boldsymbol{\gamma}_k), \tag{4.3.7}$$

where the number iteration $k \in [1, \mathcal{I}]$. In Nocedal & Wright (2006), Shanmugam & Florence (2012) the least square estimates of the $\boldsymbol{\gamma}^T$ are used as initial parameter estimates.

Since there are no analytical solutions for the log-likelihood functions, $\mathcal{L}(\boldsymbol{\gamma})$, the system of nonlinear equations are solved by taking the first partial derivative known as the gradient and the second partial derivative known as the hessian to obtain the estimates of the predictive models while maximizing the log-likelihood function in equation (4.3.3) with respect to $\boldsymbol{\gamma}$, those are respectively given by

$$\mathcal{L}'(\boldsymbol{\gamma}_k) = \frac{\partial \mathcal{L}(\boldsymbol{\gamma})}{\partial \gamma_k} = 0 \tag{4.3.8}$$

$$\mathcal{L}''(\boldsymbol{\gamma}_k) = \frac{\partial^2 \mathcal{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_k \partial \boldsymbol{\gamma}_k^T} = 0. \tag{4.3.9}$$

Equation (4.3.7) is evaluated repeatedly until a convergence criterion is achieved, i.e., $\hat{\boldsymbol{\gamma}}_{k+1} \approxeq \hat{\boldsymbol{\gamma}}_k$. Statistical consistency also holds whenever $\hat{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}$.

The NR algorithm have shown to be an excellent performance and a rapid convergence rate when eight different numerical algorithms for computing MLE's in terms of computational complexity and performance are compared (Minka 2003). It should also be noted that sometimes the method of MLE fails to converge or does not produce any results especially for the Cloglog function due to highly extreme prediction values. But the failure of the MLE procedure itself does not imply lack of MLE. This is because failure may occur due to the overflow error during computation of the cdf at the extreme values of the arguments (Demidenko 2001). Demidenko (2001) shows more reliable approximation formulae mainly for probit MLE, its gradient and hessian to deal with the overflow error problem. The approach could also be applied for Cloglog MLE failure whenever such behavior is realized. The log-likelihood function can be shown to be globally concave for the mean response functions and numerical routines converge well to the unique global maximum. We continue applying equation (4.3.7) until there is essentially no change between the elements of parameter estimates $\hat{\gamma}_{k+1}$ and $\hat{\gamma}_k$ from one iteration to the next. Therefore, when the latter outcomes are realized the MLEs are said to have converged.

4.4 Model performance and Goodness-of-fit

In this section the assessment criteria are outlined according to the measures derived from the confusion matrix. Adequate measure should not only provide the means to compare the predictive models, but be used to drive the learning of predictive models. The test statistics and measures of the goodness-of-fit (GoF) for each predictive model are calculated. Selection criteria are compared to each other and are discussed in the following Subsections. One of the most widely used criteria in the area of finance in particular is the confusion matrix. The average correct classification rate measures the proportion of the correctly classified cases as non-default and default in the dataset. The average correct classification rate to be discussed is a significant criterion in evaluating the classification capability of the proposed predictive models Abdou & Pointon (2011). Furthermore, to assess MR in terms of misspecification, information coming from the confusion matrix is considered. Some of these performance measures are the accuracy, sensitivity, specificity, precision and receiver operating characteristic (ROC), developed to evaluate the performance of predictive models (Allison 2012).

4.4.1 Goodness-of-fit (GoF) tests

To assess the significance and the adequacy of predictive models discussed in Subsections 4.2.1 and 4.2.2 given parameters of interest and a predictor variable, we consider the GoF tests. The GoF tests are purely considered to determine whether predictive model accurately describes the generated dataset or in a practical situation observed dataset.

The **Likelihood Ratio Test** (LR) which tests the null hypothesis $H_0 : F(\mathbf{X}_i, \boldsymbol{\gamma}) = \pi_i$, against the alternative hypothesis $H_a : F(\mathbf{X}_i, \boldsymbol{\gamma}) \neq \pi_i$.

The LR test statistic is given by

$$LR = \frac{L(\hat{\gamma})}{L(null)}, \tag{4.4.1}$$

where $L(null)$ is the likelihood function computed with the model that contains a constant only. $L(\hat{\gamma})$ is obtained using the saturated/full model, referred to in equation (4.3.3). The logarithm of the LR is expressed as

$$LLR = \mathcal{L}(null) - \mathcal{L}(\hat{\gamma}). \tag{4.4.2}$$

The larger the values of LLR for the model of interest, the poor the saturated model is as described by the dataset. The commonly and practically used test statistic is the **Deviance**(Dev). The Dev has a sampling distribution known as chi-square (χ^2) distribution with number of parameters, $p = df$, equal to the degrees of freedom (Nelder & Wedderburn 1972). It is defined as

$$Dev = -2[\mathcal{L}(null) - \mathcal{L}(\hat{\gamma})]. \tag{4.4.3}$$

A small $p - value$ leads to rejecting H_0 and then concluding that the predictive model coefficient is different from zero. Large values of the Dev indicates that the predicted model ($\hat{\pi}_i$) is not significantly fit. The applicable decision rule is that if $Dev \leq \chi^2(1-\alpha, df)$ where df is the degrees of freedom denoted as number of cells minus one, then the predictive model is significantly fit. Alternatively, if the $p-value$ is greater than the significance level α then the predictive model is significantly fit. Other different statistics used to assess the GoF of the chosen model include the Pearson Chi-Square (χ^2) and the Pseudo R^2 which are briefly described below.

Pearson Chi-Square (χ^2) test is one of the GoF that can be utilized simultaneously with Dev. The null hypothesis which imply that the chosen model fits the dataset also holds for the Pearson χ^2 test. The test statistic calculates the overall difference between the observed responses against those predicted by the models. The desired outcome is that large $p-value$ should be obtained to indicate that our predictive model is significantly fit. The test statistic is defined as

$$\textbf{Pearson } \chi^2 = \sum_{i=1}^{n} \left[\frac{(Y_i - \hat{\pi}_i)^2}{\hat{\pi}_i} \right]. \tag{4.4.4}$$

Pseudo R^2 statistic is slightly analogous to the coefficient of determination measure from linear regression model. This statistic takes values in the range from 0 to 1, whereby its larger values indicate a best model fit. However, it should be noted that it cannot be interpreted as the amount of variance in Y_i explained by the predictor variable. But, rather the proportional improvement in $\hat{\pi}_i$ explained by the X_i. It is defined as

$$\textbf{Pseudo } R^2 = \frac{\mathcal{L}(null) - \mathcal{L}(\hat{\gamma})}{\mathcal{L}(null)}. \tag{4.4.5}$$

4.4.2 Model selection criteria

A model selection criterion allows us to choose the most appropriate predictive model in a set of models. A predictive model should have appropriate GoF, significant parameter estimated through MLE and pure model simplicity. The simple predictive model is generally measured by including a few parameters in the model Jefferys & Berger (1991), hence the study considers only one predictor variable. The model selection criteria use Akaike information criterion (AIC) and Bayesian information criterion (BIC).

The AIC is defined as

$$AIC_p = -2\mathcal{L}(\hat{\gamma}) + 2p. \tag{4.4.6}$$

The most appropriate predictive model is chosen when AIC value is the minimum among the AIC values of alternative predictive models.

The BIC is defined as

$$BIC_p = -2\mathcal{L}(\hat{\gamma}) + p\ln(n). \tag{4.4.7}$$

The BIC generally penalizes the model more heavier for including extra parameters than the AIC, mainly due to the relative magnitude of n (Neter et al. 1996). This imply that the coefficient in the second term for AIC is always 2, while for BIC the coefficient is $\ln(n)$. The value of this coefficient is greater than 2 if $n \geq 8$, which is a singularity that always hold in practice.

4.4.3 Performance Measures

A **confusion matrix** contains the information about actual or true values and predicted $\hat{\pi}_i$ given by a Logit and the Cloglog techniques. It is a 2×2 discriminative matrix which directly compares the actual classes to those predicted by the models, as displayed in Table 4.1.

Table 4.1: Confusion Matrix for binary predictive model.

	Predicted : $\hat{\pi}_i \geq \tau$	Predicted :$\hat{\pi}_i < \tau$
Actual: $Y_i = 1$	TP	FN
Actual: $Y_i = 0$	FP	TN

The **True Positive** (TP) denote the number of positive cases that were correctly predicted by the model to be above the threshold value τ, where $\tau \in [0; 1]$. **False Positive** (FP) denote the number of positive cases that were incorrectly predicted. **True Negative** (TN) denote the number of negative cases that were correctly predicted. **False Negative** (FN) denote the number of negative cases that were incorrectly predicted.

The following performance measures are derived from a confusion matrix:

- **Accuracy:** The performance measure is one of the popular measures in PD modelling. It indicates how close the prediction of the models, $\hat{\pi}_i$, are to the actual values, $Y_i = \{0, 1\}$). It is defined as

$$\mathcal{A} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{4.4.8}$$

- **Precision:** It illustrate how many of the predicted positive values indeed belong to the actual $Y_i = 1$ values. It is the proportion of the number of TP values over the number of TP plus the number of FP. It is defined as

$$\mathcal{P} = \frac{TP}{TP + FP}. \tag{4.4.9}$$

- **Recall:** The performance measure computes the proportion of actual positives that the PD models identified correctly. It is proportion of number of TP over the number of TP plus the number of FNs, defined as

$$\mathcal{R} = \frac{TP}{TP + FN}, \tag{4.4.10}$$

also referred to as *Sensitivity* or in statistical terms the *Power*, which is referred to as the probability of $1 - (Type - II - Error)$.

- **Specificity:** It is the equivalence of the $\mathcal{R}$ but for the opposing class which evaluating the accuracy rate of the $\mathcal{PD}_i$ on the negative class, expressed as

$$\mathcal{S} = \frac{TN}{TN + FP}, \tag{4.4.11}$$

also known as *Significance level = 1 - Specificity.* Statistically referred to as the probability of $Type - I - Error$.

- **F1-Measure:** In some practical problems it is common that misclassication costs are non-uniform, therefore more magnitude might be put on Recall than on Precision. Therefore, the more general F-measure which could be interpreted as the harmonic mean of precision and recall is defined as

$$\mathcal{F}_1 = 2 \times \frac{\mathcal{P} \times \mathcal{R}}{\mathcal{P} + \mathcal{R}}. \tag{4.4.12}$$

Similar to the $\mathcal{A}, \mathcal{P}$ and $\mathcal{R}$, the best value for $\mathcal{F}_1$ measure is 1 and the worst case is 0. Whenever the FP and FN are too different, one should consider looking closely at $\mathcal{F}_1$ measure.

- **ROC:** The Receiver Operating Characteristics (ROC) curve is a 2 dimensional plot, which represents the proportion of default cases predicted by the models as non-defaults (i.e., sensitivity or recall $\mathcal{R}$ which is plotted on the vertical axis) versus the proportion of non-default cases predicted as defaults (i.e., *power = Sensitivity=* $\mathcal{R}$ which is plotted on the horizontal axis) at all chosen threshold values. A useful property is that the area under the curve equals the

Mann–Whitney statistic (Kozeny 2015).

All the above model performance measures have an optimal point of 1 for one to conclude that the predictive performance is perfect. According to Abdou & Pointon (2011), the criterion gives an evaluation of the effectiveness of the predictive model's performance. This can cause a serious problem to the financial institutions in the case of the absence of these estimations, especially with the actual defaults predicted as non-defaults events. It is generally believed that the costs associated with probability of $Type - I - error$ (proportion of non-default predicted as default) and probability of $Type - II - error$ (proportion of default predicted as non-default) are significantly different because the misspecification associated with probability of $Type - II - errors$ are much higher than those related with probability of $Type - I - errors$ (Lee & Chen 2005).

4.5 Simulation results and discussions

In this section we present the results of the simulation experiment and discuss the MR for misspecification against appropriate PD predictions. The quantification is done by assessing the performance measures of both the Logit and Cloglog techniques. The parameters of both predictive models are estimated using the MLE as discussed in Section 4.3. The statistical inference analysis, GoF and model selection criteria discussed in Section 4.4 are utilized to demonstrate the model's misspecification.

4.5.1 Simulation setup

The simulation experiment for the results of the study is given in the following chronological steps:

1. Model parameter values are initialized to be $\boldsymbol{\gamma} = (\gamma_0, \gamma_1) = (0, 0.5)$ for the purpose of generating the paired response and predictor variables, (Y_i, X_i).

2. For **scenario A**, a predictor variable, X_i where $i \in [1; n]$, is generated using a uniform distribution [-8; 8] with a step size of $s = 0.0025$, thereby retrieving a sample size, $n = 6400$. For **scenario B** the predictor variable is randomly generated using the standard normal distribution $N(0, 1)$ with the same sample size $n = 6400$. The behavior of the pdf's and cdf's for both the scenarios are illustrated in Figure 4.1, where the Logit and Cloglog techniques are constructed with parameters in step (1).

3. Y_i are drawn from a binomial distribution with $n = 6400$ cases and

$$P(Y_i = 1|X_i) = \frac{1}{1 + e^{-(0.5X_i)}} + \epsilon_i,$$

where $\epsilon_i \sim N(0, 1)$, are the PDs. Thus, given as

$$Y_i = \begin{cases} 1, & \text{if } P(Y_i = 1|X_i) \geq \tau \\ 0, & \text{if } P(Y_i = 1|X_i) < \tau \end{cases} \tag{4.5.1}$$

where τ is the threshold separating default against non-default PDs. If $\tau \geq 0.5$ then an event of default $(Y_i = 1)$ is observed otherwise the event of non-default is observed. In this case, we assume the number of defaults is the same as number of non-defaults, which imply balanced dataset.

4. The parameter estimates, GoF and their model selection criteria are generated on the paired dataset (X_i, Y_i). The results are generated from the Generalized Linear Models class of the Statsmodels library in Python 3.7.1.

5. The dataset is split into 50% training and 50% testing datasets. The training dataset is used to build the models by obtaining the parameter estimates that maximize the MLE. Parameter optimization are obtained through iterations given by $\mathcal{I} = \{1000, 10000, 30000\}$ over Equation (4.3.7) until convergence is

realized. The learning rate for training the model is set to 0.001 with zero number of random states, i.e., the data is not shuffled before being used. The testing dataset is used to evaluate the performance of the models and compared to the performance of the models generated by training dataset.

6. The Logit and Cloglog predictions are presented by examples and discussed.

7. Finally, examples to illustrate model misspecification are presented, the percentage change from one technique to the other is given as

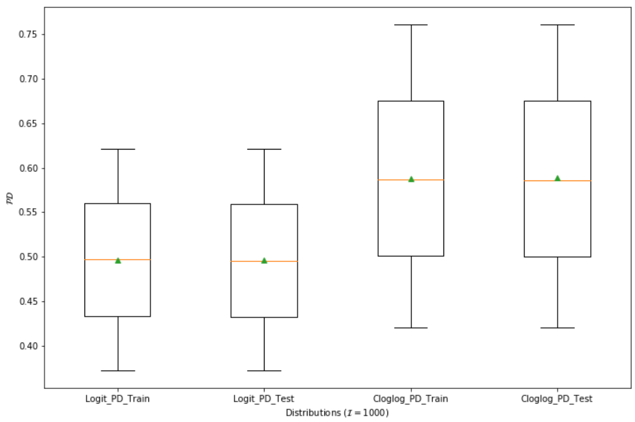
$$pc = \frac{\|\hat{\pi}_{Logit} - \hat{\pi}_{Cloglog}\|}{\|\hat{\pi}_{Cloglog}\|} \times 100\%,$$

   where $\hat{\pi}$ is the estimated PD value from a given technique.

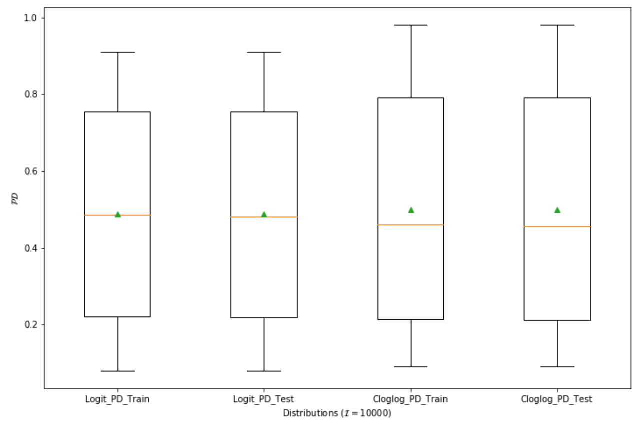
For illustrative purposes of model misspecification, we consider Analyst 1 and Analyst 2 from the same credit risk department in a financial institution. The analysts undertaking the independent process of assessing GoF, model selection and performance measurements with the same dataset. We further make an assumption that all the results generated by Analyst 1 are based on Logit technique and for Analyst 2 are based on the Cloglog technique.

4.5.2 Simulation results

Figure 4.2(a) and (b) show the boxplots, which are generated using the optimized parameter estimates at $\mathcal{I} = \{1000, 10000\}$ for both Logit PD and Cloglog PD models under scenario A and B. The diamond marker is the mean and the orange line denote the median from the distribution. Figures 4.3(a) and (b) show the boxplots with extreme values, which are denoted by the dots.

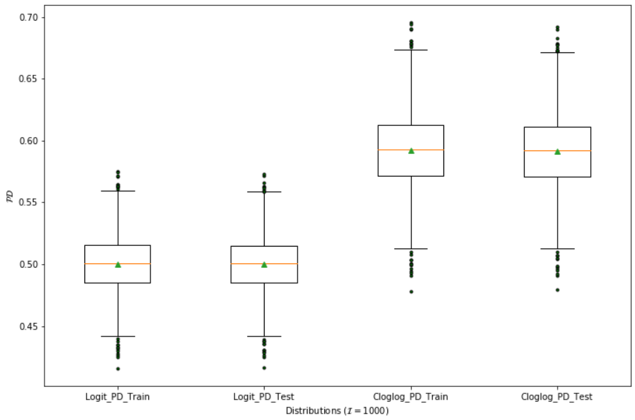


(a) Logit and Cloglog box plots with small iterations.

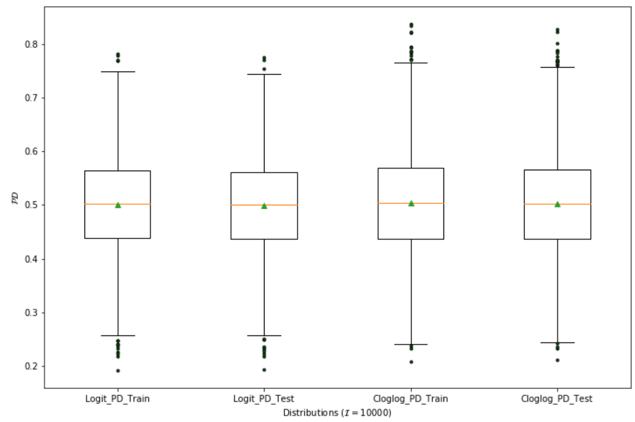


(b) Logit and Cloglog box plots with large iterations.

Figure 4.2: PD models boxplots under Scenario A.



(a) Logit and Cloglog box plots with small iterations.



(b) Logit and Cloglog box plots with large iterations.

Figure 4.3: PD models boxplots under Scenario B.

Table 4.2: Estimated parameters, standard errors and p-values for the predictive models.

PD Model	Variable	$\hat{\gamma}$	$SE(\hat{\gamma})$	z	$p > \|z\|$	2.5% LB	97.5% UB
Scenario A:	$x_i \sim U[-8;8]$,						
Logit	Constant	-0.0099	0.0340	-0.2900	0.7720	-0.0760	0.0570
	x_i	0.4344	0.0100	43.1060	0.0000	0.4150	0.4540
Cloglog	Constant	-0.5329	0.0236	-22.5795	0.0000	-0.5792	-0.4867
	x_i	0.2780	0.006	46.4018	0.0000	0.2636	0.2868
Scenario B:	$x_i \sim N(0,1)$						
Logit	Constant	0.0152	0.0256	0.5934	0.5529	-0.0349	0.0653
	x_i	0.4335	0.0264	16.4043	0.0000	0.3817	0.4853
Cloglog	Constant	-0.3696	0.0184	-20.1352	0.0000	-0.4056	-0.3336
	x_i	0.3050	0.0184	16.5928	0.0000	0.2689	0.3410

Table 4.3: Results for the GoF and the model selection criteria.

PD Model	$\mathcal{I}$	$\mathcal{L}(null)$	$\mathcal{L}(\hat{\boldsymbol{\gamma}})$	Pearson χ^2	Dev	AIC	BIC
Scenario A:	$x_i \sim U[-8;8]$,						
Logit	5	-4436.10	-2721.10	6260	5442.20	5446.1934	-50629.8343
Cloglog	7	-4436.10	-2759.00	7160	5518.00	5521.9652	-50554.4476
Scenario B:	$x_i \sim N(0,1)$						
Logit	4	-4436.10	-4292.30	6400	8584.70	8588.6692	-47487.7436
Cloglog	6	-4436.10	-4293.50	6400	8587.00	8590.9783	-47485.4345

Table 4.4: Model performance measures per given the number of iterations ($\mathcal{I}$).

PD Model	$\mathcal{I}$	TP_{tr}	TP_{ts}	FP_{tr}	FP_{ts}	TN_{tr}	TN_{ts}	FN_{tr}	FN_{ts}
Scenario A:	$x_i \sim U[-8;8]$,	$s = 0.0025$							
Logit	1 000	1335	1289	312	349	1239	1291	314	271
	10 000	1337	1290	312	350	1239	1290	312	270
	30 000	1340	1292	314	351	1237	1289	309	309
Cloglog	1 000	729	741	59	57	1492	1583	920	819
	10 000	1364	1321	344	380	1207	1260	285	239
	30 000	1381	1341	371	412	1180	1228	268	219
Scenario B:	$x_i \sim N(0,1)$,								
Logit	1 000	913	945	638	636	962	977	687	642
	10 000	918	947	644	639	956	974	682	640
	30 000	919	947	646	642	954	971	681	640
Cloglog	1 000	5	5	1	2	1599	1611	1595	1582
	10 000	913	944	638	635	962	978	687	643
	30 000	950	969	682	670	918	943	650	618

Table 4.5: Optimized parameter estimates and the PD models performance measures results for training.

PD Model	$\mathcal{I}$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\mathcal{A}_{tr}$	$\mathcal{P}_{tr}$	$\mathcal{F}_{1tr}$	$\mathcal{R}_{tr}$
Scenario A:	$x_i \sim U[-8;8]$,	$s = 0.0025$					
Logit	1 000	-0.0135	0.2932	0.8063	0.8265	0.8063	0.7871
	10 000	-0.0667	1.3843	0.8063	0.8269	0.8063	0.7866
	30 000	-0.0973	1.8499	0.8066	0.8279	0.8064	0.7860
Cloglog	1000	-0.1237	0.2779	0.7263	0.659	0.7833	0.9652
	10 000	-0.4894	1.085	0.8066	0.8406	0.8028	0.7683
	30 000	-0.5848	1.321	0.8028	0.8487	0.7956	0.7488
Scenario B:	$x_i \sim N(0,1)$,						
Logit	1 000	-0.0001	0.0881	0.6006	0.6035	0.6046	0.6057
	10 000	-0.0026	0.3707	0.6003	0.6035	0.6037	0.6038
	30 000	-0.0037	0.4105	0.5994	0.6027	0.6024	0.6020
Cloglog	1000	-0.1107	0.0825	0.505	0.5045	0.6704	0.9988
	10 000	-0.3665	0.2796	0.6006	0.6033	0.6048	0.6063
	30 000	-0.3817	0.2937	0.5975	0.6041	0.5942	0.5846

Table 4.6: Optimized parameter estimates and the PD models performance measures results for testing.

PD Model	$\mathcal{I}$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\mathcal{A}_{ts}$	$\mathcal{P}_{ts}$	$\mathcal{F}_{1ts}$	$\mathcal{R}_{ts}$
Scenario A:	$x_i \sim U[-8;8]$,	$s = 0.0025$					
Logit	1 000	-0.0135	0.2932	0.8044	0.7978	0.7983	0.7898
	10 000	-0.0667	1.3843	0.8050	0.7988	0.7988	0.7988
	30 000	-0.0973	1.8499	0.8053	0.8001	0.7988	0.7975
Cloglog	1000	-0.1237	0.2779	0.6941	0.6186	0.7530	0.962
	10 000	-0.4894	1.085	0.8034	0.8090	0.7933	0.7782
	30 000	-0.5848	1.321	0.8003	0.8149	0.7869	0.7608
Scenario B:	$x_i \sim N(0,1)$,						
Logit	1 000	-0.0001	0.0881	0.5859	0.5834	0.5922	0.6013
	10 000	-0.0026	0.3707	0.5856	0.5836	0.5905	0.5975
	30 000	-0.0037	0.4105	0.5853	0.5835	0.5898	0.5963
Cloglog	1000	-0.1107	0.0825	0.5013	0.5006	0.6671	0.9994
	10 000	-0.3665	0.2796	0.5859	0.5834	0.5922	0.6013
	30 000	-0.3817	0.2937	0.5838	0.5855	0.5795	0.5738

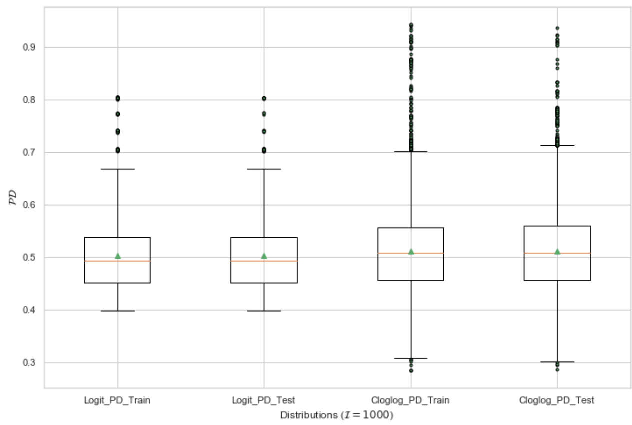
4.5.3 Credit default data set results

The MR procedure used to identify model misspecification in two credit model predictions for the simulated dataset is repeated for Taiwan default payments dataset. The credit default payments datasets with thirty thousands sample size ($n = 30000$)

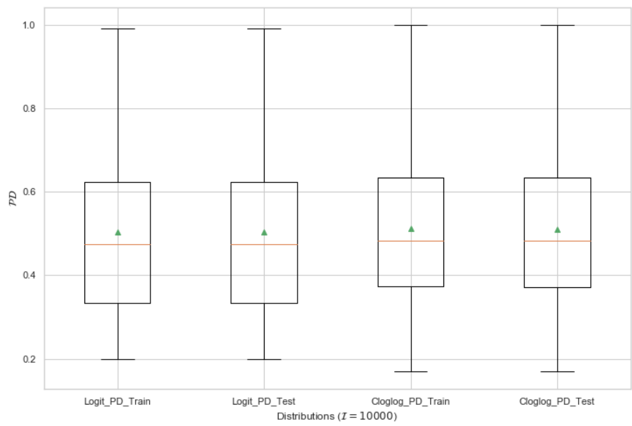
used for this experiment have been downloaded from the University of California at Irvine (UCI) machine learning repository Yeh & Lien (2009). The dataset is the Taiwan's credit card clients' default of 6636 (22.1%) cases and non-default of 23364 (77.88%) cases collected in the year 2005 over the months from April to September. For the purpose of this research, we extract the binary variable of default payments where client's defaulted is equal to 1 and non-defaulted is equal to 0 as the response variable (y_i). The predictor variables of interest extracted are the clients age in years (x_{i1}) and the past monthly payment records of September 2005 (x_{i2}). The scale used for measuring x_{i2} is given by the status of clients repayment, where {-2 = payment duly; -1 = payment delayed for one month; 0 = payment delayed for two months; 1 = payment delayed for three months; 2 = payment delayed for four months; 3 = payment delayed for five months, 4 = payment delayed for six months; and 5 = payment delayed for seven months; 6 = payment delayed for eight months and above}. Step (4) to step (7) given in subsection 4.5.1 are utilized to obtain results of the credit default dataset. To ensure high perfomance accurace between the two predictive models, we split the dataset into 50% training and 50% testing, similarly to simulated data. However the dataset contains 22.12% of client's defaulted, which imply that this in minority class of the response variable. Thus, the credit default dataset is imbalanced. The predictive models generally perform poorly for imbalance dataset problem and result in low accuracy on the defaulted minority class. As a result we implement the Synthetic Minority Oversampling Technique (SMOTE) to balance the credit default dataset Chawla et al. (2002).

Table 4.7: Estimated parameters, standard errors and p-values for the predictive models from credit default dataset.

PD Model	Variable	$\hat{\gamma}$	$SE(\hat{\gamma})$	z	$p > \|z\|$	2.5% LB	97.5% UB
Logit	Constant	-1.6319	0.0581	-28.0687	0.0000	-1.7459	-1.5180
	x_{i1}	0.0065	0.0016	4.1139	0.0000	0.0034	0.0095
	x_{i2}	0.7386	0.0143	51.7227	0.0000	0.7106	0.7665
Cloglog	Constant	-1.6462	0.0487	-33.7810	0.0000	-1.7417	-1.5507
	x_{i1}	0.0043	0.0013	3.2892	0.0010	0.0017	0.0069
	x_{i2}	0.5094	0.0107	47.7840	0.0000	0.4885	0.5303



(a) Logit and Cloglog box plots with small iterations on real data.



(b) Logit and Cloglog box plots with large iterations.

Figure 4.4: PD models boxplots under for credit default dataset.

Table 4.8: Results for the GoF and the model selection criteria for credit default data.

PD Model	$\mathcal{I}$	$\mathcal{L}(null)$	$\mathcal{L}(\hat{\boldsymbol{\gamma}})$	Pearson χ^2	Dev	AIC	BIC
Logit	5	-15853.00	-14257.00	33800	28515.00	28520.7369	-280722.9160
Cloglog	67	-15853.00	-14370.00	5780000	28741.00	28746.6098	-280497.0432

Table 4.9: Model performance measures per given number of iterations ($\mathcal{I}$) for credit default dataset.

PD Model	$\mathcal{I}$	TP_{tr}	TP_{ts}	FP_{tr}	FP_{ts}	TN_{tr}	TN_{ts}	FN_{tr}	FN_{ts}
Logit	1000	9904	10071	1728	1661	5635	5506	5997	6226
	10000	9904	10071	1728	1661	5635	5506	5997	6226
	30000	9904	10071	1728	1661	5635	5506	5997	6226
Cloglog	1000	6380	6632	5252	5100	4100	4032	7532	7700
	10000	8965	9163	2667	2569	5039	4915	6593	6817
	30000	9904	10071	1728	1661	5635	5506	5997	6226

Table 4.10: Optimized parameter estimates and the PD models performance measures results for training credit default data.

PD Model	$\mathcal{I}$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\mathcal{A}_{tr}$	$\mathcal{P}_{tr}$	$\mathcal{F}_{1tr}$	$\mathcal{R}_{tr}$
Logit	1000	-0.0029	-0.0007	0.1790	0.6835	0.7763	0.6196	0.5156
	10000	-0.0434	-0.0019	0.6063	0.6835	0.7763	0.6196	0.5156
	30000	-0.1280	0.0002	0.6459	0.6835	0.7763	0.6196	0.5156
Cloglog	1000	-0.0114	-0.0103	0.1641	0.5980	0.5892	0.6170	0.6475
	10000	-0.1178	-0.0097	0.4309	0.6688	0.7120	0.6312	0.5668
	30000	-0.3011	-0.0049	0.4409	0.6835	0.7763	0.6196	0.5156

Table 4.11: Optimized parameter estimates and the PD models performance measures results for testing credit default data.

PD Model	$\mathcal{I}$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\mathcal{A}_{ts}$	$\mathcal{P}_{ts}$	$\mathcal{F}_{1ts}$	$\mathcal{R}_{ts}$
Logit	1000	-0.0029	-0.0007	0.1790	0.6946	0.7894	0.6347	0.5307
	10000	-0.0434	-0.0019	0.6063	0.6946	0.7894	0.6347	0.5307
	30000	-0.1280	0.0002	0.6459	0.6946	0.7894	0.6347	0.5307
Cloglog	1000	-0.0114	-0.0103	0.1641	0.6108	0.6016	0.6278	0.6563
	10000	-0.1178	-0.0097	0.4309	0.6810	0.7263	0.6456	0.5811
	30000	-0.3011	-0.0049	0.4409	0.6835	0.7763	0.6196	0.5156

Figure 4.4(a) and (b) shows the boxplots, which are generated using the optimized parameter estimates at $\mathcal{I} = \{1000, 10000\}$ for both Logit PD and Cloglog PD models under scenario A and B. The diamond marker is the mean and the orange line denote the median from the distribution. Figures 4.4(a) and (b) show the boxplots with extreme values, which are denoted by the dots.

4.5.4 Discussion

The GoF tests for the Logit and Cloglog techniques from scenarios A and B clearly describe the dataset accuracy by performance measures. Table 4.3 shows the PD models description of the dataset in terms of Pearson χ^2. The Logit technique reveals Pearson $\chi^2 = 6260$ to be describing the dataset more accurately than the Cloglog with a Pearson $\chi^2 = 7160$. The deviance of the two models reveals the same results in favour of Logit technique. However, the results presented in Tables 4.3 and 4.8 reveals Logit technique to have minimum AIC and BIC. This will lead to Analyst 1 results being favoured, i.e., the Logit technique is an appropriate model against the Cloglog technique developed by Analyst 2. We consider the following examples for further discussions.

Example 1: Prediction of the PD models before parameter optimization

The Logit and Cloglog prediction techniques from Table 4.2 with a value of $x = 0.5$ from the predictor variable generated from **Scenario A**, respectively gives predictions as

$$\begin{aligned} \hat{\pi}_{Logit} &= \frac{1}{1 + e^{-(-0.0099+0.4344\times0.5)}} = 0.5516 \\ \hat{\pi}_{Cloglog} &= 1 - e^{-e^{(-0.5329+0.2780\times0.5)}} = 0.4905 \end{aligned}$$

We notice a first-hand discovery of MR due to model misspecification with the above example. The discrepancy is found between the values $x = -2$ and $x = 4$ when predicting PD for scenario A. Further discrepancy is found at the extremes for both scenarios shown in Figure 4.2 and Figure 4.3. The result shows that Analyst 1 model predict $\hat{\pi}_{Logit} = 0.5516$, which is a $pc = 12\%$ change to Analyst 2 prediction of $\hat{\pi}_{Cloglog} = 0.4905$.

The Logit and Cloglog prediction techniques from Table 4.2 with a value of $x = 0.5$ from the predictor variable generated from **Scenario B**, respectively gives predictions as

$$\begin{aligned}
\hat{\pi}_{Logit} &= \frac{1}{1+e^{-(0.0152+0.4335\times 0.5)}} = 0.5577 \\
\hat{\pi}_{Cloglog} &= 1-e^{-e^{(-0.3696+0.3050\times 0.5)}} = 0.5528
\end{aligned}$$

The result shows that both Analyst 1 and 2 models prediction given by $\hat{\pi}_{Logit} = 0.5577$ and $\hat{\pi}_{Cloglog} = 0.5528$ at $x = 0.5$ of scenario B are very close to each, with $pc = 0.9\%$ change.

For the simulated dataset, the same inputs to the two techniques yields different outcomes. Considering credit default dataset, the Logit and Cloglog prediction techniques from Table 4.7 with a middle aged client of 40 years of age ($x_1 = 40$) who has not paid an instalment for the past four months ($x_2 = 2$), respectively gives predictions as

$$\begin{aligned}
\hat{\pi}_{Logit} &= \frac{1}{1+e^{-(-1.6319+0.0065\times 40+0.7386\times 2)}} = 0.5263 \\
\hat{\pi}_{Cloglog} &= 1-e^{-e^{(-1.6462+0.0043\times 40+0.5094\times 2)}} = 0.4696
\end{aligned}$$

The results show that Analyst 1 model predict $\hat{\pi}_{Logit} = 0.5263$ for a middle aged client of 40 years of age ($x_1 = 40$) who has not paid an instalment for the past three months

$(x_2 = 1)$, which is a $pc = 12.07\%$ change to Analyst 2 prediction of $\hat{\pi}_{Cloglog} = 0.4696$.

For the simulated and credit default dataset, the results illustrated here emphasize that Analyst 1 employed an acceptable model because the Logit predictive technique does not either over-estimate or under-estimate the PD $\hat{\pi}_i$ for any distribution coming from scenario A or B, even for the inputs of real dataset. However, Analyst 2 Cloglog model under-estimate the $\hat{\pi}_i$ for scenario A and over-estimate for scenario B. Therefore, Analyst 2 results reveals a risk of using a predictive technique such as Cloglog inappropriately under the assumed predictor variables conditions and the natural distribution of Cloglog. Therefore, the models should be checked if they are statistically accurate and theoretically robust by comparing with alternative modelling techniques, approaches and theories (Mashele 2016).

Analysis of the PD model's performance measures

The simulated and credit default datasets show perfomance results for Logit and Cloglog in Tables 4.4 to 4.6 and Tables 4.9 to 4.11 respectively. These tables are produced according to the description given in Subsection 4.5.1. When the number of iterations is set to $\mathcal{I} = 1000$, convergence is not realized for scenario A and B on the parameters of both techniques as shown in Appendices C.3 and C.5. Convergence is clearly shown when the number of iterations is $\mathcal{I} = 30000$ shown in Appendices C.4 and C.6. Similar results are obtained for real dataset. In the process of optimizing parameter estimates $(\hat{\boldsymbol{\gamma}})$ for both techniques through maximizing their log-likelihood function, $\mathcal{L}(\hat{\boldsymbol{\gamma}})$, we show the performance measures. All the perfomance measure for Cloglog training and testing when $\mathcal{I} = 1000$ are under-estimated. Tables 4.4 and 4.5 for simulated dataset and 4.10 and 4.11 for real dataset of Taiwan credit card default confirms the latter with the accuracy rates. The increasing accuracy rates for testing dataset $\mathcal{A}_{ts}$ when $\mathcal{I} = 1000$ to $\mathcal{I} = 30000$, which is shown by the Logit technique when testing under scenario A and also with real dataset, confirms the increasing power and reliability of the Logit predictive technique. Thus, the performance measures of

Analyst 1 model are stable for simulated and also for credit default dataset when the Logit is used.

Example 2: Prediction of the PD models after parameter optimization

We consider the estimated parameters from the training dataset, they are used in quantifying and assessing the performance measures when testing the models. The Logit and Cloglog prediction techniques from the optimized parameter estimate when the iterations, $\mathcal{I} = 30000$, are given in Tables 4.5 - 4.10. A value of $x = 0.5$ from the predictor variable is considered for **scenario A** and the predictions are respectively given as

$$\begin{aligned}\hat{\pi}_{Logit} &= \frac{1}{1 + e^{-(-0.0973 + 1.8499 \times 0.5)}} = 0.5812 \\ \hat{\pi}_{Cloglog} &= 1 - e^{-e^{(-0.5848 + 1.3210 \times 0.5)}} = 0.4485\end{aligned}$$

Even after the MLE optimization for parameter estimates, similar result shown in Example 1 are revealed, this time with an increased $pc = 30\%$ change prediction of Analysis 1 to Analyst 2 under scenario A.

When a value of $x = 0.5$ from the predictor variable is considered for **scenario B**, the predictions are respectively given as

$$\begin{aligned}\hat{\pi}_{Logit} &= \frac{1}{1 + e^{-(-0.037 + 0.4105 \times 0.5)}} = 0.5420 \\ \hat{\pi}_{Cloglog} &= 1 - e^{-e^{(-0.3817 + 0.2937 \times 0.5)}} = 0.5465\end{aligned}$$

The results concur with results given by Example 1, for the simulated dataset.

When a 40 years of age ($x_1 = 40$) who has not paid an instalment for the past four months ($x_2 = 2$) is considered under real dataset, the predictions are respectively given as

$$\begin{aligned}
\hat{\pi}_{Logit} &= \frac{1}{1+e^{-(-0.1280+0.0002\times 40+0.6459\times 2)}} = 0.7635 \\
\hat{\pi}_{Cloglog} &= 1-e^{-e^{(-0.3011-0.0049\times 40+0.4409\times 2)}} = 0.7699
\end{aligned}$$

The bigger picture of these results are shown by Figure 4.4 (a) and (b). Thus the lower 50 percentile of predictions from logit PD are different from those from Cloglog PD predictions. Cloglog predictive technique reveals that its distributional properties are violated because the default events and the non-default events are not of the same importance, this situation is also illustrated by Arjas & Haara (1987).

Model uncertainty is not an issue in this case since the probability distribution of the Logit and Cloglog are known. The MR increases as a result of ignoring properties of the predictive models. The predictor variable has also shown to have different impact on the predictive ability of the models, confirming the results revealed by Memić (2015). When parameter estimation convergence has been attained, at number of iterations $\mathcal{I} = 30000$ for scenario A, the distribution of the Logit technique is still symmetrical, i.e., the mean is equal to the median, while the distribution of the Cloglog technique is skewed, supported by the mean which is greater than the median for both training and testing datasets. The extreme and quantiles of Cloglog prediction values are higher than those of the Logit prediction values. Boxplots given in Figure 4.3, reveal that the distribution of the two PD models are the same because the location measures are similar even though the extremes of PD are present. The analysis adheres to the findings from Example 1 and 2. Model misspecification has been revealed by Cloglog predictive technique to be inappropriately applied, also the properties about the form of the predictive model are ignored by Analyst 2. Therefore, the risk from using a predictive model which cannot capture the true nature of events corresponds to the risk of using a model inadequately with the assumed known datasets distributions.

4.6 Conclusion

The new Basel Capital Accord enforces financial institutions to develop models for the prediction of the PDs. These predictive models need to be reviewed on a continuous basis in order to minimize MR and thus avoiding model misspecification and inappropriate use of the model. A simulation study was conducted to quantify model risk which is caused by model misspecification. The results reveal that the Logit technique is significantly acceptable under any of the scenarios assumed for simulated dataset, also when the real dataset of Taiwan credit defaults is considered. The results for the Logit technique were consistent and stable with the split of training and testing dataset. However, the Cloglog predictive technique as an alternative PD predictive model to Logit technique has shown to be misspecified because its results have been under-estimating the PD when the predictor variable is simulated to follow a uniform distribution. Parameter optimization has increased the Logit technique performance measures significantly. In this chapter, the gap detected was to use different predictive techniques for modeling probability of default with the same predictor variable(s) and at the end attaining different results especially, when interpolating using the same input predictor variable(s). We fill the gap by quantifying MR through the statistical predictive techniques, Logit and Cloglog, using simulation dataset and Taiwan credit card default datasets. We choose to quantify PD model using these techniques among other credit risk models. We show that there exist model misspecification among the two techniques applied to PD. The results reveal to be favouring Logit technique when we consider the properties for a predictor variable that comes from the uniform distribution and also when it comes from a standard normal distribution for simulated dataset. The Logit predictive technique is a good fit since its performance measurements are stable. The stability of perfomance measures when Logit technique is considered for PD modelling are clearly indicated by the training and testing credit default results shown in Tables 4.10 and 4.11. Accuracy and precision perfomance measures are stable for balanced pair of observations in the simulation dataset and credit card default dataset with a threshold of 0.5. Recommendation for future research is to increase the number of predictor variables while varying the threshold. Also, a search for different parameter estimation techniques is suggested, especially for symmetrical and asymmetrical techniques, that can capture the imbalance set of events and perform independent comparison for PD binary responses that carries deficiencies.

5. Inappropriate parameter estimation

The purpose of this chapter is to assess model risk with respect to parameter estimation for a simple binary logistic regression model applied as a predictive model. The assessment is done by comparing the effectiveness of eleven different parameter estimation methods. The results from the historical credit dataset of a certain financial institution confirmed that using several optimization methods to address parameter estimation risk for predictive models is substantial. This is the case, especially when there exists a numerical optimization method that estimates the optimum parameters and minimizes the cost function among alternative methods. Our study only considers a univariate predictor with static sample size of cases. This research work contributes to the literature by presenting different parameter estimation methods for predicting probability of default through binary logistic regression model and determining optimum parameters that minimize the objective model's cost function. The Mini-Batch Gradient Descent method is revealed to be the best parameter estimator.

Keywords: Binary Logistic Regression, Model Risk, Parameter Estimation, Probability of Default.

5.1 Introduction and background

During the financial crisis and the systems reformation that took place in the years 2007 to 2009, financial risk prediction was identified as a major concern for the public afterwards. As a result, an understanding of model risk, especially the parameter estimation risk for predictive models is now of significant interest to academics, policy makers and practitioners. According to Tunaru (2015), parameter estimation risk is a problem in that the dynamic model's specification and parameter set are viewed as being known by the financial model developers and users whereas the true parameter values are basically not known with certainty. Thus, either because of uncertainty in model's specifications or properties related to the parameter estimator being used or the reliability of the estimated parameter computed through inappropriate parameter estimation method(s), such that their proxies are returned to represent the true parameter values. Alternatively, the risk that the estimated parameters used in the models are not true representative of future outcomes is parameter estimation risk.

A financial institution that provides services to customers frequently relies on financial models to ensure good service delivery of their financial products, such as personal loans, overdraft facility and mortgage loan. Their daily operations may be negatively impacted by model risk, mainly due to relying frequently on models for predicting the outcome of future events (such as defaults) and for describing the relationship between variables (e.g., probability of default and the frequency of payments during the contract). Model risk may be increased as a result of the financial crises, because the usual functions of many models are stopped post the crises, investigations and validations are performed for the purpose of reducing similar financial crises happening again in the future. Among the causes of model risk is the inappropriate parameter estimator, which is the risk related to an inappropriate numerical method used to estimate the parameter of a correct model. Developers of the models regularly estimate and change the parameters of the model post the crises without following the entire model development processes. Thus, increasing model risk as a result of inappropriate parameter estimator.

In the literature a comparison study of several statistical methods and unconstrained optimization methods for obtaining the Maximum Likelihood Estimation (MLE) or minimizing the cost function from the binary LRM, which is an objective model that has not been looked at for different organizational fields and academic problems (Minka 2003, Diers et al. 2013, Borowicz & Norman 2006, Millar 2011). Yang et al. (2016) show the use of Iteratively Reweighted Least Squares (IRLS) and Kalman

Filter with Expectation Maximization (EM) in measurement error covariance estimation. They reveal that on average the IRLS converges quickly and gives a more accurate parameter estimate for the model of interest. Despite the concavity of the objective function, literature reveals that for some data the MLE may not exist. The issue of the MLE existence in binary LRM was considered by Silvapulle (1981), Candès & Sur (2018), Wang et al. (2018) and Albert & Anderson (1984). MLE computed through IRLS method which is like the Newton-Raphson (NR) method is a widely preferred parameter estimator and has desirable properties of large-sample normal distributions, asymptotical unbiasedness, asymptotical consistency, convergence of the parameters as the sample size n increases and asymptotical efficiency, producing large-sample standard errors no greater than those from other estimation methods (Agresti & Kateri 2011). According to Diers et al. (2013), asymptotic normality approach for modelling parameter risk takes advantage of the fact that under the true distribution model, for example the binary LRM, commonly used parameter estimators are asymptotically normally distributed with zero mean and the asymptotic variance-covariance matrix of the estimator as the sample size goes to infinity. The asymptotic variance-covariance matrix may be constructed using the inverse of the Fisher information matrix (Agresti & Kateri 2011). The distribution of the MLE may be approximated using the normal distribution with the expected parameter estimator and the estimated variance-covariance matrix. Dinse (2011) adopted the EM method for fitting a four-parameter LRM to binary response data, and confirms that EM method automatically satisfies certain constraints, such as finding variance-covariance matrix of estimates, that are more complicated to implement with other parameter estimation methods. Shen & He (2015) proposed an EM test based on a small number of EM iterations toward the logistic normal mixture model likelihood and obtained the test statistic which has asymptotic representation, while Hinton et al. (2018) achieved significantly better accuracy when using the EM algorithm. The EM algorithm for these recent articles has similar implementation to be used in this chapter. Stochastic Gradient Descent (SGD) and its variants were versatile parameter estimators that have been proven invaluable as learning algorithms or step size for large datasets (Bottou 2012). Advice from the Bottou (2012), is for a successful application of these Batch Gradient Descent (BGD), Mini-Batch GD (MBGD) and SGD to be considered when one performs small-scale problems, whereas the majority of researchers allude that the methods work efficiently for large-scale problems (Nocedal & Wright 2006, Minka 2003, Robles et al. 2008, Ruder 2016). Conjugate Gradient (CG) method was applied for comparison of three Artificial Neural Network (ANN) methods in the application of bankruptcy prediction (Charalambous et al. 2000). The latter study provides superior results to ANN methods against those obtained from the LRM. The field of ANN has recently been explored and further research in this regard is eminent. The line search Newton CG methods such as Truncated Newton (TN) method have been highly effective approaches for large-scale unconstrained optimization (Dembo & Steihaug 1983), but their use for LRM has not been fully exploited, hence it has been considered in this chapter. Some of the most popular updates for minimizing the unconstrained nonlinear functions, i.e., the cost function of binary LRM, are the Broyden, Fetcher, Goldfarb and Shanno (BFGS) method and its variant Limited-Memory BFGS (LM-BFGS) methods. The LM-BFGS is mostly used to save on the memory needed for computation of the Hessian matrix that BFGS method usually waste (Nocedal & Wright 2006). The Nelder Mead (NM) simplex method developed after the Powell's (PW) method is considered to be performing efficiently for the computation of symmetrical balanced binary response in a widely used LRM (Noubiap & Seidel 2000, Powell 1964).

The question for this chapter is, given the parameter estimator that has been used extensively in the past years, is there any other optimum parameter estimator that can be utilised to ensure that model risk is significantly reduced? Hence, this research work consider the assessment of model risk using eleven parameter estimation methods for the binary LRM. These methods are chronologically given as BGD, SGD, MBGD, IRLS, EM, NM, PW, CG, TN, BFGS and LM-BFGS. Therefore, our primary interest in this chapter is to explore and compare the methods when the binary Logistic Regression Model (LRM) is used for predicting the probability of default (PD) in credit risk. The fundamental returns of the latter, will be to limit the lending exposure and reduce the risk associated with the financial institutions against counterparties. Furthermore, this is done to address suggestions made by the Banking Supervision in the Basel II framework (see Basel (2004) and Caruana (2010)) that the Financial institutions ought to gauge their model risk and model validation as one component of the Pillar 1 Minimum Capital Requirements and Pillar 2 Supervisory Review Process guiding the process. The Basel Committee on Banking Supervision (BCBS), i.e. on Basel III international regulatory framework for banks, picked some of the model risk scenarios such as high ratings (i.e., AAA's) in structured finance instruments, e.g. the mortgage-backed security, which financial institutions and investors believed to be validated by credit enhancement methods from agencies as good credit. The financial crises triggered by financial models started when the rating agencies downgraded majority of the structured finance instruments to being useless or of no value. The subprime mortgage market in the United States that developed into a full-blown international banking crisis with the collapse of the investment bank Lehman Brothers on September 15, 2008 led to the largest bankruptcy ever recorded by then (Schiereck et al. 2016). Given the latter crises, it is evident that lack of model risk management could have contributed significantly.

According to Derman (1996) and Mashele (2016), a financial model may be correct in an idealized world but incorrect when realities are taken into account. Therefore, model risk occurs because either the financial model may be used inappropriately or the financial model may have fundamental errors which can occur at any point from design through implementation and may produce inaccurate outputs when viewed against the design objective and intended business uses. There are three main reasons for model risk to be specified, given as

- the model parameters may not be estimated correctly,
- the model may be misspecified,
- the model may be incorrectly implemented.

Our focus in this chapter is drawn to the first bullet, where the model parameters may be estimated inappropriately. We refer to this situation as parameter estimation risk caused by the use of an inappropriate parameter estimator. Disregarding parameter estimation risk means that point estimates for the model given the dataset are computed, whereas, it is known from estimation theory that an estimator is a random variable by itself (Agresti 2015). Thus, a point estimator may not be asymptotically unbiased, efficient, consistent or normally distributed. Hence, techniques that are heavily reliant on using a point estimate sometimes neglect the properties of MLE and most importantly parameter estimation risk. Hence, we propose comparison of unconstrained optimization problems using the binary LRM with application of statistical parameter estimation and numerical optimization methods prior to the PD model implementation and practical considerations.

The contribution of this chapter is presenting different parameter estimation methods for predicting PD through binary LRM and determining optimum parameters that minimize the objective model's cost function. The parameter estimation method with a minimum cost function among the other methods is considered to be the better parameter estimator. Thus, the high the binary LRM cost function the more inappropriate the parameter estimator becomes.

The remainder of the chapter is organized as follows: Section 5.2 briefly describes eleven parameter estiamtion methods for determining the optimum parameter through minimizing the cost function of the binary LRM. Section 5.3 provides the simulation construction process and the experimental results. Section 5.4 presents the application results from the real dataset. Section 5.5 discusses the simulation and application results given in the tables and figures. Finally, section 5.6 summarizes and concludes the chapter.

5.2 Parameter estimation methods for predictive models

In this section, the binary LRM with its cost function are briefly described. The eleven parameter estimation methods for minimizing the cost function are all reviewed in the sub-sections. In order to examine factors influencing a decision of whether an obligor experiences a default event or not, we consider the following binary LRM to quantify the PD model, recommended by Neter et al. (1996):

$$Y_i = \mathbf{X}_{i,p}^T \boldsymbol{\gamma} + \epsilon_i, \tag{5.2.1}$$

where Y_i is a binary response variable indicating the status of the obligor, which should satisfying the following:

$$Y_i = \begin{cases} 1, & \text{if default event occurs} \\ 0, & \text{otherwise} \end{cases},$$

$\mathbf{X}_{i,p}$ is the design matrix of $p = 2$ predictor variables with the sample size $n \in \mathbb{N}$, cases $i = 1, 2, \ldots, n$, $\boldsymbol{\gamma}$ is the vector of parameters for the binary LRM and assume that the error terms ϵ_i are independent and identically logistic distributed. We let the conditional probability $\pi_i = P(Y_i = 1|\mathbf{X}_{i,p}^T)$ to be PD event given the predictor variables, denoted by the logistic function as

$$\pi_i = \frac{1}{1 + e^{(-\mathbf{X}_{i,p}^T \boldsymbol{\gamma})}}. \tag{5.2.2}$$

The model in equation (5.2.1) can be estimated by MLE techniques and the use of Logit function given by

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{X}_{i,p}^T \boldsymbol{\gamma}. \tag{5.2.3}$$

The objective of the study is to estimate the parameter vector, $\boldsymbol{\gamma}$, such that the cost function is minimized or the following log-likelihood function is maximized

$$\mathcal{L}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \left[Y_i \ln(\pi_i) + (1 - Y_i) \ln(1 - \pi_i) \right]. \tag{5.2.4}$$

Since the maximization of $\mathcal{L}(\boldsymbol{\gamma})$ is the same as minimization of $-\mathcal{L}(\boldsymbol{\gamma})$, we consider minimizing the average cost over the entire dataset, and denote it with the cost function given as

$$\mathcal{C}(\boldsymbol{\gamma}) = -\frac{1}{n} \sum_{i=1}^{n} \left[Y_i \ln(\pi_i) + (1 - Y_i) \ln(1 - \pi_i) \right]. \tag{5.2.5}$$

Therefore, the estimator of interest is shown as

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \left[\mathcal{L}(\boldsymbol{\gamma}) \right] = \arg\min_{\boldsymbol{\gamma}} \left[\mathcal{C}(\boldsymbol{\gamma}) \right]. \tag{5.2.6}$$

To find the estimates given in equation (5.2.6), we use different estimation methods described in the following sub-sections. Equation (5.2.5) which is the cost function $\mathcal{C}(\gamma)$ of the binary simple LRM represents the cost that the PD ($\mathcal{PD}_i = \hat{\pi}_i$) that a model will have to pay if it predicts a value $\hat{\pi}_i$ while the actual cost label turns out to be Y_i. For model risk mitigation the optimal parameter estimation method should ensure that the cost function is minimized among all other optimization methods.

5.2.1 Batch Gradient Descent

The BGD method is a first-order iterative optimization algorithm for finding the minimum of a nonlinear function. It minimizes the cost function $\mathcal{C}(\boldsymbol{\gamma})$ iteratively by starting from an initial random value of $\boldsymbol{\gamma}$ and update the parameter values using some step size referred to as the learning rate (Bottou 2012). For each iteration step, the parameter value $\boldsymbol{\gamma}$ is updated by

$$\gamma_{k+1} = \gamma_k - \lambda \nabla \mathcal{C}(\gamma_k), \tag{5.2.7}$$

where

$$\nabla \mathcal{C}(\gamma_k) = \frac{\partial}{\partial \gamma_k} \mathcal{C}(\boldsymbol{\gamma})$$

and λ is the learning rate. The formulation can be described as starting from some random parameter value γ_0 and then for every iteration $k \geq 0$ towards the direction of $-\nabla \mathcal{C}(\gamma_k)$ by the learning rate λ to the next parameter value γ_{k+1}, this is done recursively until converging to a stationary parameter value. The gradient of the cost function with respect to the slope γ_1 is given as

$$\nabla \mathcal{C}(\gamma_1) = \frac{1}{n} \sum_{i=1}^{n} X_i(\pi_i - Y_i). \tag{5.2.8}$$

The gradient of the cost function with respect to the intercept γ_0 is given as

$$\nabla \mathcal{C}(\gamma_0) = \frac{1}{n} \sum_{i=1}^{n} (\pi_i - Y_i). \tag{5.2.9}$$

Therefore, equations (5.2.8) and (5.2.9) respectively may be expressed for slope γ_1 as

$$\gamma_{1,k+1} = \gamma_{1,k} - \frac{\lambda}{n} \sum_{i=1}^{n} X_i(\pi_i - Y_i), \tag{5.2.10}$$

and for the intercept γ_0 as

$$\gamma_{0,k+1} = \gamma_{0,k} - \frac{\lambda}{n} \sum_{i=1}^{n} (\pi_i - Y_i). \tag{5.2.11}$$

The disadvantage of the procedure is that, starting from different γ_0 could lead to a distinct optimum γ_{k+1}, for some complicated cost function $\mathcal{C}(\boldsymbol{\gamma})$ with multiple local minima and high computing time per iteration. If the learning rate λ is very small, then the minimization procedure takes more time to converge. Otherwise it could diverge from the optimum parameter.

5.2.2 Stochastic Gradient Descent

The SGD method is an alternative and simplified version of the BGD for minimizing the differentiable cost function Bottou (2010) given in equation (5.2.5). It processes a single case chosen sequentially or randomly per iteration, resulting in the parameters being updated after one iteration in which only a selected case has been processed. The method outputs either the last iterate parameter $\gamma_{n\mathcal{I}}$ or the mean of the iterated parameters

$$\bar{\gamma} = \frac{\sum_{k=1}^{\mathcal{I}} \gamma_k}{\mathcal{I}},$$

where n denote the last case and the $\mathcal{I}$ is the number of iterations (Polyak & Juditsky 1992). Unlike the BGD, the SGD recursively computes the expression as

$$\gamma_{k+1} = \gamma_k - \lambda \nabla \mathcal{C}_i(\gamma_k). \tag{5.2.12}$$

The SGD approximates full gradient by an unbiased estimator given as

$$E\left[\nabla \mathcal{C}_i(\gamma_k)\right] = \nabla \mathcal{C}(\gamma_k).$$

It remains the preferred method when the number of cases in a dataset are too large to fit, or data cases arrive continuously. The SGD method updates the parameter estimates through the gradient of the cost function with respect to γ_1, expressed as

$$\nabla \mathcal{C}_i(\gamma_1) = X_{i_k}(\pi_{i_k} - Y_{i_k}). \tag{5.2.13}$$

The gradient of the cost function with respect to γ_0 as

$$\nabla \mathcal{C}_i(\gamma_0) = (\pi_{i_k} - Y_{i_k}), \tag{5.2.14}$$

where $i_k \in (1, \ldots, n)$ is chosen case at each iteration k. Therefore, equations (5.2.13) and (5.2.14) respectively may be expressed for γ_1 as

$$\gamma_{1,k+1} = \gamma_{1,k} - \lambda X_{i_k}(\pi_{i_k} - Y_{i_k}), \tag{5.2.15}$$

and for the γ_0 as

$$\gamma_{0,k+1} = \gamma_{0,k} - \lambda(\pi_{i_k} - Y_{i_k}). \tag{5.2.16}$$

The regular updates of parameters immediately give an insight into the performance of the model, which can result in faster cost convergence. The noises can make it hard for the method to settle on a cost function which is minimum for the model. According to Bottou (2012), SGD is a very versatile technique, especially for too large datasets.

5.2.3 Mini-Batch Gradient Descent

The MBGD is a sub-method of the BGD and SGD that partition the dataset into small batches of dataset, used to compute the model cost function given in equation (5.2.5) and update model parameter estimates in equation (5.2.6). The sum of the gradient over the mini-batch reduces the time spent for approximated convergence and the average of the gradient further reduces the variance of the SGD. The MBGD chooses a random subset size $b \subseteq (1, \ldots, n)$, such that $b \ll n$ and recursively computes the following expression at each iteration k

$$\gamma_{k+1} = \gamma_k - \lambda \nabla \mathcal{C}_j(\gamma_k). \tag{5.2.17}$$

The MBGD also approximate full gradient by an unbiased estimator given as

$$E\left[\frac{1}{b}\sum_{j=1}^{b}\nabla\mathcal{C}_j(\gamma_k)\right] = \nabla\mathcal{C}(\gamma_k).$$

The MBGD converges in fewer iterations than BGD because parameter estimates are updated more frequently. For notational simplicity, we assume that the number of cases in the dataset n is divisible by the number of mini batches m. Then we partition the cases into m mini batches, each of size b, if $b = n$ then the method is the same as BGD. BGD and SGD are traditional approaches that have high cost, but MBGD have shown to be capable of decreasing the variance in the stochastic estimates, but it also comes at a cost (Konečný et al. 2016). The MBGD method updates the parameter estimates through the gradient of the cost function with respect to γ_1 as

$$\frac{1}{b}\sum_{j=1}^{b}\nabla\mathcal{C}_j(\gamma_1) = \frac{1}{b}\sum_{j=1}^{b}X_{jk}(\pi_{jk} - Y_{jk}), \tag{5.2.18}$$

and the gradient of the cost function with respect to the γ_0 as

$$\frac{1}{b}\sum_{j=1}^{b}\nabla\mathcal{C}_j(\gamma_0) = \frac{1}{b}\sum_{j=1}^{b}(\pi_{jk} - Y_{jk}), \tag{5.2.19}$$

where $b \in (1, \ldots, n)$ and m mini batches of the dataset are chosen at each iteration k. Therefore, equations (5.2.18) and (5.2.19) respectively are substituted in equation (5.2.17) and give an expression that recursively updates γ_1 as

$$\gamma_{1,k+1} = \gamma_{1,k} - \frac{\lambda}{b}\sum_{j=1}^{b}X_{jk}(\pi_{jk} - Y_{jk}), \tag{5.2.20}$$

and an expression that recursively updates γ_0 as

$$\gamma_{0,k+1} = \gamma_{0,k} - \frac{\lambda}{b}\sum_{j=1}^{b}(\pi_{jk} - Y_{jk}). \tag{5.2.21}$$

MBGD is practically preferred by industry since it tries to balance between the robustness of SGD and the efficiency of BGD (Ruder 2016). According to Konečnỳ et al. (2016), MBGD reduces the variance of the gradient estimates by a factor of $\frac{1}{b}$, but it is also b times more expensive.

5.2.4 Iteratively Re-weighted Least Squares

The IRLS is a numerical method used to find the optimum parameter value (γ) that maximizes the log-likelihood function $\mathcal{L}(\boldsymbol{\gamma})$ given in equation (5.2.4) or gives a gradient $\nabla\mathcal{L}(\boldsymbol{\gamma}) = 0$, known as the score function. One of the fastest and most applicable methods for maximizing a function is the NR method Lindstrom & Bates (1988), which is based on approximating $\nabla\mathcal{L}(\boldsymbol{\gamma})$ by a linear function of γ in a small region. Utilizing the first-order Taylor series approximation and determining the starting estimate γ_0 (usually through the Ordinary Least Squares (OLS)), the linear approximation is given as

$$\nabla\mathcal{L}(\gamma) = \nabla\mathcal{L}(\gamma_0) - I(\gamma_0)(\gamma - \gamma_0) = 0,$$

where $I(\gamma)$ is the expected information matrix. Solving for γ results in the expression

$$\gamma = \gamma_0 + I^{-1}(\gamma_0)\nabla\mathcal{L}(\gamma_0).$$

The process is continued recursively to find other parameter estimates. For k iterations, the next parameter estimate is obtained from the previous parameter estimate using the expression

$$\gamma_{k+1} = \gamma_k + I^{-1}(\gamma_k)\nabla\mathcal{L}(\gamma_k), \tag{5.2.22}$$

where the approximate variance-covariance matrix is the inverse of the expected information matrix

$$\nabla^2\mathcal{L}(\gamma_k) = I^{-1}(\gamma_k) = (\mathbf{X}^T\mathbf{W}_k\mathbf{X})^{-1},$$

where $\mathbf{W}_k$ is the diagonal matrix with main diagonal elements given as $\pi_i(1-\pi_i)$. The score function $\nabla\mathcal{L}(\gamma_k)$ is the first derivative with respect to parameter of interest given in equations (5.2.8) and (5.2.9) for k iterations. The following cost function can easily be retrieved, instead of using the log-likelihood function, as

$$\mathcal{C}(\boldsymbol{\gamma}) = -\frac{1}{n} \times \mathcal{L}(\boldsymbol{\gamma}).$$

This is reflected from equation (5.2.4) to equation (5.2.5). For large datasets, the expected information matrix is the estimated variance-covariance matrix for parameter estimate (γ_k) (Agresti 2015). IRLS method using NR will converge to a local minimum of the cost function very consistently and reparameterization is key to ensuring consistent convergence of the NR (Lindstrom & Bates 1988).

5.2.5 Expectation Maximization

The EM method may be utilized to obtain the maximum log-likelihood expectation for the parameter of interest (Dempster et al. 1977). The OLS method is used to obtain the starting parameter, thereafter EM method recursively iterates between expectation and maximization steps until convergence. Hinton et al. (2018) and Shen & He (2015) explains the cost function that is minimized when using the EM procedure to fit a mixture of Gaussians. Similar steps are followed for minimizing the cost function of LRM. At each iteration step k, the first step calculates expectations of the sufficient statistics for the complete dataset, given the dataset and the current parameter estimates. The second step calculates the parameter estimate value that minimizes the cost function of the current complete dataset. Each EM iteration increases the likelihood of the observed dataset. According to Scott & Sun (2013), the EM method is based on Polya-Gamma data augmentation whereby at each iteration k the parameter estimates are updated as follows

$$\gamma_{k+1} = (\mathbf{X}^T\Omega_k\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - 0.5), \tag{5.2.23}$$

where

$$\Omega_k = diag\left(\frac{\tanh[(\gamma_0 + \gamma_1 x_i)/2]}{2(\gamma_0 + \gamma_1 x_i)}\right),$$

for $i = (1, \ldots, n)$ and 0.5 is the chosen probability threshold.

The EM method converges very slowly if a poor choice of initial parameter estimate values are used and its rate of convergence is generally linear (Laird et al. 1987). EM does not automatically provide an estimate for the variance-covariance matrix of the parameter estimates. However, this disadvantage can be easily dealt with by using appropriate methodology associated with the EM (McLachlan & Krishnan 2007). The EM algorithm has an unusual property that when there are no missing cases, the iterations are still computed the same way as IRLS, but the rate of convergence changes from linear to quadratic.

5.2.6 Nelder-Mead Simplex

The NM simplex method has always been the most widely used method for nonlinear unconstrained optimization (Nelder & Mead 1965). The method minimizes a scalar-valued nonlinear function of p real variables using the cost function values and disregarding any derivative information. Convergence to a minimizer is not guaranteed for general strictly convex functions when NM is used, but it requires substantially fewer function evaluations. Audet & Tribes (2018) gives the details for the mechanism of the NM simplex algorithm.

The aim of the NM method is to solve equation (5.2.6) for $p+1$ number of parameters in the function $\mathcal{C}(\gamma)$. It is based on the iterative update of a simplex made of $p+1$ points $\{\mathbf{v}_j\}_{j=1,\ldots,p+1}$, known as the *vertex*. The vertices are related to the function value $\mathcal{C}_j = \mathcal{C}_j(\mathbf{v})$ for $j = 1, \ldots, p+1$. The vertices are ordered by increasing function values in such a manner that the best vertex has index 1 and the worst vertex has index $p+1$, given as

$$\mathcal{C}_1 \leq \mathcal{C}_2 \leq \cdots \leq \mathcal{C}_p \leq \mathcal{C}_{p+1},$$

where $\mathbf{v}_1$ is the best since it relates to lowest cost function $\mathcal{C}_1$ and $\mathbf{v}_{p+1}$ is the worst since it relates to the largest cost function value $\mathcal{C}_{p+1}$. The mean of the simplex

$$\bar{\gamma}(i) = \frac{1}{p} \sum_{j=1, j\neq i}^{p+1} \mathbf{v}_j$$

The method uses one coefficient $\rho > 0$, known as the reflection factor. The standard value of this coefficient is $\rho = 1$. The method attempts to replace some vertex $\mathbf{v}_i$ by the new vertex $\gamma(\rho, i)$ on the line from the vertex $\mathbf{v}_i$ to the mean $\bar{\gamma}(i)$ by the expression

$$\gamma(\rho, i) = (p+1)\bar{\gamma}(i) - \rho\mathbf{v}_i. \tag{5.2.24}$$

The method behaviour is compared against PW method with regards to its free derivative. NM method does not require as many function evaluations as compared to most of its variants and it can become slower as the dimension increases. It converges to a non-stationary parameter point (Lagarias et al. 1998). Further, NM method is efficient in moving to the general area of a minimum point but it is not efficient in converging to a precise minimum value of the function.

5.2.7 Powell

The PW method is an optimization method that approximates the minimum value of a function by making an assumption that the partial derivatives of the cost function does not exist (Powell 1964, 1965, 2007). Let γ_0 be an initial parameter guess at the location of the minimum of the cost function $\mathcal{C}(\boldsymbol{\gamma})$. The method instinctively approximate a minimum of the cost function $\mathcal{C}(\boldsymbol{\gamma})$ given in equation (5.2.5) by generating the next approximation parameter γ_1 by proceeding successively to a minimum of cost function $\mathcal{C}(\boldsymbol{\gamma})$ along each of the λ standard base vectors. The process generates the $\mathcal{I}^{th}$ sequence of points or a set of unit vectors which are chosen to be linear independent directions $\lambda_0, \lambda_1, \ldots, \lambda_{\mathcal{I}}$, in an iteration k. The next parameter point γ_1 is determined to be the point at which the minimum of the cost function occurs, along the vector $\lambda_{\mathcal{I}} - \lambda_0$. The method recursively moves along one direction until a minimum is reached, then moves along the next direction until a minimum is reached again, and so on. The optimization procedure will stop when

$$|\gamma_{k+1} - \gamma_k| < \frac{1}{2}\varepsilon(|\gamma_{k+1}| - |\gamma_k|), \tag{5.2.25}$$

for the $(k+1)^{th}$ and k^{th} iterations. ε is the scalar parameter (tolerance) determining when the optimization procedure should stop.

The PW method is a robust direction set method and does not find the local minimum as quickly as other methods. There is no guarantee that it will find the global minimum for the cost function at the end of all iterations. More implications of this method and its conditions are discussed by Powell (1964).

5.2.8 Conjugate Gradient

The CG is a method that efficiently avoids the calculation of the inverse Hessian by iteratively descending on the conjugate directions (p_k). According to Nocedal

& Wright (2006), the CG method is among the most useful techniques for solving large linear systems of equations and can be modified to solve nonlinear optimization problems. The first nonlinear CG method was introduced by Fletcher and Reeves in 1964 (Babaie-Kafaki & Ghanbari 2015). We briefly describe the use of a nonlinear conjugate gradient method of Polak and Ribiere, which is a variant of the Fletcher and Reeves method. The method only considers the first derivative in the computation. Starting from an initial parameter point γ_0, the CG method generates a sequence of parameter points γ_k given by

$$\gamma_{k+1} = \gamma_k - \lambda_k p_k, \tag{5.2.26}$$

where $\lambda_k > 0$ is a step length obtained from a line search, the search direction p_k of the CG method is defined as

$$p_k = \begin{cases} -\nabla\mathcal{C}_k, & \text{if } k = 0, \\ -\nabla\mathcal{C}_k + \beta_k p_{k-1}, & \text{if } k \geq 1, \end{cases},$$

where $\mathcal{C}_k = \mathcal{C}(\gamma_k)$ and β_k is known as the Polak and Ribiere parameter (PRP) for CG method defined as

$$\beta_k = \frac{(\nabla\mathcal{C}_k - \nabla\mathcal{C}_{k-1})^T \nabla\mathcal{C}_k}{\|\nabla\mathcal{C}_{k-1}\|}.$$

According to Babaie-Kafaki & Ghanbari (2015), Polak and Ribiere technique showed that when the PRP formula and an exact line search are used, the CG method is globally convergent. However, other researchers suggested a non-negative value of the Polak and Ribiere CG formula to ensure global convergence since it does not gurantee global convergence in all nonlinear unconstrained problems (Alhawarat et al. 2017). The CG method uses relatively little memory for large scale problems and require no numerical linear algebra, thus each step is quite fast. It converges much more slowly than Newton or Quasi-Newton methods. There are line search and trust-region implementations of a strategy whereby CG is terminating if negative curvature is encountered, which is called the Newton–CG. Modified Newton method is the second approach consisting of modifying the Hessian matrix $\nabla^2\mathcal{C}(\gamma_k)$ during each iteration so that it becomes sufficiently positive definite.

5.2.9 Truncated Newton

The TN method is also known as the inline search Newton CG method. The TN method uses less and predictable amount of computational storage, and only requires the objective function and its gradient values at each iteration with no other information about the minimization problem. The search direction is computed by applying the CG method to the Newton equations given by

$$B_k p_k = -\nabla\mathcal{C}_k \tag{5.2.27}$$

where $B_k = \nabla^2\mathcal{C}(\gamma_k)$ is the approximation of Hessian at k^{th} iteration and $\nabla\mathcal{C}_k$ is the gradient. When B_k is positive definite, the inner iteration sequence will converge to the Newton step p_k that solves equation (5.2.27). At each iteration, the termination criteria ε is defined as $min(0.5, \sqrt{\|\nabla\mathcal{C}\|})$ known as the forcing sequence (Nocedal & Wright 2006). There are other methods in literature that can be utilized for the choice of the tolerance. If the Hessian is detected to be an indefinite matrix then the CG iteration is terminated. The approximate solution of the search direction p_k is then used in a line search to get an updated parameter point, through the expression

$$\gamma_{k+1} = \gamma_k - \lambda_k p_k, \tag{5.2.28}$$

where $\lambda_k > 0$ satisfies the Wolfe, Goldstein, or Armijo backtracking conditions and $\mathcal{C}_{k+1} < \mathcal{C}_k$ (Nash & Nocedal 1991). The TN method has some similarities to BFGS to be discussed in the next section. For a good performance of the TN method to be realized, the CG stopping criteria need to be tuned so that the method uses just enough steps to get a good search direction.

5.2.10 Broyden Fletcher Goldfarb Shanno

The BFGS method is a Quasi-Newton method also known as a variable metric algorithm. This is a nonlinear optimization method for solving unconstrained problems (Shanno 1970, Nocedal & Wright 2006). The method constructs an approximation to the second derivatives of the cost function, given in equation (5.2.5), using the difference between successive gradient vectors. By combining the first and second derivatives the method can take Newton-type steps towards the minimum value of the function. Therefore, it is a more direct approach to the approximation of Newton's update for the parameter estimates that minimizes the cost function, $\mathcal{C}(\boldsymbol{\gamma})$. The updates at each iteration to the parameter estimates are given by the expression

$$\gamma_{k+1} = \gamma_k + \lambda B(\gamma_k)\nabla\mathcal{C}(\gamma_k), \tag{5.2.29}$$

where λ is the learning rate or the step size, γ_k is the old parameter estimate for the first iteration when $k = 0$, and γ_{k+1} is the new parameter estimate. The procedure adopted by Quasi-Newton methods is to approximate the inverse with a matrix $B(\gamma_k) = \mathbf{B}_k$, which is recursively refined by the low rank updates to become better approximation of the inverse Hessian matrix, $H^{-1} = \frac{1}{\nabla^2\mathcal{C}_k}$. The latter matrix is updated by recursively computing the expression

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k\varsigma_k\varsigma_k^T\mathbf{B}_k}{\varsigma_k^T\mathbf{B}_k\varsigma_k^T} + \frac{\upsilon_k\upsilon_k^T}{\upsilon_k^T\varsigma_k},$$

where

$$\varsigma_k = \gamma_{k+1} - \gamma_k$$

and

$$\upsilon_k = \nabla\mathcal{C}_{k+1} - \nabla\mathcal{C}_k.$$

The properties for BFGS should hold so that the method is efficient. The Hessian matrix, H, is symmetric, so should its inverse. Thus, it is reasonable that at each iteration approximation H_k should be symmetric. If this holds for the update B_k then the B_{k+1} will inherit the symmetry from H_k. Somewhere during the iteration the Quasi-Newton condition given as

$$\Delta\gamma_i = B_{k+1}\Delta v_k,$$

should hold for $0 \leq i \leq k$. As a result of H_k being symmetric and Quasi-Newton condition being satisfied, then the approximation of the Hessian will be positive definite. According to Nocedal & Wright (2006), the BFGS method is the most effective among most of the Quasi-Newton updating methods for unconstrained non-linear problems. It is considered successful due to being highly independent on the line-search methods, such as PW method and others, for determining a parameter point which is very near to the true minimum along the line. The BFGS method spends less time refining each line search but needs a huge memory due to storage of the inverse Hessian matrix, making it impractical if there exist high number of parameters (Vetterling et al. 1992).

5.2.11 Limited-memory Broyden Fletcher Goldfarb Shanno

The LM-BFGS method is an extension of BFGS method that belongs to the variants of Quasi-Newton optimization problems. The method resolves the cost function minimization problem by calculating approximations to the Hessian matrix for the function. The main idea of LM-BFGS method is to use curvature information from the most recent iterations to construct the Hessian approximation. The curvature information from earlier iterations, which is less likely to be relevant to the actual behaviour of the Hessian at the current iteration, is discarded in the interests of saving storage (Nocedal & Wright 2006). The memory costs of the BFGS method can be significantly decreased by computing the approximation Hessian matrix B using the same method as the BFGS algorithm but beginning with the assumption that B_t is an identity matrix, rather than storing the approximation from one step to the next. Its strategy with no storage can be generalized to include more information about the Hessian by storing some of the vectors used to update B at each iteration step, which costs less.

5.3 Simulated Results

In the effort of trying to address Model Risk with respect to parameter estimations risk, we compare the performance of the parameter estimates when the binary LRM is applied to estimate the PD. Several optimization methods, given from Section 5.2.1 to 5.2.11, are employed to find the optimal parameter estimates through minimizing the cost function given in equation (5.2.5). For this, the true underlying parameter values determining the PD must be known. All the codes of the analysis were computed on Python version 3.7.1 with Jupyter Notebook version 5.7.4.

We set the parameter intercept $\gamma_0 = 0.0$ and the parameter slope $\gamma_1 = 0.5$. Therefore, we use the simulation to produce the balanced dataset of default and non-default events. For each of true parameters and sample size of 6 400, the dataset was simulated and analyzed. To keep our model simple, we included only one predictor variable which is uniformly distributed (i.e., $X_i \sim U[-8; 8]$) for which the cost function $\mathcal{C}(\boldsymbol{\gamma})$ is investigated. The dataset was simulated using the LRM and setting the parameter to 0.5 for a Bernoulli distribution resulting in a dichotomous response variable Y_i indicating whether an event occurred or not.

Figure 5.1 was generated using the BGD parameter estimator. The blue dots represent the variety values of the intercept γ_0 and the slope γ_1 parameters, respectively shown on the axis. These parameters are varied such that they minimize the cost function $\mathcal{C}(\boldsymbol{\gamma})$ on the y-axis using the LRM.

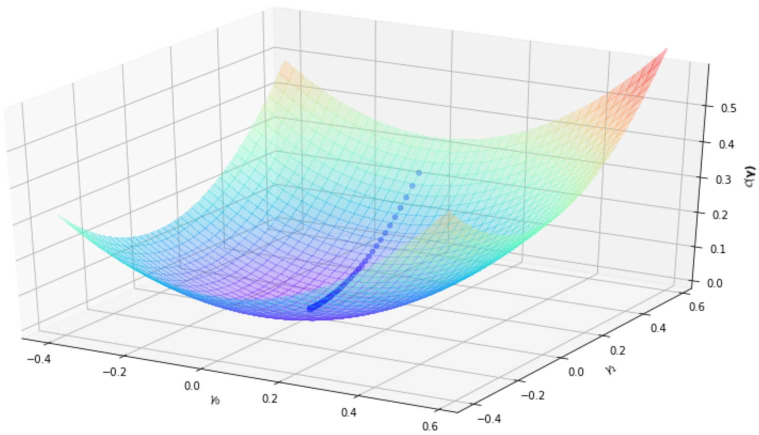


Figure 5.1: Plot of a Cost Function $\mathcal{C}(\gamma)$ for a simple binary LRM

The PD is then estimated through the use of LRM given in equation (5.2.5) as

$$\mathcal{PD}_i = \hat{\pi}_i = \frac{1}{1 + e^{(-\hat{\gamma}_0 - \hat{\gamma}_1 X_i)}}. \tag{5.3.1}$$

We use the accuracy rate $\mathcal{A}$ to assess the performance of the optimized parameters in the model given in equation (5.3.1), which is expressed as

$$\mathcal{A} = \frac{\sum_{i=1}^{n}[I(\hat{\pi}_i \geq 0.5) \equiv (Y_i = 1)] + \sum_{i=1}^{n}[I(\hat{\pi}_i < 0.5) \equiv (Y_i = 0)]}{n}, \tag{5.3.2}$$

where $I(.) = \{0; 1\}$ is the indicator function and n is the sample size.

In Table 5.1, the results $\mathcal{I}$ is the maximum number of iteration the method was set to, b is the random mini-batch size, κ is the number of iteration the method converged to and ε the tolerance level in brackets. For the gradient descent methods, the tolerance

Table 5.1: Parameter estimation method results for PD using Binary LRM on simulated dataset

Parameter estimator	$\mathcal{I}$	b	$\kappa(\varepsilon)$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\mathcal{A}$	$\mathcal{C}(\boldsymbol{\gamma})$
BGD	100	1	2(0.01)	0.0051	0.4914	0.8294	0.3881
SGD	100	1	10(0.01)	0.0075	0.4898	0.8294	0.3881
MBGD	100	1	10(0.01)	0.0099	0.4996	0.8292	0.2936
IRLS	100	1	7(1e-08)	0.0134	0.4930	0.8291	0.3881
EM	100	1	39 (1e-08)	0.0134	0.4930	0.8291	0.3881
NM	100	1	88 (1e-08)	0.0239	0.4977	0.8236	0.3853
PW	100	1	2 (1e-08)	0.0239	0.4977	0.8236	0.3853
CG	100	1	9 (1e-08)	0.0239	0.4977	0.8236	0.3853
TN	100	1	8 (1e-08)	0.0239	0.4977	0.8236	0.3853
BFGS	100	1	10 (1e-08)	0.0239	0.4977	0.8236	0.3853
LM-BFGS	100	1	9 (1e-08)	0.0239	0.4977	0.8236	0.3853

level is 0.01 and in all other methods it is smaller, i.e., 1e-08 = 0.00000008. $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are respectively the intercept and slope estimated parameters. $\mathcal{A}$ is the accuracy classification score given by equation (5.3.2). $\mathcal{C}(\gamma)$ is the cost function for LRM defined in equation (5.2.5). Note that for MBGD 40 sub-batches were generated from the sample size.

In Figure 5.2, the plots of the cost function values $\mathcal{C}(\gamma)$ against the number of iterations $\mathcal{I}$ per 100 iterations are shown, they are computed on the simulated dataset. The five parameter estimation methods are BGD, SGD, MBGD, IRLS and EM.

In Figure 5.3, the plots of the cost function values $\mathcal{C}(\gamma)$ against the number of iterations $\mathcal{I}$ per 100 iterations are shown, they are computed on the simulated dataset. The six parameter estimation methods are NM, PW, CG, TN, BFGS and LM-BFGS.

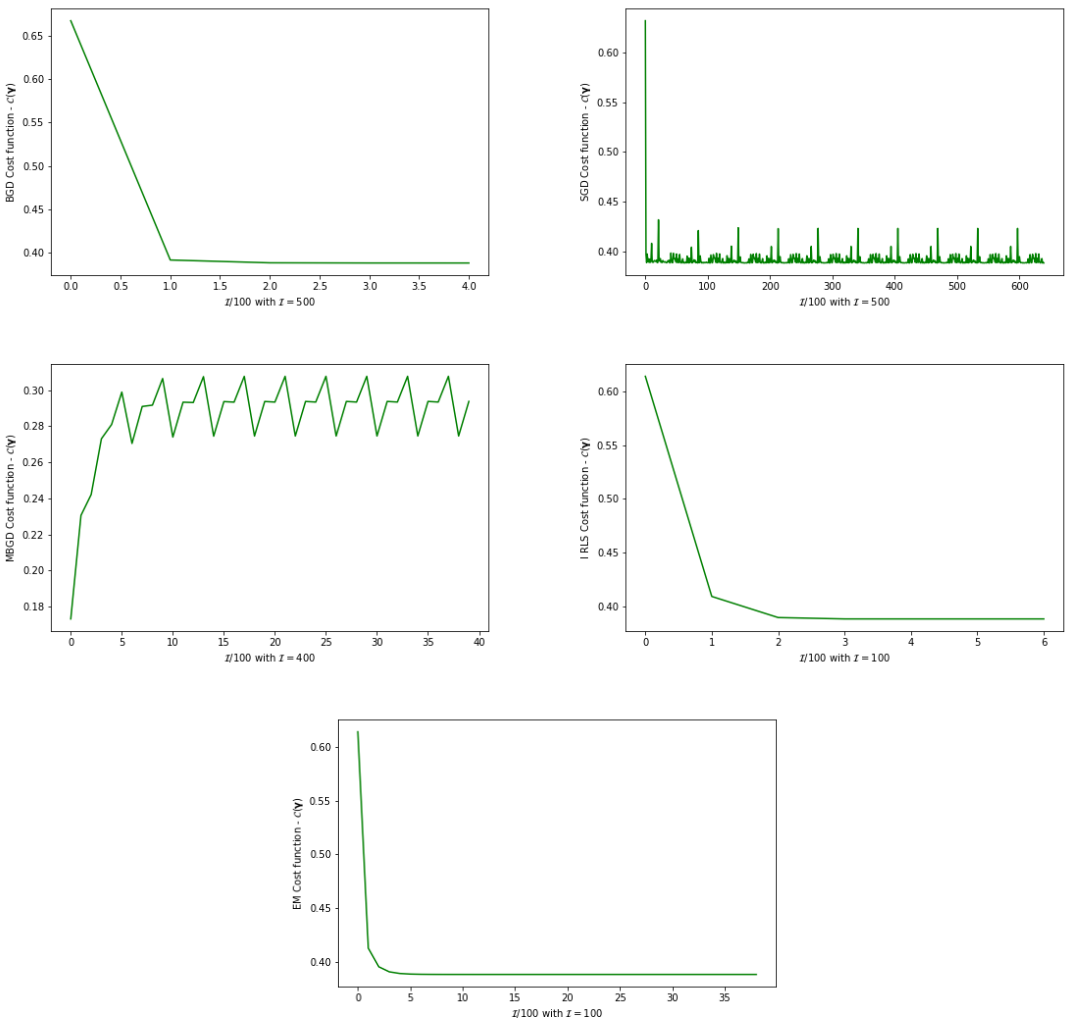


Figure 5.2: The plot of the cost function from five parameter estimators against number of iterations.

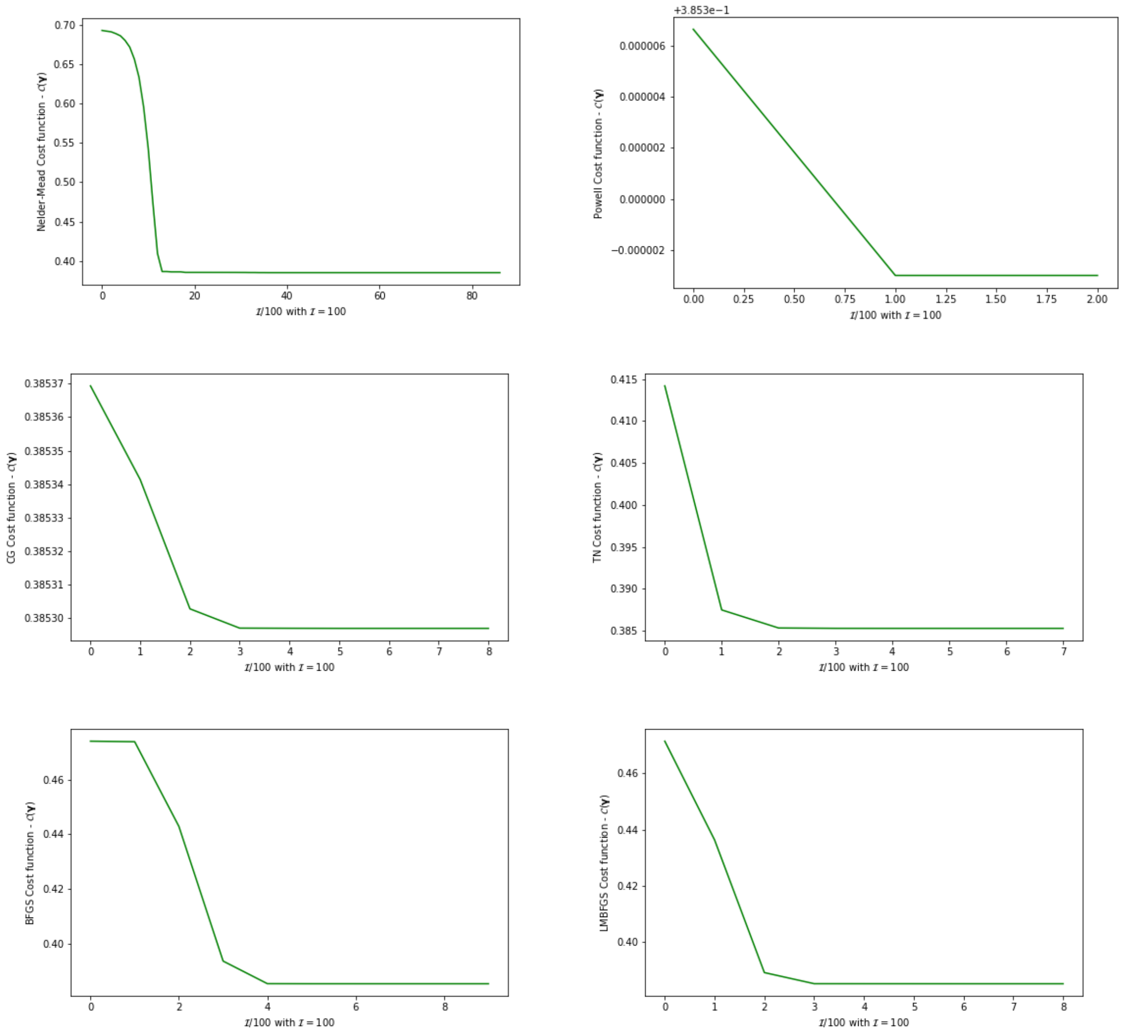


Figure 5.3: The plot of the cost function from six parameter estimators against number of iterations.

5.4 Applications to real dataset

This section is based on the application of the proposed methodology in section 5.2 to the benchmarking dataset. The anonymous dataset collected during the years 2016 to 2018 is from one of the South African financial institutions that provide loans to clients. The dataset contains the history of 1057 clients with the default indicator being the binary response variable (Y), i.e. default = 1 and non-default = 0. To empirically compare the simulation and the real data results, we only considered one predictor variable (X_i). This variable is the average percentage credit to disposable income of the clients recorded monthly over the given period.

Table 5.2: Parameter estimation method results for PD using Binary LRM on real dataset

Parameter estimator	$\mathcal{I}$	b	$\kappa(\varepsilon)$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\mathcal{A}$	$\mathcal{C}(\boldsymbol{\gamma})$
BGD	100	1	10(0.01)	0.1426	0.0883	0.6244	1.3172
SGD	100	1	10(0.01)	2.1909	0.1115	0.6244	2.3416
MBGD	100	1	10(0.01)	0.2211	0.1369	0.6244	0.2564
IRLS	100	1	4 (1e-08)	1.1281	-0.0179	0.6339	0.6559
EM	100	1	8 (1e-08)	1.1281	-0.0179	0.6339	0.6559
NM	100	1	85 (1e-08)	1.1281	-0.0179	0.6339	0.6559
PW	100	1	60 (1e-08)	1.1281	-0.0179	0.6339	0.6559
CG	100	1	14 (1e-08)	1.1281	-0.0179	0.6339	0.6559
TN	100	1	15 (1e-08)	1.1281	-0.0179	0.6339	0.6559
BFGS	100	1	8 (1e-08)	1.1281	-0.0179	0.6339	0.6559
LM-BFGS	100	1	17 (1e-08)	1.1281	-0.0178	0.6339	0.6559

The MBGD method shown in Table 5.2 used 12 sub-batches for computations of the output.

In Figure 5.4, the plots of the cost function values $\mathcal{C}(\gamma)$ against the number of iterations $\mathcal{I}$ per 100 iterations is shown, computed using the real dataset. The five parameter estimation methods are BGD, SGD, MBGD, IRLS and EM.

In Figure 5.5, the plots of the cost function values $\mathcal{C}(\gamma)$ against the number of iterations $\mathcal{I}$ per 100 iterations is shown, computed using the real dataset. The six parameter estimation methods are NM, PW, CG, TN, BFGS and LM-BFGS.

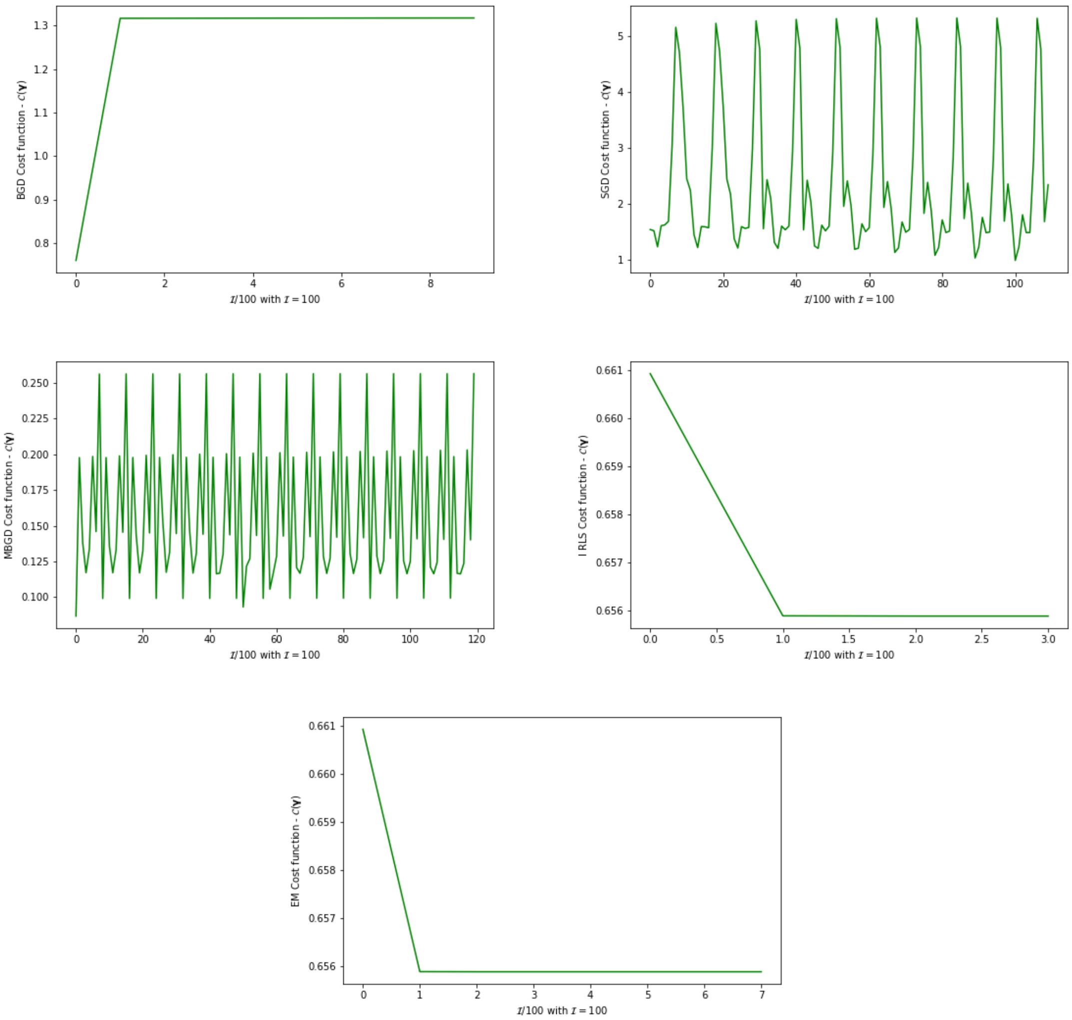


Figure 5.4: The plot of the cost function from five parameter estimators against number of iterations for real dataset.

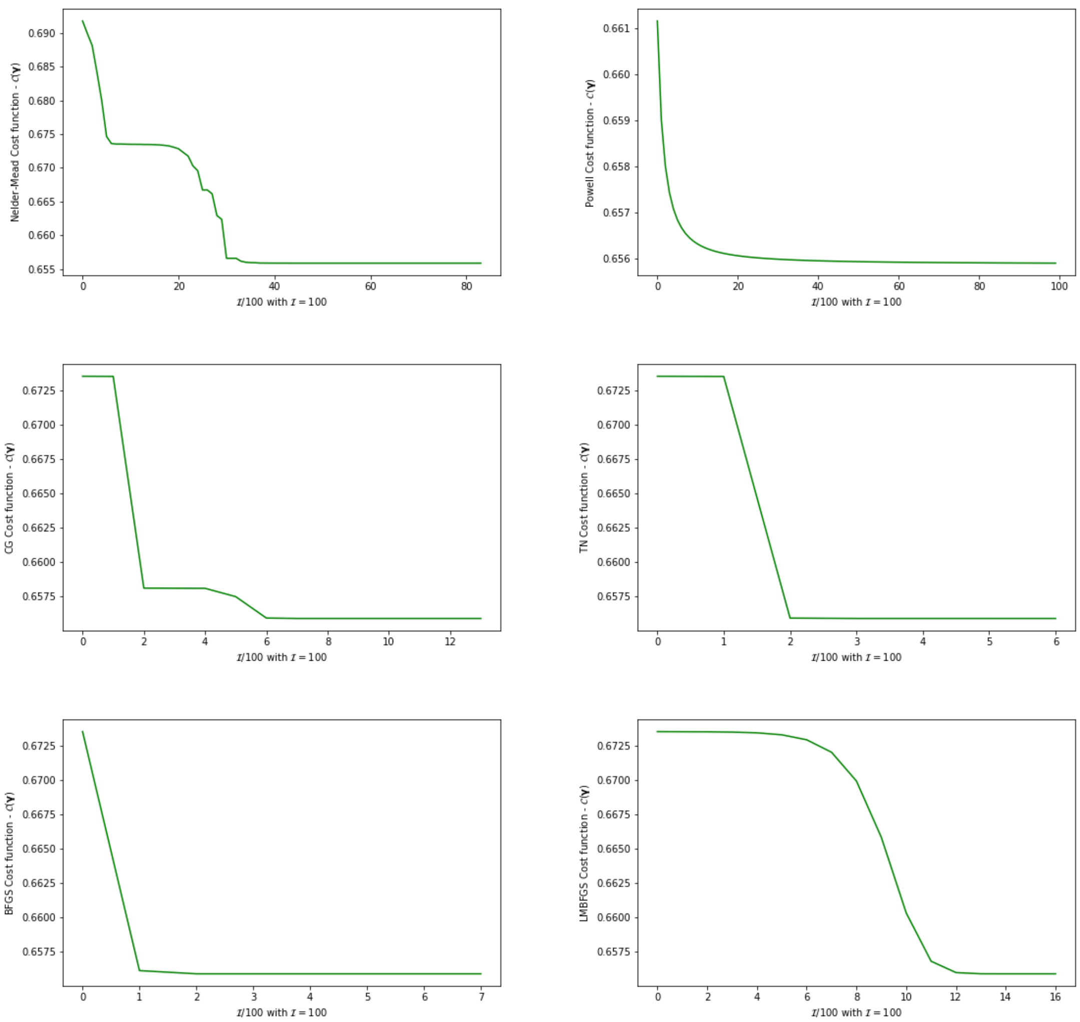


Figure 5.5: The plot of the cost function from six parameter estimators against number of iterations for real dataset.

5.5 Results discussion

Tables 5.1 and 5.2 present the results of the optimized parameters for the binary LRM that minimizes the cost function, computed using 11 different parameter estimation methods. The corresponding graphical representations about the convergence of the parameter estimations cost function are presented in Figures 5.2, 5.3, 5.4 and 5.5.

The BGD method described in Sub-Section 5.2.1 was configured to run 100 iterations for the simulated dataset but reveals convergence of the cost function $\mathcal{C}(\gamma) = 0.3881$ in only two iterations (i.e. $\kappa = 2$). For the real unbalanced dataset the method of BGD cost function value is very high. Figures 5.2 and 5.4 show the nature of convergence and its termination. The asymptotic rate of convergence can be inferior to alternative methods when there are many variables to work with. The method produced relatively reasonable cost function value with the learning rate of 0.01 while the parameter estimates are optimized, as illustrated by Figure 5.1. The SGD which is the simplified version of BGD method has a premature convergence, as it is revealed in the real dataset shown in figure 5.4 and the highest cost function of 2.3416. This is mainly due to the regular updates of the parameters through simulation (i.e., resampling without replacement for real dataset) and huge memory utilization for their storage, thus computationally expensive. Figure 5.2 for the SGD show that the method converges very fast but not with the exact optimum parameters, this is shown by the randomness of the cost function until all the configured iterations are executed.

IRLS and EM methods described in Sub-section 5.2.4 and Sub-section 5.2.5 reveals similar optimized parameters and minimized cost function values for the simulated and the real dataset. The only difference observed from the simulated dataset is on the convergence rate of the cost function, where IRLS method reaches convergence in only $\kappa = 7$ iterations and terminate due to the tolerance $\varepsilon = 1e-08$, which is similar for the real dataset. EM method reaches convergence in 39 iterations and terminates with a tolerance $\varepsilon = 1e-08$. We observe that the tolerance criterion used for the EM method to be the size of the change in the cost function or parameter estimates from one iteration to the next. This is a measure for lack of progress but not of the actual convergence. We view this as a major drawback of the EM method. MBGD method computed with simulated dataset show optimum parameters very close to the true parameters and the minimized cost function $\mathcal{C}(\gamma) = 0.2936$ among all the alternative optimization methods. MBGD method shows to have reduced the variance of the parameter updates, which leads to more stable convergence of the cost function value between 0.25 and 0.31. Similarly, Figure 5.4 - MBGD method for the real dataset the cost function obtained are between 0.1 and 0.25. According to Ruder (2016), MBGD does not always guarantee good convergence, but has few challenges that need to be taken into consideration when implemented. In this study MBGD shows very clear convergence and the challenge of the learning rate of 0.01 that we choose is efficient for the method. Also, the choice of the mini-batch size for the given sample size is important for the model. For the simulated dataset the mini-batch size is 40 and for the real dataset the mini-batch size is 12. The remainder of numerical optimization methods, i.e. NM, PW, CG, TN, BFGS and LM-BFGS, show the same optimized parameters and cost functions with slightly different convergence information presented in Figures 5.3 and 5.5. The measure for lack of progress but not of the actual convergence is again observed for NM method. The progress in convergence of PW method is revealed to be fast, just like that of BGD for the simulated dataset. The progress in convergence of IRLS method is revealed to be fast for the given real dataset.

5.6 Conclusion

In Section 5.2, the binary LRM is proposed as a default model to assess model risk with respect to parameter estimation risk, that is inappropriate parameter estimation method. The respective Sub-sections 5.2.1 to 5.2.11 describe recommended numerical optimization methods for estimating the model parameters through minimization of the binary LRM cost function. It is revealed that parameter risk is important and essential through the comparison of numerical experiments and simulation done in section 5.3. The MBGD method is shown to outperform the alternative optimization methods. MBGD estimators are accurate, since the bias is smaller among alternative methods, i.e.

$$E(\hat{\gamma}_1) - \gamma_1 = 0.4996 - 0.5 = -0.0004$$

Disregarding parameter risk can lead to a significant under-estimation of risk capital requirements, depending on the size of the underlying datasets. Therefore, we conclude that predicting PD using the binary LRM with the known varying thresholds will lead to substantially different results when parameter risk is taken into consideration. That is, when several optimization methods are employed. Numerical optimization estimation methods are identified as being the ones that have parameters which minimize the cost function or maximizes the log-likelihood function of the simple binary LRM. The impact of parameter estimation risk is depicted for an optimization method that yield the lowest cost function. Our experimental results support the need for further research of parameter estimation risk for binary LRM and other family of exponential models. Binary LRM with high order of predictor variables and interaction terms with different distributions may exhibit high parameter estimation risk implications. Therefore, it can be explored for further research of parameter estimation risk. Model risk management researchers and practitioners are therefore encouraged to consider parameter estimation risk through exploring different optimization methods as opposed to using the same traditional estimation methods repeatedly.

6. Conclusion and recommendations

The thesis suggested the valuation of initial margin for over-the-counter derivative (OTCD) and the quantification of model risk (MR) for models utilized mostly by financial institutions such as banks. In Chapter 2, global OTCDs were identified as a major cause of systemic risk to the financial system. Therefore, the valuation of initial margin (IM) calculation using the resampling method was suggested following the request by BCBS and IOSCO regulatory bodies that all standardised OTCDs be cleared through central counterparties (CCPs) and margins be posted by participants. The issues about determining IM for OTCDs remain a topic of interest. In Chapters 3, 4 and 5, MR quantification for credit risk models was discussed in detail. The MR assessment was approached in three forms: as distribution model risk (DMR), model misspecification and parameter estimation risk. The credit risk models of interest used are the Vasicek single factor model for determining the upper confidence levels of DMR, the binary logistic regression model (BLRM) and complementary log-log regression model (Cloglog) were utilised to quntify model misspecification. Lastly, BLRM was considered for determining the optimum parameter estimation method. Each of these topics in the three chapters came to some conclusions and also suggested recommendations for further research, which will be discussed on the following paragraphs.

In **Chapter 2** , the proposed parametric bootstrap method gave reasonable initial margin amounts under normal OTCDs market conditions, which will not leave the buyers or sellers of OTCD overexposed in the event of either counterparty defaulting. However, IM amounts may change and become highly inflated under stressed financial markets. The bootstrap method was applied to the real dataset of FTSE100 on swaps time series and conformed very well in relation to the simulated data. The posting of IM for OTCDs has been phased in since September 2016 for outstanding notional amounts greater than or equal to 3 trillion euros. The phase-in threshold of notional amounts as shown in Appendix A - Figure A.5, will continue until 1 September 2021 in order to ensure smooth and consistent implementation of IM across jurisdictions. Further research work to be considered will be to assess the impact and correlation of large markets against small markets with regards to posting of IM for OTCDs, perhaps using Monte-Carlo method to formulate the stress testing scenarios. The bootstrap methodology applied will enable counterparties to meet regulatory and risk management margin requirements under small OTCD outstanding notionals. However, it remains a critical question for small market participants that posting an initial margin for OTCDs does indeed reduce systemic risk and contributes significantly in making the entire financial markets safe.

The objectives of **Chapter 3** were to determine the upper confidence level for a risk measure called credit VaR in credit risk management and to investigate different model risk measures existing in credit risk management. These objectives were attained by using statistical methods such as the bootstrap confidence interval from Efron & Tibshirani (1986) and Efron & Tibshirani (1994). A newly proposed approach called modified hybrid percentile bootstrap (MHP) method was introduced. The MHP method was identified as a superior method for determining the credit VaR that can be used in capital requirement. Furthermore, MHP method was shown to be a good method in terms of distinguishing between the symmetric and asymmetric distribution, also for respectively under-estimating and over-estimating the credit VaR. Therefore, because of this performance it is critical to understand the statistical asymptotic properties of the MHP bootstrap method. This opens up further research studies.

Chapter 4 focused on two predictive techniques, the BLRM and Cloglog, for estimating the probability of default (PD). The assumptions and properties of the two predictive techniques were highlighted. Model misspecification was identified as the danger of working with a potentially *not well-suited model* to the given dataset and also as the observational indistinguishable predictive models that are having consequences. These consequences were reflected by the estimated values of the Cloglog technique under-estimating PDs while for BLRM the estimated values were consistent with high accuracy performances. However, our study was only limited to a predictor variable that comes from uniform and normal distributions, also excluding the presence of extreme and missing values. We also considered default events to be having the same importance as non-default events, i.e., balanced pair of observations in the dataset with a threshold of 0.5. Recommendation for future research is to increase the number of predictor variables while varying the threshold. Also, further research for different parameter estimation techniques is required, especially for symmetrical and asymmetrical models, that can capture the imbalance set of events and perform independent comparison for PD binary responses that carry deficiencies.

Finally, in **Chapter 5**, quantification of model risk was identified by determining the inappropriate parameter estimators. The impact of parameter estimation risk was depicted for a parameter estimation method that yielded the lowest cost function. The results support the need for possible further research of estimation parameter risk using family of exponential models and statistical learning techniques to estimate the risk measurements.

Model risk is dynamic in financial institutions and regulatory bodies that continue to seek improved models to better handle their big data and understand the trends for better decision making. Model risk will continue to grow as we move on to the 4^{th} , 5^{th}, and i^{th} industrial revolutions. The thesis only scratches the surface of the OTCDs initial margin and financial model risk. The future research work will quantify and analyse the global OTCDs markets and model risk using advanced mathematical, statistical and machine learning techniques.

Appendix A. Theory, tables and graphs from Chapter 2

A.1 Global OTCD trends

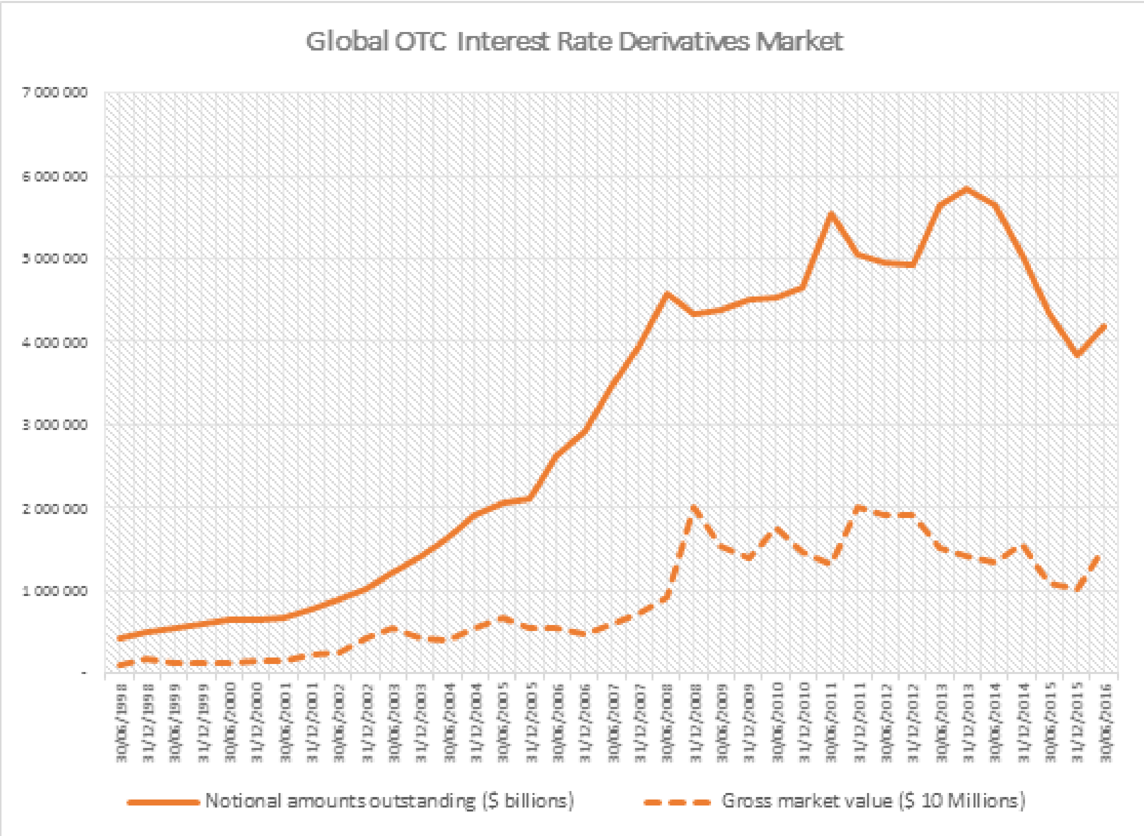


Figure A.1: Semi-annual global OTCD outstanding notional amount in US dollars.

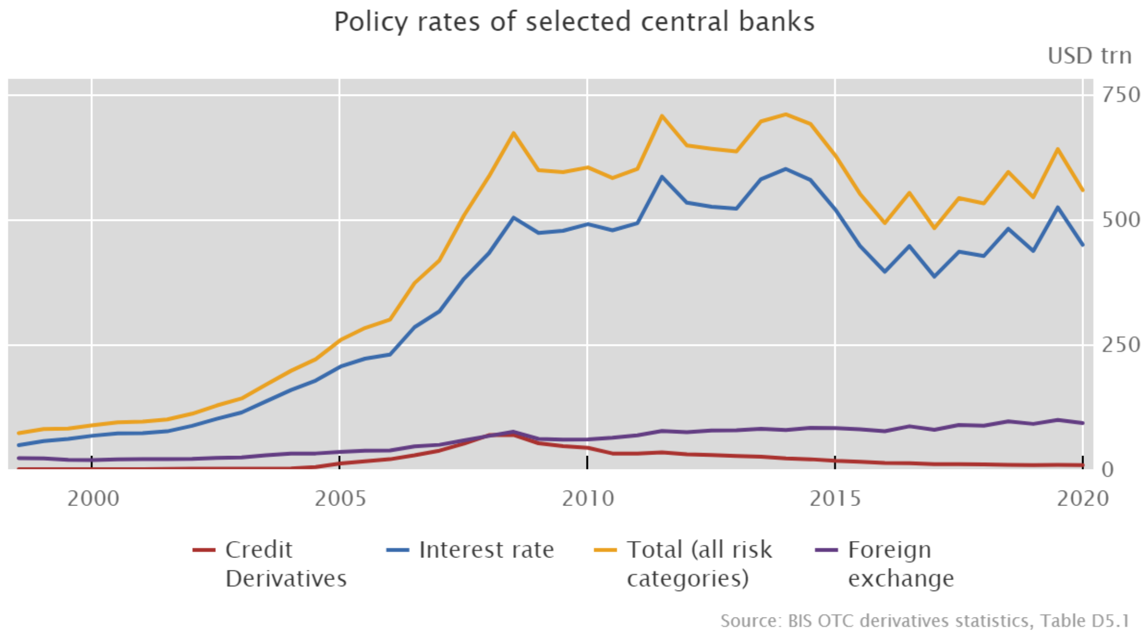


Figure A.2: OTCD outstanding notional amount in US dollars by risk category.

Source: BIS Statistics Explorer (https://www.bis.org/statistics/about_derivatives_stats.htm?m=6%7C32%7C639)

Exchange-traded futures and options, by currency

Notional principal, in billions of US dollars

	Open interest			Daily average turnover						
	Dec 2018	Sep 2019	Dec 2019	2018	2019	Aug 2019	Sep 2019	Oct 2019	Nov 2019	Dec 2019
Interest rate	94,454	109,459	96,152	8,775	8,984	10,482	9,604	8,102	6,686	5,900
Australian Dollar	1,352	1,569	1,572	142	180	148	196	139	144	157
Brazilian Real	979	1,447	2,342	54	71	76	66	81	96	75
Canadian Dollar	753	981	1,041	102	103	107	113	115	80	115
Swiss Franc	290	259	224	29	26	34	38	21	21	15
Renminbi	12	18	21	6	9	9	9	12	15	13
Danish Krone	0	0	0	0	0	0	0	0	0	0
EUR	15,023	12,845	13,257	1,467	1,206	1,178	1,306	964	858	979
Pound Sterling	7,406	11,352	9,111	717	642	674	829	693	542	504
Hong Kong Dollar	0	0	0	0	0	0	0	0	0	0
Forint	0	0	0	0	0	0	0	0	0	0
Indian Rupee	0	1	1	0	0	0	0	0	0	0
Yen	203	164	132	47	42	40	62	30	38	48
Won	38	37	34	14	16	16	22	18	17	16
Mexican Peso	0	0	0	0	0	0	0	0	0	0
Norwegian Krone	3	2	3	0	0	0	0	0	0	0
New Zealand Dollar	109	120	116	5	6	7	7	5	6	5
Zloty	0	0	0	0	0	0	0	0	0	0
Russian rouble	0	1	1	0	0	0	0	0	0	0
Swedish Krona	184	84	72	5	2	1	2	2	1	1
Singapore Dollar	0	0	0	0	0	0	0	0	0	0
New Turkish Lira	0	0	0	0	0	0	0	0	0	0
New Taiwan Dollar	0	0	0	0	0	0	0	0	0	0
US Dollar	68,093	80,571	68,216	6,186	6,680	8,190	6,951	6,021	4,869	3,970
Rand	7	7	9	0	0	0	0	1	0	0
Other currencies	-0	-0	-0	0	-0	-0	-0	-0	0	0
Foreign exchange	396	387	388	162	140	152	158	132	123	150
Australian Dollar	14	16	17	9	7	6	7	7	6	8
Brazilian Real	120	104	105	39	36	43	34	36	39	33
Canadian Dollar	18	15	19	7	6	6	7	7	5	7
Swiss Franc	10	11	10	4	4	4	5	4	3	5
Renminbi	7	9	8	3	5	6	5	4	5	4
Danish Krone	0	0	0	0	0	0	0	0	0	0
EUR	119	122	121	47	34	31	45	28	23	41
Pound Sterling	35	34	31	12	11	9	14	14	8	14
Hong Kong Dollar	0	0	0	0	0	0	0	0	0	0
Forint	1	1	1	0	0	0	0	0	0	0
Indian Rupee	6	12	13	12	12	15	14	10	12	12
Yen	42	36	36	18	16	20	19	14	14	18
Won	7	8	8	3	4	4	4	4	4	3
Mexican Peso	7	10	13	2	2	2	2	1	2	2
Norwegian Krone	0	0	1	0	0	0	0	0	0	0
New Zealand Dollar	2	5	3	2	2	2	2	1	2	2
Zloty	0	0	0	0	0	0	0	0	0	0
Russian rouble	5	6	7	2	2	2	2	2	2	2
Swedish Krona	0	0	0	0	0	0	0	0	0	0
Singapore Dollar	0	0	0	0	0	0	0	0	0	0
New Turkish Lira	2	2	1	0	0	0	0	0	0	0
New Taiwan Dollar	0	0	0	0	0	0	0	0	0	0
US Dollar	382	372	374	159	138	150	154	129	121	147
Rand	8	6	3	1	0	0	0	0	0	0
Other currencies	7	7	6	1	1	2	1	1	1	1

Figure A.3: Exchange-traded futures and options.

Source: BIS Statistics Explorer (https://stats.bis.org/statx/srs/table/d2?f=pdf)

Turnover of OTC foreign exchange instruments, by currency

"Net-net" basis, April 1989-2019 daily averages, in billions of US dollars

Currency	1989		1992		1995		1998		2001		2004		2007		2010		2013		2016		2019	
	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%	Amount	%
USD	485	90	668	82	981	83	1,325	87	1,114	90	1,702	88	2,845	86	3,371	85	4,662	87	4,437	88	5,824	88
EUR	...	...	...	...	...	...	...	...	470	38	724	37	1,231	37	1,551	39	1,790	33	1,590	31	2,129	32
JPY	151	28	190	23	291	25	332	22	292	24	403	21	573	17	754	19	1,235	23	1,096	22	1,108	17
GBP	78	15	112	14	110	9	168	11	162	13	319	16	494	15	512	13	633	12	649	13	844	13
AUD	13	2	20	2	31	3	46	3	54	4	116	6	220	7	301	8	463	9	349	7	447	7
CAD	8	1	27	3	40	3	54	4	56	4	81	4	143	4	210	5	244	5	260	5	332	5
CHF	52	10	69	8	85	7	108	7	74	6	117	6	227	7	250	6	276	5	243	5	327	5
CNY	...	...	...	...	...	...	0	0	0	0	2	0	15	0	34	1	120	2	202	4	285	4
HKD	5	1	8	1	13	1	15	1	28	2	34	2	90	3	94	2	77	1	88	2	233	4
NZD	...	...	2	0	3	0	3	0	7	1	21	1	63	2	63	2	105	2	104	2	137	2
SEK	6	1	9	1	7	1	5	0	31	2	42	2	90	3	87	2	94	2	112	2	134	2
KRW	...	...	...	...	...	...	2	0	10	1	22	1	38	1	60	2	64	1	84	2	132	2
SGD	2	0	2	0	5	0	17	1	13	1	18	1	39	1	56	1	75	1	91	2	119	2
NOK	1	0	2	0	3	0	4	0	18	1	27	1	70	2	52	1	77	1	85	2	119	2
MXN	...	...	...	...	...	...	7	0	10	1	21	1	44	1	50	1	135	3	97	2	114	2
INR	...	...	...	...	...	...	1	0	3	0	6	0	24	1	38	1	53	1	58	1	114	2
RUB	...	...	...	...	...	...	5	0	4	0	12	1	25	1	36	1	86	2	58	1	72	1
ZAR	...	...	2	0	4	0	6	0	12	1	14	1	30	1	29	1	60	1	49	1	72	1
TRY	...	...	...	...	...	...	...	...	0	0	2	0	6	0	29	1	71	1	73	1	71	1
BRL	...	...	...	...	...	...	3	0	6	0	5	0	13	0	27	1	59	1	51	1	71	1
TWD	...	...	...	...	...	...	2	0	3	0	8	0	12	0	19	0	24	0	32	1	60	1
DKK	3	0	4	0	6	1	4	0	15	1	17	1	28	1	23	1	42	1	42	1	42	1
PLN	...	...	...	...	...	...	1	0	6	0	7	0	25	1	32	1	38	1	35	1	41	1
THB	...	...	...	...	...	...	2	0	2	0	4	0	6	0	8	0	17	0	18	0	32	0
IDR	...	...	...	...	...	...	1	0	1	0	2	0	4	0	6	0	9	0	10	0	27	0
HUF	...	...	...	...	...	...	1	0	0	0	4	0	9	0	17	0	23	0	15	0	27	0
CZK	...	...	...	...	...	...	4	0	2	0	3	0	7	0	8	0	19	0	14	0	26	0
ILS	...	...	...	...	...	...	...	...	1	0	2	0	5	0	6	0	10	0	14	0	20	0
CLP	...	...	...	...	...	...	1	0	2	0	2	0	4	0	7	0	16	0	12	0	19	0
PHP	...	...	...	...	...	...	0	0	1	0	1	0	4	0	7	0	8	0	7	0	19	0
AED	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	14	0
COP	...	...	...	...	...	...	...	...	0	0	1	0	2	0	4	0	6	0	8	0	12	0
SAR	...	...	1	0	...	...	1	0	1	0	1	0	2	0	3	0	5	0	15	0	12	0
MYR	...	...	...	...	...	...	1	0	1	0	1	0	4	0	11	0	21	0	18	0	9	0
RON	...	...	...	...	...	...	...	...	...	...	...	...	2	0	3	0	7	0	5	0	6	0
PEN	...	...	...	...	...	...	...	...	0	0	0	0	1	0	1	0	3	0	4	0	5	0
ARS	...	...	...	...	...	...	2	0	...	...	1	0	1	0	2	0	1	0	1	0	4	0
BHD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
BGN	...	...	...	...	...	...	...	...	...	...	...	...	0	0	1	0	1	0	1	0	2	0
DEM	141	26	323	40	430	36	465	30	...	...	...	...	...	...	...	...	...	...	...	...	...	...
FRF	9	2	32	4	94	8	76	5	...	...	...	...	...	...	...	...	...	...	...	...	...	...
XEU	5	1	24	3	26	2	21	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ITL	4	1	10	1	14	1	16	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...
NLG	6	1	8	1	8	1	14	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ESP	3	0	6	1	9	1	9	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...
FIM	1	0	4	0	1	0	2	0	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BEF	3	0	3	0	7	1	9	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ATS	...	...	1	0	3	0	2	0	...	...	...	...	...	...	...	...	...	...	...	...	...	...
PTE	...	...	1	0	1	0	2	0	...	...	...	...	...	...	...	...	...	...	...	...	...	...
LUF	...	...	0	0	1	0	1	0	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Other	103	19	108	13	189	16	314	21	81	7	127	7	253	8	184	5	83	2	103	2	129	2
Total	539	200	817	200	1,182	200	1,527	200	1,239	200	1,934	200	3,324	200	3,973	200	5,357	200	5,066	200	6,595	200

Figure A.4: Global OTC derivatives market

Source: BIS Statistics Explorer (https://stats.bis.org/statx/srs/table/d11.3?f=pdf). Turnover of OTC foreign exchange instruments, by currency.

Summary of changes to the implementation of the margin requirements for non-centrally cleared derivatives		
	March 2015 framework	**July 2019 revisions**
	Initial margin	
Covered entities belonging to a group whose aggregate month-end average notional amount of non-centrally cleared derivatives exceeds:		
€3.0 trillion	1 September 2016 to 31 August 2017 (based on average notional amounts for March, April and May 2016)	1 September 2016 to 31 August 2017 (based on average notional amounts for March, April and May 2016)
€2.25 trillion	1 September 2017 to 31 August 2018 (based on average notional amounts for March, April and May 2017)	1 September 2017 to 31 August 2018 (based on average notional amounts for March, April and May 2017)
€1.5 trillion	1 September 2018 to 31 August 2019 (based on average notional amounts for March, April and May 2018)	1 September 2018 to 31 August 2019 (based on average notional amounts for March, April and May 2018)
€0.75 trillion	1 September 2019 to 31 August 2020 (based on average notional amounts for March, April and May 2019)	1 September 2019 to 31 August 2020 (based on average notional amounts for March, April and May 2019)
€50.0 billion	Not applicable	1 September 2020 to 31 August 2021 (based on average notional amounts for March, April and May 2020)
Covered entities belonging to a group whose aggregate month-end average notional amount of non-centrally cleared derivatives exceeds €8.0 billion	From 1 September 2020 onwards (based on average notional amounts for March, April and May that year)	From 1 September 2021 onwards (based on average notional amounts for March, April and May that year)

Figure A.5: Summary of changes to the implementation of the margin requirements for OTCDs

The source for further infomation: https://www.bis.org/bcbs/publ/d475.pdf

A.2 Basic statistics for Value-at-Risk

Our general fundamental assumption for the Value at Risk (VaR_α) is that the asset returns (R) are normally distributed. Thus, denote the realised h- day returns for the asset price (S_t) as

$$R_{t,h} = \ln\left(\frac{S_t}{S_{t-h}}\right).$$

We assume that

$$R_{t,h} \sim N(\mu_t, \sigma_t^2).$$

The 100α% h-day Value-at-Risk is the value called $VaR_{\alpha,h}$ such that

$$VaR_{\alpha,h} = G^{-1}(\alpha) = inf\{r|G_R(r) \geq \alpha\},$$

where $G_R(r)$ is the Gaussian probability distribution function for the random variable r and the significance level $\alpha \in [0,1]$ is the probability level for the distribution of the returns (also referred to as lower percentile).

Therefore, from the statistical Gaussian probability distribution theory, the VaR can be written in terms of the density function $g_R(r)$, given as follows

$$VaR_\alpha = \int_{\inf}^{\alpha} g_R(r)dr,$$

where

$$g_R(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(r-\mu)^2}{2\sigma^2}, r \in \Re,$$

μ is the mean of the returns and σ is standard deviation of the asset return.

We can further represent the VaR as a probability

$$P(R_{t,h} \leq VaR_{\alpha,h}) = \alpha.$$

Employing the standard normal transformation to the probability, give us the following expression

$$P\left(\frac{R_{t,h} - \mu_t}{\sigma_t} \leq \frac{VaR_{\alpha,h} - \mu_t}{\sigma_t}\right) = \alpha.$$

Note: the left hand term follow the standard normal probability distribution

$$\frac{R_{t,h} - \mu_t}{\sigma_t} \sim N(0,1),$$

on which we will refer to as the standard normal variate z_t. Then the probability can be expressed in terms of the variates

$$P(z_t \leq z_\alpha) = \alpha.$$

Because the z_t distribution is symmetrical, it imply that

$$P(z_t \geq z_\alpha) = 1 - \alpha.$$

Therefore, z_α is the 100α percentile of the standard normal density, given in terms of the expression

$$z_\alpha = \frac{VaR_{\alpha,h} - \mu_t}{\sigma_t}.$$

Rearranging the expression, we have the following

$$VaR_{\alpha,h} = \mu_t + z_\alpha \sigma_t.$$

The VaR for h-day holding period can be estimated as:

$$\widehat{VaR}_{\alpha,h} = \bar{R}_t + \Phi^{-1}(\alpha) s_t.$$

where

- $\bar{R}_t$ is the sample mean of the asset returns estimating μ_t,
- $s_t = \sqrt{\frac{\sum_{i=0}^{n}(R_i - \bar{R})^2}{n-1}} \times \sqrt{h}$ is the volatility or standard deviation of the asset returns estimating σ_t, and
- $\Phi^{-1}(\alpha) = z_\alpha$ is the α-quantile of the standard normal distribution.

A.3 Basic VaR illustrations

The considered log returns are the effective daily returns with monthly continuous compounding interest. The VaR in this case is the amount of the financial position (OTCD) times the log returns.

The Bootstrap Monte Carlo Simulation (BMCS) is the generation of the returns (frequently generated from imitating asset price paths) by use of random numbers, drawn with replacement from an assumed Gaussian distribution. As such the VaR is calculated as a lower percentile of the standard normal probability distribution of the returns given. As an example, suppose we simulate 1250 different sample points of the returns from the OTCD price as the first bootstrap sample, then the 1 day 95% and 99.5% VaR's respectively are the values of the return for the 62.5th and 6.25th worst outcomes from an array of returns.

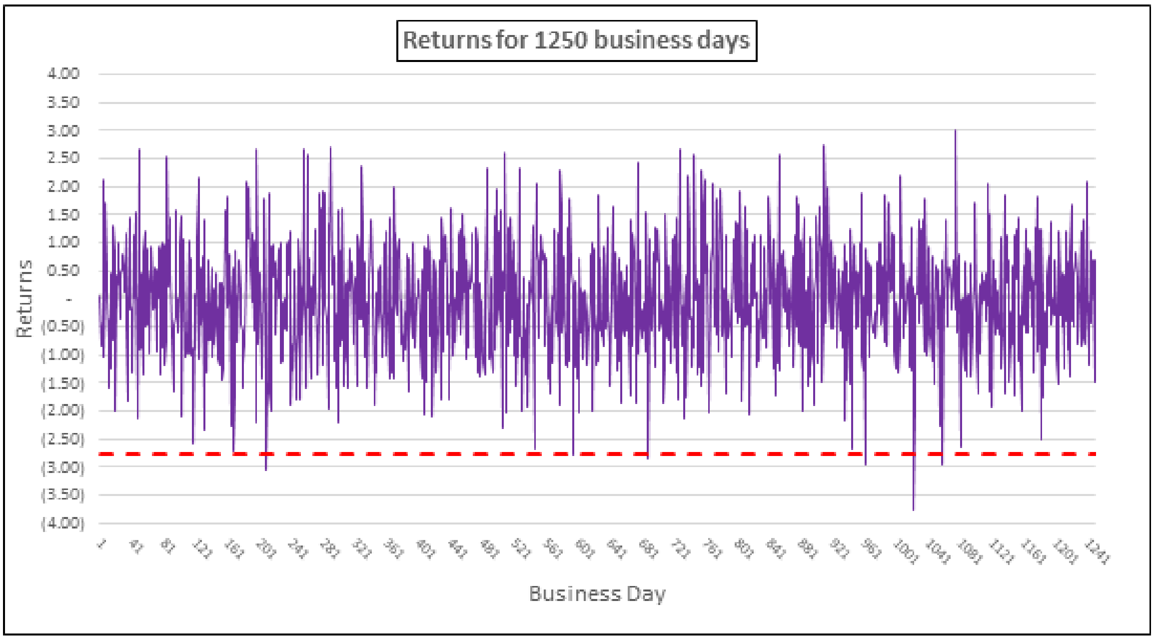


Figure A.6: Standard Gaussian daily returns for history of 3 years.

Figure A.6, show the red dashed line to be 6.25th worst outcome given by 2.76 (i.e. VaR). The purple line show the returns from the assumed distribution which imitate the returns from an OTCD. For the BMCS, we select sample sizes $n = 250, 750, 1250, 1750, 2250, 3250$ as a representation of daily OTCD returns generated using the random number which are drawn from the standard Gaussian distribution with the mean zero and variance of one. The re-sample generated repeatedly has no restrictions concerning the distribution function. If the number of bootstrap replications and the number of Monte-Carlo replications are high enough (say, $B \geq 200$, $MC \geq 1000$), then the method provides acceptable results as an approximation method (Müller et al. 2017).

Figure A.7 is the 99.5% VaR's from the 500 bootstrap samples. Note that the figure show that the 95% VaR (Yellow bar), 99.5% VaR (Orange bar) and 99.9% VaR (Blue bar) are given by 2.76, 2.85 and 2.90 respectively. Thus, we chose the worst replicate

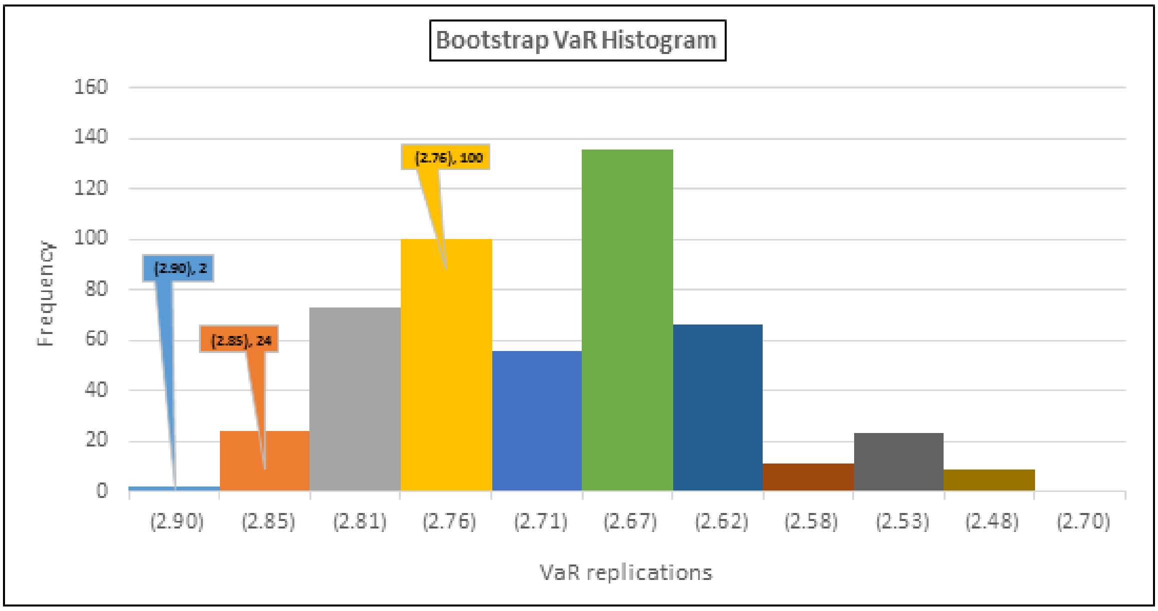


Figure A.7: Histogram of the 99.5% VaR from 500 bootstrap samples.

(at 99.9% VaR) to come up with the Bootstrap Initial Margin(BIM) to be used for determining the amount of collateral required, i.e., the Bootstrap Initial Margin, to open an OTCD position for counterparties.

A variance swap is a contractual agreement between two parties to exchange cash flows based on the observed variance or volatility of the equity index. The fixed leg of the variance swap will be a pre-determined volatility level agreed upon at the beginning, sometime determined by the trader on OTC market. The floating leg of the variance swap is the observed variance or volatility of the equity index. The contract is arranged such that its value at commencement is zero, making the value of the fixed-leg equal to the value of the floating leg.

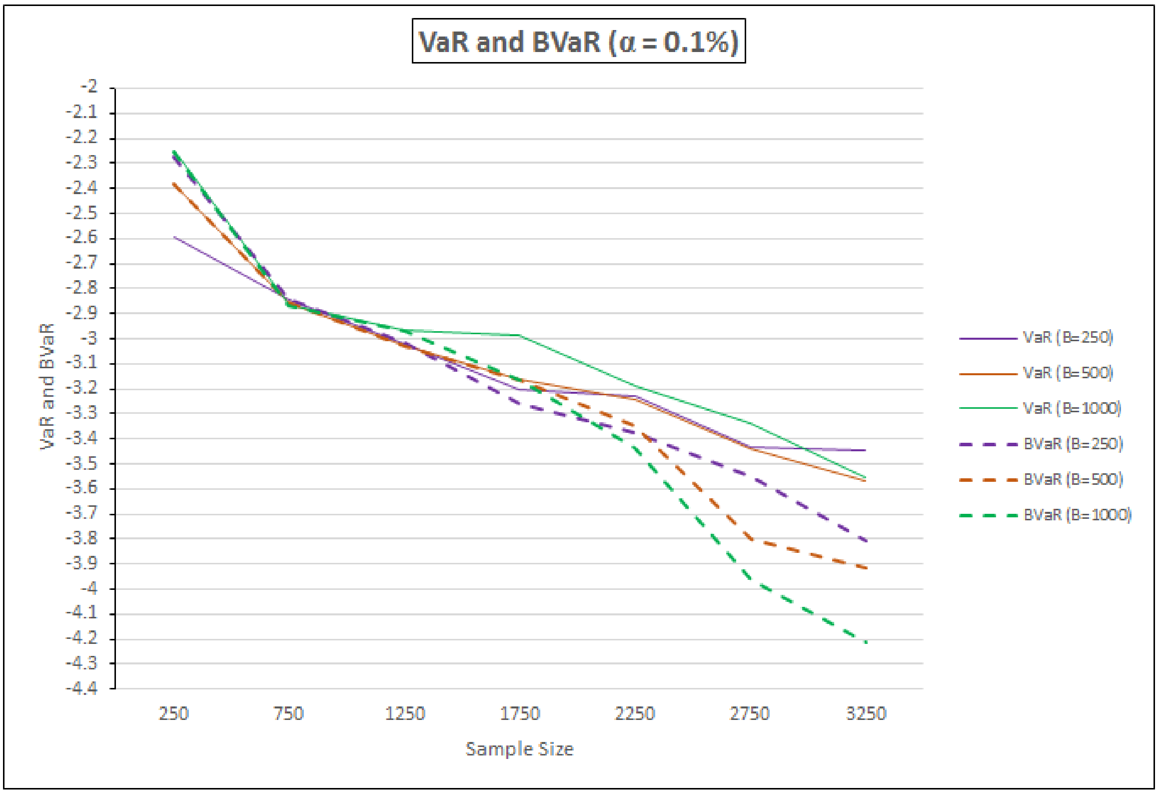


Figure A.8: VaR and BVaR at 0.1% significance level.

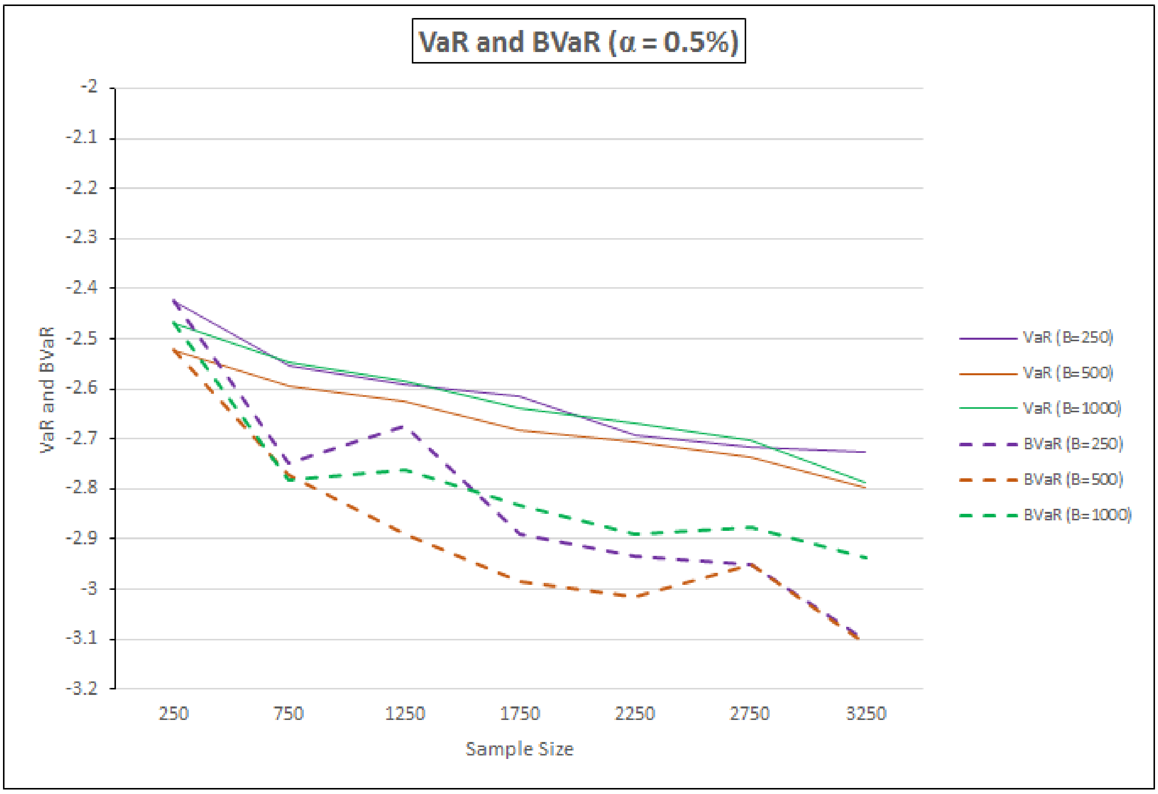


Figure A.9: VaR and BVaR at 0.5% significance level.

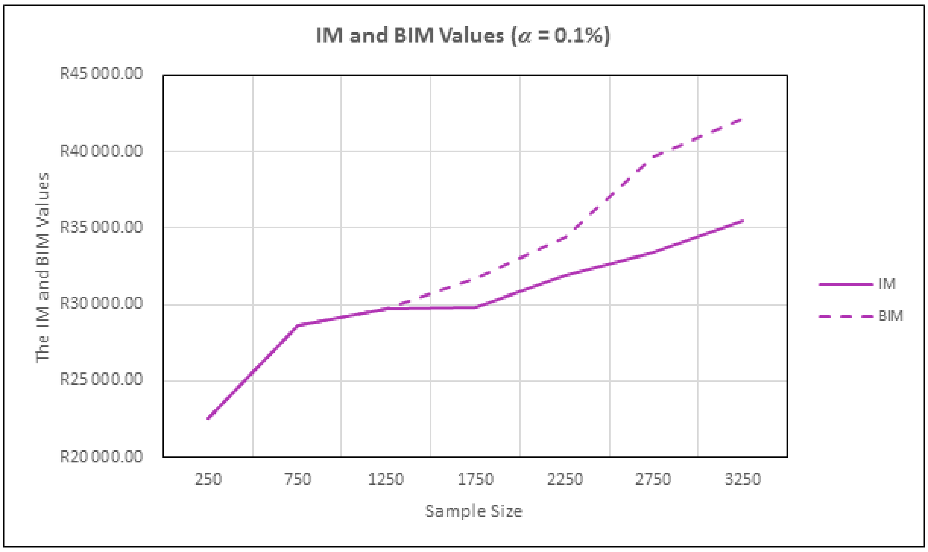


Figure A.10: IM and BIM at 0.1% significance level.

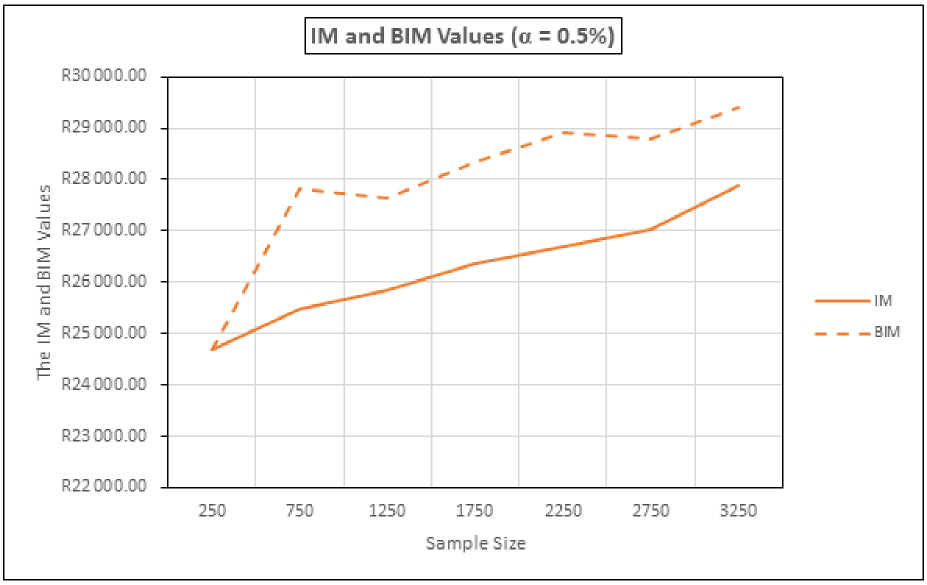


Figure A.11: IM and BIM at 0.5% significance level.

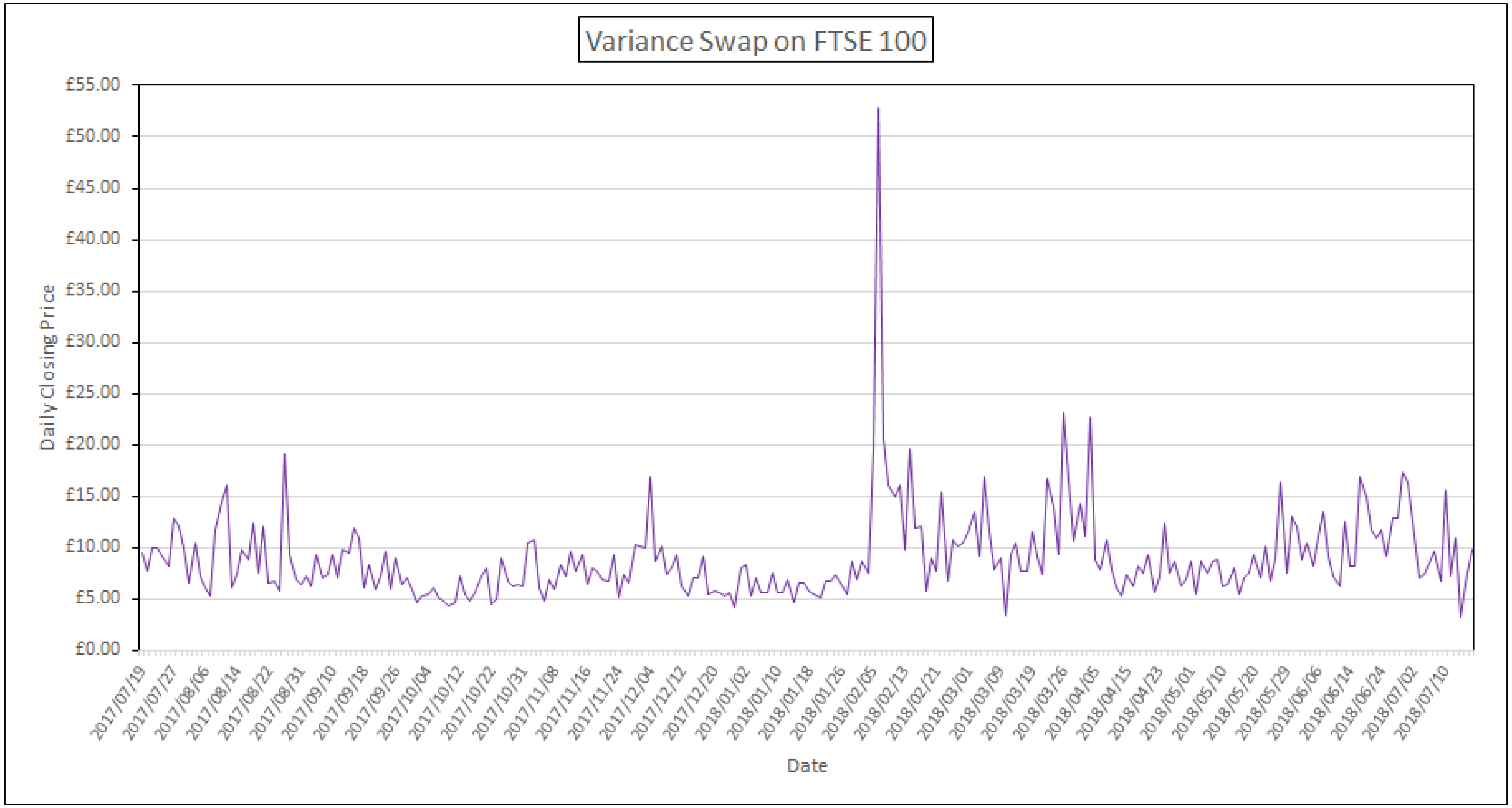


Figure A.12: Closing Price of the Variance Swap on FTSE100.

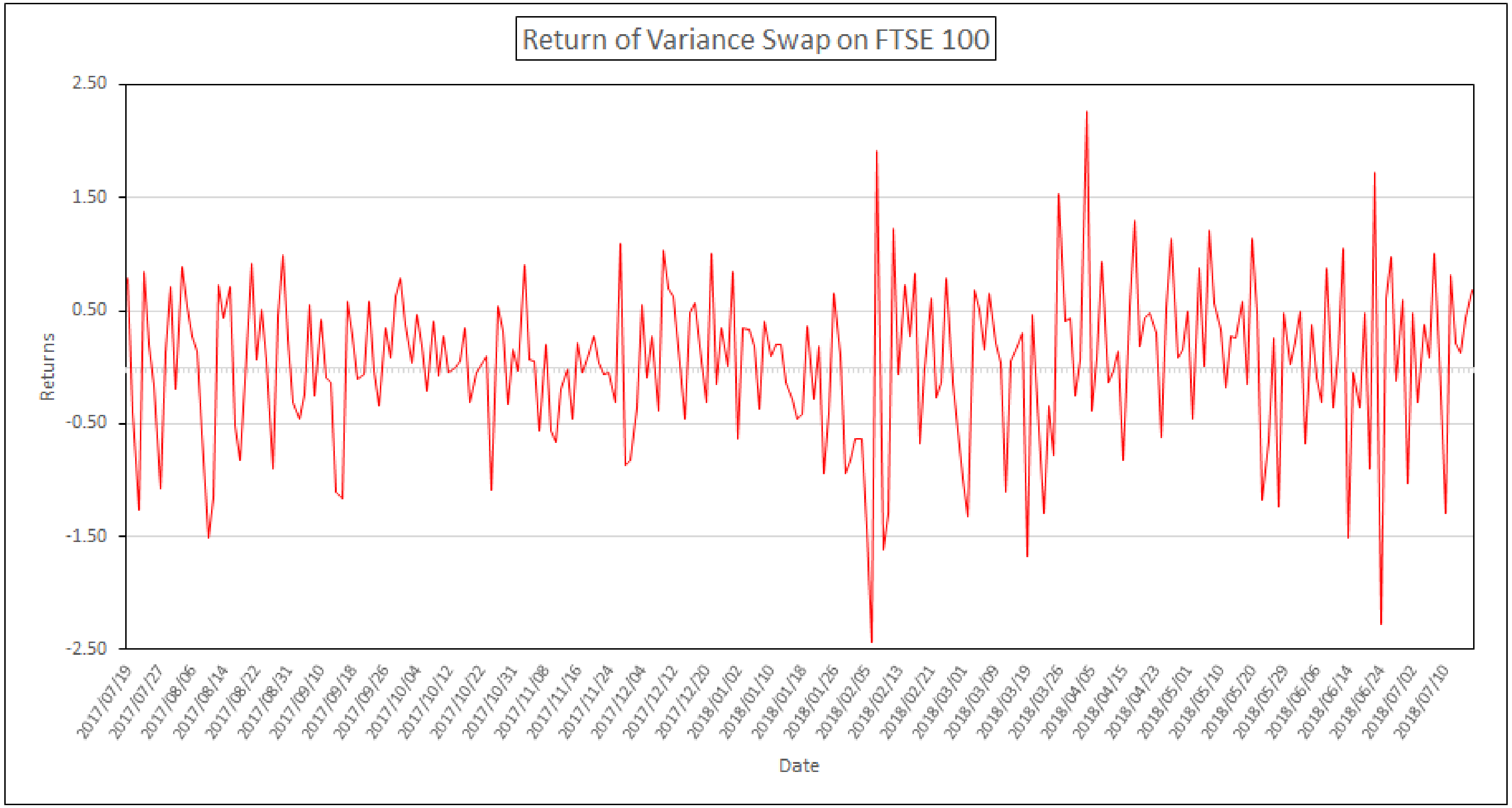


Figure A.13: Returns of the Variance Swap on FTSE100.

Appendix B. Statistical distributions and graphs from Chapter 3

B.1 Normal Distribution

If the normal distribution $N(\mu, \sigma^2)$ is given by r, then the Value at risk (VaR) is denoted by :

$$VaR_\alpha = -\mu - \Phi^{-1}(\alpha)\sigma$$

where Φ^{-1} is the inverse cdf of the standard normal distribution. The corresponding density function is the well-known given as

$$\phi(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{r-\mu}{\sigma}\right)^2\right],$$

where r represent the returns on the portfolio or an asset, μ is the mean and σ is the standard diviation of r. Both the parameters can be estimated using the mean of n sample of returns and sample standard deviation which are respectively denoted by $\bar{r}$ and S_r.

B.2 Log-Normal Distribution

Suppose the $X = \ln(L)$ then $X \sim \log N(\mu, \sigma^2)$. The pdf for X is given as

$$\phi(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right].$$

The pdf for $L = \exp(X)$ is given as

$$\theta(L) = \frac{1}{\sqrt{2\pi\sigma^2}L} \exp\left[-\frac{1}{2}\left(\frac{\ln(L)-\mu}{\sigma}\right)^2\right].$$

Therefore, the n^{th} moment generation function of L can be determined through

$$M_n(L) = \int_0^\infty \exp nL\theta(L)dL = \exp(n\mu + \left(\frac{n\sigma}{\sqrt{2}}\right)^2).$$

B.3 Student t-distribution

Suppose the profit-and-loss (i.e., PnL) or the returns are denoted by R, such that $R = \mu + \sigma t$, where t has a standardised Student-t distribution. Thus, VaR is expressed as

$$VaR_\alpha = -\mu - \sqrt{\frac{\nu-2}{\nu}} t_\nu^{-1}(\alpha)\sigma$$

where $t_\nu^{-1}(\alpha)$ is the inverse cdf of a Student-t distribution with ν degrees of freedom and α significance level, which has density function given as

$$f_\nu(r) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{r^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

B.4 Assumed graphical representation of parametric distributions

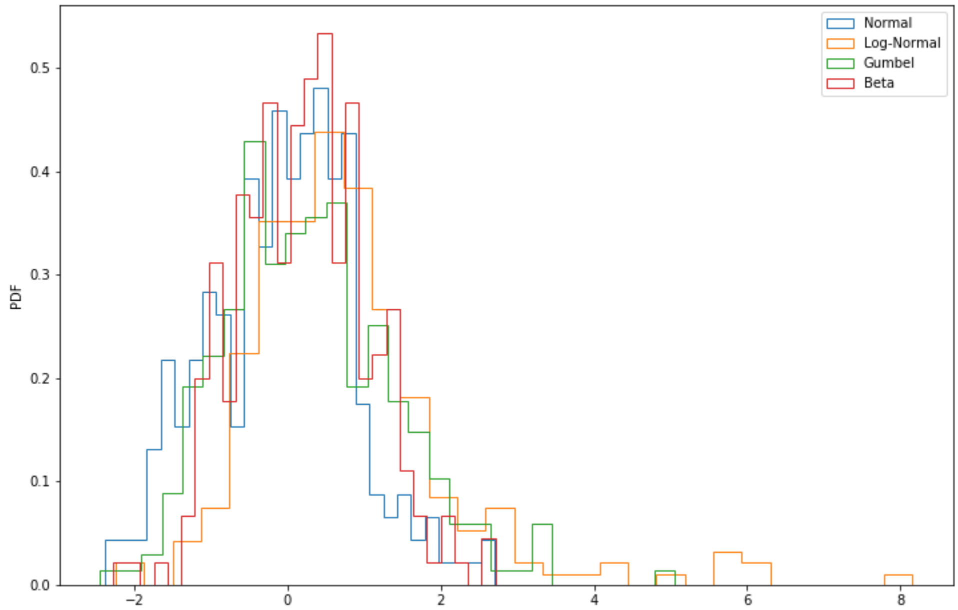


Figure B.1: Distributions of the standardized asset log-returns X_{ik}

In Figure B.1, X_{i1}, X_{i4}, X_{i6} and X_{i8} respectively include systemic common factors from Standard Normal, Log-Normal, Gumbel and Beta distributions.

In Figure B.2, X_{i2} and X_{i3} respectively include systemic common factors from Standard Student-t and Cauchy distributions.

In Figure B.3, X_{i5}, X_{i7}, X_{i9} and X_{i10} respectively include systemic common factors from Exponential, Gamma, Weibull and Exponweib distributions.

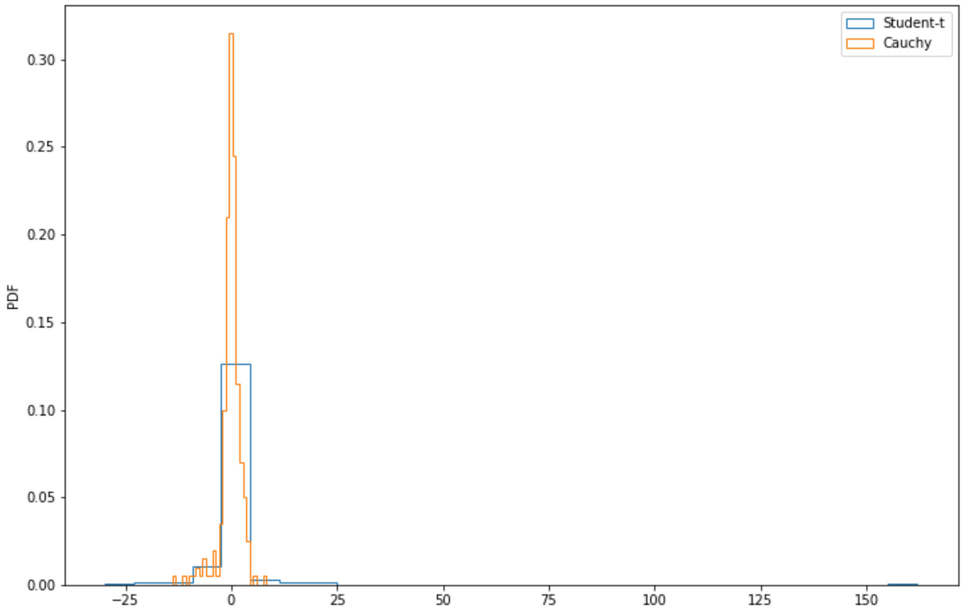


Figure B.2: Distributions of the standardized asset log-returns X_{ik}

In Figure B.4, X_{iS} include systemic common factor from the sum of all the assumed distributions.

In Figure B.5, X_{iP} include systemic common factor from the product of all the assumed distributions.

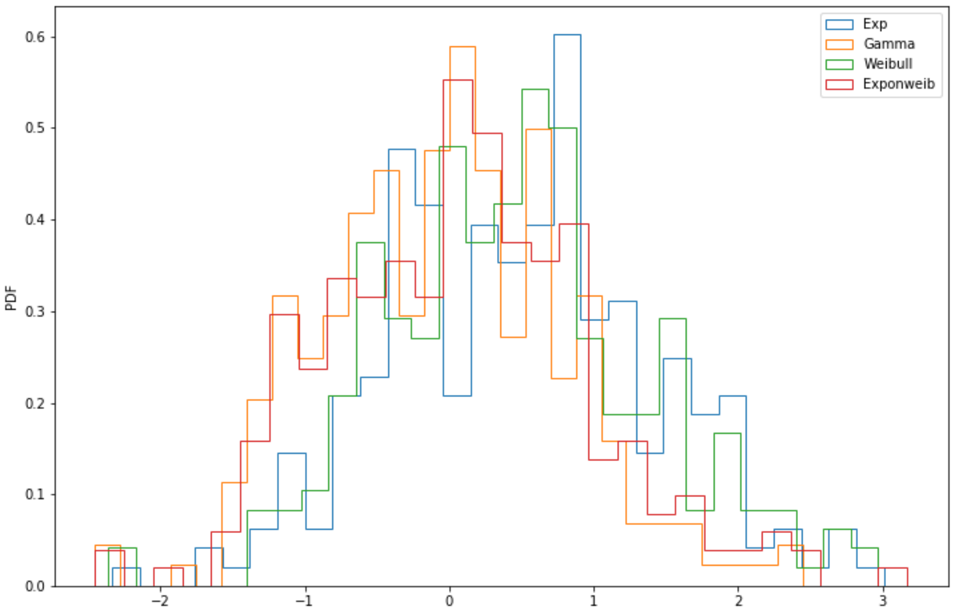


Figure B.3: Distributions of the standardized asset log-returns X_{ik}

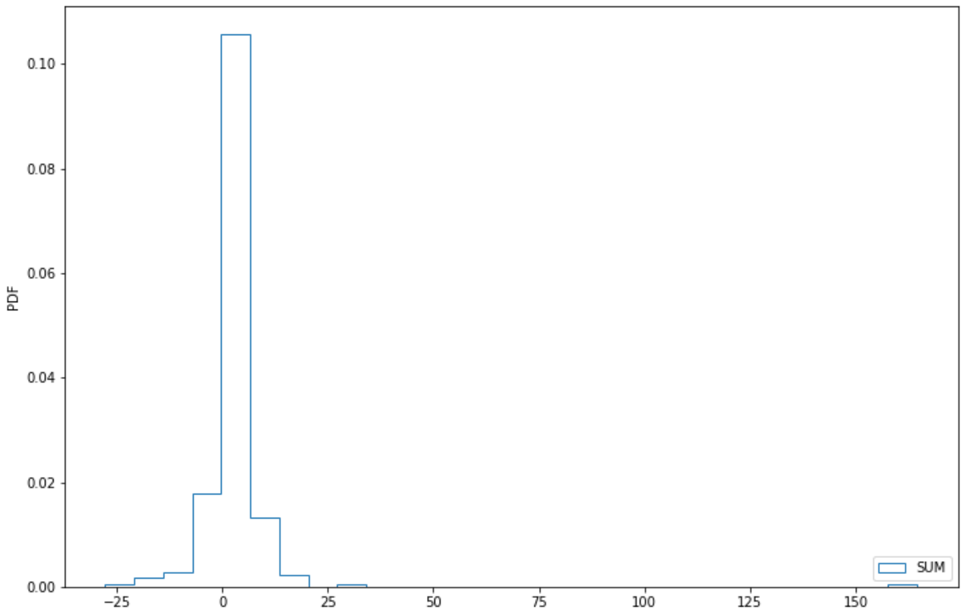


Figure B.4: Sum of the standardized asset log-returns distributions X_{ik}

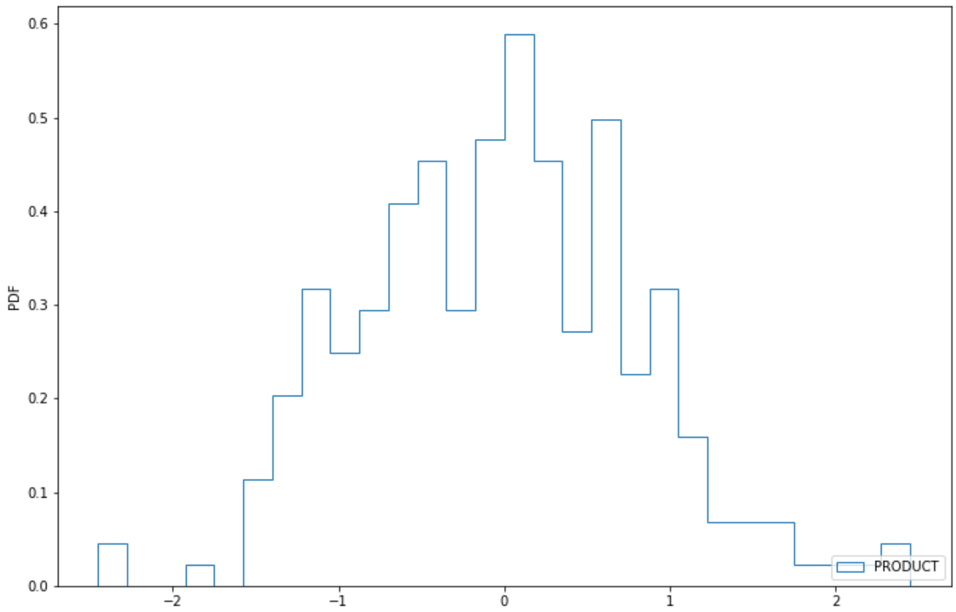


Figure B.5: Product of the standardized asset log-returns distributions X_{ik}

B.5 Bootstrap VaR and Normal VaR from the bootstrap MC samples

In Figures B.6 and B.7, X_{is} include systemic common factor from the sum of all the assumed distributions. The bootstrap replicates at 99.00%, 99.90% and 99.99% confidence levels are respectively shown by the blue shaded highlights. The blue lines are the lower and upper bound of the bootstrap replicates confidence intervals. The green dashed line is the bootstrap credit-VaR upper confidence levels. The red dashed line is the average of the replicated for bootstrap credit-VaR upper confidence levels. And orange dashed line is the original RV's credit-VaR. In Figures B.8 and B.9, X_{ip} include systemic common factor from the product of all the assumed distributions. The bootstrap replicates at 99.00%, 99.90% and 99.99% confidence levels are respectively shown by the blue shaded highlights. The blue lines are the lower and upper bound of the bootstrap replicates confidence intervals. The green dashed line is the bootstrap credit-VaR upper confidence levels. The red dashed line is the average of the replicated for bootstrap credit-VaR upper confidence levels. And orange dashed line is the original RV's credit-VaR.

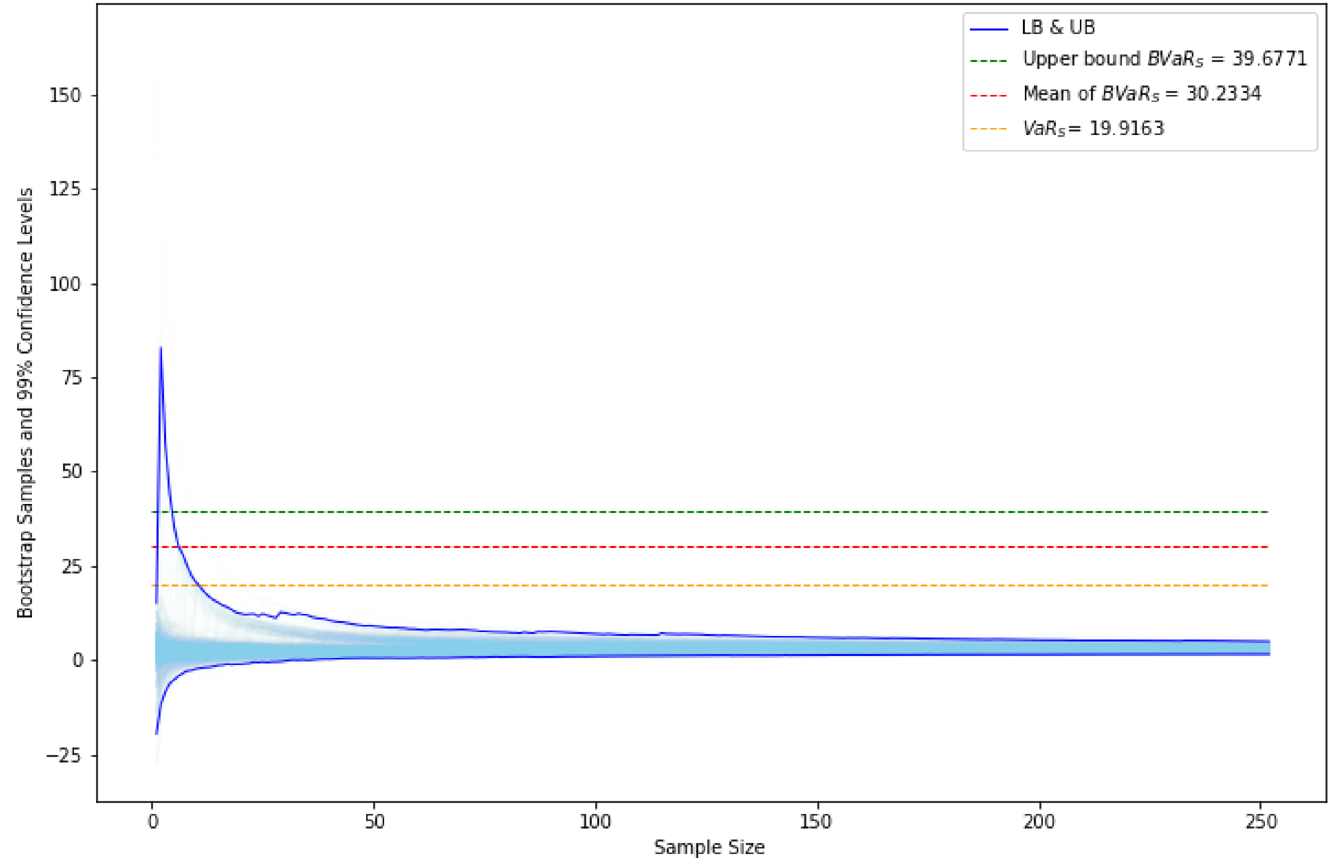
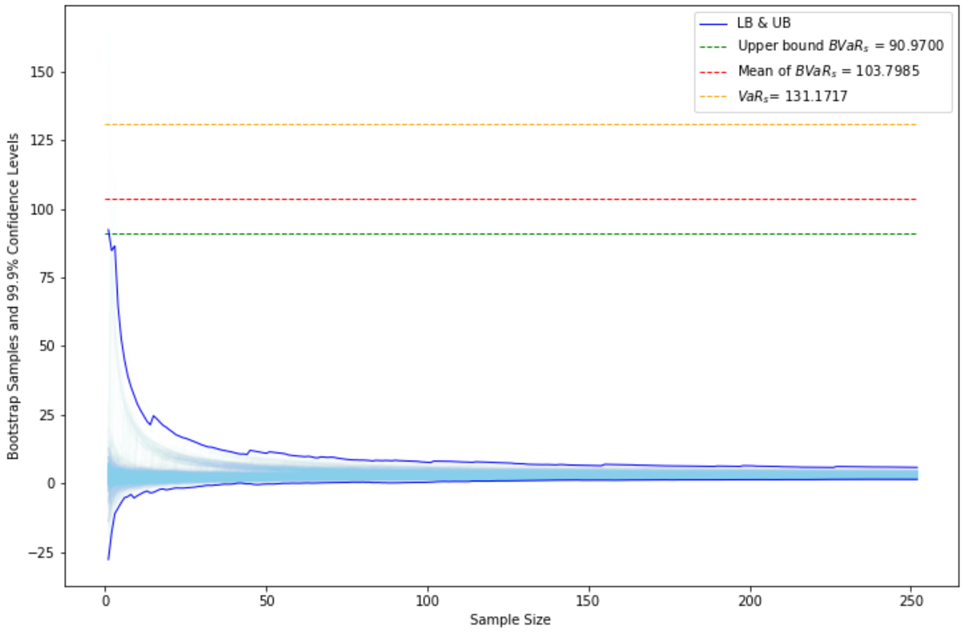


Figure B.6: Bootstrap MC of X_{is} for credit-VaR.



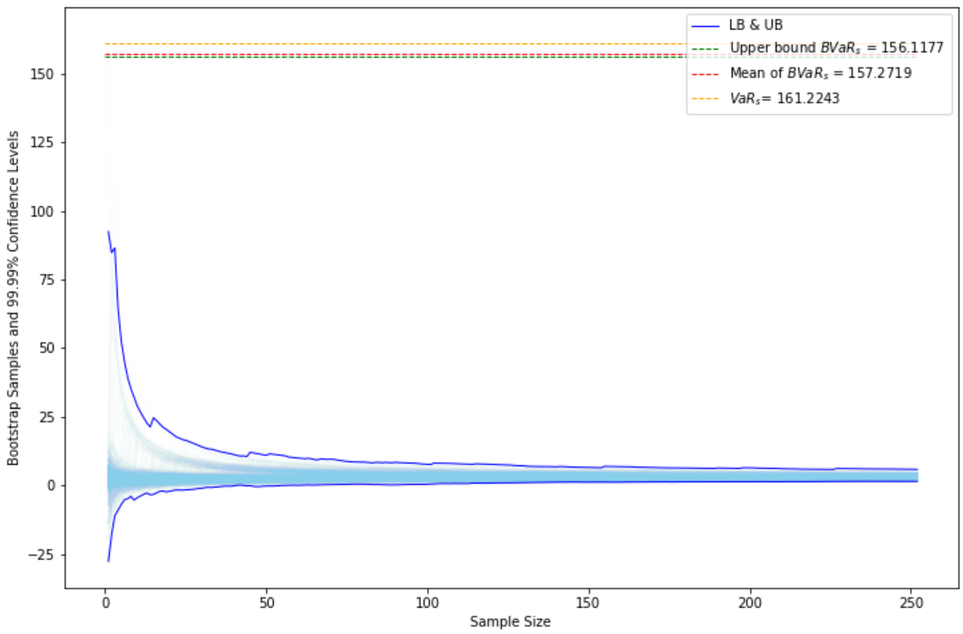


Figure B.7: Bootstrap MC of X_{is} for credit-VaR.

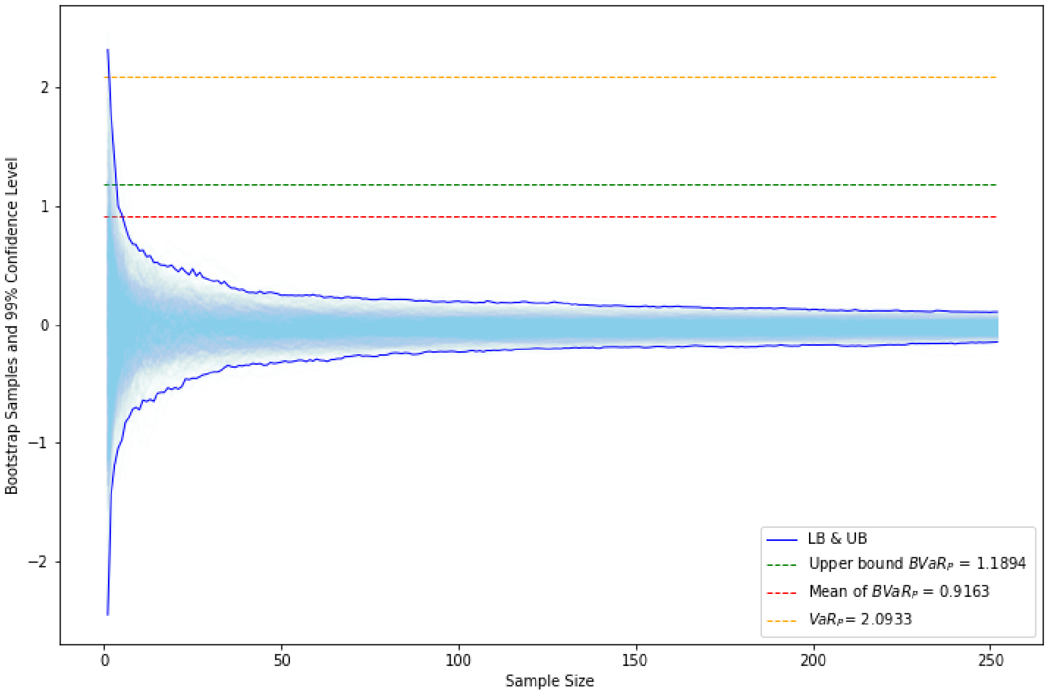


Figure B.8: Bootstrap MC of X_{ip} for credit-VaR.

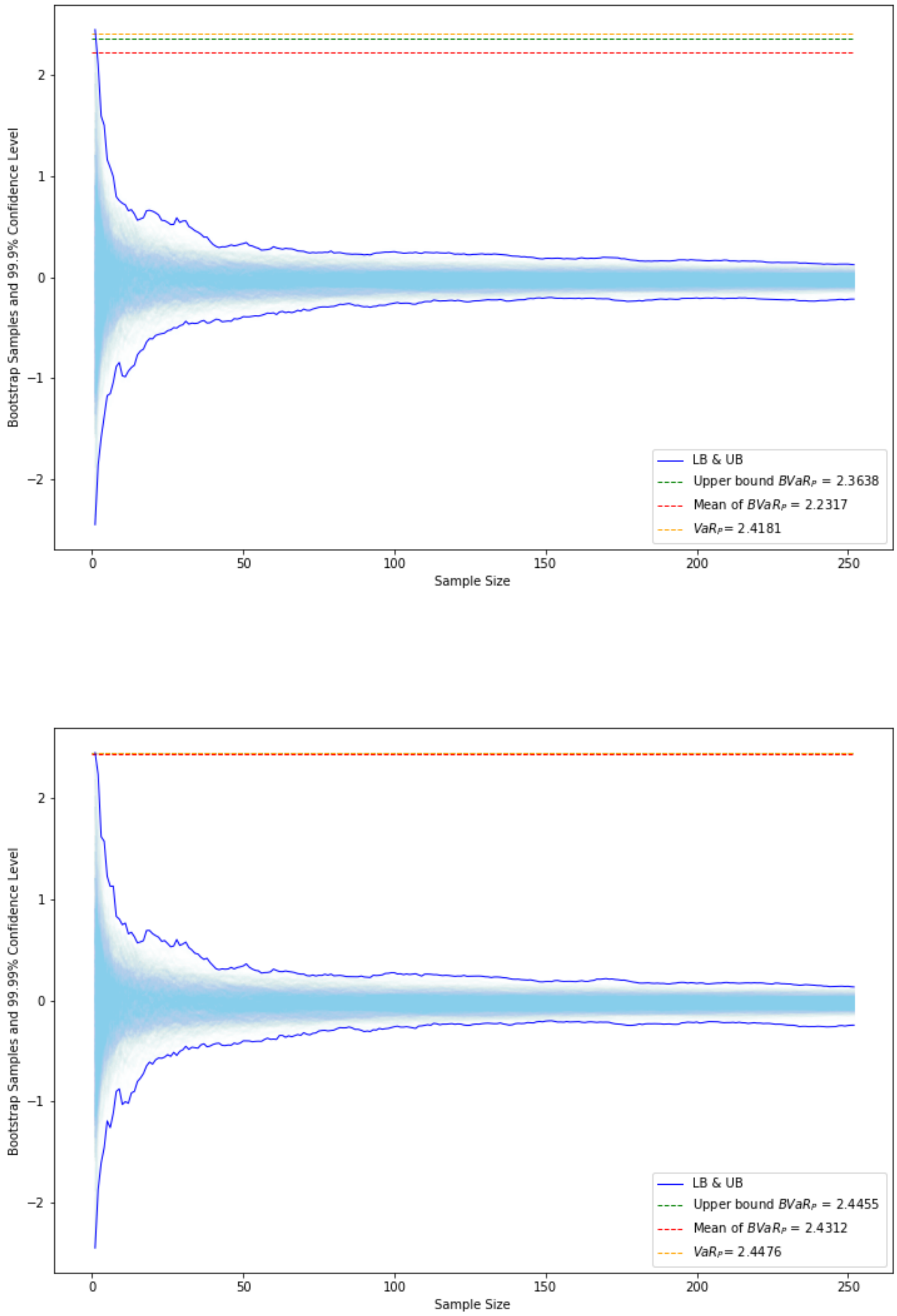


Figure B.9: Bootstrap MC of X_{ip} for credit-VaR.

B.6 Confidence levels from the bootstrap methods

In Figures B.10 and B.12, bootstrap replicates of credit-VaR at significance level 1% is denoted with a black line, where bootstrap size is $B = 1000$. Bootstrap 99% upper confidence level (99%BUCL) using Basic Percentile (BP) is denoted by green dashed line. 99%BUCL usng Standard Hybrid Percentile (SHP) is denoted by gold dashed line. 99%BUCL using Bias corrected percentile (BCP) is denoted by red dashed line. 99%BUCL using Modified hybrid percentile (MHP) is denoted by blue dashed line. The purple line the systemic common factor credit-VaR for the sum of RV's at 99% level.

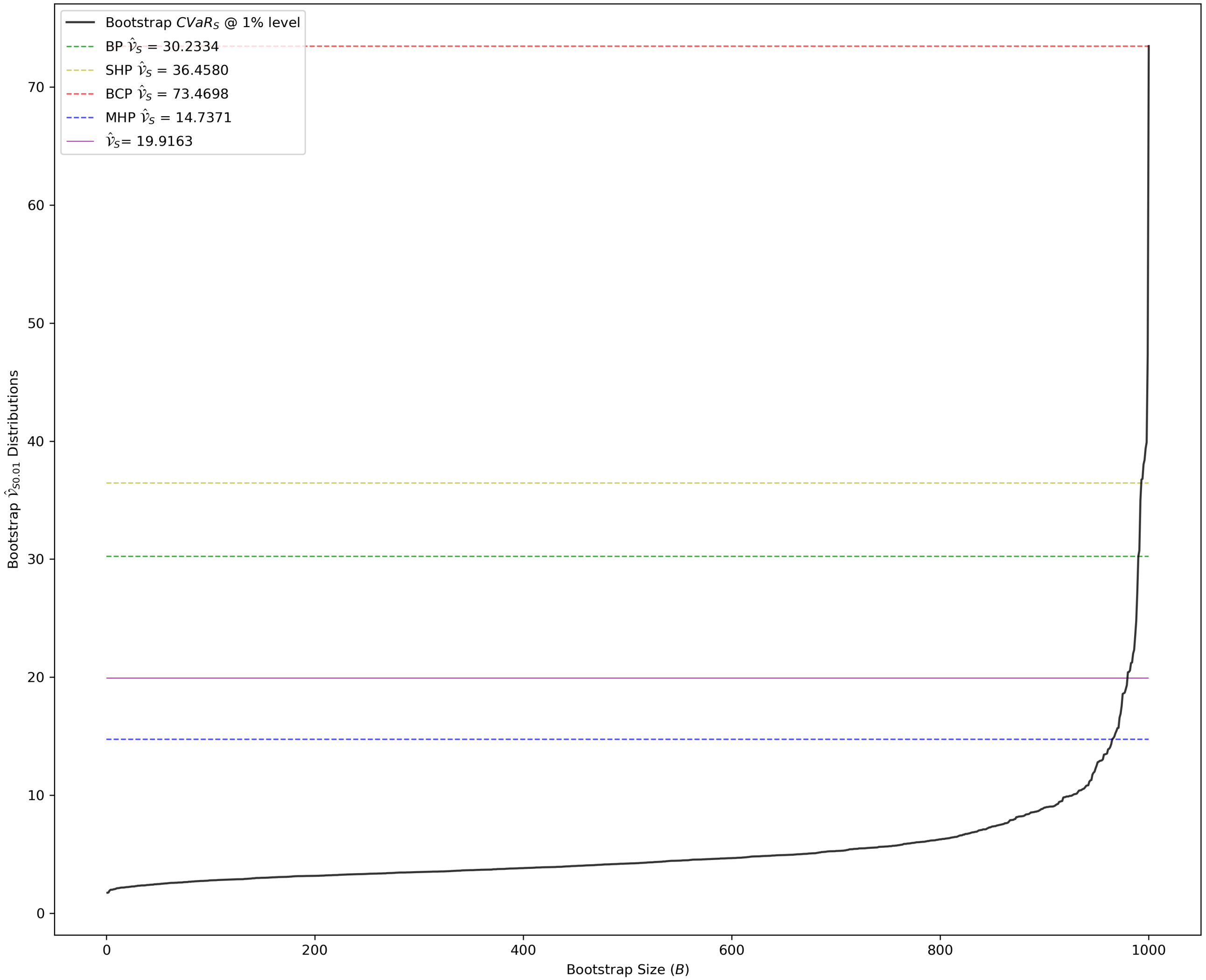


Figure B.10: Bootstrap confidence levels for X_{iS}.

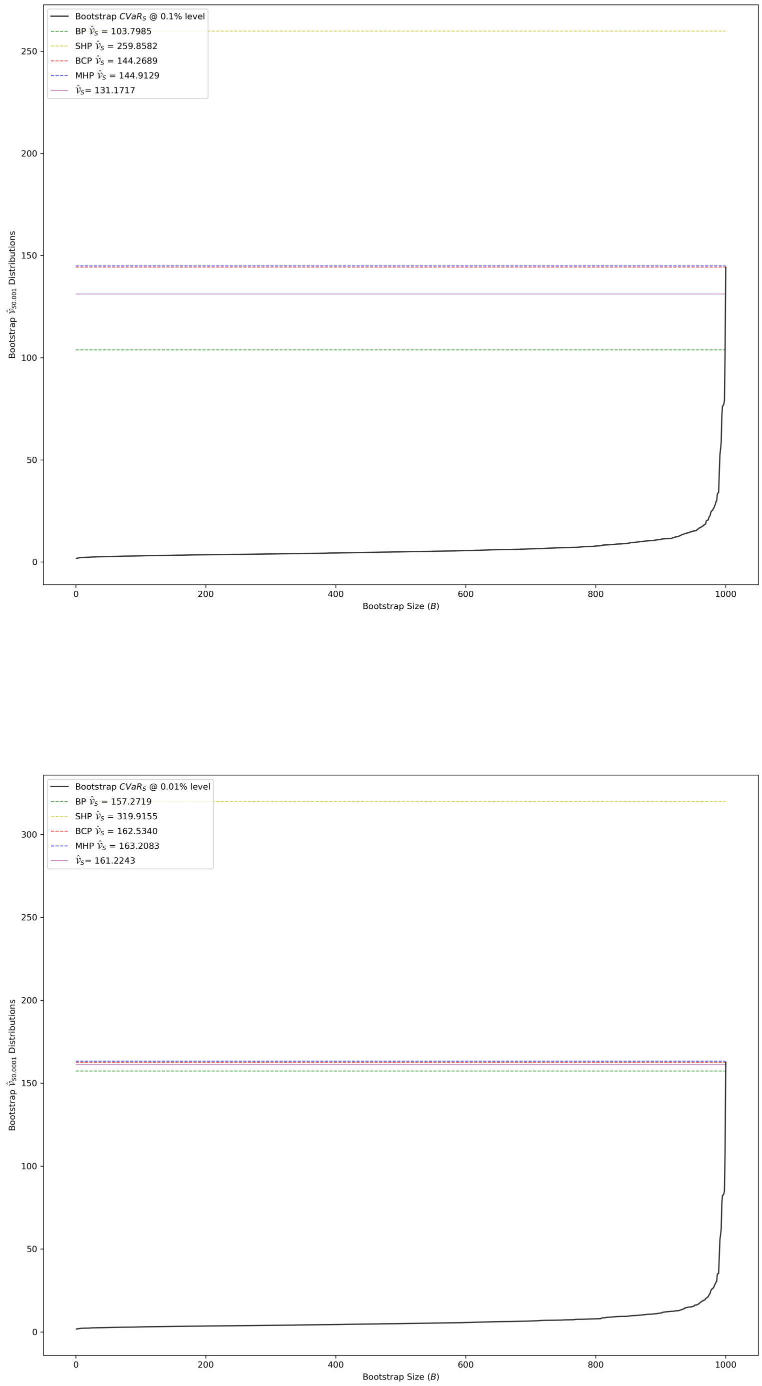


Figure B.11: Bootstrap confidence levels for X_{iS}.

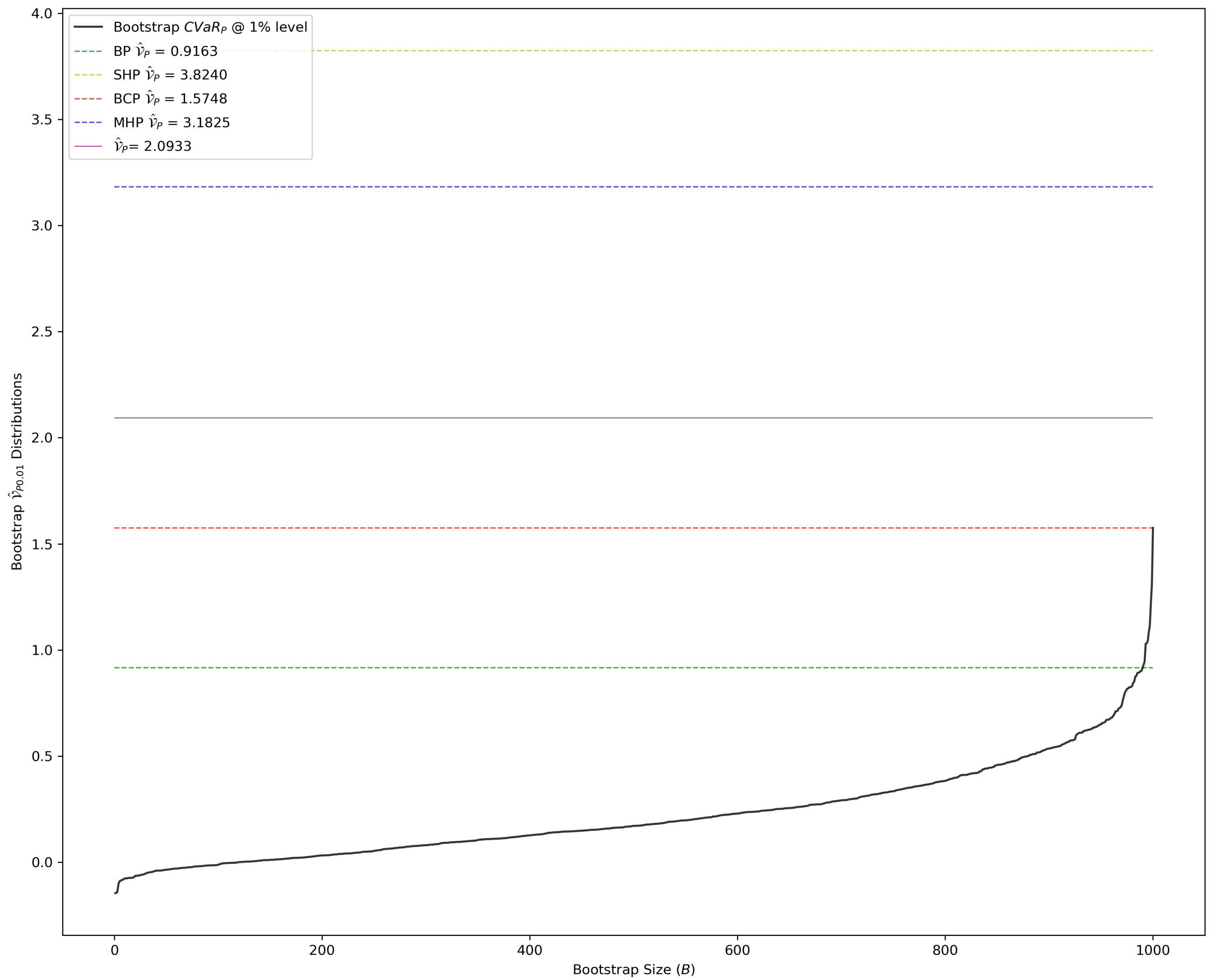
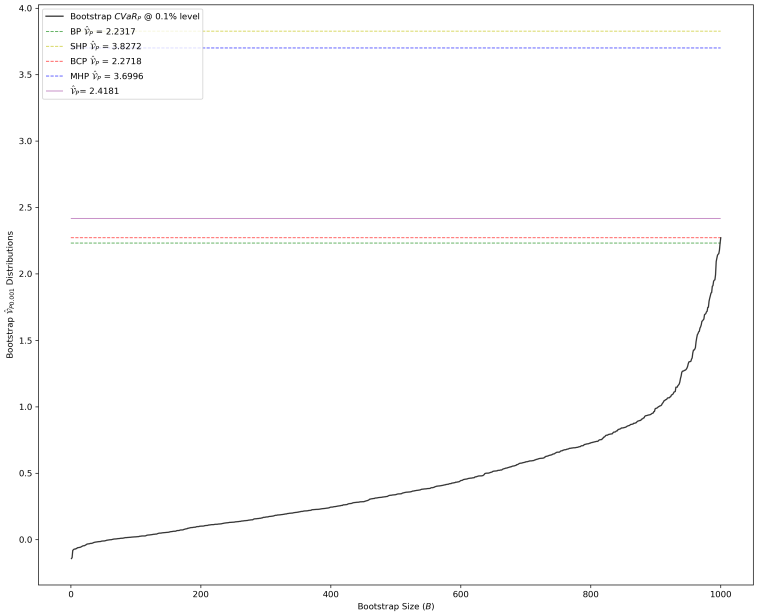


Figure B.12: Bootstrap confidence levels for X_{iP}.



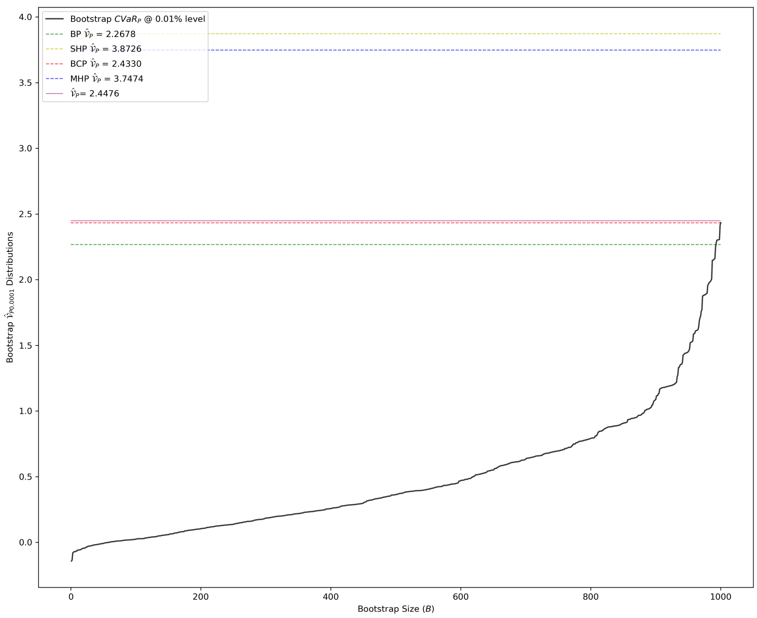


Figure B.13: Bootstrap confidence levels for X_{iP}.

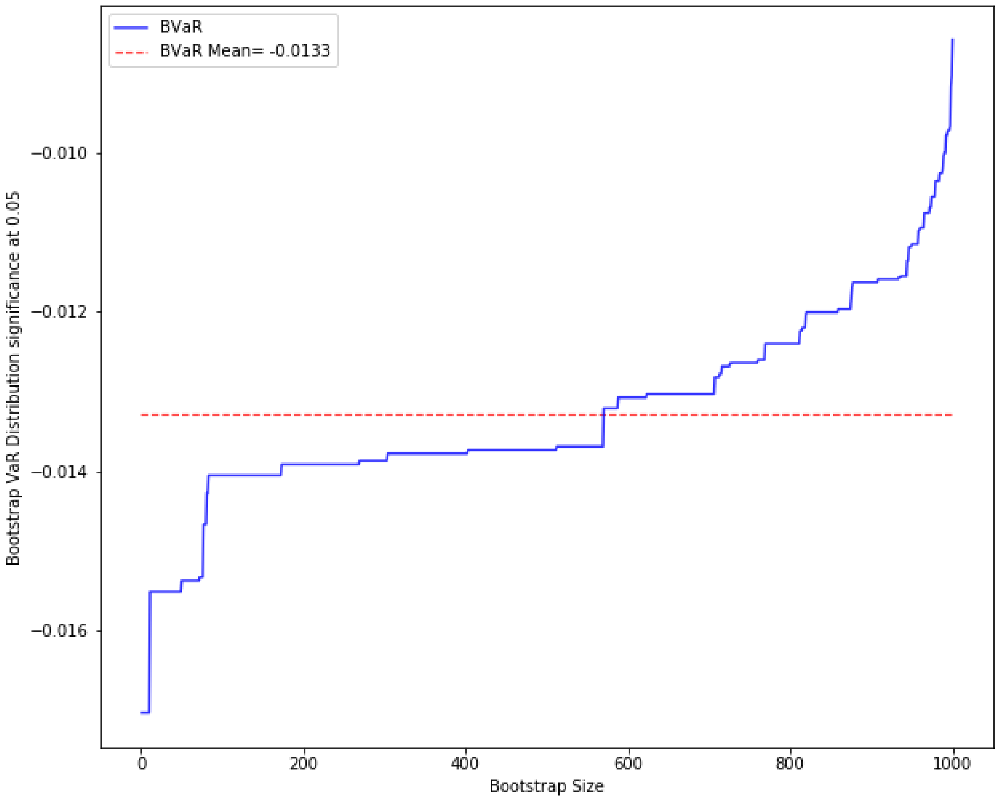


Figure B.14: Confidence levels as bootstrap sample size increases.

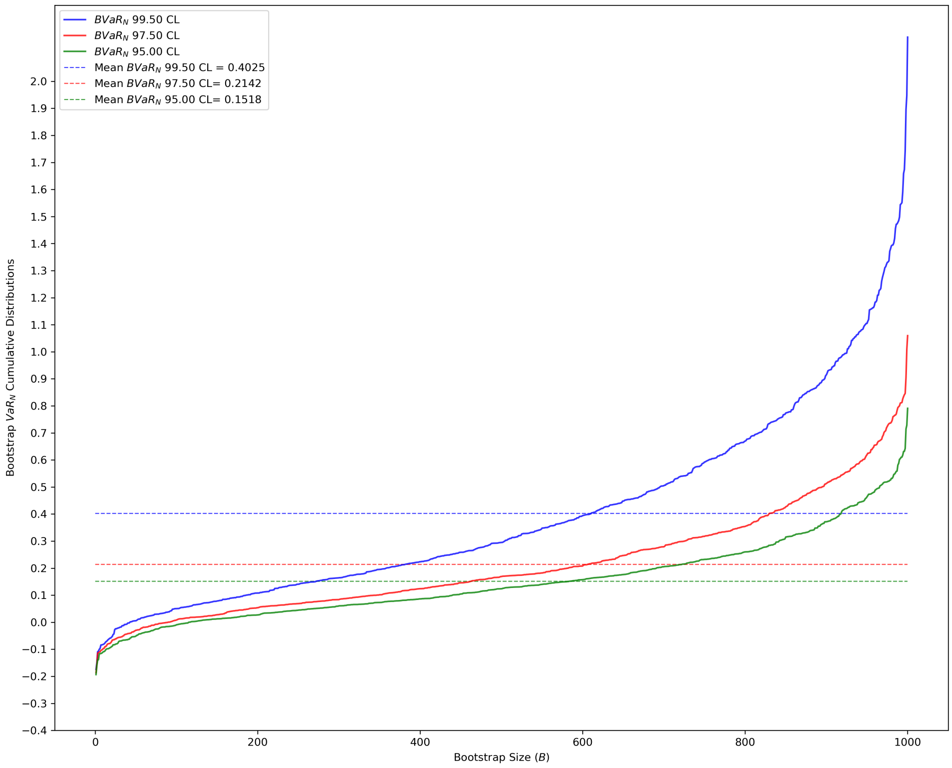


Figure B.15: Normal confidence levels as bootstrap sample size increases.

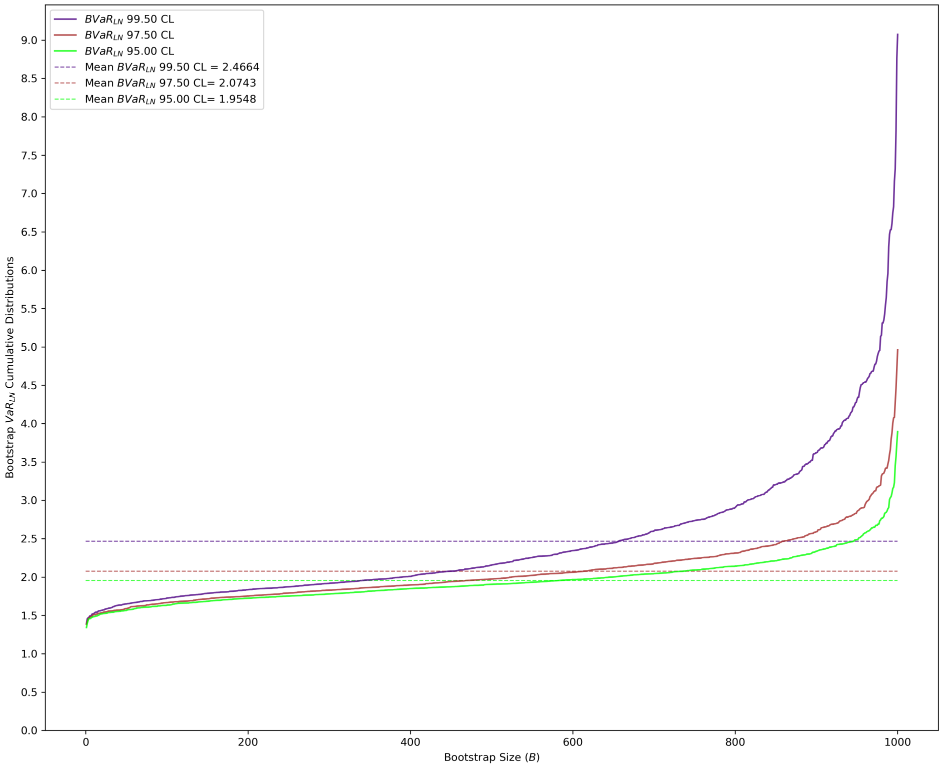


Figure B.16: Log-Normal confidence levels as bootstrap sample size increases.

Appendix C. More theory and graphs from Chapter 4

C.1 Loss function for Logit function

The Logit probability density function (PDF) is given as

$$f_1(\mathbf{X}_i, \gamma) = \frac{e^{\mathbf{X}_i^T \gamma}}{[1 + e^{(\mathbf{X}_i^T \gamma)}]^2}$$

and the Logit cumulative distribution function (CDF) is given as

$$F_1(\mathbf{X}_i, \gamma) = \pi_{1i} = \frac{e^{\mathbf{X}_i^T \gamma}}{1 + e^{(\mathbf{X}_i^T \gamma)}}.$$

The MLE function is

$$\mathcal{L}_1(\gamma) = \sum_{i=1}^{n} [Y_i \ln(\pi_{1i}) + (1 - Y_i) \ln(1 - \pi_{1i})]$$

Substituting π_{i1} in $\mathcal{L}_1(\gamma)$ we have

$$\begin{aligned}
\mathcal{L}_1(\gamma) &= \sum_{i=1}^{n}\left[Y_i \ln\left(\frac{e^{\mathbf{X}_i^T\gamma}}{1+e^{(\mathbf{X}_i^T\gamma)}}\right) + (1-Y_i)\ln\left(1-\frac{e^{\mathbf{X}_i^T\gamma}}{1+e^{(\mathbf{X}_i^T\gamma)}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(\frac{e^{\mathbf{X}_i^T\gamma}}{1+e^{\mathbf{X}_i^T\gamma}}\right) + (1-Y_i)\ln\left(\frac{e^{-\mathbf{X}_i^T\gamma}}{1+e^{-\mathbf{X}_i^T\gamma}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(\frac{e^{\mathbf{X}_i^T\gamma}}{1+e^{\mathbf{X}_i^T\gamma}}\right) + (1-Y_i)\ln\left(\frac{e^{-\mathbf{X}_i^T\gamma}}{1+e^{-\mathbf{X}_i^T\gamma}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(\frac{e^{\mathbf{X}_i^T\gamma}}{1+e^{\mathbf{X}_i^T\gamma}}\right) + \ln\left(\frac{e^{-\mathbf{X}_i^T\gamma}}{1+e^{-\mathbf{X}_i^T\gamma}}\right) - Y_i\ln\left(\frac{e^{-\mathbf{X}_i^T\gamma}}{1+e^{-\mathbf{X}_i^T\gamma}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(\frac{e^{\mathbf{X}_i^T\gamma}}{1+e^{\mathbf{X}_i^T\gamma}}\right) - Y_i\ln\left(\frac{1}{1+e^{\mathbf{X}_i^T\gamma}}\right) + \ln\left(\frac{1}{1+e^{\mathbf{X}_i^T\gamma}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(\frac{e^{\mathbf{X}_i^T\gamma}}{1+e^{\mathbf{X}_i^T\gamma}}\times\frac{1+e^{\mathbf{X}_i^T\gamma}}{1}\right) + \ln\left(\frac{1}{1+e^{\mathbf{X}_i^T\gamma}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(e^{\mathbf{X}_i^T\gamma}\right) + \ln\left(1+e^{\mathbf{X}_i^T\gamma}\right)^{-1}\right] \\
&= \sum_{i=1}^{n}\left[Y_i\left(\mathbf{X}_i^T\gamma\right) - \ln\left(1+e^{\mathbf{X}_i^T\gamma}\right)\right]
\end{aligned}$$

The gradient $\mathcal{L}_1'(\gamma)$ of the log-likelihood function for the Logit function is given as

$$\begin{aligned}
\mathcal{L}_1'(\gamma_k) &= \mathbf{X}_i^T\mathbf{Y} - \frac{\mathbf{X}_i^T}{\left(1+e^{-\mathbf{X}_i^T\gamma}\right)} \\
\mathbf{X}_i^T\left[\mathbf{Y} - \frac{1}{1+e^{-\mathbf{X}_i^T\hat{\gamma}}}\right] &= 0 \\
\mathbf{X}_i^T(\mathbf{Y} - \mathcal{PD}_{1i}) &= 0.
\end{aligned}$$

The Hessian $\mathcal{L}_1''(\gamma)$ of the log-likelihood function for the Logit function is given as

$$\begin{aligned}
\mathcal{L}_1''(\gamma_k) &= -\mathbf{X}_i^T \frac{1}{\left(1+e^{-\mathbf{X}_i^T\gamma}\right)\left(1+e^{\mathbf{X}_i^T\gamma}\right)}\mathbf{X}_i \\
-\mathbf{X}_i^T\left[\frac{1}{\left(1+e^{-\mathbf{X}_i^T\hat{\gamma}}\right)\left(1+e^{\mathbf{X}_i^T\hat{\gamma}}\right)}\right]\mathbf{X}_i &= 0 \\
-\mathbf{X}_i^T\left[\mathcal{PD}_{1i}(1-\mathcal{PD}_{1i})\right]\mathbf{X}_i &= 0.
\end{aligned}$$

Therefore, it is clear that Equation 4.4.8 for the Logit function simplifies to

$$\hat{\gamma}_{k+1} = \hat{\gamma}_k + [\mathbf{X}_i^T\mathbf{W}_k\mathbf{X}_i]^{-1}\mathbf{X}_i^T(\mathbf{Y}-\mathcal{PD}_{1i}), \tag{C.1.1}$$

where

- $\hat{\gamma}_k$ is the old estimate and $\hat{\gamma}_{k+1}$ is the new estimate in each iteration,
- $\mathbf{X}_i$ is the design matrix,
- $\mathbf{Y}$ is the response vector,
- $\mathcal{PD}_{1i} = [1+e^{-\mathbf{X}_i^T\hat{\gamma}}]^{-1}$ is the predicted probability of default values at the current iteration, and
- $\mathbf{W}_k$ is the diagonal matrix with elements given as $\mathcal{PD}_{1i}(1-\mathcal{PD}_{1i})$.

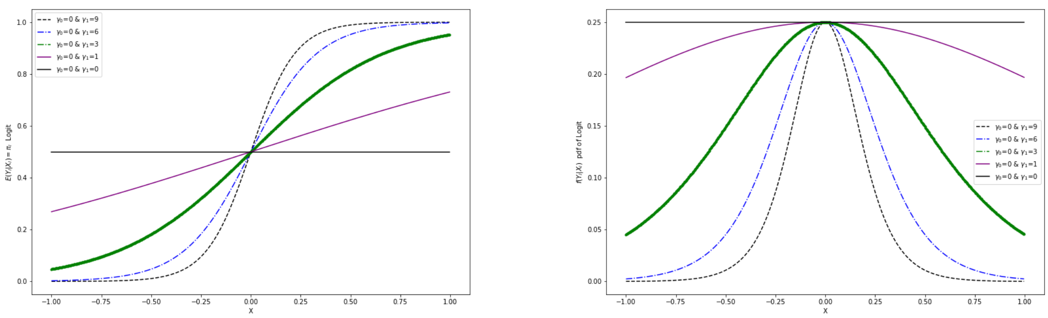


Figure C.1: CDF and PDF plot of BLRM

C.2 Loss function for Cloglog function:

The Cloglog probability density function (PDF) is given as

$$f_2(\mathbf{X}_i, \gamma) = e^{[\mathbf{X}_i^T \gamma - e^{\mathbf{X}_i^T \gamma}]}$$

and the Cloglog cumulative distribution function (CDF) is given as

$$F_2(\mathbf{X}_i, \gamma) = \pi_{2i} = 1 - e^{-e^{(\mathbf{X}_i^T \gamma)}}.$$

The MLE function is

$$\mathcal{L}_2(\gamma) = \sum_{i=1}^{n} [Y_i \ln(\pi_{2i}) + (1 - Y_i) \ln(1 - \pi_{2i})]$$

Substituting π_{i2} in $\mathcal{L}_2(\gamma)$ we have

$$\begin{aligned}
\mathcal{L}_2(\gamma) &= \sum_{i=1}^{n}\left[Y_i \ln\left(1 - e^{-e^{(\mathbf{X}_i^T\gamma)}}\right) + (1-Y_i)\ln\left(1 - 1 + e^{-e^{(\mathbf{X}_i^T\gamma)}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(1 - e^{-e^{(\mathbf{X}_i^T\gamma)}}\right) + (1-Y_i)\ln\left(e^{-e^{(\mathbf{X}_i^T\gamma)}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(1 - e^{-e^{(\mathbf{X}_i^T\gamma)}}\right) + \ln\left(e^{-e^{(\mathbf{X}_i^T\gamma)}}\right) - Y_i \ln\left(e^{-e^{(\mathbf{X}_i^T\gamma)}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \left(\ln\left(1 - e^{-e^{(\mathbf{X}_i^T\gamma)}}\right) - \ln\left(e^{-e^{(\mathbf{X}_i^T\gamma)}}\right)\right) + \ln\left(e^{-e^{(\mathbf{X}_i^T\gamma)}}\right)\right] \\
&= \sum_{i=1}^{n}\left[Y_i \left(\ln\left(1 - e^{-e^{(\mathbf{X}_i^T\gamma)}}\right) - \ln\left(e^{-e^{(\mathbf{X}_i^T\gamma)}}\right)\right) - e^{(\mathbf{X}_i^T\gamma)}\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(\frac{1 - e^{-e^{(\mathbf{X}_i^T\gamma)}}}{e^{-e^{(\mathbf{X}_i^T\gamma)}}}\right) - e^{(\mathbf{X}_i^T\gamma)}\right] \\
&= \sum_{i=1}^{n}\left[Y_i \ln\left(e^{e^{(\mathbf{X}_i^T\gamma)}} - 1\right) - e^{\mathbf{X}_i^T\gamma}\right]
\end{aligned}$$

The gradient $\mathcal{L}_2'(\gamma)$ and hessian $\mathcal{L}_2''(\gamma)$ of the log-likelihood function for the Cloglog function may be numerically determined.

C.3 Graphical represantation of the cost functions, ROC and predictions

In Figures C.3 through to C.10, the green line represent the chanage in $\mathcal{L}(\gamma)$, while the number of iterations per hundreds (i.e., $\mathcal{I}/100$) increases for $\mathcal{I} = (1000, 10000, 30000)$.

In Figures C.11 through to C.18, the red line denotes the accuracy rate from training

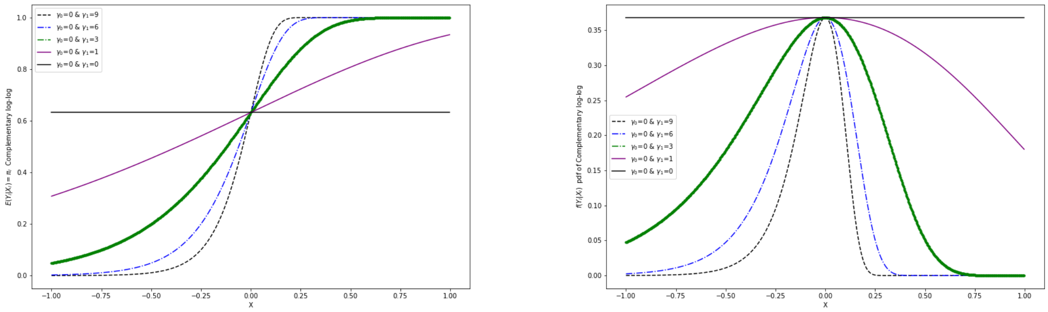


Figure C.2: CDF and PDF plot of Cloglog

dataset $\mathcal{A}_{tr}$ and green line denotes accuracy rate from testing dataset $\mathcal{A}_{ts}$ for models produced for $\mathcal{I} = (1000, 10000, 30000)$. The blue dotted line denotes the threshold.

In Figures C.19 through to C.22, the boxplots were generated using the optimised parameter estimates at $\mathcal{I} = (1000, 10000, 30000)$ for both Logit $\mathcal{PD}_i$ and Cloglog $\mathcal{PD}_i$ models, i.e., when the predictor variable come from the uniform distribution and standard normal distribution. The diamond marker is the mean and the line denote the median from the distribution. The extreme values from the distribution are denoted by dots.

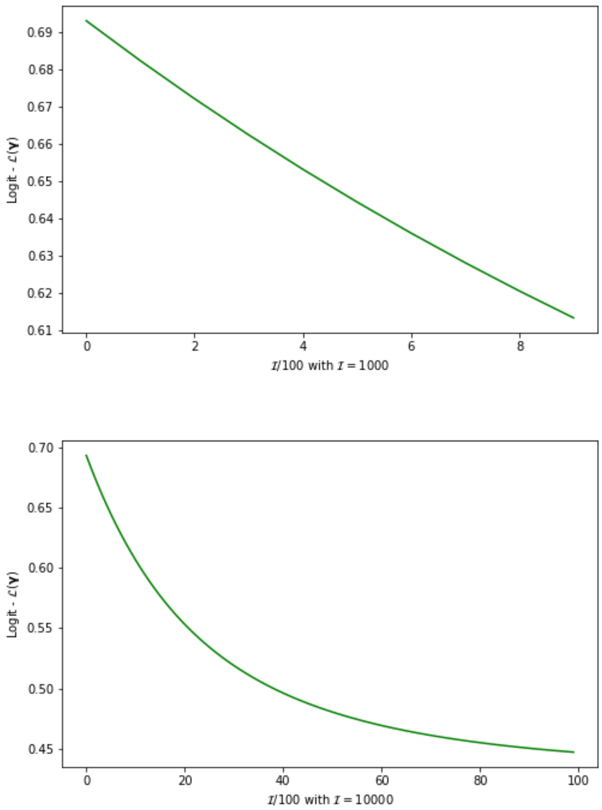


Figure C.3: Plots of the Logit cross-entropy cost function $\mathcal{L}(\gamma)$ with Scenario A.

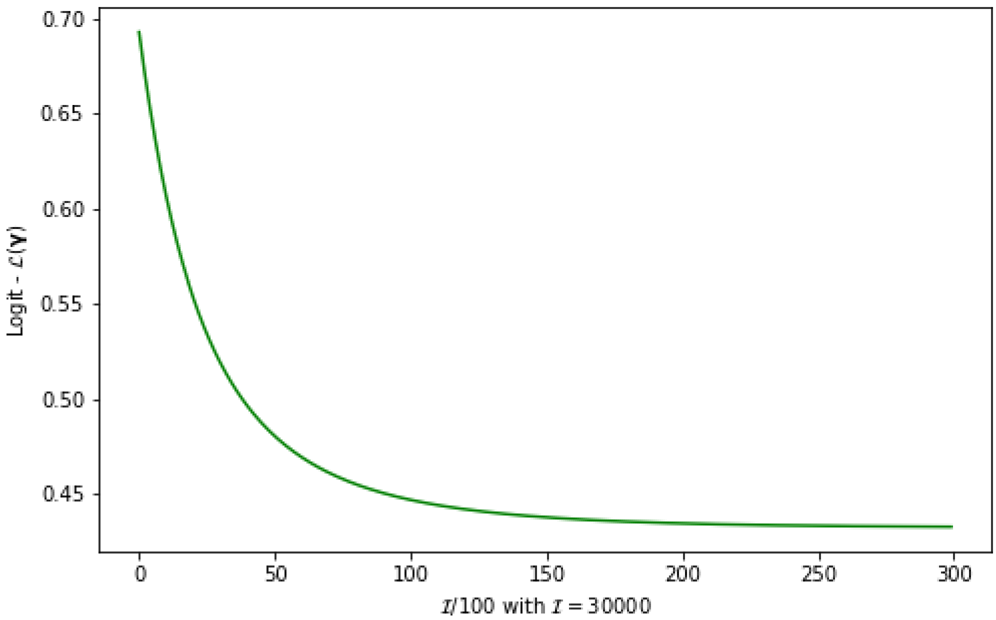


Figure C.4: Plots of the Logit cross-entropy cost function $\mathcal{L}(\gamma)$ with Scenario A with 30000 iterations.

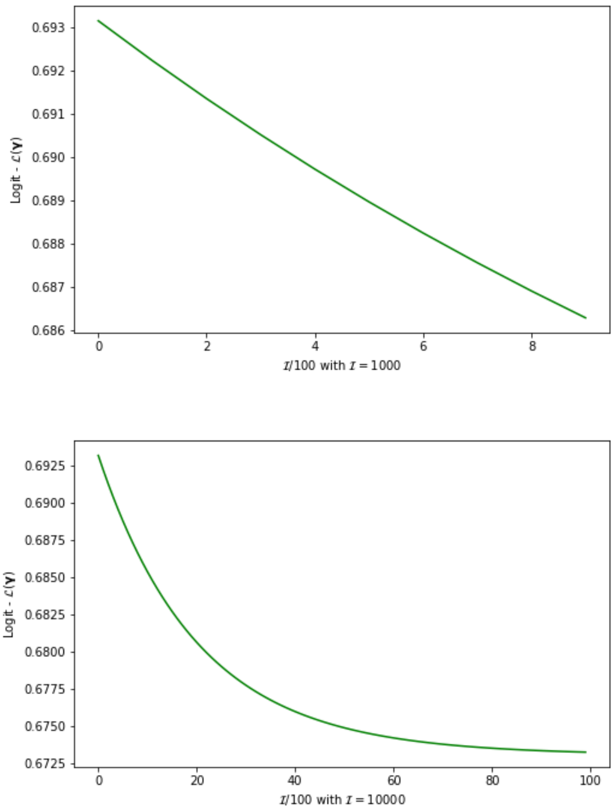


Figure C.5: Plots of the Logit cross-entropy cost function $\mathcal{L}(\gamma)$ with Scenario B.

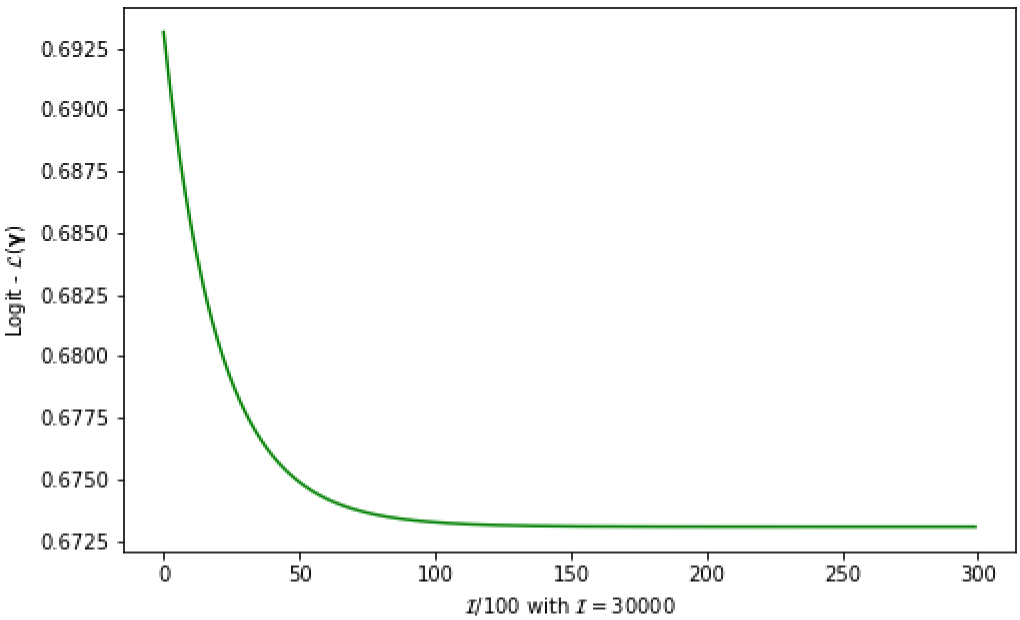


Figure C.6: Plots of the Logit cross-entropy cost function $\mathcal{L}(\gamma)$ with Scenario B with 30000 iterations.

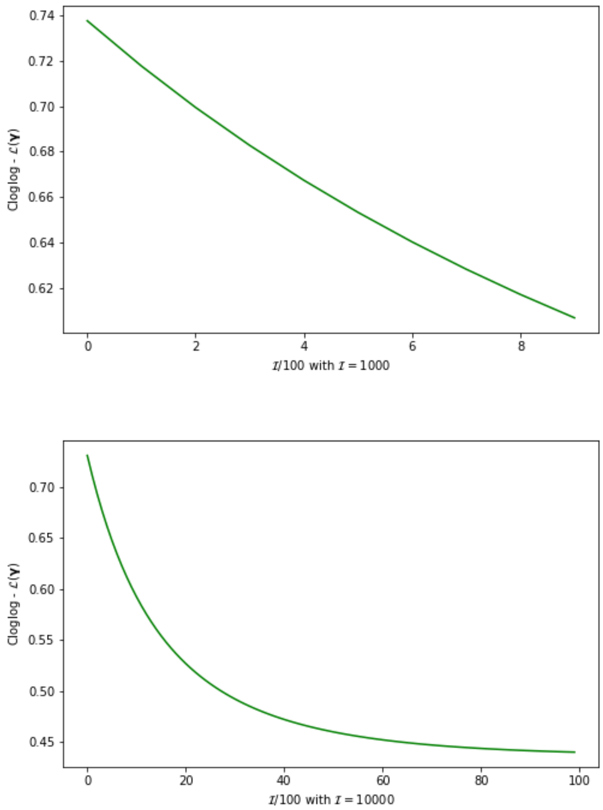


Figure C.7: Plots of the Cloglog cross-entropy cost function $\mathcal{L}(\gamma)$ with Scenario A.

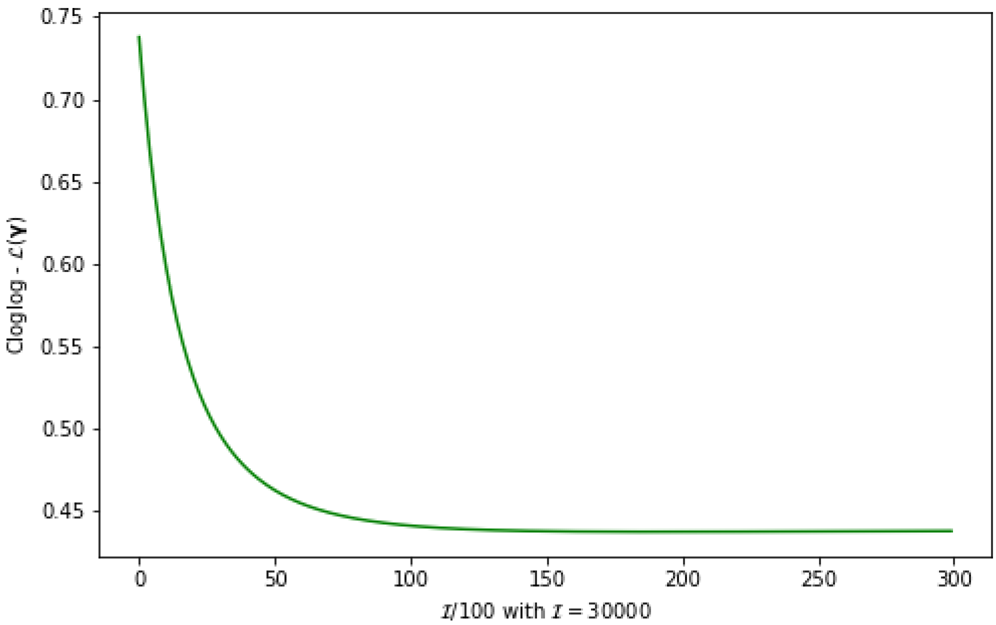


Figure C.8: Plots of the Cloglog cross-entropy cost function $\mathcal{L}(\gamma)$ with Scenario A with 30000 iterations.

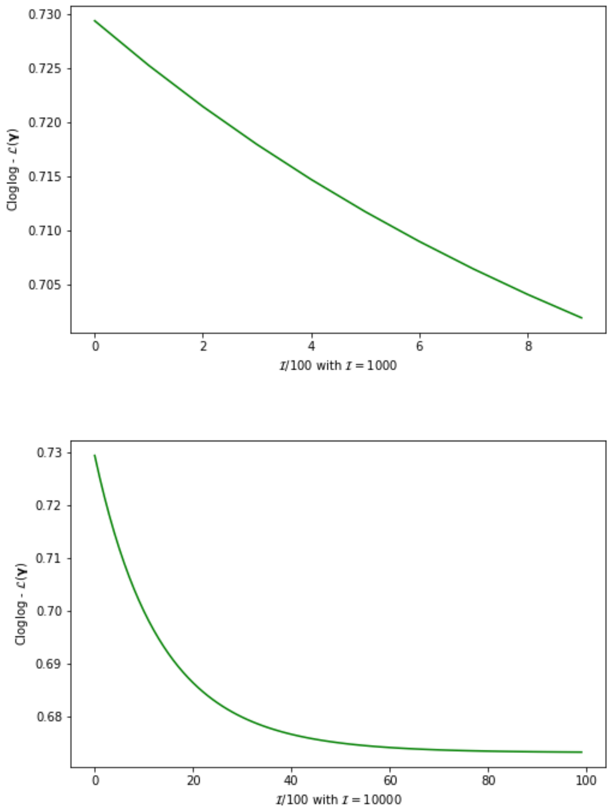


Figure C.9: Plots of the Cloglog cross-entropy cost function $\mathcal{L}(\gamma)$ with Scenario B.

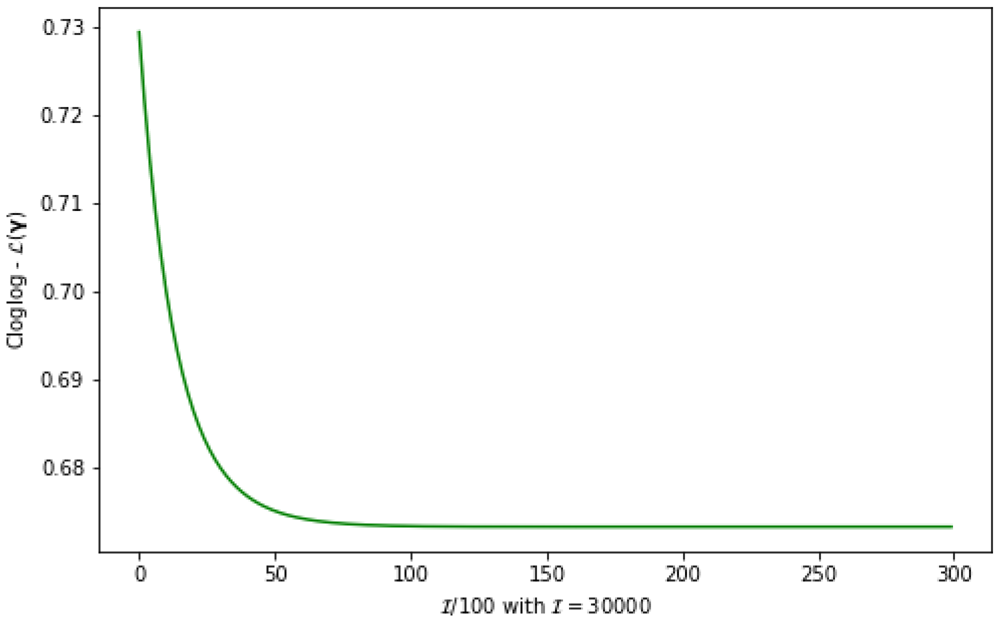


Figure C.10: Plots of the Cloglog cross-entropy cost function $\mathcal{L}(\gamma)$ with Scenario B with 30000 iterations.

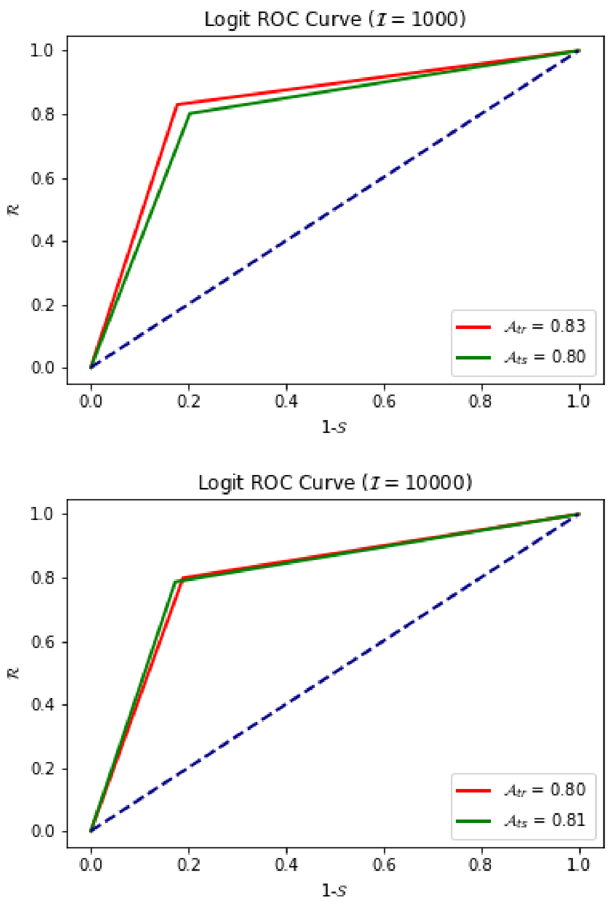


Figure C.11: Plots of the Logit ROC curve with Scenario A.

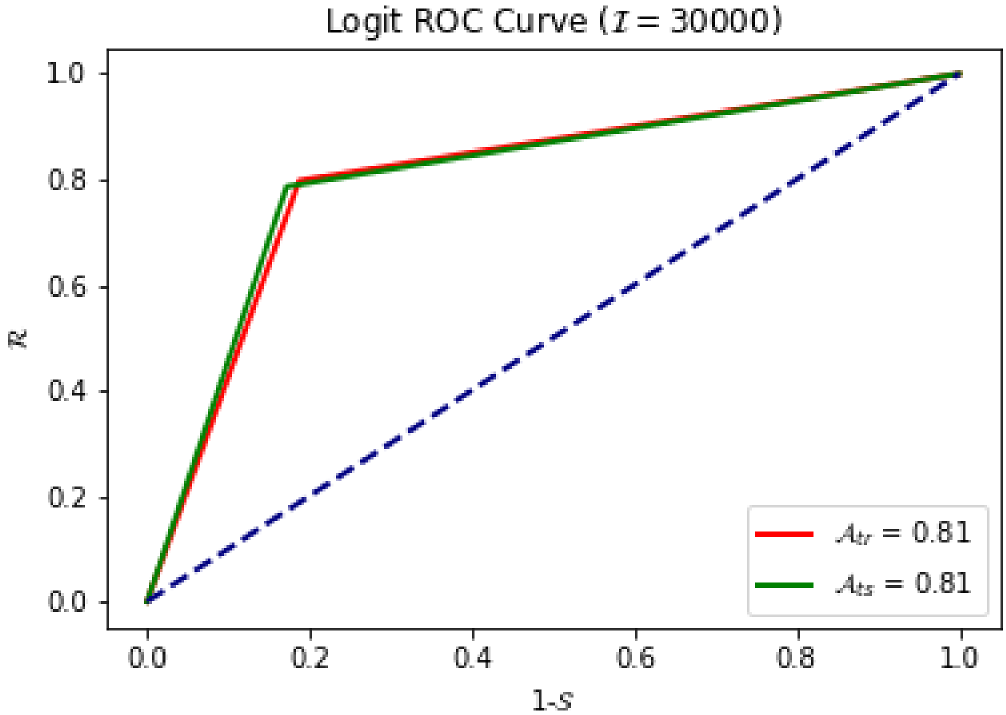


Figure C.12: Plots of the Logit ROC curve with Scenario A using 30000 iterations.

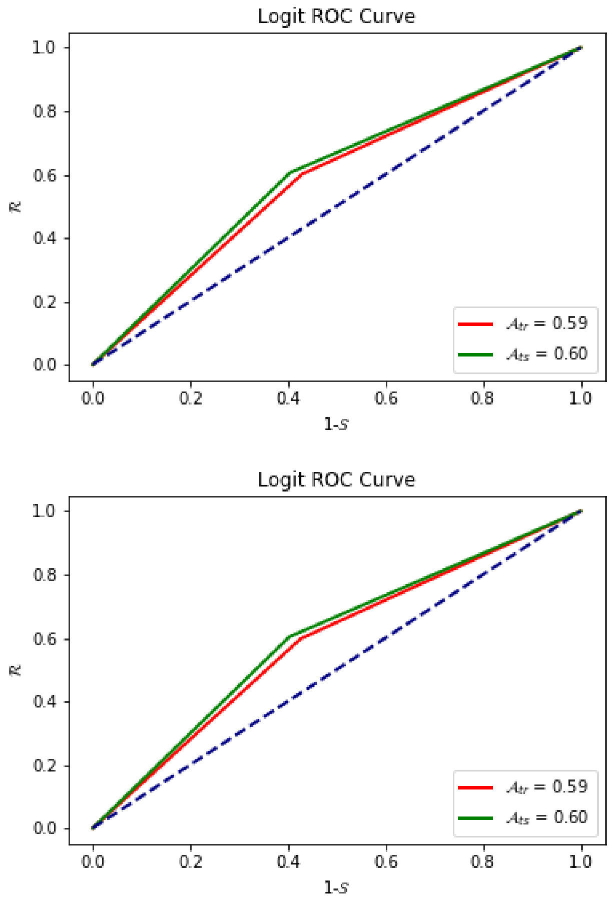


Figure C.13: Plots of the Logit ROC curve with Scenario B.

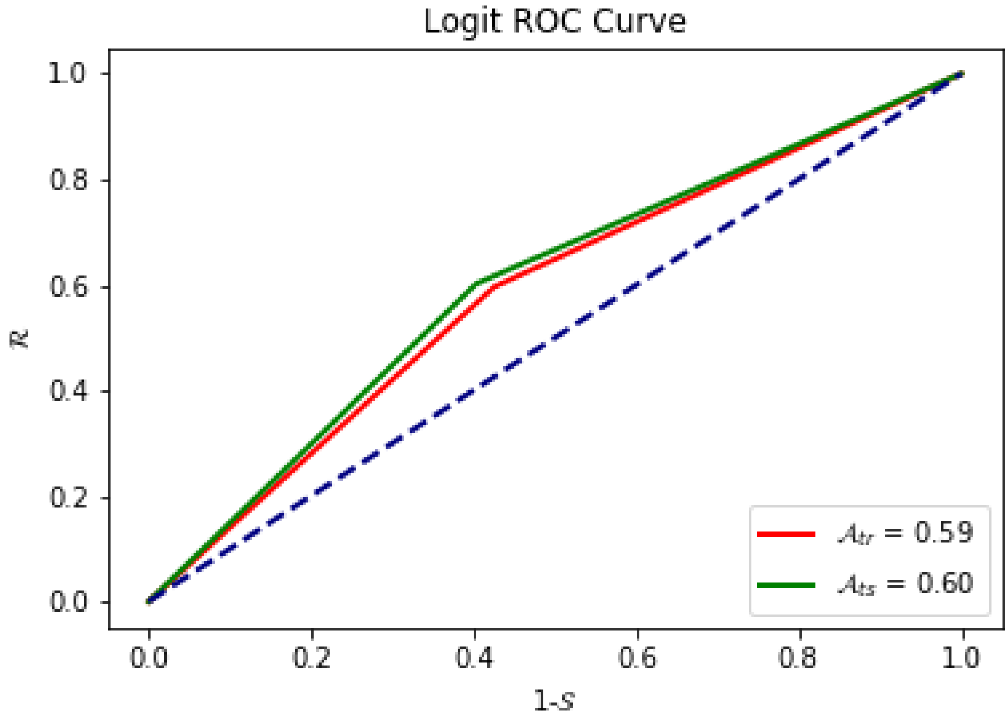


Figure C.14: Plots of the Logit ROC curve with Scenario B using 30000 iterations.

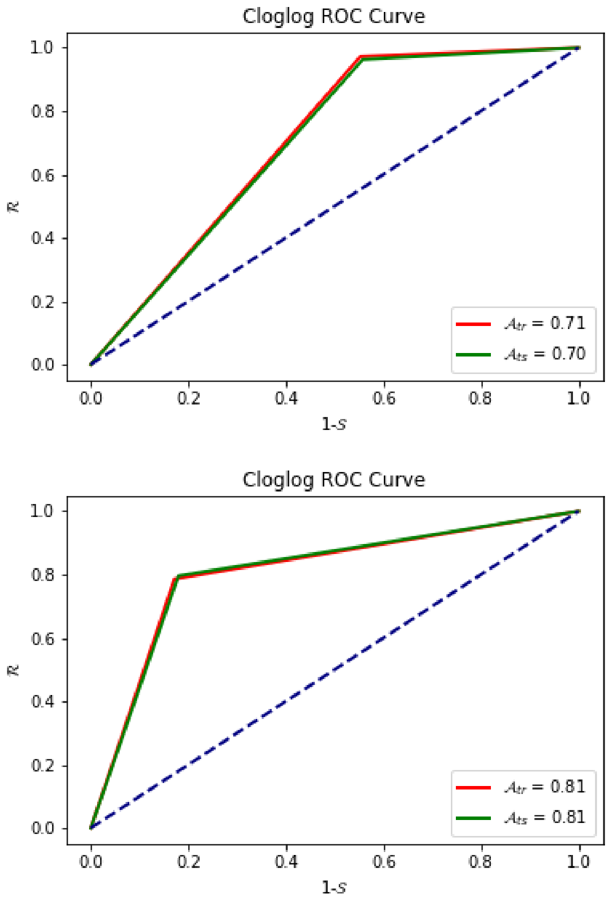


Figure C.15: Plots of the Cloglog ROC curve with Scenario A.

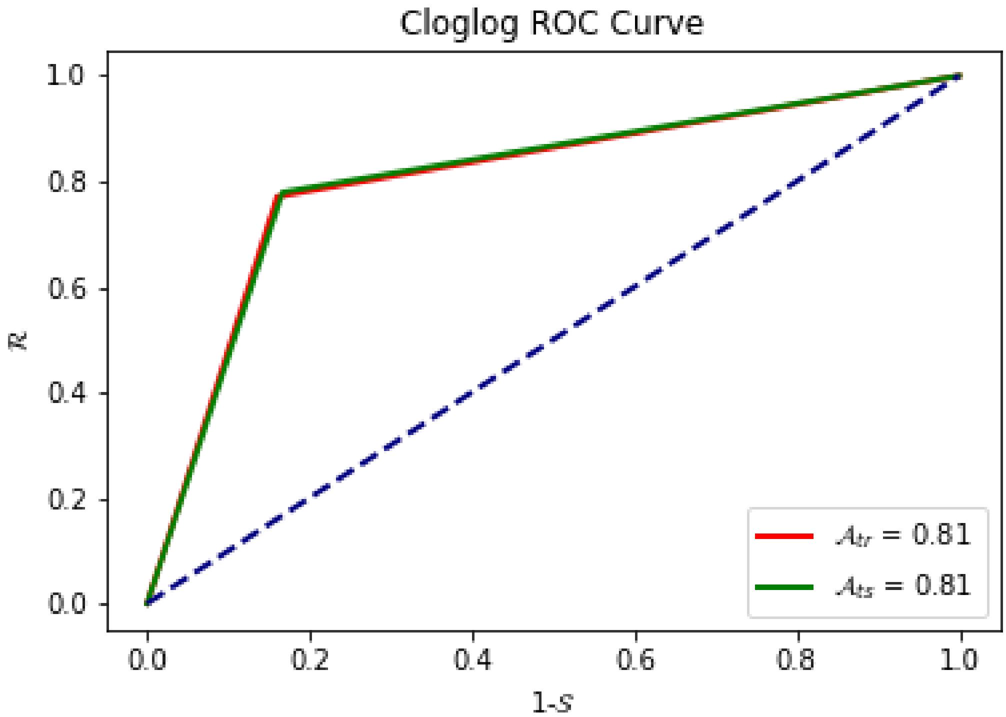


Figure C.16: Plots of the Cloglog ROC curve with Scenario A using 30000 iterations.

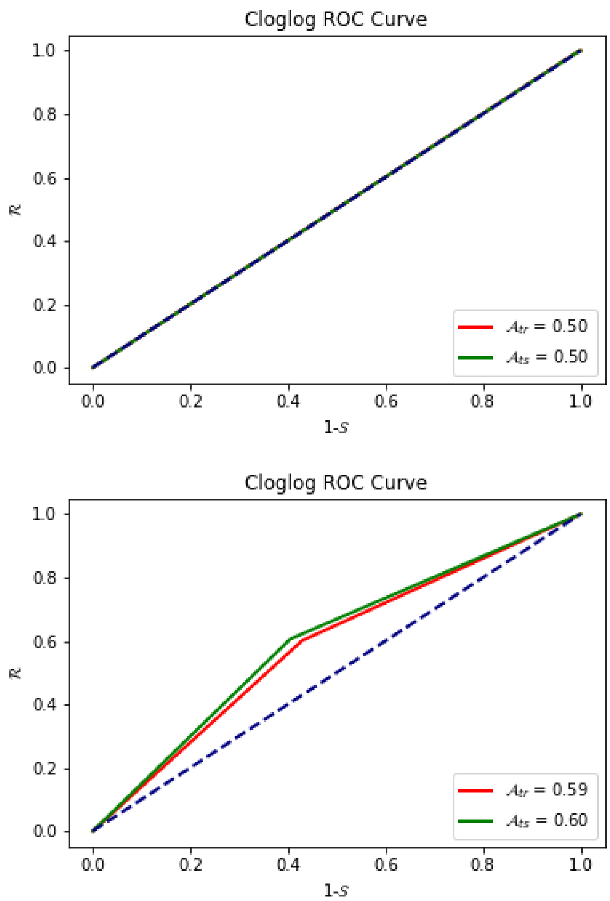


Figure C.17: Plots of the Cloglog ROC curve with Scenario B.

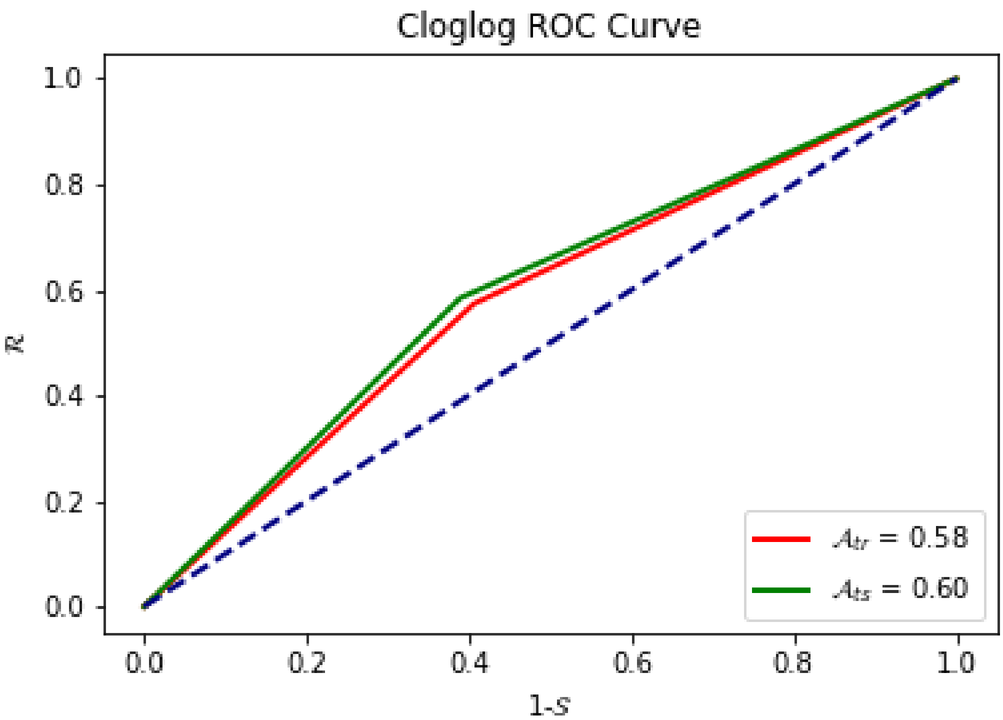
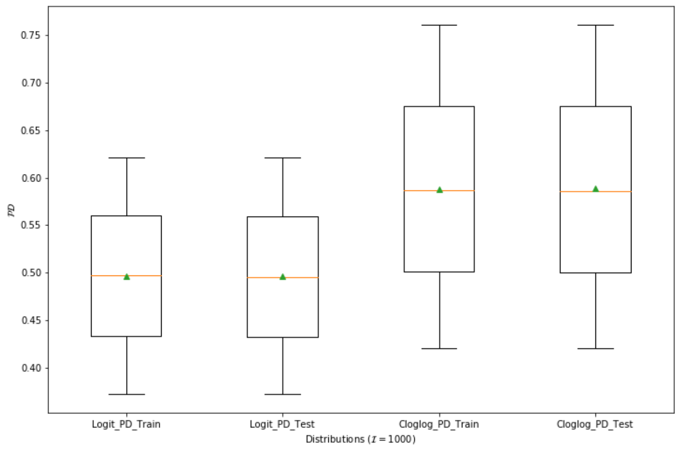


Figure C.18: Plots of the Cloglog ROC curve with Scenario B using 30000 iterations.



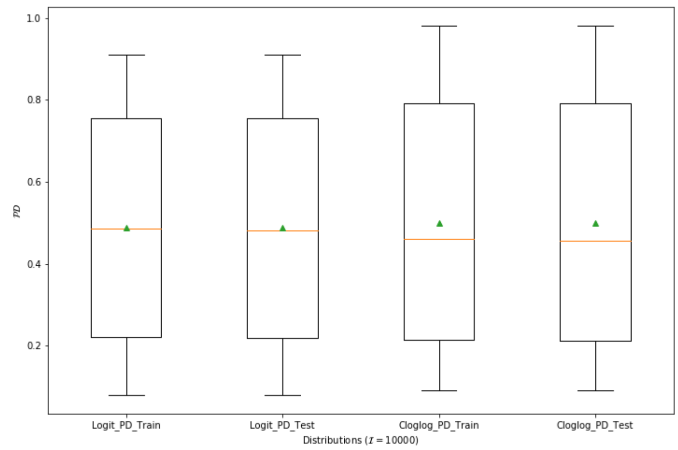


Figure C.19: Comparison "$\mathcal{PD}$" models boxplots with Scenario A.

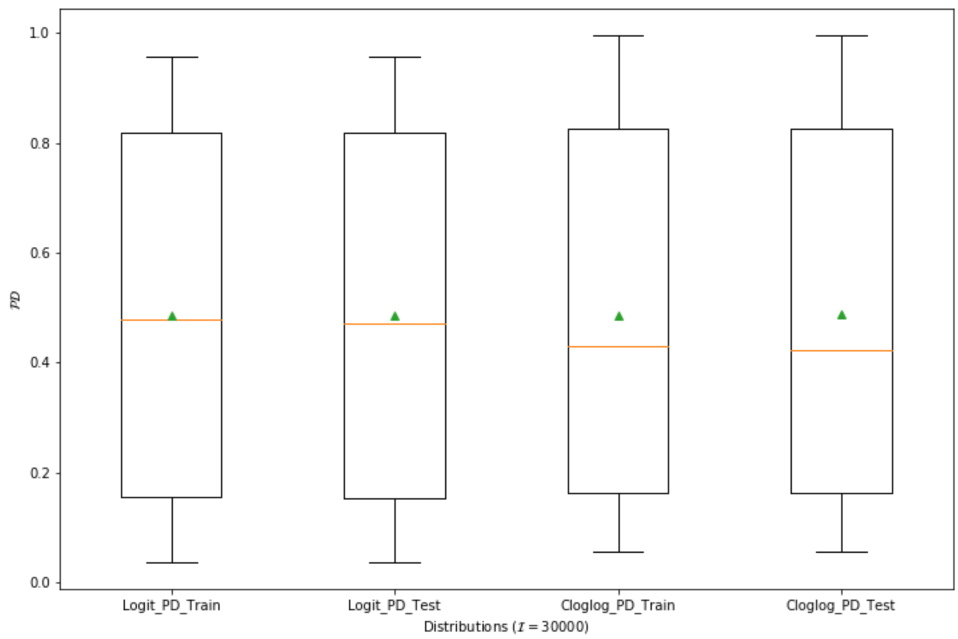


Figure C.20: Comparison "$\mathcal{PD}$" models boxplots with Scenario A generated using 30000 iterations.

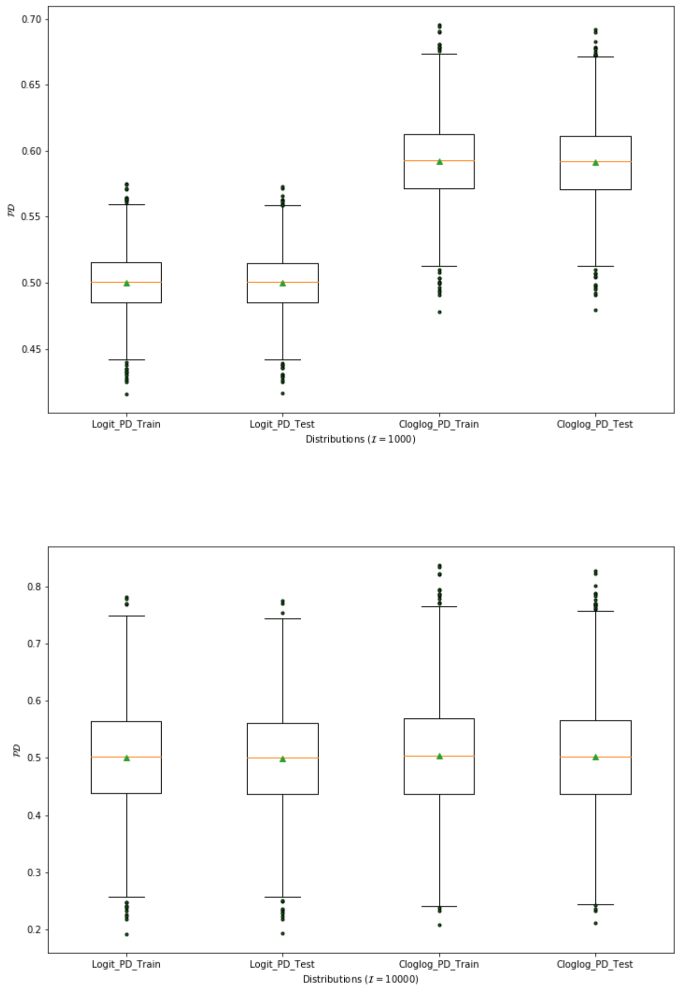


Figure C.21: Comparison "$\mathcal{PD}$" models boxplots with Scenario B.

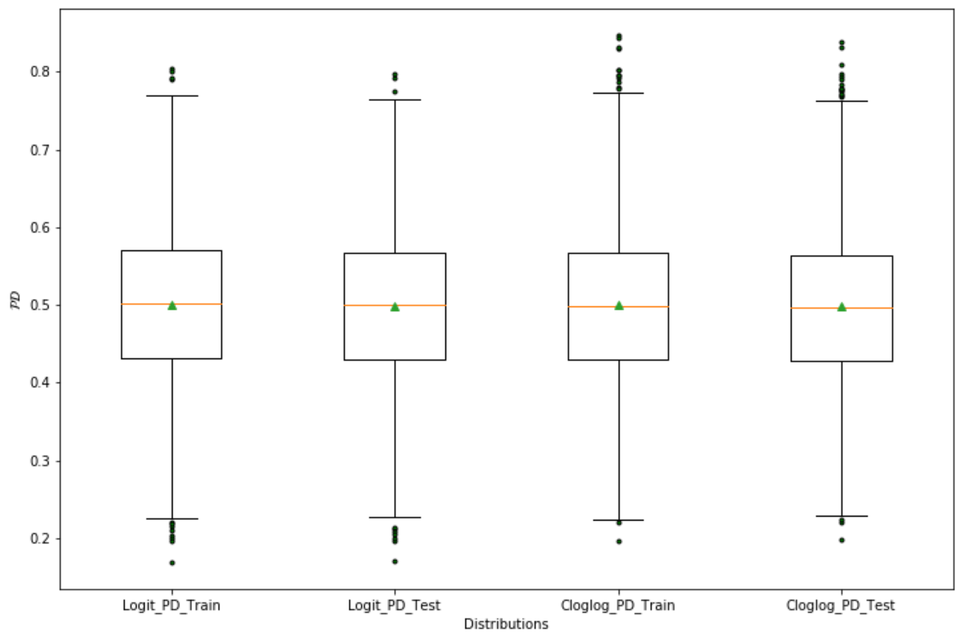


Figure C.22: Comparison "$\mathcal{PD}$" models boxplots with Scenario B generated using 30000 iterations.

Bibliography

Abdou, H. A. & Pointon, J. 2011. 'Credit scoring, statistical techniques and evaluation criteria: a review of the literature', *Intelligent Systems in Accounting, Finance and Management* **18**(2-3), 59–88.

Agresti, A. 2003. *Categorical data analysis*, Vol. 482, John Wiley & Sons.

Agresti, A. 2015. *Foundations of linear and generalized linear models*, John Wiley and Sons.

Agresti, A. & Kateri, M. 2011. *Categorical data analysis*, Springer.

Albanese, C., Caenazzo, S. & Frankel, O. 2016. 'Regression sensitivities for initial margin calculations', *Available at SSRN* .

Albert, A. & Anderson, J. A. 1984. 'On the existence of maximum likelihood estimates in logistic regression models', *Biometrika* **71**(1), 1–10.

Alemany, R., Bolancé, C. & Guillén, M. 2013. 'A nonparametric approach to calculating value-at-risk', *Insurance: Mathematics and Economics* **52**(2), 255–262. **URL:** *http://www.sciencedirect.com/science/article/pii/S0167668713000048*

Alhawarat, A., Salleh, Z., Mamat, M. & Rivaie, M. 2017. 'An efficient modified polak–ribière–polyak conjugate gradient method with global convergence properties', *Optimization Methods and Software* **32**(6), 1299–1312.

Allen, S. L. 2012. *Financial risk management: a practitioner's guide to managing market and credit risk*, Vol. 119, John Wiley & Sons.

Allison, P. D. 2012. *Logistic regression using SAS: Theory and application*, SAS Institute.

Arjas, E. & Haara, P. 1987. 'A logistic regression model for hazard: Asymptotic results', *Scandinavian Journal of Statistics* pp. 1–18.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. 1999. 'Coherent measures of risk', *Mathematical finance* **9**(3), 203–228.

Audet, C. & Tribes, C. 2018. 'Mesh-based nelder–mead algorithm for inequality constrained optimization', *Computational Optimization and Applications* **71**(2), 331–352.

Babaie-Kafaki, S. & Ghanbari, R. 2015. 'A hybridization of the polak-ribière-polyak and fletcher-reeves conjugate gradient methods', *Numerical Algorithms* **68**(3), 481–495.

Ball, J. & Fang, V. 2006. 'A survey of value-at-risk and its role in the banking industry', *Journal of Financial Education* **32**, 1–31.
URL: *http://www.jstor.org.nwulib.nwu.ac.za/stable/41948541*

Barrieu, P. & Scandolo, G. 2015. 'Assessing financial model risk', *European Journal of Operational Research* **242**(2), 546–556.

Basel, I. 2004. 'International convergence of capital measurement and capital standards: a revised framework', *Bank for international settlements* .

Batiz-Zuk, E., Christodoulakis, G. & Poon, S.-H. 2013. 'Basel ii credit loss distributions under non-normality', *Journal of Fixed Income* **23**(1), 56–75.

BCBS, I. 2015. 'Margin requirements for non-centrally cleared derivatives', *Basel Committee on Banking Supervision* .

Beck, T., De Jonghe, O. & Mulier, K. 2018. 'Bank sectoral concentration and (systemic) risk: Evidence from a worldwide sample of banks', *Available at SSRN 2959273* .

Begley, J., Ming, J. & Watts, S. 1996. 'Bankruptcy classification errors in the 1980s: An empirical analysis of altman's and ohlson's models', *Review of accounting Studies* **1**(4), 267–284.

Berg, D. 2007. 'Bankruptcy prediction by generalized additive models', *Applied Stochastic Models in Business and Industry* **23**(2), 129–143.

Berkowitz, J., Christoffersen, P. & Pelletier, D. 2011. 'Evaluating value-at-risk models with desk-level data', *Management Science* **57**(12), 2213–2227.
URL: *https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1080.0964*

Bignozzi, V. & Tsanakas, A. 2016. 'Parameter uncertainty and residual estimation risk', *Journal of Risk and Insurance* **83**(4), 949–978.
URL: *http://dx.doi.org/10.1111/jori.12075*

Bluhm, C., Overbeck, L. & Wagner, C. 2016. *Introduction to credit risk modeling*, Chapman and Hall/CRC.

Borowicz, J. M. & Norman, J. P. 2006. The effects of parameter uncertainty in the extreme event frequency-severity model, *in* '28th International Congress of Actuaries, Paris', Vol. 28, Citeseer.

Bottou, L. 2010. Large-scale machine learning with stochastic gradient descent, *in* 'Proceedings of COMPSTAT'2010', Springer, pp. 177–186.

Bottou, L. 2012. Stochastic gradient descent tricks, *in* 'Neural networks: Tricks of the trade', Springer, pp. 421–436.

Braione, M. & Scholtes, N. 2016. 'Forecasting value-at-risk under different distributional assumptions', *Econometrics* **4**(1), 3.

Breuer, T. & Csiszár, I. 2016. 'Measuring distribution model risk', *Mathematical Finance* **26**(2), 395–411.

Brigo, D., Morini, M. & Pallavicini, A. 2013. *Counterparty credit risk, collateral and funding : with pricing cases for all asset classes*, John Wiley and Sons Inc., Chichester, West Sussex. 2012051762 Damiano Brigo, Massimo Morini, Andrea Pallavicini. illustrations ; 26 cm Includes bibliographical references (pages [415]-421) and index.

Brigo, D. & Pallavicini, A. 2014. 'Ccp cleared or bilateral csa trades with initial or variation margins under credit, funding and wrong-way risks: A unified valuation approach'.

Calabrese, R. 2014. 'Predicting bank loan recovery rates with a mixed continuous-discrete model', *Applied stochastic models in business and industry* **30**(2), 99–114.

Calabrese, R. & Giudici, P. 2015. 'Estimating bank default with generalised extreme value regression models', *Journal of the Operational Research society* **66**(11), 1783–1792.

Calabrese, R. & Osmetti, S. A. 2013. 'Modelling small and medium enterprise loan defaults as rare events: the generalized extreme value regression model', *Journal of Applied Statistics* **40**(6), 1172–1188.

Calabrese, R., Osmetti, S. A. et al. 2011. 'Generalized extreme value regression for binary rare events data: an application to credit defaults', *Bulletin of the International Statistical Institute LXII, 58th Session of the International Statistical Institute* pp. 5631–5634.

Candès, E. J. & Sur, P. 2018. 'The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression', *arXiv preprint arXiv:1804.09753* .

Caruana, J. 2010. 'Basel iii: Towards a safer financial system', *Basel: BIS* .

Charalambous, C., Charitou, A. & Kaourou, F. 2000. 'Comparative analysis of artificial neural network models: Application in bankruptcy prediction', *Annals of operations research* **99**(1-4), 403–425.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. 2002. 'Smote: synthetic minority over-sampling technique', *Journal of artificial intelligence research* **16**, 321–357.

Chuang, C.-S. & Lai, T. L. 2000. 'Hybrid resampling methods for confidence intervals', *Statistica Sinica* pp. 1–33.

Clementi, F. 2016. Heavy-tailed distributions for agent-based economic modelling, *in* 'Economics with Heterogeneous Interacting Agents', Springer, pp. 157–190.

Cotter, J. & Dowd, K. 2006. 'Extreme spectral risk measures: An application to futures clearinghouse margin requirements', *Journal of Banking and Finance* **30**(12), 3469–3485.
URL: *http://www.sciencedirect.com/science/article/pii/S0378426606001373*

Cox, C. & Matheson, M. 2014. 'A comparison of the generalized gamma and exponentiated weibull distributions', *Statistics in medicine* **33**(21), 3772–3780.

Cox, D. R. 2018. *Analysis of binary data*, Routledge.

Danielsson, J. & De Vries, C. G. 2000. 'Value-at-risk and extreme returns', *Annales d'Économie et de Statistique no. 60* pp. 239–270.
URL: *http://www.jstor.org/stable/20076262*

Danielsson, J., James, K. R., Valenzuela, M. & Zer, I. 2016. 'Model risk of risk models', *Journal of Financial Stability* **23**, 79–91.

Davison, A. C. & Hinkley, D. V. 1997. *Confidence Intervals*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, pp. 191–255.

De Jongh, P. & Venter, J. 2015. 'A framework for normal mean variance mixture innovations with application to garth modelling', *South African Statistical Journal* **49**(2), 139–152.

Del Brio, E. B., Mora-Valencia, A. & Perote, J. 2014. 'Semi-nonparametric var forecasts for hedge funds during the recent crisis', *Physica A: Statistical Mechanics and its Applications* **401**, 330–343.
URL: *http://www.sciencedirect.com/science/article/pii/S0378437114000491*

Dembo, R. S. & Steihaug, T. 1983. 'Truncated-newtono algorithms for large-scale unconstrained optimization', *Mathematical Programming* **26**(2), 190–212.

Demidenko, E. 2001. 'Computational aspects of probit model', *Mathematical Communications* **6**(2), 233–247.

Dempster, A. P., Laird, N. M. & Rubin, D. B. 1977. 'Maximum likelihood from incomplete data via the em algorithm', *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22.

Derman, E. 1996. 'Model risk, quantitative strategies research notes', *Goldman Sachs, New York* .

Desai, V. S., Conway, D. G., Crook, J. N. & Overstreet Jr, G. A. 1997. 'Credit-scoring models in the credit-union environment using neural networks and genetic algorithms', *IMA Journal of Management Mathematics* **8**(4), 323–346.

Dey, S., Saha, M., Maiti, S. S. & Jun, C.-H. 2018. 'Bootstrap confidence intervals of generalized process capability index cpyk for lindley and power lindley distributions', *Communications in Statistics-Simulation and Computation* **47**(1), 249–262.

Diers, D., Eling, M. & Linde, M. 2013. 'Modeling parameter risk in premium risk in multi-year internal models', *The Journal of Risk Finance* **14**(3), 234–250.

Dinse, G. E. 2011. 'An em algorithm for fitting a four-parameter logistic model to binary dose-response data', *Journal of agricultural, biological, and environmental statistics* **16**(2), 221–232.

Donald, W. K. A. 2000. 'Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space', *Econometrica* **68**(2), 399–405.
URL: *http://www.jstor.org.nwulib.nwu.ac.za/stable/2999432*

Duffie, D. & Pan, J. 1997. 'An overview of value at risk', *Journal of derivatives* **4**(3), 7–49.

Efron, B. 1979. 'The 1977 rietz lecture', *The Annals of Statistics* **7**(1), 1–26.

Efron, B. & Hastie, T. 2016. *Computer age statistical inference*, Vol. 5, Cambridge University Press.

Efron, B. & Tibshirani, R. 1986. 'Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy', *Statistical science* pp. 54–75.

Efron, B. & Tibshirani, R. J. 1994. *An introduction to the bootstrap*, CRC press.

Fan, J. & Gu, J. 2003. 'Semiparametric estimation of value at risk', *The Econometrics Journal* **6**(2), 261–290.

Fan, X. & Wang, L. 1999. 'Comparing linear discriminant function with logistic regression for the two-group classification problem', *The Journal of experimental education* **67**(3), 265–286.

Flowers-Cano, R., Ortiz-Gómez, R., León-Jiménez, J., López Rivera, R. & Perera Cruz, L. 2018. 'Comparison of bootstrap confidence intervals using monte carlo simulations', *Water* **10**(2), 166.

Fontana, R., Luciano, E. & Semeraro, P. 2019. 'Model risk in credit risk', *arXiv preprint arXiv:1906.06164* .

Freue, G. V. C. 2007. 'The pitman estimator of the cauchy location parameter', *Journal of Statistical Planning and Inference* **137**(6), 1900–1913.

FSB, F. S. B. 2015. Global shadow banking monitoring report 2015, Technical report, Technical Report, November.

Garcia Trillos, C. A., Henrard, M. & Macrina, A. 2016. 'Estimation of future initial margins in a multi-curve interest rate framework', *Available at SSRN 2682727* .

Glasserman, P. & Ruiz-Mata, J. 2006. 'A comparison of approximation techniques for portfolio credit risk', *Journal of Credit Risk* **2**, 33–66.

Glasserman, P. & Xu, X. 2014. 'Robust risk measurement and model risk', *Quantitative Finance* **14**(1), 29–58.

Green, A. D. & Kenyon, C. 2015. 'Mva: initial margin valuation adjustment by replication and regression'.

Gregory, J. 2015. *The XVA Challenge : Counterparty Credit Risk, Funding, Collateral and Capital*, John Wiley & Sons, New York.
URL: *http://ebookcentral.proquest.com/lib/northwu-ebooks/detail.action?docID=4043028*

Gregory, J. 2016. 'The impact of initial margin', *Available at SSRN 2790227* .

Gro, xdf, ma, xdf & Lidan 2014. 'Liquidity and the value at risk', *Journal of Economics and Statistics* **234**(5), 572–602.
URL: *http://www.jstor.org.nwulib.nwu.ac.za/stable/24614922*

Hall, P. 1988. 'Theoretical comparison of bootstrap confidence intervals', *The Annals of Statistics* pp. 927–953.

Harrell, F. E. & Lee, K. L. 1985. 'A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality', *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences', North-Holland, New York, United States* pp. 333–343.

Hastie, T., Tibshirani, R. & Friedman, J. 2001. 'The elements of statistical learning'.

Heber, G., Lunde, A., Shephard, N. & Sheppard, K. 2018. 'Oxford-man institute's realized library, version 0.3', *Oxford-Man Institute* .

Hedegaard, E. 2011. 'How margins are set and affect asset prices', *Job Market Paper* .

Heilbron, D. C. 1994. 'Zero-altered and other regression models for count data with added zeros', *Biometrical Journal* **36**(5), 531–547.

Hinton, G. E., Sabour, S. & Frosst, N. 2018. Matrix capsules with EM routing, *in* 'International Conference on Learning Representations'.
URL: *https://openreview.net/forum?id=HJWLfGWRb*

Huang, X., Oosterlee, C. W. & Mesters, M. 2007. 'Computation of var and var contribution in the vasicek portfolio credit loss model: a comparative study', *Journal of Credit Risk* **3**(3), 75–96.

Islam, S. R., Eberle, W. & Ghafoor, S. K. 2018. 'Credit default mining using combined machine learning and heuristic approach', *arXiv preprint arXiv:1807.01176* .

Jefferys, W. H. & Berger, J. O. 1991. 'Sharpening ockham's razor on a bayesian strop', *Technical Report* .

Jorion, P. 1996. 'Risk2: Measuring the risk in value at risk', *Financial analysts journal* **52**(6), 47–56.

Kerkhof, J., Melenberg, B. & Schumacher, H. 2010. 'Model risk and capital reserves', *Journal of Banking & Finance* **34**(1), 267–279.

Khindanova, I., Rachev, S. & Schwartz, E. 2001. 'Stable modeling of value at risk', *Mathematical and Computer Modelling* **34**(9), 1223–1259.

URL: *http://www.sciencedirect.com/science/article/pii/S0895717701001297*

Kim, K. A. & Oppenheimer, H. R. 2002. 'Initial margin requirements, volatility, and the individual investor: Insights from japan', *Financial Review* **37**(1), 1–15.

Kitali, A. E., Kidando, E., Sando, T., Moses, R. & Ozguven, E. E. 2017. 'Evaluating aging pedestrian crash severity with bayesian complementary log–log model for improved prediction accuracy', *Transportation Research Record* **2659**(1), 155–163. **URL:** *https://doi.org/10.3141/2659-17*

Konečnỳ, J., Liu, J., Richtárik, P. & Takáč, M. 2016. 'Mini-batch semi-stochastic gradient descent in the proximal setting', *IEEE Journal of Selected Topics in Signal Processing* **10**(2), 242–255.

Kozeny, V. 2015. 'Genetic algorithms for credit scoring: Alternative fitness function performance comparison', *Expert Systems with Applications* **42**(6), 2998–3004.

Krishnamurthy, S. 2013. 'Ten best practices for an effective model risk management program', *Quant University White paper, available at: http://www. quantuniversity. com/EffectiveModel RiskManagement. pdf (last accessed on 25th February, 2015)* .

Lagarias, J. C., Reeds, J. A., Wright, M. H. & Wright, P. E. 1998. 'Convergence properties of the nelder–mead simplex method in low dimensions', *SIAM Journal on optimization* **9**(1), 112–147.

Laird, N., Lange, N. & Stram, D. 1987. 'Maximum likelihood computations with repeated measures: application of the em algorithm', *Journal of the American Statistical Association* **82**(397), 97–105.

Lee, T.-S. & Chen, I.-F. 2005. 'A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines', *Expert Systems with Applications* **28**(4), 743–752.

Lin, L. & Surti, J. 2015. 'Capital requirements for over-the-counter derivatives central counterparties', *Journal of Banking and Finance* **52**, 140–155.
URL: *http://www.sciencedirect.com/science/article/pii/S0378426614002829*

Lindsten, F., Wahlström, N., Svensson, A. & Schön, T. B. 2018. *Statistical Machine Learning*, Uppsala University: Lecture note - Department of Information Technology.

Lindstrom, M. J. & Bates, D. M. 1988. 'Newton—raphson and em algorithms for linear mixed-effects models for repeated-measures data', *Journal of the American Statistical Association* **83**(404), 1014–1022.

Löeffler, G. & Posch, P. N. 2011. *Credit risk modeling using Excel and VBA*, John Wiley & Sons.

Longin, F. M. 2000*a*. 'From value at risk to stress testing: The extreme value approach', *Journal of Banking & Finance* **24**(7), 1097–1130.

Longin, F. M. 2000*b*. 'From value at risk to stress testing: The extreme value approach', *Journal of Banking and Finance* **24**(7), 1097–1130.

Lou, W. 2016. 'Mva transfer pricing', *Available at SSRN 2673310* pp. 72–77.

Lyon, A. 2013. 'Why are normal distributions normal?', *The British Journal for the Philosophy of Science* **65**(3), 621–649.

Martin, J.-L. & Wu, D. 2018. 'Pedestrian fatality and impact speed squared: Cloglog modeling from french national data', *Traffic injury prevention* **19**(1), 94–101.

Mashele, H. P. 2016. *Aligning the economic capital of model risk with the strategic objectives of an enterprise*, Potchefstroom: North-West University, Master of Business Administration-dissertation.

McBride, P. M. 2010. 'The dodd-frank act and otc derivatives: The impact of mandatory central clearing on the global otc derivatives market', *The International Lawyer* **44**(4), 1077–1122.
URL: *http://www.jstor.org.nwulib.nwu.ac.za/stable/41806613*

McLachlan, G. & Krishnan, T. 2007. *The EM algorithm and extensions*, Vol. 382, John Wiley & Sons.

McNeil, A. J. 1999. 'Extreme value theory for risk managers', *Departement Mathematik ETH Zentrum* .

McNeil, A. J., Frey, R., Embrechts, P. et al. 2005. *Quantitative risk management: Concepts, techniques and tools*, Vol. 3, Princeton university press Princeton.

McNeil, A. J. & Wendin, J. P. 2007. 'Bayesian inference for generalized linear mixed models of portfolio credit risk', *Journal of Empirical Finance* **14**(2), 131–149.

Memić, D. 2015. 'Assessing credit default using logistic regression and multiple discriminant analysis: Empirical evidence from bosnia and herzegovina', *Interdisciplinary Description of Complex Systems: INDECS* **13**(1), 128–153.

Memmel, C., Gündüz, Y. & Raupach, P. 2015. 'The common drivers of default risk', *Journal of Financial Stability* **16**, 232–247.

Mileris, R. & Boguslauskas, V. 2011. 'Credit risk estimation model development process: main steps and model improvement', *Inžinerinė ekonomika* pp. 126–133.

Millar, R. B. 2011. *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB*, Vol. 111, John Wiley & Sons.

Minka, T. P. 2003. 'A comparison of numerical optimizers for logistic regression', *Microsoft Research - Unpublished draft* pp. 1–18.
URL: *https://tminka.github.io/papers/logreg/minka-logreg.pdf*

Müller, F., Zeiler, P. & Bertsche, B. 2017. 'Bootstrap monte carlo simulation of reliability and confidence level with periodical maintenance', *Forschung im Ingenieurwesen* **81**(4), 383–393.
URL: *https://doi.org/10.1007/s10010-017-0220-6*

Morini, M. 2011. *Understanding and Managing Model Risk: A practical guide for quants, traders and validators*, John Wiley & Sons.

Morris, T. P., White, I. R. & Crowther, M. J. 2019. 'Using simulation studies to evaluate statistical methods', *Statistics in medicine* **38**(11), 2074–2102.

Mullahy, J. 1986. 'Specification and testing of some modified count data models', *Journal of econometrics* **33**(3), 341–365.

Müller, M. 2012. Generalized linear models, *in* 'Handbook of Computational Statistics', Springer, pp. 681–709.

Murphy, D., Vasios, M. & Vause, N. 2014. 'An investigation into the procyclicality of risk-based initial margin models', *Bank of England Financial Stability Paper No. 29* .
URL: *http://dx.doi.org/10.2139/ssrn.2437916*

Nash, S. G. & Nocedal, J. 1991. 'A numerical study of the limited memory bfgs method and the truncated-newton method for large scale optimization', *SIAM Journal on Optimization* **1**(3), 358–372.

Nelder, J. A. & Mead, R. 1965. 'A simplex method for function minimization', *The computer journal* **7**(4), 308–313.

Nelder, J. A. & Wedderburn, R. W. 1972. 'Generalized linear models', *Journal of the Royal Statistical Society: Series A (General)* **135**(3), 370–384.

Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W. 1996. *Applied linear statistical models*, Vol. 4, Irwin Chicago.

Nocedal, J. & Wright, S. 2006. *Numerical optimization*, Springer Science & Business Media.

Norgren, C. 2010. 'The causes of the global financial crisis and their implications for supreme audit institutions', *Auditor General of the Swedish National Audit Office Stockholm* .

Noubiap, R. F. & Seidel, W. 2000. 'A minimax algorithm for constructing optimal symmetrical balanced designs for a logistic regression model', *Journal of Statistical Planning and Inference* **91**(1), 151–168.

Office of the Comptroller of the Currency, F. 2011. *Supervisory guidance on model risk management*, FRB/OCC Supervisory letter SR 11-7, Board of Governors of the Federal Reserve System, Washington, DCC.

Ohlson, J. A. 1980. 'Financial ratios and the probabilistic prediction of bankruptcy', *Journal of accounting research* pp. 109–131.

Ortiz-Gracia, L. & Masdemont, J. 2011. 'Credit risk contributions under the vasicek one-factor model: a fast wavelet expansion approximation', *Journal of Computational Finance* .

Pararai, M., Warahena-Liyanage, G. & Oluyede, B. O. 2017. 'Exponentiated power lindley–poisson distribution: Properties and applications', *Communications in Statistics-Theory and Methods* **46**(10), 4726–4755.

Pekasiewicz, D. A. 2016. 'Bootstrap estimation methods of value at risk', *WSB University in Wroclaw Research Journal* **16**(3), 123–132.

Penman, A. D. & Johnson, W. D. 2009. 'Complementary log–log regression for the estimation of covariate-adjusted prevalence ratios in the analysis of data from cross-sectional studies', *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **51**(3), 433–442.

Pérez, F. & Granger, B. E. 2007. 'IPython: a system for interactive scientific computing', *Computing in Science and Engineering* **9**(3), 21–29.
URL: *https://ipython.org*

Pohar, M., Blas, M. & Turk, S. 2004. 'Comparison of logistic regression and linear discriminant analysis: a simulation study', *Metodoloski zvezki* **1**(1), 143.

Polyak, B. T. & Juditsky, A. B. 1992. 'Acceleration of stochastic approximation by averaging', *SIAM Journal on Control and Optimization* **30**(4), 838–855.

Powell, M. 1965. 'A method for minimizing a sum of squares of non-linear functions without calculating derivatives', *The Computer Journal* **7**(4), 303–307.

Powell, M. J. 1964. 'An efficient method for finding the minimum of a function of several variables without calculating derivatives', *The computer journal* **7**(2), 155–162.

Powell, M. J. 2007. 'A view of algorithms for optimization without derivatives', *Mathematics Today-Bulletin of the Institute of Mathematics and its Applications* **43**(5), 170–174.

Premaratne, G. & Bera, A. K. 2017. 'Adjusting the tests for skewness and kurtosis for distributional misspecifications', *Communications in Statistics-Simulation and Computation* **46**(5), 3599–3613.

Rice, J. A. 2006. *Mathematical statistics and data analysis*, Cengage Learning.

Robles, V., Bielza, C., Larrañaga, P., González, S. & Ohno-Machado, L. 2008. 'Optimizing logistic regression coefficients for discrimination and calibration using estimation of distribution algorithms', *Top* **16**(2), 345.

Ruder, S. 2016. 'An overview of gradient descent optimization algorithms', *arXiv preprint arXiv:1609.04747* .

Samart, K., Jansakul, N. & Chongcheawchamnan, M. 2018. 'Exact bootstrap confidence intervals for regression coefficients in small samples', *Communications in Statistics-Simulation and Computation* **47**(10), 2953–2959.

Schiereck, D., Kiesel, F. & Kolaric, S. 2016. 'Brexit:(not) another lehman moment for banks?', *Finance Research Letters* **19**, 291–297.

Scott, J. G. & Sun, L. 2013. 'Expectation-maximization for logistic regression', *arXiv preprint arXiv:1306.0040* .

Seitshiro, M. B. 2006. *Two-sample comparisons for serially correlated data*, Potchefstroom: North-West University, Master of Science-Dissertation.

Seitshiro, M. B. & Mashele, H. P. 2020. 'Assessment of model risk due to the use of an inappropriate parameter estimator', *Cogent Economics & Finance* **8**(1), 1710970.

Shanmugam, L. & Florence, L. 2012. 'A comparison of parameter best estimation method for software reliability models', *International Journal of Software Engineering & Applications* **3**(5), 91.

Shanno, D. F. 1970. 'Conditioning of quasi-newton methods for function minimization', *Mathematics of computation* **24**(111), 647–656.

Shen, J. & He, X. 2015. 'Inference for subgroup analysis with a structured logistic-normal mixture model', *Journal of the American Statistical Association* **110**(509), 303–312.

Shi, M., Yu, X. X. & Zhang, K. 2018. 'Initial margin with risky collateral', *Journal of Risk* **20**(3), 49–68. ISI Document Delivery No.: FV7BX Times Cited: 0 Cited Reference Count: 15 Shi, Ming Yu, Xinxin Zhang, Ke 0 4 INCISIVE MEDIA LONDON J RISK.
URL: <GotoISI>://WOS:000424737500004

Shi, X., Tang, Q. & Yuan, Z. 2017. 'A limit distribution of credit portfolio losses with low default probabilities', *Insurance: Mathematics and Economics* **73**, 156–167.

Shirreff, D. 1999. 'Lessons from the collapse of hedge fund, long-term capital management', *Case Study des IFCI Risk Institute* .

Siegl, T. & West, A. 2001. 'Statistical bootstrapping methods in var calculation', *Applied Mathematical Finance* **8**(3), 167–181.

Silvapulle, M. J. 1981. 'On the existence of maximum likelihood estimators for the binomial response models', *Journal of the Royal Statistical Society. Series B (Methodological)* **43**(3), 310–313.
URL: *http://www.jstor.org/stable/2984941*

Singh, M. M. 2010. *Collateral, netting and systemic risk in the OTC derivatives market*, International Monetary Fund - Working paper.

Swanepoel, J. W. H. & De Beer, C. F. 1993. 'A modified bootstrap estimator for the mean of an asymmetric distribution', *The Canadian Journal of Statistics* **21**(1), 79–87.
URL: *http://www.jstor.org.nwulib.nwu.ac.za/stable/3315660*

Szegö, G. 2002. 'Measures of risk', *Journal of Banking & finance* **26**(7), 1253–1272.

Tanner, M. A. & Wong, W. H. 1987. 'The calculation of posterior distributions by data augmentation', *Journal of the American statistical Association* **82**(398), 528–540.

Tunaru, R. 2015. *Model risk in financial markets: From financial engineering to risk management*, World Scientific.

Tunaru, R. et al. 2015. 'Model risk in financial markets: From financial engineering to risk management', *World Scientific Books* .

Vasicek, O. 2002. 'The distribution of loan portfolio value', *Risk* **15**(12), 160–162.

Vetterling, W. T., Teukolsky, S. A., Press, W. H. & Flannery, B. P. 1992. *Numerical recipes: the art of scientific computing.*, Vol. 2, Cambridge university press Cambridge.

Wackerly, D., Mendenhall, W. & Scheaffer, R. L. 2014. *Mathematical statistics with applications*, Cengage Learning.

Wang, H., Zhu, R. & Ma, P. 2018. 'Optimal subsampling for large sample logistic regression', *Journal of the American Statistical Association* **113**(522), 829–844.

West, G. 2004. 'Risk measurement for financial institutions', *Financial Modelling Agency* .

Wooldridge, P. 2016. 'Central clearing predominates in otc interest rate derivatives markets', *BIS Quarterly Review* pp. 22–26.

Wu, D. & Olson, D. L. 2010. 'Enterprise risk management: coping with model risk in a large bank', *Journal of the Operational Research Society* **61**(2), 179–190.

Yang, S., Zhang, H. et al. 2018. 'Comparison of several data mining methods in credit card default prediction', *Intelligent Information Management* **10**(05), 115.

Yang, Y., Brown, T., Moran, B., Wang, X., Pan, Q. & Qin, Y. 2016. A comparison of iteratively reweighted least squares and kalman filter with em in measurement error covariance estimation, *in* '2016 19th International Conference on Information Fusion (FUSION)', IEEE, pp. 286–291.

Yeh, I.-C. & Lien, C.-h. 2009. 'The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients', *Expert Systems with Applications* **36**(2), 2473–2480.