



DSPACE

<https://dspace.org/>

A new internal data measure for operational risk: a case study of a South African bank

Hooхло, Mphekeleli

2015-03-12

<http://hdl.handle.net/10539/17271>

A New Internal Data Measure For Operational Risk: A Case Study Of A South African Bank

Mphekeleli Hoohlo

Supervisor: Professor Eric Schaling

A dissertation submitted in partial fulfillment of the requirements
for the degree of Master of Management in Finance and Investment, in the
School of Law, Commerce and Management
University of Witwatersrand,
Johannesburg, South Africa



November 13, 2014

ABSTRACT

This study examines the effect of automation on operational risk (**OpRisk**) measurement in a South African bank. It uses historical process risk loss data from the first quarter (2013Q1) derived from the automated trade amendment tracker (**ATAT**) database – a computerised tool designed to automate the collection of internally generated **OpRisk** events at the bank in question. The results indicate that a Value-at-Risk (VaR) estimate for **OpRisk** largely depends on the accuracy of the loss data. Capital adequacy is determined using this estimate of VaR, suggesting that the **ATAT** device used in operational risk measurement improves on investment services activity in South Africa. Finally, it appears that risk management practices in the South African banking industry are more concerned about traditional operational risk management (ORM) rather than the determination of **OpRisk** VaR as it becomes a matter of great concern for financial institutions (FI's) across the globe.

DECLARATION

I, Mphekeleli Hoohlo declare that this report is my own, unaided work. It is submitted in partial fulfillment of the requirements for the degree of Master of Management in Finance and Investment at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Mphekeleli Hoohlo

November 13, 2014

DEDICATION

In memory of my mother

Sophy Mafini Mamphekeleli Hoohlo

1947 - 2014

ACKNOWLEDGMENTS

I would like to express my sincere thanks to my supervisor, Eric Schaling, for his time and patience, my appreciation to my parents, Setsomi and Sophy, for their tireless assistance, encouragement, and insight. I would also like to thank everyone at CIBW MO who made this research report possible.

ABBREVIATIONS

AIC	Akaike Information Criterion
AMA	Advanced Measurement Approach
ATAT	Automated Trade Amendment Tracker
BIA	Basic Indicator Approach
Basel Committee	Basel Committee of Banking Supervision
Basel I	The Capital Adequacy Accord
Basel II	The New Capital Adequacy Accord
BL	Business Line
BO	Back Office
c.d.f.	Cumulative distribution Function
CIBW	Corporate Investment Banking and Wealth
d.f.	Distribution Function
EC	Economic Capital
ET	Event Type
FI	Financial Institution
FO	Front Office
FX	Foreign Exchange

HQIC	Hannan–Quinn Information Criterion
IB	Investment Banking
i.i.d.	Independent and Identically Distributed
IMA	Internal Measurement Approach
KRI	Key Risk Indicator
LDA	Loss Distribution Approach
m.l.e	Maximum Likelihood Estimate
MO	Middle Office
NDA	Non–disclosure Agreement
OpRisk	Operational Risk
ORX	Operational RiskData Exchange
p.d.f.	Probability Distribution Function
P&L	Profit and Loss
RC	Regulatory Capital
SA	Standardised Approach
SIC	Schwartz Bayes Information Criterion
VaR	Value–at–Risk

CONTENTS

<i>Abstract</i>	i
<i>Declaration</i>	ii
<i>Dedication</i>	iii
<i>Acknowledgments</i>	iv
<i>Abbreviations</i>	v
<i>1. INTRODUCTION</i>	5
1.1 Purpose of the Study	5
1.2 Context of the Study	8
1.3 Problem statement	16
1.3.1 Main Problem	16
1.4 Significance of the Study	16
1.5 Delimitations of the Study	21
<i>2. LITERATURE REVIEW</i>	23
2.1 Introduction	23
2.2 Definition of Operational Risk	25

2.3	Capital Adequacy	26
2.3.1	A Synthesis of Concepts	26
2.3.2	Theoretical Framework	27
2.3.3	Hypothesis 1	29
2.4	Conclusion of Literature Review	29
3.	<i>METHODOLOGY</i>	32
3.1	Research Paradigm	32
3.2	Research Design	33
3.2.1	Computing the Frequency Distribution	34
3.2.2	Computing the Severity Distribution	35
3.2.3	Formal Results	37
3.2.4	Dependence Effects (Copulae)	38
3.3	Population and Sample	40
3.3.1	Population	40
3.3.2	Sample and Sampling method	41
3.4	The Research Instrument	44
3.5	Procedure for data collection	46
4.	<i>RESULTS</i>	48
4.1	Data Analysis and Interpretation	48
4.1.1	Fitting Distributions to Data	49
4.1.2	Information criterion	52
4.1.3	Goodness-of-fit tests	53

4.1.4	Further manipulation of data	54
4.1.5	Distribution Fitting for VaR	55
4.1.6	The Generalised Extreme Value Distribution	59
4.1.7	VaR Analysis	67
5.	<i>Discussion</i>	69
5.1	Limitations of the Study	69
5.2	Validity and Reliability	70
5.3	Conclusion	72
5.4	Future Research	74
	<i>References</i>	76
	<i>Appendices</i>	80
A.	<i>Goodness-Of-Fit Plots and Statistics</i>	81
A.1	Fitting a Discrete Parametric Distribution to BBL Loss Data: Geometric Variate $\mathbf{G} : p, p = 0.4179105$	82
A.2	Fitting a Discrete Parametric Distribution to TBL Loss Data: Polya Variate $Polya(\alpha, \beta), \alpha = 6.557774, \beta = 0.7924074$	83
A.3	Fitting a Continuous Parametric Distribution to BBL Loss Data: Lognormal $(\alpha, \beta), \alpha = 411736.7, \beta = 2271727$	84
A.4	Fitting a Continuous Parametric Distribution to TBL Loss Data: Lognormal $(\alpha, \beta), \alpha = 1043960, \beta = 6121955$	85

A.5	Displaying information criterion for the optimally ranked (according to SIC) estimated best fit frequency and severity distributions	86
B.	<i>Descriptive Statistics Tables</i>	87
B.1	Table B.1 generated by Excel2LaTeX from excel sheet '1.Statistics'	87
B.2	Table B.2 generated by Excel2LaTeX from sheet '1.Statistics'	87
C.	<i>Miscellaneous figures</i>	90
C.1	A sample ATAT database	90
C.2	Data request letter	90
C.3	The Aggregate Multi Monte Carlo Loss Model for Process Risk	90
C.4	Raw Data for Process Risk	90

LIST OF FIGURES

1.1	Sample process diagram	17
2.1	The LDA model	28
3.1	A fixed income blotter template	44
3.2	Developing an OR filter	45
4.1	Vose software	49
4.2	Scatter Plot	50
4.3	Histograms of process risk loss events: 02 Jan 2013 – 20 March 2013	55
4.4	Graphs of process risk loss events: 02 Jan 2013 – 20 March 2013	56
4.5	Combining the Frequency and Severity Distributions	58
4.6	Estimating the Aggregate Loss Distribution	61
4.7	GEV Density Function	63
4.8	GEV Cumulative Probability Distribution	64
4.9	GEV P–P Plot	65
4.10	GEV Q–Q Plot	66
5.1	Statistical analysis of data by ATAT	72
5.2	Root cause analysis of data by ATAT	72

A.1	Geometric Distribution	82
A.2	Polya Distribution	83
A.3	Lognormal Distribution	84
A.4	Lognormal Distribution	85
A.5	Information Criteria	86
C.1	Letter of motivation in support of data request	90
C.2	The automated trade amendment tracker (ATAT) loss database platform . . .	91
C.3	Model–Process–Risk.xlsx	91
C.4	Process–Risk–Data.xlsx	91

LIST OF TABLES

3.1	The BL/ET Matrix for SA Bank	41
3.2	The BL/ET Matrix for a SA Investment Bank	42
B.1	Descriptive Statistics of Input Distributions	88
B.2	Descriptive Statistics of Output Distributions	89

Chapter 1

INTRODUCTION

1.1 Purpose of the Study

The advent of the computer in the 1960s, worldwide, brought about the biggest change in banking since ledgers were mechanised, and set off a revolution in financial products/service offerings that still continues today. Automation via computers emerged a decade later in South Africa. The longer it took for computers to arrive in South Africa, the more there was to gain from the experience of others and from advances made in computer construction and techniques.

In recent years it has turned out that many phenomena in finance can be described successfully by mathematical models. A model is a small replica of real-life; through modeling we attempt to mimic a real-life scenario to a given degree of certainty, subject to limitations and/or given assumptions. The assumptions must make sense, if not, nobody will use it (the model). Take the use of normal distributions in modeling the behaviour of stock prices in financial markets which turned out to be a big problem as stock prices can rarely be predicted by the models used to forecast their future prices.

Computers are now an integral part of the banking industry due to their power and ability to

automate processes. The more sophisticated and complex the industry becomes, the greater the need for sophisticated computer hardware and software packages to handle the day to day running of a financial institution (FI) such as a bank. In this study, historical operational loss data is taken, and through the use of mathematics and the computer, frequency and severity distributions are fitted to the data, then summed up or aggregated to find maximum loss due to operational risk (**OpRisk**) at a South African bank. It goes without saying that this procedure is easier said than done.

More often than not there is no exact analytic solution representing the total aggregate loss distribution. Numerical approximation techniques (computer algorithms) successfully bridge the divide between theory and implementation for the problems of mathematical analysis. It has been amply documented that the Monte Carlo¹method, Panjer's recursive approach, and the inverse of the characteristic function can be used to provide an approximation of the compound loss distribution (Franchot et al (2001); Aue & Kalkbrener (2007); Panjer (2001); & others). Monte Carlo simulation is popular in the literature, and has been proposed in this study as it has received most attention.

Broadly speaking, banking operations consist of traditional concepts such as, authorisation to commit a firms money, protection of assets, complete and accurate recording of transactions, efficient due diligence operations when opening new relationships, orderly and timely processing of and clearance of transactions and reconciliation of individual trade details to a firms records (King, 1999). A FI relying on simplistic and traditional techniques in managing their operations stands the risk of the potential loss of competitiveness in the marketplace

¹ A random sampling method whereby computer software simulation packages generate random values representing samples from a theoretical distribution

due to inadequate technology, both for maintaining the bank and servicing its customers. Also this risk is often associated with the need to protect systems and the data contained in them from unauthorised access and tampering. In fact this area of risk control could set one bank apart from the rest; effective maintenance of banking procedures could unleash optimal performance in teams and individuals whilst business objectives are achieved, realising the full potential of the organisation. An imbalance of skills in the labour force and adequate technological advances can lead to a breakdown in banking services activity.

It is important to note that the risks associated with operational errors are wide ranging and may or may not bear profit and loss (P&L) implications. Ultimately, we want to measure the impact of operational errors upon an FI's P&L, in a FI these risks are always initiated at the dealing phase. Figure 1.1 is a detailed diagram of the operational tasks, controls and reports performed throughout the life cycle of a deal/trade in an typical investment banking process. The reporting lines from deal origination to settlement are formed through correct relationships between the staff and duties of front/middle/back office. In our study the middle office (MO)/back office (BO) will to a large extent be instrumental in the process of managing the risk, hence are only involved after the risk has been assumed, therefore front office (FO)/dealing staff is responsible for operational events to be included in the classification of the losses that directly affect the P&L, such as booking errors and/or some but not all, system problems.

Computers have to a great degree removed the extent to which errors are generated through the banks operations, and enabled organisations to efficiently meet and adapt to internal operational risk practices as well as external regulations. However, they have also opened up new ways in which errors can be generated, with arguably more severe downstream effects

than in the past, such as in the potential of accumulated losses during system downtime or in the generation of multitudes of **P&L** losses when unauthorised trades are booked due to computer bugs, notably observed in modern arbitrage trading platforms. A computer trading error was revealed in March 1997 in an investigation led by the Securities and Futures Authority in the UK (the former City of London Regulator, since superseded by the Financial Services Authority) in rogue trading in a program trade at SBC Warburg. (A program trade is a transaction where one agent, generally a fund, choses another agent, generally a bank or a broker, to sell part of its shares in the market at a day and hour determined by market prices). The program trading error that made SBC Warburg the subject of the investigation is thought to have cost it no more than £5mil (Cruz, 2002, pg. 21).

Downstream effects of the post automation phase differ in magnitude to those in the past, in that there was less computation then and the processing of information leading to operational losses were carried out manually.

1.2 Context of the Study

Since the mid-1980s, there have been some spectacular losses in financial markets. Of all risks associated with these losses, operational risk (**OpRisk**) can be the most devastating and at the same time, the most difficult to anticipate. Managerial and regulatory focus on **OpRisk** has been heightened, following a number of very costly and highly publicized operational events (Cummings, Christopher & Ran, 2006).

The most prominent examples of losses due to **OpRisk** have been “rogue trader ”cases. These are thought to be intelligent, articulate and charming individuals, but may prove to

be criminally minded insiders acting on their own, who stand out due to their extreme past and/or on-going successes, enabling them to engage in their underhanded practices for years without raising any suspicion. When these losses materialise, the appearance can result in sudden and dramatic reductions in the value of the firm. Examples of famous losses (representing a broad range of losses) below contextualise rogue trading activities and the impact of the realised losses on the FI's, therefore giving impetus to the need of containing OpRisk.

- * Toshihide Iguchi, a star bond trader at Daiwa Bank confessed in 1995, that he had forged more than 30,000 trading slips which made it appear as though he was generating half of his employers trading profits. In his confession, the reality was that he was hiding more than \$1,1bn in losses accumulated over a decade.
- * In the same year, Nick Leeson of Barings Bank, made a name for himself by taking bad bets and parking his losses in a fabricated separate account. By February, this account was hiding losses of \$827m. Systems within Barings were so inadequate that nobody knew what he was doing. This eventually wiped out the 200-year-old British merchant bank due to an accumulated total trading loss of \$1.4bn.
- * A year later, Yasuo Hamanaka, Sumitomo Corporation's chief copper trader was caught hiding losses through unauthorised trades after he tried to manipulate the market. His trades helped send the copper price to record highs at the time, only for prices to tumble down once his secret was uncovered. His was the biggest trading loss at the time amounting to \$2.6bn.
- * John Rusnak, a US-based foreign exchange trader for Allied Irish Bank lost money to

the bank. Upon realizing his loss and due to fear of losing his reputation as a solid performer, he spent five years trying to cover his losses by creating fictitious trades. He would end up costing the lender \$700m when his unauthorised speculative activities were exposed in 2002.

- * Vince Ficarra and David Bullen of the National Australian Bank, were in 2003 found guilty of unauthorised trading and falsely inflating profits. Their team had broken their trading limits on about 800 occasions and the two had been found to have falsely inflated profits just minutes before the end of the bank's financial year enabling them to trigger bonuses.
- * Jerome Kerviel began in Societe Generale's MO and was well versed in the company policies in approving and regulating trading among its brokers. He was soon promoted to trading. He had extensive knowledge of the computers and systems used in the company. Through this knowledge, he conducted tens of billions of dollars' worth of unauthorised trades. He hid his early gains by creating fake offsetting trades in the computers systems and logs. When the fraud appeared and managers tried to close out the uncovered bets in January 2008, he was to cost SocGen a record \$7.2bn.
- * Kweku Adoboli, an ETF trader at UBS began covering up his trading losses in 2008 using his knowledge of the back office to sidestep compliance controls. When he made a \$400,000 loss on a trade, he opted to hide his loss rather than tell his manager. By June 2011 he had racked up \$2.3bn in trading losses which he confessed to via email sending shock waves through the bank.

The key lesson to be learned from these losses is the importance of internal controls. A trader in a FI should be allowed to take positions on the future direction of relevant market variables. However, Hull (2012) correctly argues that the sizes of the positions that can be taken should be limited and the systems in place should accurately report the risks being taken. The risks taken by traders, the models used, and the amounts of different types of business done, should all be controlled.

It is essential that all FI's define in a clear and unambiguous way, limits to the financial risks that can be taken and it is equally important to have procedures and practices that ensure that the limits are adhered to. Limits are usually established on the basis of the amounts of exposure, measured in *value-at-risk* (VaR), that a FI is prepared to risk for, say the amount of foreign exchange (FX) paper allowed to be held over a certain period.

Suppose a trader sees an arbitrage opportunity (i.e. an opportunity to enter into a contract at no cost, no chance of making a loss and at a greater than or zero probability of making a profit) on a 1 year forward on the price of gold and the annual interest charged on borrowed funds (suppose the cost of borrowed funds is much lower than the return on the 1 year forward). Suppose it were such a profitable opportunity that he/she wanted to buy as many ounces of gold as possible and go short futures contracts on the same number of ounces, realising a guaranteed profit in a years' time. The trader will have borrowing limits based on the trading book i.e. the amount of borrowed funds.

The opportunity may appear to be so attractive that he may be tempted to exceed his daily borrowing limit in order to realise a considerable profit. If the trader were to exceed this limit without being given prior approval to do so, their actions would effectively constitute a breach of company financial policy and hence a potential operational loss event. Warning signs

such as the exceeding of risk limits or considerable increases in profit should be investigated alongside each other. An automated **OpRisk** framework should be able to flag this issue, providing management with an early warning signal underscoring those areas where pre-defined limits are breached and thus highlighting potential operational (e.g. rogue trader) losses in a timely fashion.

FI's operating as intermediaries, such as those in the banking industry, are special because of their ability to efficiently transform financial claims of household savers into claims issued to corporations, individuals, and governments. A FI's ability to evaluate information and to control and monitor borrowers, allows it to transform claims at the lowest possible cost to all parties. This comes at cost in the form of risk. One of the specific types of risk is based on the creditworthiness = credit risk of the borrower: It is the risk that the borrower may default on repayment of a loan, otherwise known as a credit default. Another risk type of major concern is market risk: The risk related to the uncertainty of a FI's earnings on its trading portfolio caused by changes in market conditions, such as the price of an asset, interest rates, market volatility, and market liquidity.

Large operational errors highlighted by "rogue trader" events, brought awareness of risks that significantly impact the performance of a FI other than credit and market risks. A recent survey (Cruz, 2002) showed that banks estimate their risks are divided into market risk (50%), credit risk (15%) and operational risk (35%). A common misconception amongst risk professionals is to take the complement of losses attributed to market and credit risk and regard them as losses resulting from **OpRisk**, stating: "it is every risk not included in market and credit risk". This definition is too broad and makes the task of creating an **OpRisk** database unmanageable.

According to the Basel Committee on Banking Supervision (2001) a universally agreed definition of operational risk is: “**The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events**”. Therefore, OpRisk is related to losses originated from operational errors of any sort that affect the P&L of the FI.

Any type of FI will often face OpRisk long before it embarks on its first market trade or credit transaction. Practitioners of risk management have attempted to identify and measure OpRisk in the past. One of the more common approaches has been to try to apply techniques from areas of credit and market risk to the problem of fitting and modeling OpRisk, with limited degrees of success. One of the reasons for this is the general lack of publicly available data for operational losses, in direct contrast to market and credit risk, for which data of publicly traded companies is widely available.

Considering the inherent scarcity of OpRisk data, qualitative aspects such as the generation of scenarios, the development of a program for key risk indicator’s (KRI’s) and/or expert judgement of reasonableness have been proposed to make up for the shortfall in OpRisk databases. This, besides the fact that several banks have developed state-of-the-art market and credit risk measurement techniques which are quantitative in nature, and clearly more superior to qualitative techniques.

It is important to note, as illustrated by Cruz (2002), that when a dealer incorrectly states a valuation parameter such as the volatility of a deal, the reason behind the operational event is known, which is incorporated in calculating the risk, measured in VaR, but the qualitative reason will not be incorporated into the measure, except as a technical note. Not all operational risk loss data lends itself easily to quantitative analysis (Chavez–Demoulin,

Embrechts & Neslehova, 2006). For example, if a client were to open a legal suit against the bank, who upon investigation find that the root cause of risk results from lack of experience of staff in dealing with certain products, i.e. legal risk defies a precise quantitative analysis. The qualitative approach by nature lacks an effective relationship for control, and has a high degree of subjectivity. The study focuses on the quantitative aspects of modeling **OpRisk** using internal loss data alone as the objective is just to arrive at an operational **VaR** measure, hence qualitative aspects are beyond our scope.

The credit and financial crisis of 2007–2009 has focused attention on current practices of risk measurement and management in financial markets. Regulatory authorities throughout the world have adopted **VaR** to measure the exposure faced by FI's (Lehman, Groenendaal, & Nolder, 2012). The Basel Committee on Banking Supervision (Basel Committee) is one such authority: It is comprised of a group of officials who collectively form a regulatory structure in the financial services. Their role is to set out guidelines on the amount of regulatory capital required of financial companies to form capital reserves to cover risks in the financial services sector. The Capital Adequacy Accord (**Basel I**) was implemented in 1988, and an update to it came into effect in June 2006. The New Capital Adequacy Accord (**Basel II**) focused primarily on the definition of risk-weighted assets, and introduced a framework of risk management techniques with specific emphasis on **OpRisk**, as a portion of the requirements to the accord, which specifically deal with credit and market risks, were already addressed in **Basel I**.

The three pillar approach was introduced with **Basel II**. Pillar I is a minimum capital requirement, calculated in different ways for institutions of differing sophistication. **OpRisk** modeling is of concern here; central to this and to the content of this study is the perceived

triumph of complex quantitative procedures over respect for and reliance on seasoned professional judgement. Exciting new approaches to **OpRisk** have been introduced in the early days of **Basel II**. The first formal system adopted by the Basel Committee required certain institutions to calculate a 99.9th percentile, left-tail confidence interval in deriving the **VaR** measure. The advanced measurement approach (AMA) had been thought to be the more advanced. Management is mostly free to choose among various methods for estimating this, including historical simulations and Monte Carlo simulations. The idea was that the most complex banks would adopt the AMA, which would then provide an incentive for less sophisticated banks to follow suit and make progress to improve their risk management. Capital adequacy is determined using this estimation of **VaR**.

Pillar II is a supervisory review process through which additional capital requirements can be imposed and Pillar III relates to market discipline and assumes transparency in risk disclosures will help keep banks in line by enabling investors to reward or punish FI's on the basis of their risk profile.

This is a study toward the determination of whether the current regulatory scrutiny on **OpRisk** is justified based on the use of a quantitative approach to measure **OpRisk** where an internal **OpRisk** management database, namely the automated trade amendment tracker (**ATAT**), is used to collect internal loss data in a South African bank. The idea of using a computerised device to obtain an automated data feed to collect operational loss data is central in determining the accuracy of **OpRisk** in that considerable challenges exist, which are reduced due to automation, in collating large volumes of data into a central depository. The development of a risk management platform which has the ability to capture loss data with a high degree of reliability/accuracy can be decisive in making an **OpRisk** project successful.

1.3 Problem statement

1.3.1 Main Problem

Although there are processes within banks whose objectives are to put detailed procedures and practices in place, through which the risks of the business are kept within acceptable bounds; operational events leading to unexpected losses regularly occur, defeating the very objectives of the internal control frameworks set by senior management necessary to send an accurate message to staff.

Furthermore, the foundation of internal control rests on managements vision of the importance of controls. Without a proper definition and tone from the top it is difficult and perhaps even impossible to set up a good control structure. **OpRisk** is the oldest type of risk facing FI's and is a key source of risk, leading to a breakdown in investment services activity.

1.4 Significance of the Study

The majority of banking institutions in South Africa rely on the manual capture of records for the tracking of trade adjustments to treasury systems. The records are usually captured in the form of an email, telephone call, or by verbal instruction which can lead to less than ideal choices of managing **OpRisk**. This seems to be quite outrageous! Even more surprising is that automated tracking devices are not common. It is clear that using an email to track records becomes a cumbersome process as the number of records increases, not to mention through telephone calls or verbal instruction. The most preferred manner in which to track these records is clearly not being prioritised.

Originally, MO/BO would have been responsible for the deal capture of all details as automated FO systems did not exist. Thus, once the dealer had manually entered the basic data, MO/BO staff would manually enter all the other details. Today many banks have complete software solutions to capturing the life of a deal, hence deal capture is executed in the FO via the FO system. Some of the FO systems are so advanced and sophisticated that the system can be trusted to capture the whole life cycle of the deal. Figure 1.1 is a detailed diagram of process tasks and controls used to compile a list of control points and related risks in the trading process.

Trading in financial securities in any financial market is carried out in the FO. Here the deal is originated and captured in the FO trade entry systems by FO staff to reflect what has been agreed. The purpose of the MO/BO is to ensure that the trade details are checked and confirmed to ensure accuracy and validity. The proper application and implementation of MO duties is to ensure that accurate P&L numbers are generated, reported and agreed with FO. In cases where the MO/BO experiences (P&L) issues relating to incompleteness or inaccuracies in a deal as a result of a mismatch between the trade booked (booking in trade feed) and the details agreed by the trader, a query arises. MO/BO should be able to respond quickly to communicate the query, then be able to go “under the bonnet ” of the querying process and fix the problem.

Suppose a deal is booked in the trade entry system and on matching the trade details and P&L reports, MO discover that the trade had been incorrectly booked, say the notional is booked the wrong way round (i.e. booked on the sell side when records show that it should have been booked on the buy side) but somehow managed to slip through the preventative web of internal controls. The actual deal P&L reported in the transaction data system will be

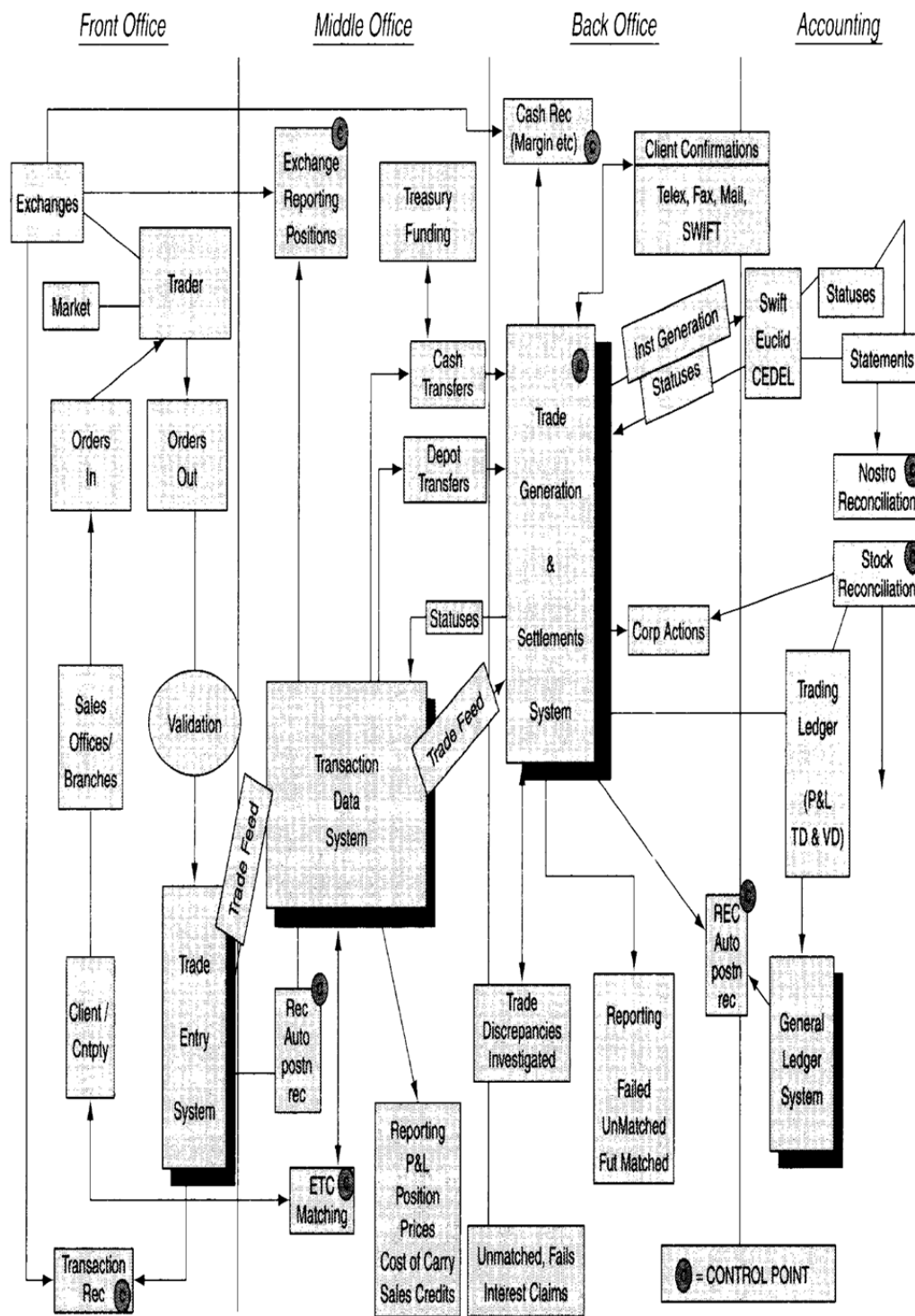


Fig. 1.1: Process diagram for investment banking operations

(King, J, 2001)

inverted, i.e. if the P&L of the deal reflects a profit of \$100 in the system, it should in reality be reflecting a loss of \$100 and vice versa, a \$200 discrepancy!. This poses a significant potential loss to the bank that will be keen to see the discrepancy resolved timeously. However, it is imperative that control procedures are always adhered to and that contact between FO and MO/BO is strictly in accordance with established lines of communication.

The trade will need to be fixed by “amending ”or manually changing the notional on the FO system from a “sell”to a “ buy”(in forward FX swaps, amendments have to very carefully monitored, for if the first leg had already matured and the swap details are subsequently amended, the true P&L will be affected). Once the trade is amended (manually fixed), a record of the amendment is stored and precautions are taken to avoid the risk of unauthorised payments being made. The trade generation and settlement system is the last line of control there to protect the bank, its staff and its clients generally.

Records of trade amendments reflect the chain of events leading up to and on executing the change to trade details prior to booking. The record of the amendment usually consists of an instruction through an approved line of communication, e.g. telephone call or email, and a tracking log which describes the change to the trade in question. An increase in the volume of trade amendments involves many individual records which need to be manually recorded by MO. Keeping track of these changes quickly turns into a nightmare and the accuracy of amendment records is easily compromised. Automated recording of the process of tracking of trade amendment data will benefit from a computerised technology that monitors risks associated with manual processing errors.

This study fills a gap in that it focuses on exploring the relevance of automating the function of the tracking of changes to trade details of deals captured in FO systems, otherwise called

trade amendments, whilst the FO system itself is adjusted for life cycle events throughout the life of a deal. In the process, we hope to uncover sources of operational errors and build up a database of processing errors large enough to be used in advanced techniques of measuring the levels of **OpRisk** with respect to a set of deals. The method of data collection will consequently prove that loss data is flowing with a high degree of reliability and accuracy, as the process is automated and the data feed comes directly from the source system. Interestingly enough, the integrity (i.e. reliability and accuracy) of internal data is central in providing the basis for a good quantitative measure of ex ante **OpRisk**, which in turn, determines capital allocation for expected operational events.

This study will provide guidance as it will demonstrate the ability to root out and record sources of operational errors in order to assess components leading to “best practice ”in **OpRisk** control. The **ATAT** is a tool which is used to automate the manual processing of amendment records and monitors processing risks in one robust software architecture. Through designing and implementing an **ATAT**, it will be possible to demonstrate the relevance of this tool, which comes with all the benefits of manual tracking systems where reports on details of amended deals were carried out manually, and has further applications unique to the tool itself.

Reports to management on details of amended trades is a good control. The message from senior management and the board of directors through written policies and training programmes should set up a good control structure for an organisation. Banking supervision is a major part of the industry which must be observed by all stakeholders. The message set from the top is only received as far as the reports to management show, hence the quality of these reports inform bank supervisors whether or not bank managers comply with their

mandate; mandates are documented along with their implications for OpRisk. King, J (2001, pg. 37) found that the most important document relating to OpRisk at the time was the *The Framework for the Evaluation of Internal Control Systems* (January 1998) wherein a mandate for bank supervision was given.

“Principle 14: Supervisors should require that all banks, regardless of size, have an effective system of internal controls that is consistent with the nature, complexity, and risk of their on-and-off-balance sheet activities and that responds to changes in the bank’s environment and conditions, in those instances where supervisors determine that a bank’s internal control system is not adequate (for example, does not cover all the principle contained in this document), they should take action against the bank to ensure that the internal control system is improved immediately.”

The ATAT is hopefully a step in the right direction at ensuring that the lead set from top management is adhered to and fraudulent activities quickly uncovered where control philosophies fall short, or are inconsistent with company policy. Finally, the concept of monitoring a treasury system through an automated tracking device is a need in the market as amongst other things it ensures that systems of internal control are operating as intended - “quis custodiet custodes? ”(or who guards the guardians?) (King, 1999). The project will also give an indication of how the market is changed through innovation, and the impact of information technology on the speed of doing business.

1.5 Delimitations of the Study

Owing to the multidimensionality of **OpRisk**, loss events are categorised into seven event-type categories (Internal Fraud, External Fraud, Employment Practices and Workplace Safety, Business disruption and System failures, Execution, Delivery & Process management, Clients, Products and Business practices & Damage to physical assets) (Cruz, Coleman & Salkin, 1998).

Process risk is defined as: **“Losses resulting from failed transaction processing or process management, from relations with trade counterparties and vendors (Cummings et al, 2006), such as booking errors, failure to comply with the transaction terms, fraud, etc.”**. Again the ATAT records processing error event types, such as amendments to incorrectly booked interest rates, start/end/value dates, or deal cancellations, to mention a few. Our study attempts to measure **OpRisk** losses under the process risk category only. The author argues that it is a convenient loss category in our case as data captured via the treasury system is process orientated and impartial; automation provides an electronic trail which facilitates the ease of data capture. According to ORX (2012), the 2012 edition of the Operational Riskdata Exchange (ORX) report on operational risk loss data findings from the 2006–2010 and 2011 surveys indicate that the process risk category consistently accounts for the majority of operational problems, i.e. 35% of the time, followed by cases of external fraud at a close 30%. It is therefore a convenient and useful risk category to select as a sample as it tends to present the largest number of events and subscribes well to the style used in the loss data extraction method ATAT platform, which will be used in experimentation.

The database of historical data loss events is built up over a 3 month (beginning January 2013) time span. The minimum acceptable historical observation period, proposed by the EU directive [2000/12/EC] requires:

“Internally generated operational risk measures used for regulatory capital calculation must be based on a minimum historical observation of 5 years for internal loss data. When a credit institution moves to an AMA, a three year data series is acceptable ”.

It suffices to say that the available data in our case falls short of this requirement. However, our aim is to demonstrate that the treatment of available data lends itself to the AMA method of calculation of **VaR** owing to the method of data collection, i.e. through the use of the **ATAT**. The sample is drawn from operational events often related to the internal process and may suffer from data deficiencies, such as data missing from business lines (**BL**) and/or event types (**ET**).

Chapter 2

LITERATURE REVIEW

2.1 Introduction

Pre-1988, banks were regulated using balance sheet measures often applying unique metrics and thresholds to inform decision making. The process required to identify the appropriate metric or portfolio of metrics can be challenging, given the unique attributes of each bank. Definitions of terms and required financial ratios varied due to differing regulatory enforcement practices from country to country and many banks find themselves falling into the trap of selecting the most common, rather than the most appropriate risk metric. A common debate focuses on economic capital (EC) and risk adjusted return on capital (RAROC), or FI's that are heavily engaged in traditional lending activities could consider return on asset (ROA) as the performance measure rather than the more common return on equity (ROE), believed to be a more useful indicator (KPMG International, 2013). Clearly, there is no worldwide response in finding the right divide between debt and equity used in project financing as these may differ from one region to another. In practice, what the results show is that the higher the debt the greater the chances of good return on equity but also a higher likelihood of solvency problems, i.e. due to interest expenses on earnings

or depreciation costs over the passage of time.

Bank leverage increased in the 1980s and the Bank of International Standards (BIS) was set up. The Capital Adequacy Accord (**Basel I**) was established so that banks could buffer against international solvency. **Basel I** intended to open up global competition by minimising rules that favoured local FI's over potential foreign competitors. **Basel I** meant that banks were required to keep a portion of the depositors money of what has been lent out to protect against credit default. According to **Basel I**, it is mandatory to comply with an industry standard in calculating the required economic capital buffer for credit risk.

With off-balance sheet trading on the increase, bank trading in equity, interest rate and exchange rate derivative products escalated throughout the 1990's. Many of these new innovations (later known to be derivatives) were not even in existence when **Basel I** was drafted. Consequently, even if **Basel I** was satisfactory in safeguarding bank depositors from traditional credit risks, their capital adequacy requirements were not sufficient to safeguard against the market risk from derivatives trading (Eun, Resnick & Sabherwal, 2012).

An amendment to **Basel I** was introduced to cater for these shortcomings which came into effect in 1998. Regulation dictates that under this amendment to **Basel I**, to ward against engaging in market activities which are prone to market risk, banks were to set aside additional capital to guard against market risk.

OpRisk, which includes matters such as computer failure, poor documentation, and fraud, was becoming evident as a significant risk category. This increased focus on the significance of **OpRisk** as a justifiable regulatory capital requirement area emanates from the key developments in the list below:

- An enhanced emphasis on transparency in firm financial reporting
- Rising levels of exposure to **OpRisk** driven by increasingly complex production technologies used by financial services firms,

alluded to by Cummings et al (2006) and maintained in more recent reports viz., the Chartis Research (2013):

- **OpRisk** is seen as the responsibility of every individual at the FI
- The influence of the chief risk officer (CRO) has increased and they are more often a part of board level decisions
- **OpRisk** is accepted as a risk that affects other risk types.

This expanded view of risk reflects the type of business in which banks now engage and the business environment in which banks operate. The framework for a New Capital Adequacy Accord (**Basel II**) was prepared in June 2004, a paper on its application was endorsed in November 2005, leading to its implementation in June 2006. In contrast to **Basel I**, it cuts the opportunity for local and regional banks to grow into viable competitors to current bigger FI's. According to (**Basel II**) there is a capital charge for **OpRisk**.

2.2 *Definition of Operational Risk*

OpRisk can be regarded as a consequence of the possibility of a realised loss in value due to an operational event. Operations are a necessary part of doing the business; therefore there is a systematic expectation that one bears this risk. The problem faced is to be able to distinguish what constitutes an operational loss event and how much should be attributed

to the risk of such a loss, in the event a loss is experienced due to these failures. As the definition implies it is not trivial to quantify **OpRisk** and is also a very difficult area to manage. The scope of **OpRisk** data includes actual realised losses and operational events that have the potential to lead to an operational loss. The greatest concern when dealing with **OpRisk** is that historical data is often very scarce and of poor quality.

2.3 Capital Adequacy

2.3.1 A Synthesis of Concepts

It is important to note that there are two types of capital that play a role in safeguarding banks against **OpRisk**, namely regulatory capital (RC) and economic capital (EC). RC is the amount of capital a regulator requires a bank to hold to safeguard against **OpRisk** and is based on the proposals of the Basel Committee with **Basel II**. EC is the amount of capital a FI itself deems necessary to operate normally, given its risk profile and its state of controls. According to the Basel Committee on Banking Supervision (Basel Committee), there are three procedures a FI can use in dealing with the capital requirement for their **OpRisk**. The first two are top down approaches and simplistic in nature and the third approach, the internal measurement approach (IMA) is bottom up and a lot more complex. The resulting operational losses are used in calculating regulatory/economic capital.

The first two, the basic indicator approach (BIA) and the standardised approach (SA) are as follows; The BIA assigns a percentage to the annual gross income of the firm as a whole to determine the annual capital charge, while in the SA the firm is split into eight business lines and assigned a different percentage of a three year average gross income per business

line, the summation of which is the capital charge.

In the IMA, the bank uses its own internal loss data to calculate capital charge. Unique to the IMA is that regulators have resolved not to put stringent controls on what IMA process to adopt, as each bank's internal framework has its own design. This adds flexibility in deciding on which model a bank could use to suit their method of analysis. The more effectively a bank manages its **OpRisk**, the less capital it is required to reserve for that risk. The advanced measurement approach (AMA) is an IMA method which applies estimation techniques such as Monte Carlo or historical simulations. As a result, a bank that undertakes an AMA should fulfill the regulatory need of obtaining a more superior capital adequacy requirement measure and may be left with more available funds, which in turn has a positive impact on the bank's competitiveness.

2.3.2 Theoretical Framework

Following the financial crisis, banks have begun to spend a lot of money on risk management practices. One of the significant areas of impact has been on AMA modeling. One of the causes is that there has been a significant increase in the number of operational events, hence improving data points in loss data which aids in the implementation of advanced quantitative methods such as the Loss Distribution Approach (LDA). The LDA is an AMA method whose main objective is to provide realistic risk estimates for the bank and its business units based on loss distributions that accurately reflect the underlying data. Using the **ATAT** dataset, we estimate the Loss Distribution Approach (LDA) model for **OpRisk** for each business line (BL) or event type (ET) combination, assuming probability distributions of the independent identically distributed (i.i.d) severity (single event P&L impact) and one-in-a-quarter event

frequency:

Figure 2.1 is a flowchart illustrating the LDA to **OpRisk**. The LDA models two primary components of **OpRisk** loss data:

- * Loss frequency
- * Loss severity

Monte Carlo simulation is utilised to bring the two distributions together

- * A large number of simulations must be run to observe a sufficient number of losses to reasonably assess what a 1-in-1000 year event might look like

The key benefits an organisation can enjoy from deploying the **ATAT** are proposed to be:

- Increased accuracy and visibility of risk information. Data is collected directly from source and a transparent view of users embedded in the system is enabled.
- The bank can apply AMA methods such as and not limited to, the Loss Distribution Approach (LDA).
- More quickly identify and remediate deficiencies. A controlled list of reasons for amendments and remedial action plans to reduce the time required for reconciliation or other cross-checking requirements.
- Increased management insight by providing sources of internal loss data, risk and control self-assessment, and KRI's.
- Optimisation of business performance; through the analysis of information by creating correlations between the different sources of loss data.

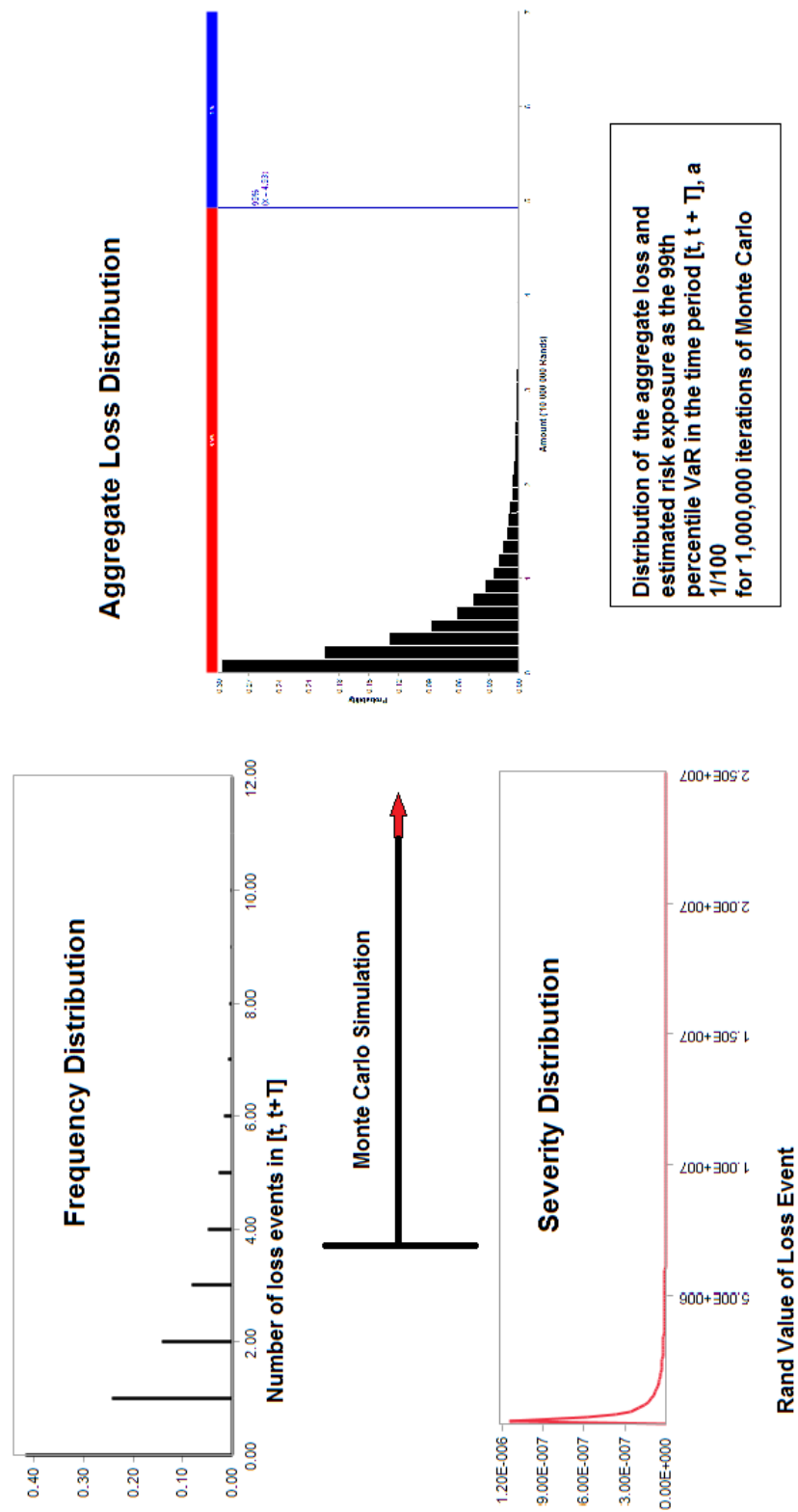


Fig. 2.1: Flowchart of the Loss Distribution Model (LDA)

- Reduce the cost and complexity of your OpRisk platform by integrating all process risk management components on a single, coherent platform.
- Incorporate robust software architecture to contribute effectively and efficiently in building and maintaining a strong control environment.

2.3.3 Hypothesis 1

A bank with sufficient internal data at its disposal for the implementation of quantitative criteria for its operational risk capital requirement will benefit from the use of an AMA, as it gives the bank more decision making options and permits the bank authorities to determine the capital adequacy requirement themselves. This serves as the ultimate goal of a regulatory incentive based process.

2.4 Conclusion of Literature Review

Fundamentals of risk areas in the financial services industry infer that FI's make a conscious decision to take on a certain amount of credit risk, market risk, and other types of risk such as OpRisk. In managing risks, it is understandable that they take on the former two types of risk (credit and market), and that here exposure can be effectively (ex ante) controlled, e.g. be risk averse in providing credit, and hedge yourself against expected adverse market movements. However, opponents of Basel II argue that OpRisk is more part of the process and thereby somewhat inherent in conducting business. According to this interpretation, the rationale for managing OpRisk is not compelling. OpRisk differs in that it always causes losses and not gains. Negative losses are positive gains, but because our focus is on losses,

we will not consider, for example, errors that result in gains. Hence to the extent that **OpRisk** reflects the burden of things always leading to a loss, it makes sense for a FI to make expenditures on managing **OpRisk**.

Accordingly, Franchot et al (2001) acknowledge **OpRisk** is being explicitly concerned by **Basel II** meaning it is now receiving and will receive the same regulatory treatment as those imposed on credit and market risks. It is now becoming a market standard for banks to demand standard quantitative computations of losses for these risk types. The AMA is a methodology whereby a bank uses its own internal data to compute the economic capital necessary to cover against potential operational losses. Banks are given freedom to calculate their **OpRisk** capital as part of the **Basel II** framework – but only if they receive regulatory approval to use the AMA. It is desirable for a bank to compute their own economic capital requirement since regulatory capital adequacy standards almost surely produce an uneven distribution of capital among various banks, creating an uneven playing field.

The LDA is (and will be) a powerful method for banks that want to better align their minimum regulatory capital with economic capital, in so doing, it would also increase the perception of fairness across banks as the capital numbers should more accurately reflect the **OpRisk** component due to its strong quantitative nature. Justifying capital adequacy numbers for **OpRisk** required the explanation of complicated models to senior management, who found it difficult to rationalise the capital amounts. The wake of the financial crisis has raised awareness of the risks, and now management sees the need to improve their assessment of risk and have better risk management. The use of proven quantitative approaches to **OpRisk** management, such as the LDA, can result in the lowering of the risk capital, hence is an attractive return-on-investment option.

Chapter 3

METHODOLOGY

The following section describes the methodology to be followed in addressing the hypothesis and will proceed as follows: Subsection 3.1 outlines the quantitative procedure, subsection 3.2 provides for the definition of the mathematical model, population sample and sampling methods are presented in subsection 3.3, in 3.4 the instrument of experimentation is described and the data gathering technique follows in 3.5.

3.1 Research Paradigm

This study uses a quantitative design to identify, analyse and describe factors contributing to operational loss events through the use of the amendment tracking loss database framework. The research maintains, that through this framework, an accurate association between variables can be achieved that minimises bias and non-representativeness of the data collected and analysed. Data is usually captured with a systematic bias. This problem is more pronounced with **OpRisk** loss data. More precisely, one would expect high frequency low severity losses to have an immaterial **P&L** impact hence possibly ignored, and low frequency high severity losses to have a low probability of being reported because of the negative attention such reports may attract to the business. A data set obtained from an automated

feed eliminates bias and the data captured is highly representative of the actual losses, versus their manual counterparts.

Very few banks are able to implement the LDA, due to scarcity of data. The data from the loss database is exceptional in that it will investigate the full nature of process risk orientated loss events and is exploratory in that it attempts to identify new knowledge and gain new insight into the downstream effects of the factors driving these events. These effects are calibrated through the underlying mathematical model. Through the model, it is possible to compute the value of the two distributions, i.e. frequency and severity, contributing to the value of the dependent variable, namely the aggregate operational loss (hence the total required capital charge), based on a VaR measure in a controlled environment, where the ATAT is the instrument of experimentation/control.

3.2 Research Design

The Loss Distribution Approach

We begin by defining some concepts:

- In line with **Basel II**, and according to Franchot et al (2001), we consider a matrix consisting of business lines BL and (operational) event types ET . The bank estimates, for each business line/event type (BL/ET) cell, the probability functions of the single event impact and the event frequency for the next three months. More precisely, in each cell of the BL/ET matrix separate distributions for loss frequency and severity are modeled and aggregated to a loss distribution at the group level. The aggregated operational losses can be seen as a sum S of a random number N of individual

operational losses $(X_1, \dots, X_N)$. This sum can be represented by:

$$S = X_1, \dots, X_N, \quad N = 1, 2, \dots \quad (3.1)$$

Three month daily statistics are taken of the time series of internal processing errors (frequency data) and their associated severities and used in each cell of the BL/ET matrix. Frequency refers to the number of events that occur within the specified time period (daily buckets) T and $T + \tau$ and severity refers to the P&L impact resulting from the frequency of events. The time (1 day bucket) period is chosen in order to ensure that the number of data points is sufficient for statistical analysis.

3.2.1 Computing the Frequency Distribution

- Let N_{ij} be variable in random selection, representing **the number of times of process risk event failures** between times T & $T + \tau$. Suppose subscript i refers to the BL which ranges from $1, \dots, k$ and subscript j to ET ($j = 1$ for process risk). We have taken a random sample implying that the observations N_{ij} , where $i, j = (1, 1), \dots, (k, 1)$ are independent and identically distributed (i.i.d). The random variable N_{i1} ¹ has distribution function² (d.f.) $\mathbf{P}_{i1}(n/\theta_0)$, where θ_0 is an unknown parameter of the estimated distribution. The unknown parameter θ_0 may be a scalar or a vector quantity θ_0 , for example, The Poisson distribution depends on one parameter called λ whereas the univariate normal distribution depends on two parameters, μ and σ^2 , the mean and variance. These parameters are to be estimated in some way. We

¹ N_{ij} where subscript $j = 1$ since we are only dealing with 1 event type i.e. process risk

² The term distribution function is monotonic increasing function of n which tends to 0 as $n \rightarrow -\infty$, and to 1 as $n \rightarrow \infty$

use the Maximum Likelihood Estimate (m.l.e) which is that value of θ that makes the observed data “most probable ” or “most likely ”. The d.f. $\mathbf{P}_{i1}(n/\theta_0)$, is the probability that N_{i1} takes a value less than or equal to n , where n is a small sample from the entire population of observed frequencies, i.e.

$$\mathbf{P}_{ij}(n) = Pr(N_{ij} \leq n) \quad i, j = (1, 1), \dots, (k, 1) \quad (3.2)$$

The probability density function³(p.d.f) of the discrete random variable N_{i1} takes discrete values of n with finite probabilities. In the discrete case the term for p.d.f. is the probability function (p.f.) also called the probability mass function, i.e. N_{i1} is given by the probability that the variable takes the value n , i.e.

$$p_{ij}(n) = Pr(N_{ij} = n), \quad i, j = (1, 1), \dots, (k, 1) \quad (3.3)$$

The r.h.s of equation (3.2) is the summation of the r.h.s of equation (3.3), we derive a relation for the **loss frequency distribution** in terms of the (p.f):

$$\mathbf{P}_{ij}(n) = \sum_{k=1}^{n_k} p_{ij}(n) \quad i, j = (1, 1), \dots, (k, 1) \quad (3.4)$$

3.2.2 Computing the Severity Distribution

- Suppose X_{ij} is a random variable representing **the amount of one loss event** in a cell of the BL/ET matrix. Define next period's loss in each cell (i, j) , where i is the number of business line cells, $L^{T+1}_{i,j}$: Operational loss for loss type $j = 1$ (process risk). One models the amount of the total operational loss of type j at a given time T

³ A density function is a non-negative function $p(n)$ whose integral, extended over the entire x axis, is equal to 1 for a given continuous random variable X . i.e. it is the area under the probability density curve.

& $T + 1$, over the future (say 3 months), as

$$L^{T+1} = \sum_{i=1}^k L_{i1}^{T+1} = \sum_{i=1}^2 \sum_{l=1}^{N_{i1}^{T+1}} X_{i1}^l \quad l = 1, 2, \dots, N_{i1} \quad (3.5)$$

Let $N_1, N_2, \dots, N_m$ (where m is the number of combinations in the BL/ET matrix) be random variables that represent the loss frequencies. It is usually assumed that the random variables X_{i1} are independently distributed and independent of the number of events N_m . A fixed number of a particular loss type would be denoted by X_{i1}^1 , i.e. the random variable X_{i1}^l , represents random samples of the severity distribution (Aue & Kalkbrener, 2007). The **loss severity distribution** is denoted by $\mathbf{F}_{i1}$. Since loss severity variate X is continuous (i.e. can take on any real value), we define a level of precision h such that the probability of X being within $\pm h$ of a given number x tends to zero. The loss severity, X_{i1} has a (d.f.) $\mathbf{F}_{i1}(x/\theta_1)$, where θ_1 is an unknown parameter and x is a small sample from the entire population of loss severity: We define probability density in the continuous case as follows:

$$\begin{aligned} f_X(x) &= \lim_{h \rightarrow 0} \frac{Pr[x < X \leq x + h]}{h} \\ &= \lim_{h \rightarrow 0} \frac{F_X(x + h) - F_X(x)}{h} \\ &= \frac{dF_X(x)}{dx} \end{aligned} \quad (3.6)$$

operate with $\int dx$ on both sides of 3.6

$$\mathbf{F}_{X_{ij}}(x) = \int_{k=1}^{\infty} f_{X_{ij}}(x) dx \quad i, j = (1, 1), \dots, (k, 1) \quad (3.7)$$

where $f_{X_{ij}}(x)$ is the probability density function (p.d.f.). Once again, the subscript X identifies the random variable for severity (P&L impact) of one loss event while the argument x is an arbitrary sample of the severity events.

3.2.3 Formal Results

Having calculated both the frequency and severity process we need now to combine them in one aggregate loss distribution that allows us to predict an amount for the operational losses to a degree of confidence. We now introduce the aggregate loss variable at time t given by $\vartheta(t)$. This new variable represents **the loss for business line i and event type j** . The aggregate loss is defined by $\vartheta(t) = \sum_{n=1}^{N(t)} X_n$ (where X represents individual operational losses).

- Once frequency and severity distributions are estimated, the compound loss distribution $\mathbf{G}(t)$ can be derived. Taking the aggregated losses we obtain:

$$\mathbf{G}_{\vartheta(t)}(x) = Pr[\vartheta(t) \leq x] = Pr\left(\sum_{n=1}^{N(t)} X_n \leq x\right) \quad (3.8)$$

The derivation of an explicit formula for $\mathbf{G}_{\vartheta(t)}(x)$ is, in most cases impossible. Again we implicitly assume that the processes $\{N(t)\}$ and $\{X_n\}$ are independent and identically distributed (i.i.d). Deriving the analytical expression for $\mathbf{G}_{\vartheta(t)}(x)$, we see a fundamental relation, corroborated by Franchot et al (2001); Cruz (2002); Embrechts, Kluppelberg & Mikosch (1997); & others given by:

$$\mathbf{G}_{\vartheta(t)}(x) = \begin{cases} \sum_{n,k=0,1}^{\infty} p_k(n) \mathbf{F}_X^{k\star}(x) & x > 0 \\ p_k(0) & x = 0 \end{cases} \quad (3.9)$$

where $\star$ is the *convolution* operator on d.f.'s, $\mathbf{F}^{k\star}$ is the k -fold convolution of $\mathbf{F}$ with itself⁴, i.e. $\mathbf{F}_X^{k\star}(x) = Pr(X_1 + \dots + X_k \leq x)$, the d.f. of the sum of k independent random variables with the same distribution as X .

⁴ the convolution of two functions $f(x)$ and $g(x)$ is the function

$$\int_0^x f(t)g(x-t)dt \quad (3.10)$$

- The aggregate loss distribution $\mathbf{G}_{\vartheta(t)}(x)$ cannot be represented in analytic form, hence approximations, expansions, recursions of numerical algorithms are proposed to overcome this problem. For purposes of our study, an approximation method will do. One such method consists of taking a set $\langle \vartheta_1, \dots, \vartheta_s \rangle$, otherwise known as the ideal generated by elements $\vartheta_1, \dots, \vartheta_s$ which are s simulated values of the random variable ϑ_{ij} for $s = 1, \dots, S$ (Fraleigh, 2000). This method is popularly known as Monte Carlo simulation coined by physicists in the 1940's, it derives its name and afore-mentioned popularity to its similarities to games of chance. The way it works in layman's terms is; in place of simulating scenario's based on a base case, any possible scenario through the use of a probability distribution (not just a fixed value) is used to simulate a model many times. In the LDA separate distributions of frequency and severity are derived from loss data then combined by Monte Carlo simulation.

3.2.4 Dependence Effects (Copulae)

The standard assumption in the LDA is that frequency and severity distributions in a cell are independent and the severity samples are i.i.d. According to **Basel II**, dependence effects in **OpRisk** are not considered. Economic capital allocation however, could benefit if it were determined in a way that recognises the risk-reducing impact of correlation effects between the risks of the BL/ET combinations. Concluding remarks from a study by Urbina & Guillen (2014) allude that failure to account for correlation may lead to risk management practices that are unfair, as evidenced in an example using data from the banking sector.

One of the main issues we are confronted with in **OpRisk** measurement is the aggregation of individual risks (in each BL/ET element). A powerful concept to aggregate the risks – the *copula* function – has been introduced in finance by Embrechts, McNeil & Straumann (2000). Copulas have been used extensively in finance theory lately and are sometimes held accountable for recent global financial failures, e.g. the global credit crunch of 2008 - 2009. They are nevertheless still applicable and in use for **OpRisk** as operational risk models follow a different stochastic process to other areas of risk, e.g. operational **VaR** is subject to more jumps than market **VaR** and is thought to be discrete whereby market **VaR** is continuous.

- Copulas are functions which conveniently incorporate correlation into a function that combines each of the frequency (marginal) distributions to produce a single bivariate cumulative distribution function. Our model is used to determine the aggregate (bivariate) distribution of a number of correlated random variables through the use a Clayton copula. Dependence matters due to the effect of the addition of risk measures over different risk classes (cells in the BL/ET matrix).
- More precisely, the frequency distributions of the individual cells of the BL/ET matrix are correlated through a Clayton copula in order to replicate observed correlations in the observed data. Let m be the number of cells, $\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_m$ the distribution functions of the frequency distributions in the individual cells and $\mathbf{C}$ the so-called copula. Abe Sklar proved in 1959 through his theorem (Sklar's Theorem) that for any joint distribution $\mathbf{G}$ the copula $\mathbf{C}$ is unique. $\mathbf{C}$ is a distribution function on $[0, 1]^m$ with uniform marginals. We refer to a recent article by Chavez-Demoulin et al (2006) for

further information: It is sufficient to note that $\mathbf{C}$ is unique if the marginal distributions are continuous.

$$\mathbf{G}(x_1, \dots, x_m) = \mathbf{C}(\mathbf{G}_1(x_1), \dots, \mathbf{G}_m(x_m)) \quad (3.11)$$

Conversely, for any copula $\mathbf{C}$ and any distribution functions $\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_m$, the functions $\mathbf{C}(\mathbf{G}_1(x_1), \dots, \mathbf{G}_m(x_m))$ is a joint distribution function with marginals $\mathbf{G}_1(x_1), \dots, \mathbf{G}_m(x_m)$. Moreover, combining given marginals with a chosen copula through Equation 3.11 always yields a multivariate distribution with those marginals. The copula function has then a great influence on the aggregation of risk.

3.3 Population and Sample

3.3.1 Population

Annex 2 of Basel Committee on Banking Supervision (2001) identifies the main activities of a FI along three standard business units: Investment Banking (IB), Banking & Other which are further subdivided into eight levels: Corporate Finance, Trading and Sales, Retail Banking, Commercial Banking, Payment and Settlement, Retail Brokerage, Asset Management, and Agency Services. Furthermore, potential losses resulting from a bank's operational events are decomposed into a number of sub risks using business lines and risk categories defined by the bank. According to the Basel Committee (2001), there are generally seven loss event type categories: Internal Fraud, External Fraud, Employment Practices and Workplace Safety, Business Disruption and System failures, Execution, Delivery & Process Management, Clients, Products and Business Practices, and Damage to Physical Assets. Accordingly, banks must estimate VaR for each of the business lines per event type combinations, this

works out to $7 \times 8 = 56$ risk types. Table 3.1 illustrates the formation of a $7 \times 3 = 21$, BL/ET matrix.

Event Type Category	$[X_i, X_j]$	Business Lines			
		IB	Banking	Other	FIRM
Internal Fraud	ET_1	1	2	3	AG_1
External Fraud	ET_2	4	5	6	AG_2
Damage	ET_3	7	8	9	AG_3
Business Disruption	ET_4	10	11	12	AG_4
Clients, products,..	ET_5	13	14	15	AG_5
Execution, delivery,..	ET_6	16	17	18	AG_6
Employment practices..	ET_7	19	20	21	AG_7

Tab. 3.1: The BL/ET Matrix for 7 event types and 3 business lines

3.3.2 Sample and Sampling method

Sample

In the study, the bank under consideration consists of an organisational structure unique to itself; nevertheless the business line split from our sample is structured to align as closely as possible to the standard adopted by the Basel approach. Our sample consists of two business lines and one event type category. The two are: $i = 1$, Banking BL_1 , & $i = 2$, Trading BL_2 and the $j = 1$, Process Risk ET_1 : The Execution, Delivery & Process Management category of **OpRisk**. Historical losses arising from failed transaction processing or process

management, from relations with trade counterparties and vendors which give rise to process risk at the bank in question have been identified for the period beginning 02 January, 2013 to 20 March, 2013.

Table 3.2 below demonstrates the two way business line split per one row event type matrix to be adopted. Each process risk event is linked to a unique trade reference or trade number: Each unique trade number is classified under either one of the two business lines forming the BL/ET matrix, i.e. if a trade falls under the Banking business line it cannot fall under the Trading business line and vice versa. A consistent data set is guaranteed by reference to the trade field entry which specifies it's business line BL_i . Trades are unique, and so are the numbers that reference them. This way repetition of trades is avoided which serves to eliminate repetition of process risk loss across cells. Permission to conduct the study

Internal Event Type	$[X_i, X_j]$	Business Lines		
		Banking	Trading	FIRM
Process risk	ET_1	1	2	AG

Tab. 3.2: The BL/ET Matrix for a South African Investment Bank:

was sought and granted by the legal, compliance and risk management authorities at the bank in question. Data for frequency of losses attributed to process risk operational errors is obtained from the ATAT database, the loss data is internally generated and reflects the investment banks loss profile for processing errors most accurately.

Completeness of data capture is ensured due to the trade by trade nature of reporting of operational fallouts, which is essential for frequency calibration. Severity loss data associated with the frequency loss data is obtained to coincide with the period noted. Furthermore,

the frequency and severity datasets coincide, hence the losses suffered through operational fallouts are specific to the trade level events that led to the losses, hence to process risk events exclusively. It must follow that sub-samples of the two distributions, i.e. frequency and severity distribution samples can be generated for the entire population.

Sampling method

Figure 3.1 is a snapshot of a blotter for vanilla interest rate derivatives, matched trades are confirmed and trades with mismatching details require amending prior to confirmation. When a deal is agreed, e.g. a fixed income trader agrees to buy a 4yr, ZAR100 million interest rate swap at a rate of 7.98% (TRD#5 in blotter Figure 3.1) between his bank and a counterparty fixed income dealer at another bank, B , a trade blotter is received and the captured trade details are compared to the details on the blotter.

The post trade booking service continues throughout the life of each deal. If one follows the chain of processing events (refer to Figure 1.1), the trader will then book the deal by capturing it in the FO system, it is checked for accuracy by MO and then confirmed against a deal confirmation, which is also agreed to by BO. The deal is now captured within the relevant portfolio in the bank's FO/treasury system. The treasury system is the control framework which enables this process to flow seamlessly. The control unit is reliant on a myriad systems (of which the treasury system is a part) controlling the reporting and processing function within an organisation. It is up to each bank's control unit to choose an automated, information technology solution that is appropriate for its specific purpose. This is extremely important as automation serves as the key to a fast, reliable and accurate trading service.

Buy/Sell	Notional	Tenor	Start	Rate	C/Party	Broker	Sales	Comments	StartDate	EndDate	DV01	Delta	FwdRate	Pnl	Capture
PAY	169	3y	0d	8.330	X				06-Apr-09	06-Apr-18	632.25	101 160	8.390	606 962	Trd#1
SELL	-1000	9x12	9m	7.040	Y				06-Jan-10	06-Apr-10	22.89	22 891	7.020	45 782	Trd#2
SELL	-2500	9x13	9m	6.940	Z				06-Jan-10	06-Apr-10	22.89	57 224	7.020	-457 818	Trd#3
BUY	2500	1x4	1m	8.100	A				06-Jan-09	06-Aug-09	24.52	-61 305	8.140	245 218	Trd#4
PAY	100	4y	0d	7.980	B				06-Apr-09	08-Apr-13	341.52	34 152	8.003	76 842	Trd#5

Fig. 3.1: Sample of a blotter template for trading in fixed income securities

Trade amendments consist of putting in manual adjustments to trade details in the trade ticket when necessary. Many times these adjustments are due a manual update, such as correcting for an incorrectly booked rate, e.g. suppose the 4yr swap was incorrectly booked at rate of 7.988% in the FO system when it should have been 7.98% (as reflected in the blotter). If this so happens, an instruction is given and the trade ticket will be amended to reflect the correct rate (of 7.98%). The amendments to details are compiled in a report at the end of the month for statistical purposes and are used as a source of KRI's.

3.4 The Research Instrument

As mentioned previously, the ATAT database will be used to collect data. The ATAT is an innovative new development used as a vital internal control tool providing an electronic solution to the management of the amendment function. As mentioned before, the transaction flow starts at the FO system which registers the transaction. The success of the tool is in its ability to automate data collection directly from the FO/treasury system where the transaction is registered, the ATAT works like a filter, collecting every cancellation, and/or amendment made to a transaction. The filter will highlight the OpRisk loss event eliminating the reliance on individuals to account for cancellations or amendments made to a transaction by printing out emails and keeping a manually updated record of events. As part

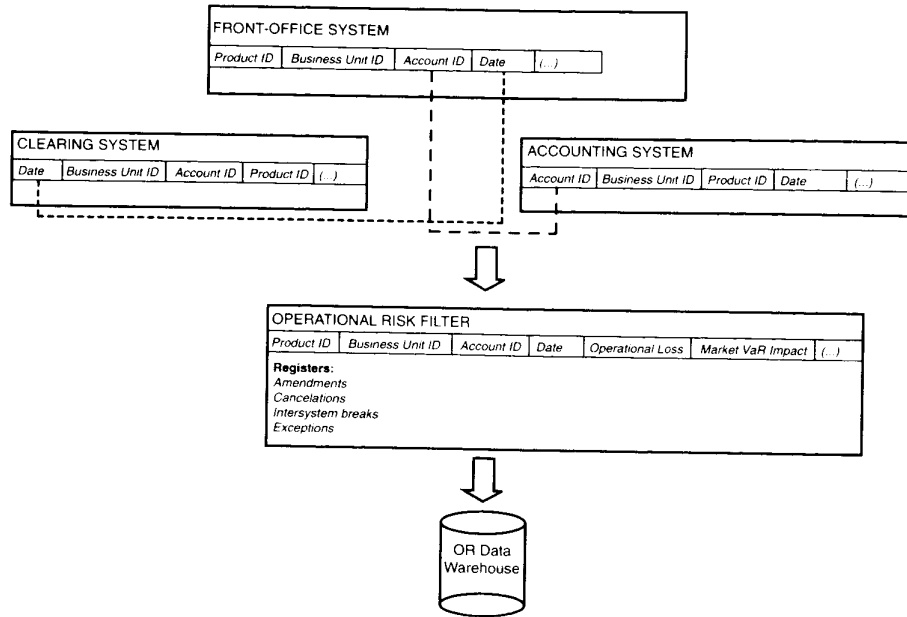


Fig. 3.2: Design of a data filter for an automated trade amendment tracker (ATAT)

of its functionality, users are able to subscribe to a controlled list of reasons for amendments, ensuring explanations around amendments are clear and easily analysed.

A typical design of a filter is shown in Figure 3.2. Data quality of loss reporting is often a major concern in many organisations. The ATAT simplifies data collection for loss reporting by ensuring that the following processes are built-in to workflow capabilities:

- Daily
 1. Each amendment in the ATAT is reviewed and the event type and business line fields are populated
 2. All supporting documentation is properly referenced to the tracker and filed in a dedicated amendment file
- Monthly

1. Extract, print and review all amends and ensure support file is complete
2. Provide the detail to the desk head of the business line and obtain physical sign off to evidence his/her review.

Displayed in Appendix C.1 is an example of what a **ATAT** might look like. A trade level view of pre-populated descriptive elements of trade details captured in tailor made organisational fields are pictured in the amendment loss database form. In the snapshot users are able to obtain a list of information to serve as a guideline of the nature of the amendment made to each and every trade by the business line desk, and then to enrich this data with a user defined description of why the amendment was required. There are certain standard entries which cannot be tampered with as they are hard coded in, a hard coded entry field is one such as the person who amended the trade, or in which business line the trade belongs. There are mandatory fields, additional fields which can be edited/updated and require no programming. These fields are where the person responsible for updating the field categorises the reasons for the amendment (operational event type), responsible persons and the nature of the trade amendment, i.e. rate, nominal, etc. A guideline is provided which is designed to give the user a step by step procedure to deduce the nature of the amendment as tabulated and displayed in the embedded view (in the bottom left corner) labeled “Amendment Detail”.

3.5 Procedure for data collection

In this study, time series of internal frequency and severity data was obtained on request from a South African bank. A letter of motivation was provided in support of the data request, (Appendix C.2) following which confidentiality agreements (NDAs) needed to be

signed and final approval issued from the bank to ensure protocol was complied with. The researcher, assisted by volunteers in the Corporate and Investment Banking Wealth Middle Office (CIBW MO) team, prepared the data on specific days when he was present at the investment bank. Their cooperation was requested and promised. The complete set of data was handed to the researcher for data analysis, who also undertook not to cause any disruption to functions of the bank.

Confidentiality of client information was maintained, no names were disclosed in the research report. The bank in question will be provided with the research report from the researcher who would supply such a report.

The following steps/protocol needed to be followed:

1. Confidentiality Agreement signed (NDAs)
2. Data to be prepared by CIBW MO
 - a. Request criteria to be defined e.g. period/type etc.
 - b. Client names removed
3. Final product signoff
 - a. No inference to be drawn from research of the bank in question
 - i. Reference made to a South African bank
 - b. The research paper should be submitted to Legal and Compliance prior to being released

Chapter 4

RESULTS

4.1 Data Analysis and Interpretation

We model the loss distribution in each cell in the BL/ET matrix through implementation of a high performing simulation based computer software package which uses calibration and simulation algorithms for a range of distributions in order to determine the appropriate distribution class for a particular cell. This package which specialises in quantitative risk analysis is an excel tool called *ModelRisk* from Vose Software BVBA.

ModelRisk was first released in the mid-2000s. The most recent version is a comprehensive Monte Carlo simulation tool that contains a number of advanced and unique capabilities and techniques. It was created by David Vose; he provided guidance on the use of *ModelRisk* and played a key role in conducting the risk analysis in this research paper. For more information on *ModelRisk* visit www.vosesoftware.com. Figure 4.1 is a snapshot of the ModelRisk Add-in. It contains a number of consistent features with excel, which makes it particularly useful in estimating OpRisk VaR. Excel is the medium of choice due to its applicability to the analysis.

We present cross-sectional data from a study of 369 process risk losses and accompanying

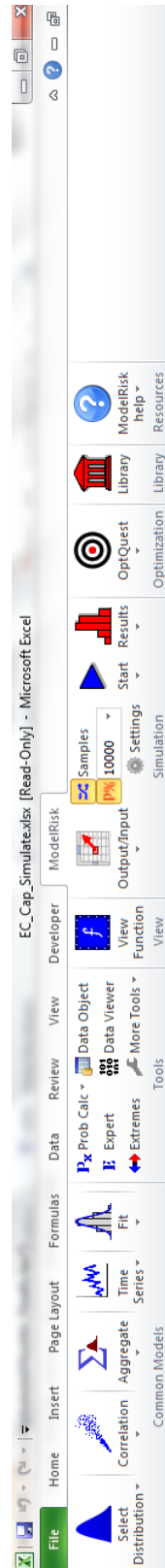


Fig. 4.1: The ModelRisk toolbar with choices of consistent features in excel

P&L per event impact of each loss amounting to R 61 534 745 over a period of three months (January 2013 - March 2013). Figure 4.2 is a scatterplot of the empirical dataset (number of loss events vs single event P&L loss impact) which provides a convenient summary showing the wide range of losses. On examination of the scatterplot we find no indication of any particular trend between the two variables. The loss values are depicted as positive in the scatter plot. The data has not been adjusted for inflation over this period. Our task was to estimate the 1 – day 99.9% OpRisk VaR for process risk in the South African (ZA) Bank.

4.1.1 Fitting Distributions to Data

Apart from generated scenarios LDA models rely on loss data and are inherently backward looking. It is therefore important as a first step, to fit distributions to the historically observed losses and then simulate what the losses might look like over the specific time interval. On the *ModelRisk* toolbar, under the drop down menu on the “Fit ” tab, shown in Figure 4.1, choose “Fit Distribution ” and follow the commands.

Once again, David Vose’s experience and skill assisted in selecting appropriate distribution classes to test against, i.e. displaying the severity loss data in descending order (from high P&L impact to low) which optimises *ModelRisk*’s data handling capabilities. Zero P&L impact entries have been excluded from the data to be tested hence have no effect in the resulting fit. Estimates of best fit distributions for the separate frequency and severity data elements are obtained and the optimal fits listed below.

1. Frequency distribution

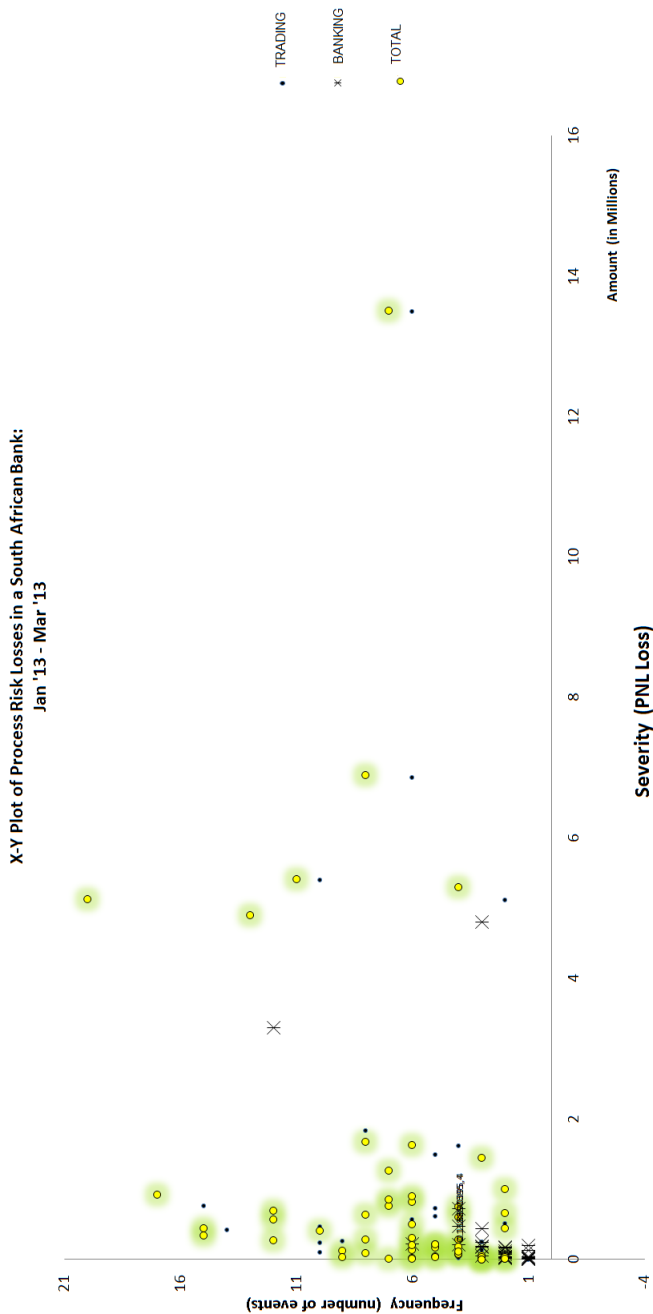


Fig. 4.2: (a)Scatterplot of the separate business lines consisting of 78 banking & 291 trading (frequency) loss events amounting to R 12 015 265 P&L loss impact and R49 519 479 P&L loss impact (severity) respectively; (b)Scatterplot of the combined business lines consisting of 369 (frequency) loss events amounting to R 61 534 745 (severity) P&L loss impact.

- The Geometric distribution

$$G : p, p = 0.4179105$$

is the optimal fitting discrete distribution for the Banking BL according to SIC.

- The Polya distribution

$$\text{Polya} : (\alpha, \beta), \quad \alpha = 6.557774, \beta = 0.7924074$$

is the optimal fitting discrete distribution for the Trading BL according to SIC.

2. Severity distribution

- The Lognormal distribution

$$\text{Lognormal}(\mu_1, \sigma_1), \quad \mu_1 = 411736.7, \sigma_1 = 2271727$$

is the optimal fitting continuous distribution for the Banking BL according to SIC.

- The Lognormal distribution

$$\text{Lognormal}(\mu_2, \sigma_2), \quad \mu_2 = 1043960, \sigma_2 = 6121955$$

is the optimal fitting continuous distribution for the Trading BL according to SIC.

4.1.2 Information criterion

Since it is not obvious by visual inspection which parametric combinations $(p, \alpha, \beta, \mu_1, \mu_2, \sigma_1, \text{ and } \sigma_2)$ to use, they are compared to goodness-of-fit tests/statistics. According to Lehman, Groenendaal, & Nolder (2012), *ModelRisk* will fit each appropriate distribution to the

data and compute several values, called information criterion, to describe how well the distribution fit, fits the data. There are several information criterion available to determine the estimated distributions. They are the Akaike Information Criterion (AIC), Schwarz Bayes Information Criterion (SIC) and the Hannan-Quinn Information Criterion (HQIC). They consist of computed values based on estimating the *likelihood* that the observed data could have come from each postulated distribution. For example the AIC is defined as:

$$AIC = -\frac{2l}{Y} + \frac{2}{Y} \times (\text{number of parameters}) \quad (4.1)$$

where the *likelihood* is evaluated at the m.l.e and Y is the sample size, i.e.

$$\begin{aligned} l &= \log(\text{likelihood}) \\ &= -\frac{Y}{2}(1 + \log 2\pi) + \log\left(\frac{\hat{\epsilon}\hat{\epsilon}}{Y}\right) \end{aligned} \quad (4.2)$$

One computes the $AIC(k), k = 1, 2, \dots, p$ and selects the one with the maximum value. Higher numbers (lower absolute numbers if they are negative) indicate better fitting distributions. This is borne out by Appendix A.5: Different distributions ranked according to SIC¹ are observed in the snapshot of computed values of the SIC, AIC, and HQIC showing the optimal fit at the top. All fits are performed parameters estimated by (m.l.e) methods.

In the current environment the selection of a copula function for modeling frequency correlations, is enabled by *ModelRisk*'s "Fit" function followed by the "Bivariate Fit Copula

¹ The SIC is an alternative to the AIC that imposes a larger penalty for additional coefficients

$$SIC = -\frac{2l}{T} + \frac{(k \log T)}{T}, \quad k = 1, 2, \dots, p \quad (4.3)$$

". More precisely, the frequency distributions of the individual cells are correlated through a copula - VoseSoftware picks the technique from each approach depending on how they fit our problem and how easy they are to implement. Based on statistical tests (information criterion) *ModelRisk* determines the Clayton copula estimate.

4.1.3 Goodness-of-fit tests

The histogram of the results are shown in Appendices A.1, A.2, A.3, and A.4 along with the visual representation of the goodness-of-fit plots. These are visual tests which offer an intuitive way of comparing the data and the fitted distributions.

- Overlaying a histogram plot of the data with a density function is perhaps the most informative comparison as it's easier to see that the general shape of the data compares relatively well given the very few x values.
- An overlay of the cumulative frequency plots of the data and the fitted distribution provides a summary of the variability of measurements. The common S-shape curves will only show large differences between the data and fitted distributions.
- Probability-probability (P-P) plot: A plot of the cumulative distribution of the fitted curve $F(x)$ against the cumulative frequency $F_n(x) = i/(n + 1)$ for all values of x_i for a sample of size n . If the underlying fitted distribution is derived from the theoretical distribution, the plot should look roughly linear.
- Quantile-quantile plot: A plot of the observed data x_i against the x values where $F(x) = F_n(x)$, i.e. $= i/(n + 1)$.

4.1.4 Further manipulation of data

The next step is to reproduce the basic data by generating Monte Carlo simulated values from the estimated distributions of the fitted data of observed losses. Figure 4.3 & Figure 4.4 are the resulting histogram and line plot of one million iterations of MC simulated loss values.

Seen in the result displayed in Figure 4.3 is a graphical overlay representation of both the banking and trading frequency loss data. As can be seen from Figure 4.3, the histograms depict the discrete nature of frequency data. By visual inspection, one can see from the overlay chart that the two histograms appear to come from separate distribution classes.

Likewise, Figure 4.4 is a graphical overlay representation of the severity loss data. As can be seen through the naked eye these line graphs appear to be somewhat similar; i.e. they appear to come from the same distribution class, i.e. we think of the simulated data as random samples from an underlying process (which follows the Lognormal distribution), with somewhat different parameters based on a different set of random values. In both cases the sets of random samples came from the same underlying process, but the fitted distribution for each set will produce different parameters for the Lognormal distribution. The line graphs capture the continuous nature of the severity data; the aggregate loss model is characterized by the differing compositions of the frequency and severity data.

4.1.5 Distribution Fitting for VaR

Whereas the previous steps dealt with the data sources that are used in the modeling process, the next step is devoted to the specification of the LDA model. The LDA involves the

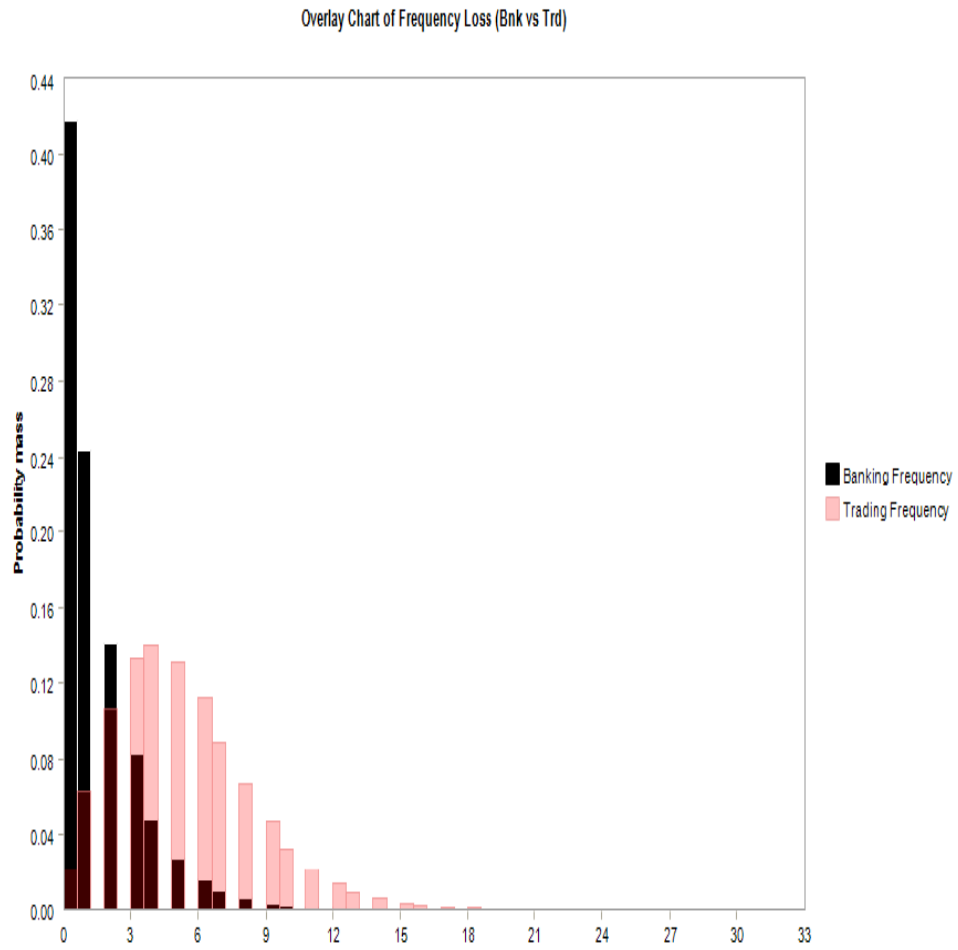


Fig. 4.3: Overlay histogram of the frequency (number of times) of process risk loss events in a ZA bank: 02 January, 2013 – 20 March, 2013

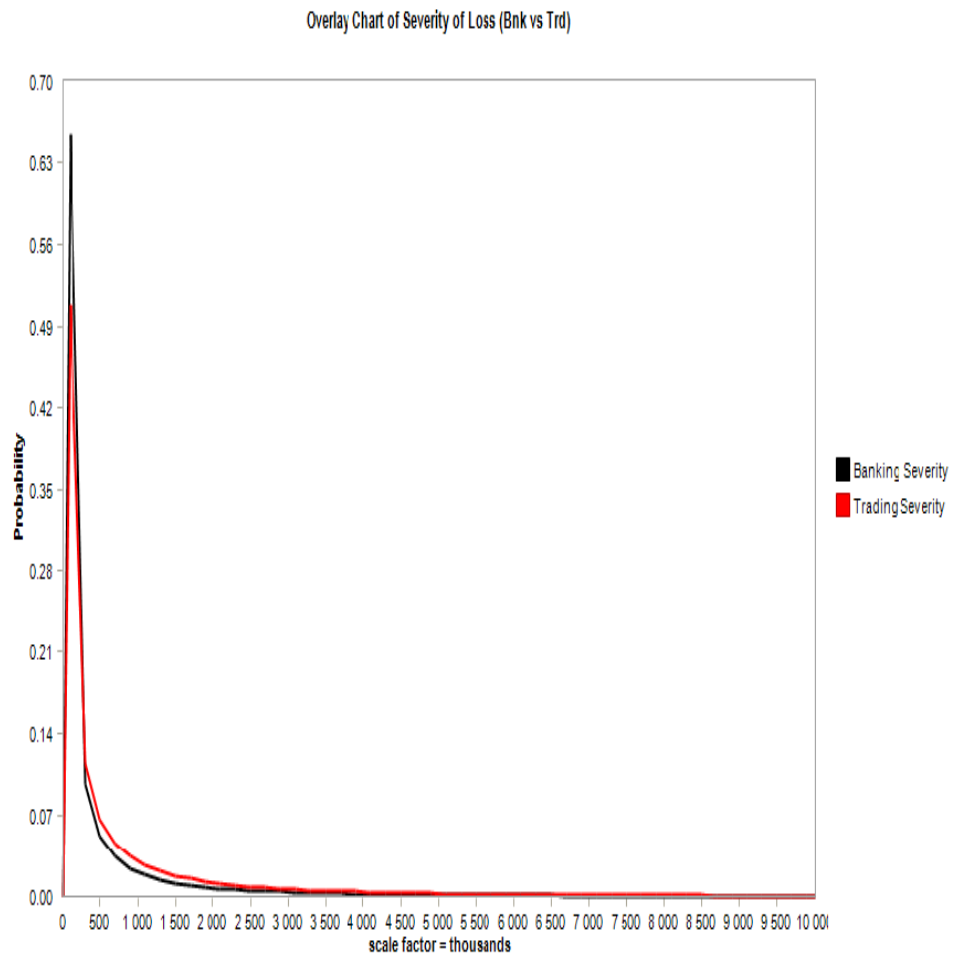


Fig. 4.4: Overlay Line Graph of the severity (single event P&L impact) of process risk loss events in a ZA bank: 02 January 2013 - 20 March 2013

analysis of multiple distributions of event frequency (correlated through a Clayton copula) and severity at the same time. The spreadsheet displayed in Appendix C.3 shows the excel model for the LDA. In this model we feature the Aggregate Multi Monte Carlo tool “VoseAggregateMultiMC ”found under *ModelRisk’s* “Aggregate ”menu which conducts a true simulation “behind the scenes ”to generate a sum (aggregate) of several frequency and severity distributions simultaneously. There are three general steps for implementing MC simulation:

1. Build a Monte Carlo model in excel
2. Run a Monte Carlo simulation for a large number of iterations
 - Sample from frequency distribution to determine the number of loss events ($= N$)
 - Sample N times from the loss severity distribution to determine the loss severity for each loss event
 - Sum loss severities to determine total loss
3. Review and present the results of the Monte Carlo simulation.

Aggregating the estimated frequency distributions and estimated severity distributions simulated through one million iterations by Monte Carlo gives the output aggregate loss distribution shown in Figure 4.5; i.e. the correlated frequency distributions – the Geometric(0.4179105) and Polya(6.557774,0.7924074), and corresponding severity distributions modeled as Lognormal(411736,2271724) and Lognormal(1043960,6121955). The random value drawn from this estimate made up of the sum of individual losses provides one estimate

of a potential aggregate operational loss to the bank and becomes one data point of the loss distribution.

4.1.6 The Generalised Extreme Value Distribution

Since the aggregate loss at any moment is a continuous random variable, it follows that the loss over the next period is also a continuous random variable. In **OpRisk** we attempt to model not only the entire range a variable might take, but also to withstand the highest loss (extreme loss). Modeling the extremes of loss events makes sense since these make the greatest impact (e.g. losses due to “rogue trader’s ”). People have put a lot of effort into determining the distributions of extremes to model data that is extreme and rare.

Since we are interested in loss values that determine whether a system will potentially fail, we undertake one million iterations of the resulting aggregate loss distribution as per Figure 4.5, by generating MC simulations through the excel model *Model-Process-Risk.xlsx* (Appendix C.3). The resulting dataset consisting of one million entries represents the total aggregate loss output probability distribution the simulation is trying to achieve. This is now estimated to the output aggregate loss probability distribution (using the “Fit ” function choose “Fit Distribution ” and follow the prompts in *ModelRisk*) as before (Section 4.1.1).

Figure 4.6 is a histogram of the output aggregate loss distribution of a million random aggregate loss samples of process risk in the ZA bank. *ModelRisk* provides the use of the Generalised Extreme Value distribution: $GEV(a,b,c)$ as an adequate fit for the operational loss data set using the single best fitting set of parameters; $a = 1917993, b = 2329948, c = 0.8124032$.

The $GEV(a,b,c)$ distribution is a continuous probability distribution developed in the theory

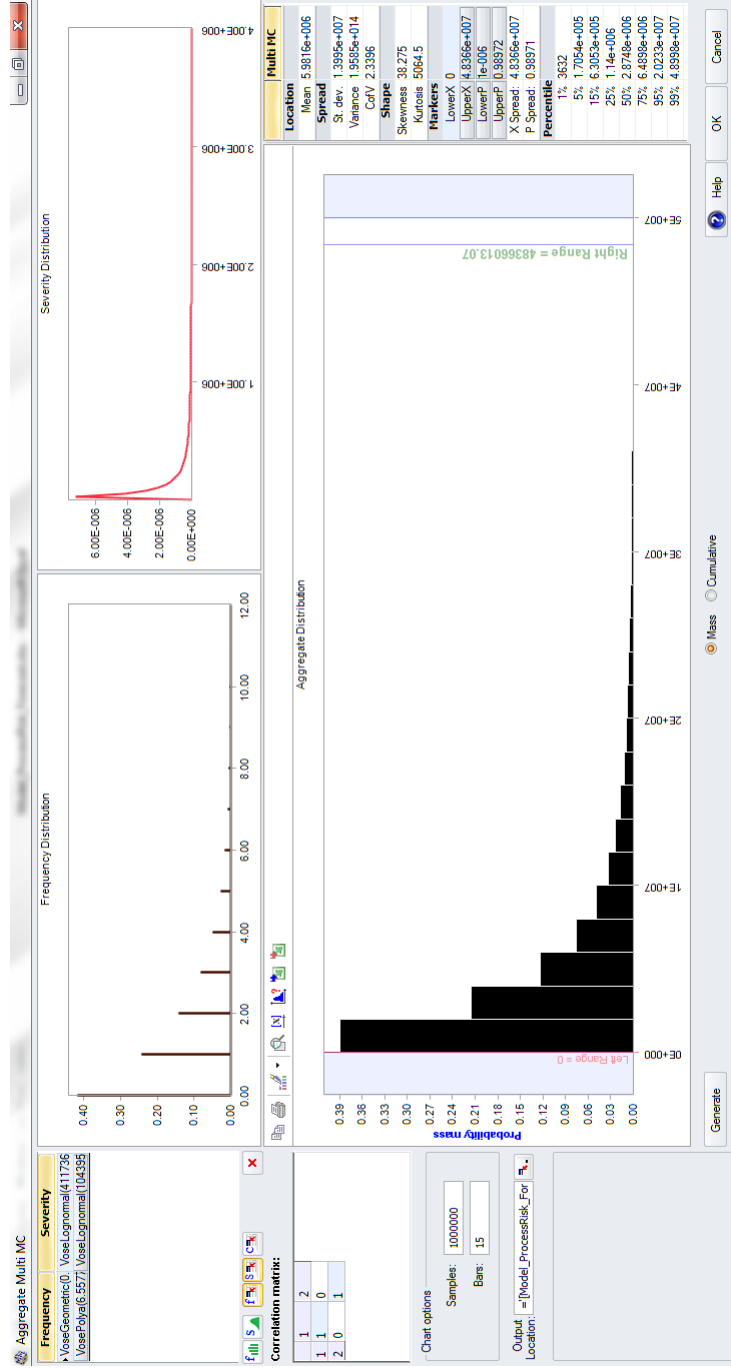


Fig. 4.5: Ordered losses for Monte Carlo simulation for Geometric G-p, $p = 0.4179105$; Polya: (α, β) , $\alpha = 6.557774$, $\beta = 0.7924074$ frequencies distributions and Lognormal (μ, σ) , where $\mu_1 = 411736.7$, $\sigma_1 = 2271727$, $\mu_2 = 1043960$, & $\sigma_2 = 6121955$ severity distributions. The mean of the cumulative distribution is R5 998 306 and standard deviation R14 106 490

behind determining extreme value distributions. Accordingly, the distribution of extreme values for large samples is given by one of three distributions that form the special cases of the GEV distribution, i.e. the GEV distribution is equivalent to a Gumbel, Frechet or Weibull distribution dependent on whether $c = 0$, $c > 0$, or $c < 0$ respectively.

ModelRisk has estimated the parameters of this function that best fit the data. As a summary of these losses, descriptive statistics are presented in Table B.2, and the visual representation in goodness-of-fit plots labeled Figures 4.7, 4.8, 4.9, and 4.10 respectively.

The standardised d.f. of the GEV distribution is written as:

$$\mathbf{H}_{a,b,c}(x) = \exp \left[- \left[1 + c \left(\frac{x-a}{b} \right) \right]^{-\frac{1}{c}} \right] \quad (4.4)$$

where a , b , and c are location, scale and shape parameters. Expressions of equations are shown below along with the boundary conditions that satisfy them.

$$\mathbf{H}_{a,b,c}(x) = \left\{ \begin{array}{ll} \exp \left[- \exp \left(\frac{x-a}{b} \right) \right] & \text{if } -\infty < x < +\infty, \quad -\infty < a < +\infty, \quad c = 0 \\ \exp \left[- \left(\frac{x-a}{b} \right)^{-c} \right] & \text{if } x \geq a, \quad b > 0 \quad c > 0 \\ \exp \left[- \left(\frac{a-x}{b} \right)^c \right] & \text{if } x \leq a, \quad b > 0 \quad c > 0 \end{array} \right\} \quad (4.5)$$

The GEV(a,b,c) contains the Frechet distribution as a special case (i.e. when $c > 0$). To generate a Frechet distribution, notice that the c.d.f for the Frechet is given by;

$$\mathbf{H}_{a,b,c}(x) = \exp \left[- \left(\frac{x-a}{b} \right)^{-c} \right] \quad \text{if } x \geq a, \quad b > 0 \quad c > 0 \quad (4.6)$$

$\mathbf{H}(x)$ is a uniform random variable, so inverting the equation we get

$$x = b \left(\frac{1}{-\ln U(0,1)} \right)^{\frac{1}{c}} + a \quad (4.7)$$

so we use this equation to generate the Frechet distribution.

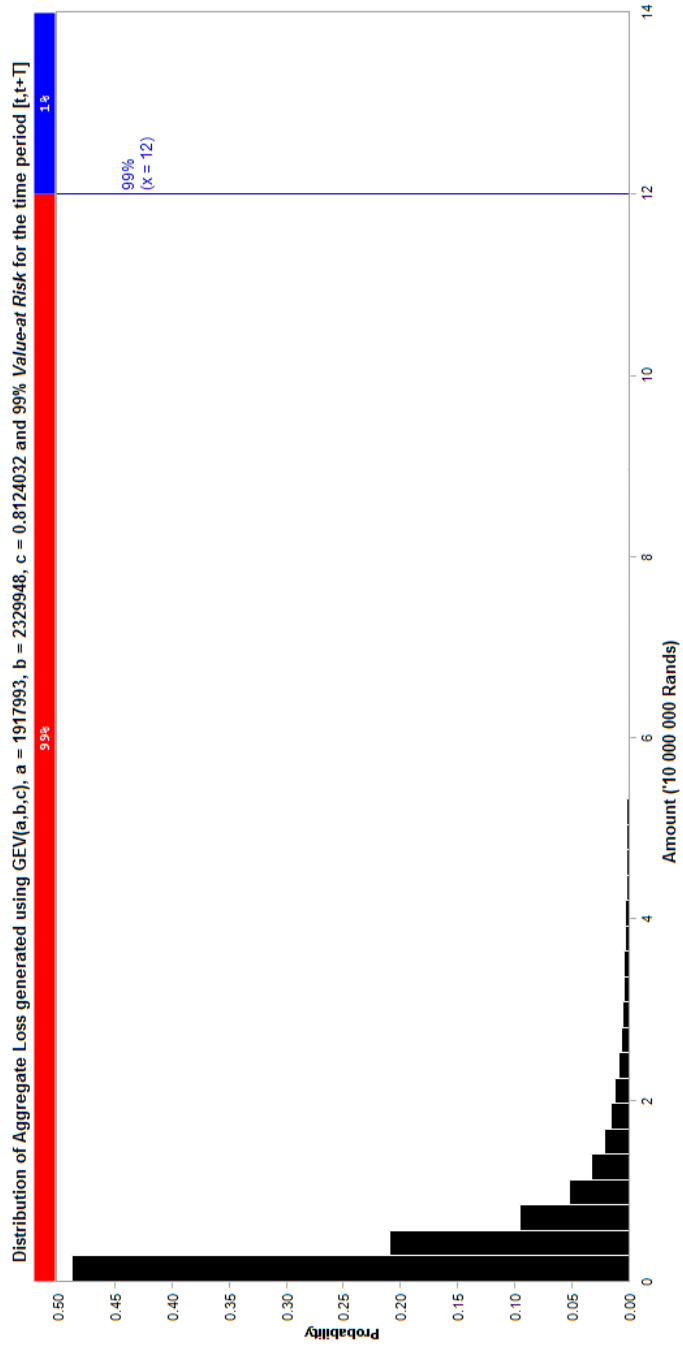


Fig. 4.6: Ordered losses for Monte Carlo simulation for Geometric $G : p, p = 0.4179105$; Polya: $(\alpha, \beta), \alpha = 6.557774, \beta = 0.7924074$ frequencies distributions and Lognormal(μ, σ), where $\mu_1 = 411736.7, \sigma_1 = 2271727, \mu_2 = 1043960, \& \sigma_2 = 6121955$ severity distributions. The mean of the cumulative distribution is R13 139 000

Comparison of probability density

It can be seen from Figure 4.7 that the estimated $GEV(a, b, c)$ line graph compares well with the input distribution histogram (probability density), i.e. it covers both the initial data and the tails. However from visual inspection alone, we cannot determine if the fit is adequate or not.

Comparison of cumulative probability distributions

Figure 4.8 is a comparison of the cumulative probability distributions (c.d.f) of the input distribution and the estimated $GEV(a, b, c)$ distribution. Again, from visual inspection the plot does not show very large differences between data and the $GEV(a, b, c)$ distribution. This graph conveniently summarizes the variability in aggregate operational risk losses. For example, we can see from the graph that about 90% of the samples had losses less than R16.9 million and about 10% had losses less than an amount of R500 000.

P-P & Q-Q plots

Figure 4.9 is a probability–probability (P–P) plot comparison of the input distribution and the estimated $GEV(a, b, c)$ distribution, and Figure 4.10 is a quantile – quantile (Q–Q) plot showing the quantiles of the internal process risk losses on the x -axis and the quantiles of a million aggregate loss sample scenarios on the y -axis. The P–P plot and Q–Q plot both indicate better fits, the closer they resemble a straight line. The plots are less sensitive to discrepancies in fit than the comparison of probability density plot, however they each have their uses. The P–P graph in Figure 4.9 clearly shows a linear relationship with no indication of any deviation of the fitted distribution to the theoretical one. It is close to the ideal

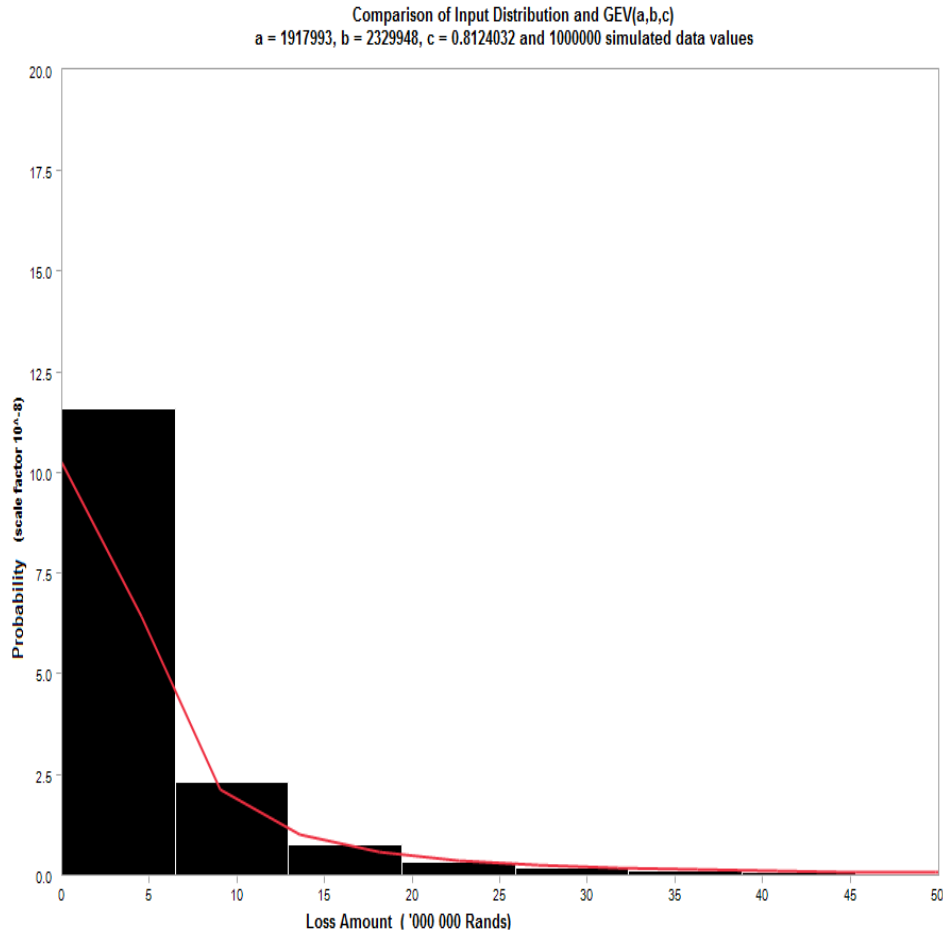


Fig. 4.7: Comparison of probability density of Monte Carlo simulations of aggregate loss data to the $GEV(a, b, c)$ distribution, where $a = 1917993, b = 2329948, c = 0.8124032$

straight line indicating a good fit. The Q-Q graph, Figure 4.10 shows a clear deviation from linearity; note that in the left tail of the plotted distribution the observations are collinear but as the x -axis values increase the y -axis values are bigger than expected, i.e. the right tails of the distribution increase more quickly (are “heavier”). From an examination of Figure 4.10 we see a concave departure from the ideal linear shape. This indicates a heavier tailed distribution whereas convexity would indicate a shorter tailed distribution.

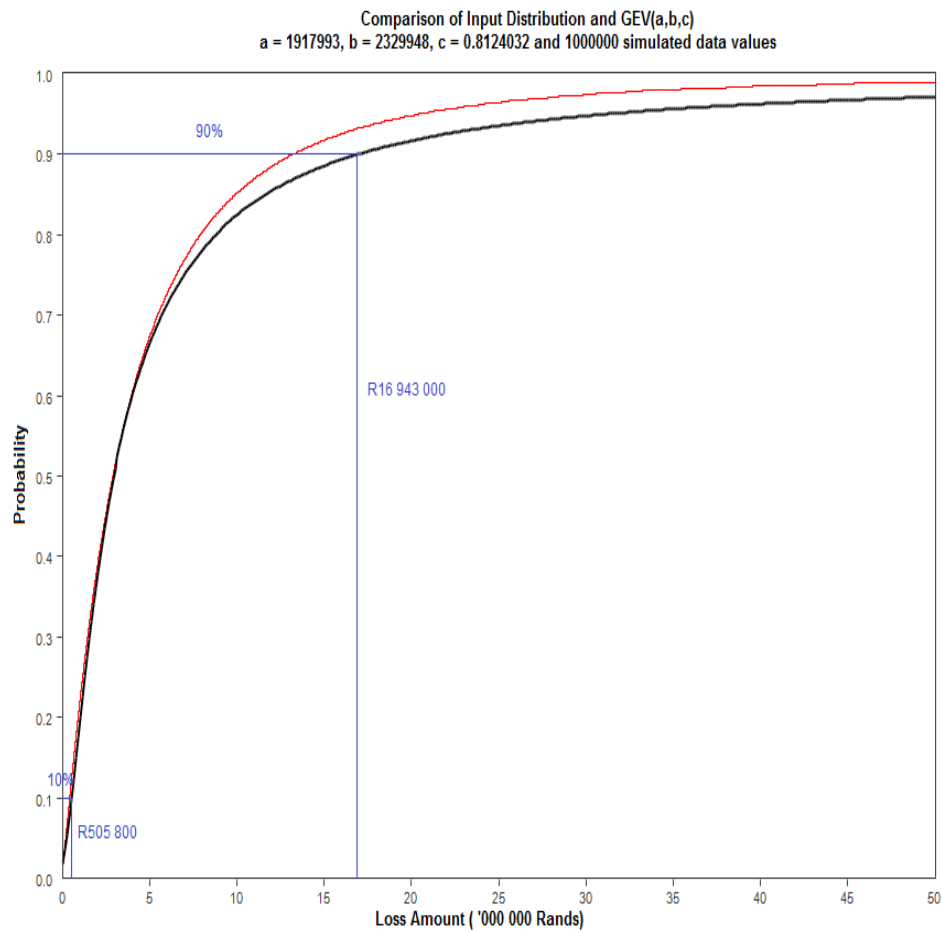


Fig. 4.8: Comparison of cumulative probability distributions of Monte Carlo simulations of aggregate loss data to the $GEV(a,b,c)$ distribution, where $a = 1917993$, $b = 2329948$, $c = 0.8124032$

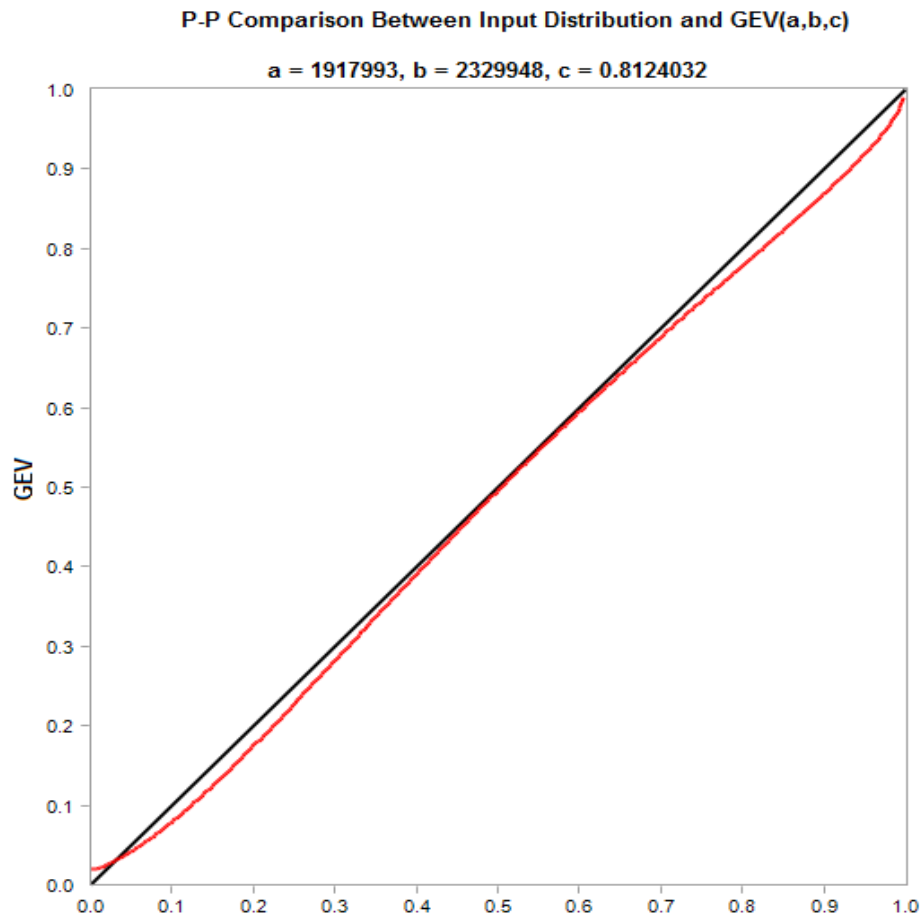


Fig. 4.9: Probability-probability plot comparison between Monte Carlo simulations of aggregate loss data and the $GEV(a,b,c)$ distribution, where $a = 1917993$, $b = 2329948$, $c = 0.8124032$

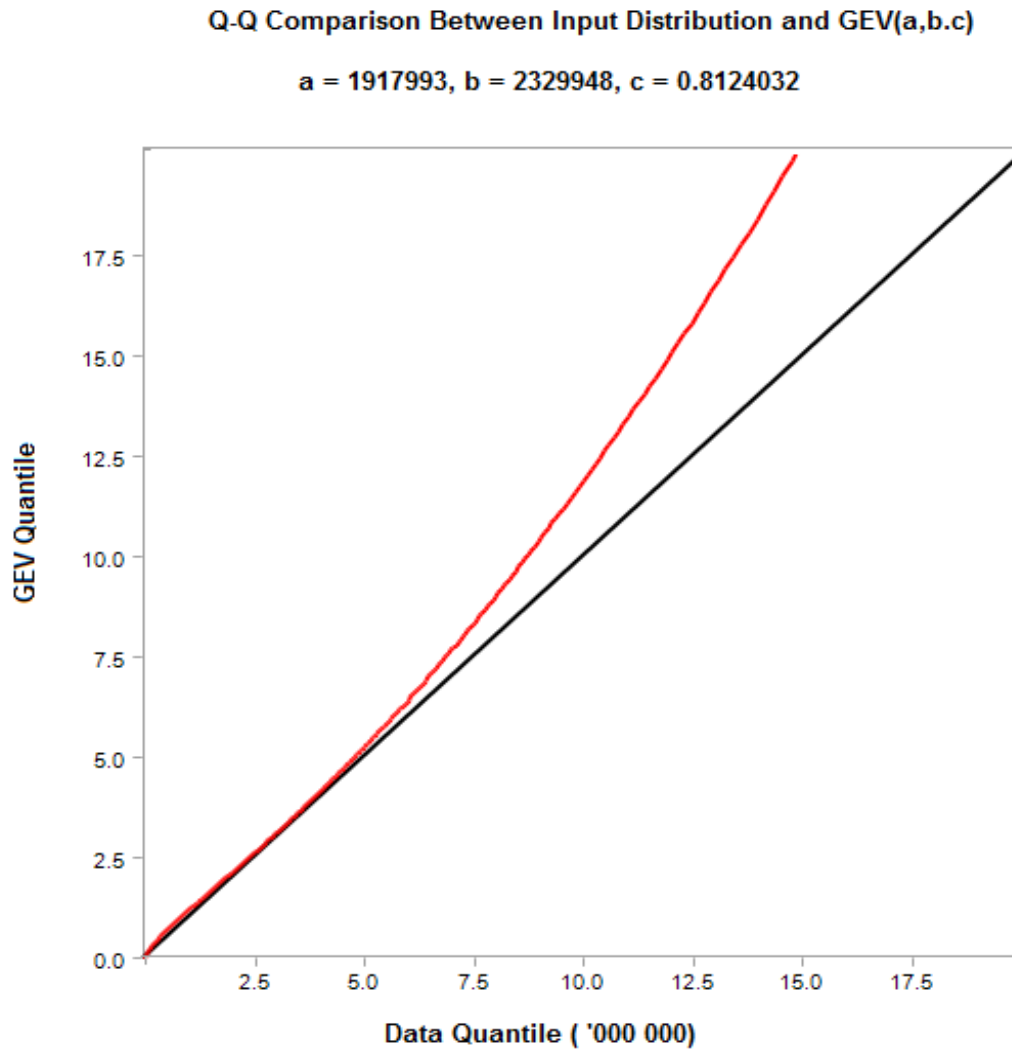


Fig. 4.10: Quantile-quantile plot comparison between Monte Carlo simulations of aggregate loss data and the $\text{GEV}(a,b,c)$ distribution, where $a = 1917993$, $b = 2329948$, $c = 0.8124032$

4.1.7 VaR Analysis

Among many other tools, **ModelRisk** has a highly technical, powerful and unique approach to extreme-value modeling, making it possible to calculate directly the probability that the largest of a million claims following a certain distribution will not exceed some value X with some defined degree of confidence. The question being asked: “What loss level is such that we are $100\%(1-\alpha)$ confident it will not be exceeded in t business days?”. We complete a VaR analysis at this stage using the $GEV(a,b,c)$ where $a = 1917993$, $b = 2329948$, $c = 0.8124032$. To complete the VaR estimation, recall that once a static distribution has been fit to the data, we can compute the 99^{th} percentile of the histogram 4.6 for the corresponding loss amount calibrated on the x -axis, i.e., **1% of the time, over the next three months the bank will not exceed a loss greater than the x -axis value (viz., R118 329 510 as depicted in the figure²).**

Our risk measure will be based on a 3 – month (quarter-of-a-year) time horizon, which is then divided by $\sqrt{(\text{no. of business days})}$ to compute the daily VaR. We calculate the VaR from a defined $100\%(1-\alpha)$ quantile, e.g. 99th percentile ($\alpha = 0.01$) left-tail confidence interval, i.e. that a 1 in 100th quarter-of-a-year event might occur. For convenience we have simulated the 95^{th} , 99^{th} , and 99.9^{th} percentiles and placed them in the model spreadsheet as seen in Appendix C.4 (using the *ModelRisk* function “VoseSimPercentile”). The results are depicted below

² The x amount in the figure (R120 000 000) is rounded up to the next 10^{th} of a million

1-day VaR from Static GEV(a,b,c) for 99.9th percentile

$$VaR_{\alpha=0.001} = \frac{R773\ 938\ 918}{\sqrt{(56)days}} = R103\ 421\ 938 \quad (4.8)$$

1-day VaR from Static GEV(a,b,c) for 99th percentile

$$VaR_{\alpha=0.01} = \frac{R118\ 329\ 510}{\sqrt{(56)days}} = R15\ 812\ 446 \quad (4.9)$$

1-day VaR from Static GEV(a,b,c) for 95th percentile

$$VaR_{\alpha=0.05} = \frac{R30\ 908\ 206}{\sqrt{(56)days}} = R4\ 130\ 283 \quad (4.10)$$

The VaR estimates decrease when the confidence interval is increased. Under **Basel II**, RC for **OpRisk** is based on a 1-year 99.9% VaR, while the confidence level consistent with EC is 99.95% or higher, e.g., VaR for economic capital:

$$R190\ 216\ 975 \times \sqrt{(252\ \text{business days})} = R3\ 019\ 600\ 866$$

The standard procedure for operational risk is to specify the economic capital as VaR - Expected Loss, i.e.,

$$\mathbf{R3.02} \text{billion} - \text{Expected Loss} = \text{Economic Capital}.$$

Chapter 5

DISCUSSION

The LDA method of AMA analysis is a fitting means of calculating the **OpRisk** VaR measure and the **ATAT** is an efficient tool used to provide sufficient internal data as it most accurately reflects internal process risk loss events. The results indicate that the $\text{GEV}(1917993, 2329948, 0.8124032)$ maximises the SIC, suggesting that it is the most appropriate model and provides the most adequate fit. Furthermore, the special case (where $c > 0$) would provide a better estimate for VaR - providing an answer to the simple question: “How bad can things get?”.

5.1 Limitations of the Study

During the peak of the financial crisis, there were numerous examples of banks providing inaccurate information to regulators. Regulators need to do more to encourage institutions to share their loss data. Aggregation of high quality data from several sources is a prerequisite if sound risk management is to be implemented within a bank or FI.

The LDA would be a successful tool but has not quite managed to be regarded as one. Historically, sources of **OpRisk** loss data are few and of low quality – the introduction of the **ATAT** is the key in our experimentation. It would be difficult to implement the LDA where data is scarce and of low quality, and where a wealth of data exists, such as where

you have a data consortium - for example the Operational RiskData eXchange Association (ORX) - which is clearly quantitative in that there are lots of different losses by the region, the product line information, the event type and the actual loss amount, bank loss data is anonymised before it is sent out, so there is very little information about the context of why the loss occurred and where the information came from.

5.2 *Validity and Reliability*

Regulators have been slow to realise the importance of effective IT systems. The **ATAT** data gathering technique attempts to address the gap experienced where FI's cannot prove loss data is flowing with a high degree of accuracy. It attempts to address a key lesson from the financial crisis - that banks', which rely on spreadsheets and manual controls to pull numbers together, are inadequate to support the broad management of financial risks. The **OpRisk** database (the **ATAT**) consists of modeled data at transaction level instead of being aggregated on a daily or monthly level. This way, the **ATAT** is capable of employing a more advanced technique of modeling the levels of risk. In June 2012, for the first time in banking regulation, the Basel Committee released a paper set out to put explicit requirements for accuracy, completeness and timeliness of risk reporting. The paper titled "Principles of Effective Risk Data Aggregation and Risk Reporting "paves the way for data integrity to drive the quality of decision making. The **ATAT** automates data validation and reliability for reduced costs and risks, while living by the principles set in the documentation of the paper. The process risk data consists of 369 losses over sixty million rand from January 2013 to March 2013, unadjusted for inflation. This period is when the loss events were recorded

through the ATAT database, hence were actual losses in real time. The dataset contains a wide range of losses with a minimum being R1 102 and the maximum being R8 752 614. The losses are posted at the most granular level (on a trade by trade basis) then aggregated on a daily level, this further indicates the wide range of losses hence the elimination of bias and non-representativeness. The ATAT is instrumental due to its ability to collect data at the most granular level. The data also comes with supporting documentation and commentary provided by end users and can therefore be traced back to uncover root causes. The data can be analysed in various formats, such as in excel, define follow up action items with due dates, and assign responsible persons for those items not yet explained by the user, providing business continuity with the benefit of an electronic solution to the management of amendments.

Figure 5.1 is a pivot chart (histogram) constructed from the input data observed in the attached excel table labeled "Input Data" in Appendix C.4. The proportion of amendments to terminations for the given period are shown and analysed. Further analysis is portrayed in Figure 5.2: Pivoted data is extracted from a source file in excel; a sample consisting of original root causes from the ATAT database is displayed. It consists of documented trade amendment details on a trade by trade basis extracted from the ATAT into excel for amendments done to trades on the 19 March, 2013. The granularity of the data is owed to the unique trade identifier (Trade field), which comes with other additional fields useful in root cause analysis. For example, "Dealer B " may have been responsible for most processing errors on this day, notably due to the reason "client request to amend economics of deal " resulting in the loss of R8 752 614. At the level of individual loss events it is fundamental that the bank knows when they happened, be able to identify the root causes of losses arising from operational

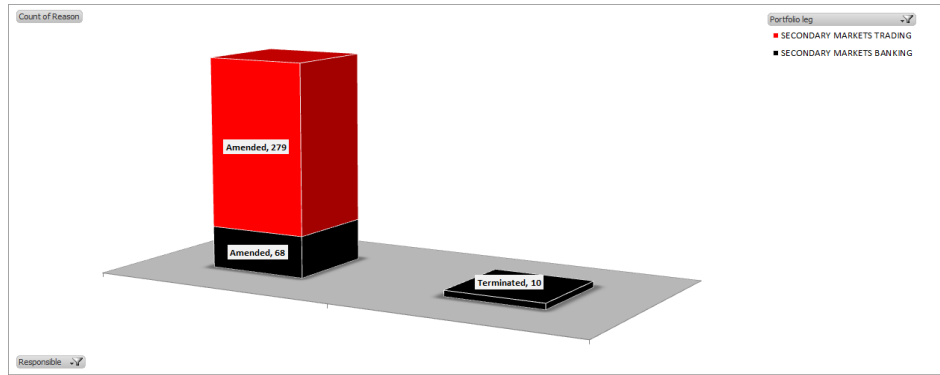


Fig. 5.1: Histogram of frequencies of loss events for amended vs terminated deals

Trade	Updated Date	Updated Time	Desk	DESK	Instrument
23154064	2013/03/19	02:05:00 PM	PCG_Credit Derivatives	CREDIT DERIVATIVES	CASH PAYMENT
26767684	2013/03/19	11:33:00 AM	PCG_Equity	EQ Derivatives	EQUITY SWAP
3635227	2013/03/19	05:39:00 PM	PCG_Rates	NLD	COMMERCIAL PAPER
25831881	2013/03/19	11:13:00 AM	PCG_Rates	IRD	INTEREST RATE SWAP
18547809	2013/03/19	08:16:00 AM	PCG_Commodities	Agris	FEES
24251679	2013/03/19	08:01:00 AM	PCG_Equity	EQ Derivatives	FEES
26549423	2013/03/19	08:27:00 AM	PCG_Prime Services	SECURITY LENDINGS	SECURITY LOAN
Trader Id	Responsible	Comment Type	Reason	Portfolio leg	SEVERITY (Loss)
Dealer A	Amended	Economic	Payments Related	BANKING	-15 145.31
Dealer B	Amended	Economic	Client Request to Amend Economics of Deal	TRADING	-8 752 614.03
Dealer F	Amended	Economic	Rate Amend	TRADING	-4 724 911.52
Dealer B	Amended	Economic	Payments Related	TRADING	-15 000.01
Dealer B	Amended	Economic	Fees/Commissions Related	TRADING	-3 277.30
Dealer H	Amended	Economic	Notional Related	TRADING	-3 038.43
System	Amended	Economic	Calendar Related	TRADING	-1 180.40

Fig. 5.2: Pivoted data of frequencies of loss events per associated severity for amended deals done on 19 March, 2013

events and attribute the losses appropriately. In addition, there are intrinsic benefits to managing OpRisk, as it maximises the certainty of meeting business objectives.

5.3 Conclusion

- OpRisk modeling is based on two underlying processes, the frequency and the severity of losses. The LDA model, which estimates the shape of the frequency of loss events and

the severity of individual events, provides a realistic measure of **OpRisk** VaR provided the input data is reliable. This measure of **OpRisk** VaR is used to determine economic capital for the bank in question. A bank's internal operations are unique to the bank, hence employing the **ATAT** in targeting their risk measure goes a long way in addressing their competitiveness as it opens up more decision making options which in turn may raise the banks' investability.

- The investment banking environment is an action orientated industry where those whose skills and experience are highly treasured (and likewise remunerated) whose decisive thinking and seasoned judgement very often materialises into profits. The **ATAT** through the LDA not only allows for an action orientated risk unit which can quickly communicate risk events, but can also quantify an important aspect of risk in a single number – VaR.
- It can be seen that the $GEV(a,b,c)$ distribution, maximises the SIC, suggesting that it is the most appropriate model for the aggregate loss distribution for process risk. From visual inspection and the goodness of fit test, it would seem that the $GEV(a,b,c)$ distribution; $a = 1917993, b = 2329948, c = 0.8124032$, does provide an adequate fit to the data, i.e., it covers both the initial loss data and the tails. The P-P plot closely resembles a straight line which indicates that the $GEV(a,b,c)$ distribution assumption is an appropriate one. The Q-Q plot has a general departure from the fourth quartile which suggests that the loss dataset would be a heavier tailed distribution. This is consistent with the theory behind extreme value modeling and suggests that operational risk data can be modeled as such.

- Previous studies have focused the use of extreme value modeling in a wide variety of applications due to its ability to model data that is extreme in severity as well as rare. They did this in a manner which would benefit from a distribution that overestimates rather than underestimates the capital adequacy requirement, as it would be better to have excess capital provisions in the event of a catastrophic loss. More specifically, previous practitioners had to look externally for good quality datasets as data was rare and of low quality, and even then, they had to determine thresholds or optimal cut off points, where data would be truncated. The mathematical reasons why optimal threshold selection is very difficult is best appreciated by Chavez–Demoulin et al (2006). I am improving on this by virtue of the use of an automated data gathering technique, in the form of the **ATAT** device; supporting the logic pointed out by Zhaoyang (2013), that recent advances in risk analysis and management should be able to better distinguish the “bads ”from the “goods ”in the discussion of **OpRisk** measurement, which eliminates the reliance on external loss databases and professional guesses. The **ATAT** dataset comes from an automated feed therefore has arguably higher reliability and is more complete. This argument is evidenced through the consistent findings for loss distributions in this study and those found in theory.

5.4 Future Research

This study has demonstrated how an automated device (the **ATAT**), in conjunction with the Loss Distribution Approach (LDA) as a measurement method for operational risk, can be used to determine VaR with a high degree of certainty, which in turn determines

the amount of capital needed to absorb unexpected losses. What we haven't done is to determine the exact meaning of "high degree of certainty", in fact, it differs as it depends on the individual risk tolerance of each bank. It is obvious that this problem becomes particularly challenging for OpRisk and is one of the reasons why internal data needs to be supplemented by external data (hence the inclusion of the term "external events" in the definition of OpRisk). Practically, it means an appropriate mix between internal and external data must be imposed, in order to enhance statistical efficiency. This process will almost surely benefit from experienced professionals in the qualitative area of OpRisk, stressing a crucial element: **Quantitative measures of fit cannot substitute for the use of judgment in modeling.**

The benefits of future research in the suggested area would be the determination of an optimal mix of internal and external data as the current Basel Committee document does not provide any solution to this issue.

BIBLIOGRAPHY

- Aue, F., & Kalkbrener, M.(2001). “*LDA at work: Deutsche bank’s approach to quantifying operational risk*”. Journal of Operational Risk 1(4), 49-93.
- Basel Committee on Banking Supervision.(2001) “*Operational risk*”. Basel: Bank fo
- Basel Committee on Banking Supervision. (2001, 4 May). “*QIS 2 - Operational risk loss data*” Retrieved 09 January, 2014, from <http://www.bis.org/bcbs/qisoprisknote.pdf>
r International Settlements.
- Chartis Research Ltd.(2013, 8 July). “*Operational risk management systems for financial services 2013: Thomson Reuters vendor highlights*” Retrieved 14 January, 2014, from url<http://www.risklibrary.net/download/chartis-report-operational-risk-management-systems-financial-services-2013-17676>
- Chavez–Demoulin, V., Embrechts, P., & Neslehova, J.(2006). “*Quantitative models for operational risk: Extremes, dependence and aggregation*”. Journal of Banking & Finance 30(10), 2635-2658.
- Chavez–Demoulin, V., Embrechts, P., & Neslehova, J.(2006). “*Quantitative models for operational risk: Extremes, dependence and aggregation*”. Journal of Banking & Finance 30(10), 2635-2658.

- CheckPoint Technologies (2011, 25 November 2011). “*How to effectively manage operational risk: For Basel II, Solvency II and arrow*” Retrieved 9 February 2013, 2011 from <http://www.checkpoint.com/grc/downloads/whitepapers/wp-manage-oper-risk.pdf>
- Cruz, M.(2002). “*Modeling, measuring and hedging operational risk*”. (First ed.).West Sussex: John Wiley & Sons, Ltd.
- Cruz, M.,Coleman, R., & Salkin, G.(1998). “*Modelling and measuring operational risk*”. Journal of Risk 1(4), 63-72.
- Cummings, D., Christopher, M., & Ran, W.(2006). “*The market value loss of operational loss events for US bankers and insurers*”. Journal of Banking and Finance 30, 2605-2634.
- Cummings, D., Christopher, M., & Ran, W.(2006). “*The market value loss of operational loss events for US bankers and insurers*”. Journal of Banking and Finance 30, 2605-2634.
- Embrechts, P., Kluppelberg, C.,& Mikosch, T. “*Modelling extremal events for insurance and finance*”. Berlin: Springer.
- Embrechts, P.,McNeil, A.,& Straumann, D.(2000). “*Correlation and dependency in risk management: properties and pitfalls* ”. Great Britain: Cambridge University Press.
- Eun, C., Reswick, B., & Sabwerhal, S.(2012). “*International Financial Management*”. (Sixth ed.).New York: John Wiley & Sons, Inc.
- Feller, W.(1950). “*An introduction to probability theory and its applications*”. (Third ed.).New York: John Wiley & Sons, Inc.

-
- Frachot, A., Georges, P., & Roncalli, T.(2001). “*Loss distribution approach for operational risk*”. Journal of Operational Risk 2(4), 103-109.
- Frachot, A., Georges, P., & Roncalli, T.(2001). “*Loss distribution approach for operational risk*”. Journal of Operational Risk 2(4), 103-109.
- Fraleigh, J.(2000). “*A first course in abstract algebra*”. (Sixth ed.).Massachusetts: Addison-Wesley Publishing Company, Inc.
- Hull, J.(2012). “*Risk management and financial institutions*”. (Third ed.).New Jersey: John Wiley & Sons, Inc.
- King, J.(2001). “*Operational risk: measuring and modelling*”. West Sussex: John Wiley & Sons, Ltd.
- King, M.(1999). “*Back office and beyond: A guide to procedures, settlements and risk in financial markets*”. (Second ed.).Great Britain: Gilmour Drummond Publishing.
- Klugman, S., Panjer, H., & Willmot, G.(2004). “*Loss models: from data to decisions*”. (Second ed.).New Jersey: John Wiley & Sons, Inc.
- KPMG International Cooperative. (2014, 13 January). “*Developing a strong risk management framework – challenges and solutions*”Retrieved 14 January, 2014, from [urlhttp://www.risklibrary.net/download/developing-strong-risk-appetite-program-challenges-solutions-19609](http://www.risklibrary.net/download/developing-strong-risk-appetite-program-challenges-solutions-19609)
- Lehman, D., Groenendaal, H., & Nolder, G.(2012). “*Practical spreadsheet risk modeling for management*”. (First ed.).Boca Raton: Taylor & Francis Group, LLC.

-
- ORX(2012). “*2012 ORX report on operational risk loss data*”. Operational Riskdata eXchange, Zurich.
- Panjer, H.(2006). “*Operational risk: modeling analytics*”. New Jersey: John Wiley & Sons, Inc.
- Urbina, J., & Guillen, M.(2014). “*An application of capital allocation principles to operational risk and the cost of fraud*”. Journal of Banking & Finance 41(16), 7023-7031.
- Zhaoyang, L.(2013). “*Measuring the capital charge for operational risk of a bank with the large deviation approach*”. Mathematical and Computer Modelling 58(9–10), 1634-1647.

Appendices

Appendix A

GOODNESS-OF-FIT PLOTS AND STATISTICS

A.1 Fitting a Discrete Parametric Distribution to BBL Loss Data:

Geometric Variate $\mathbf{G} : p, p = 0.4179105$

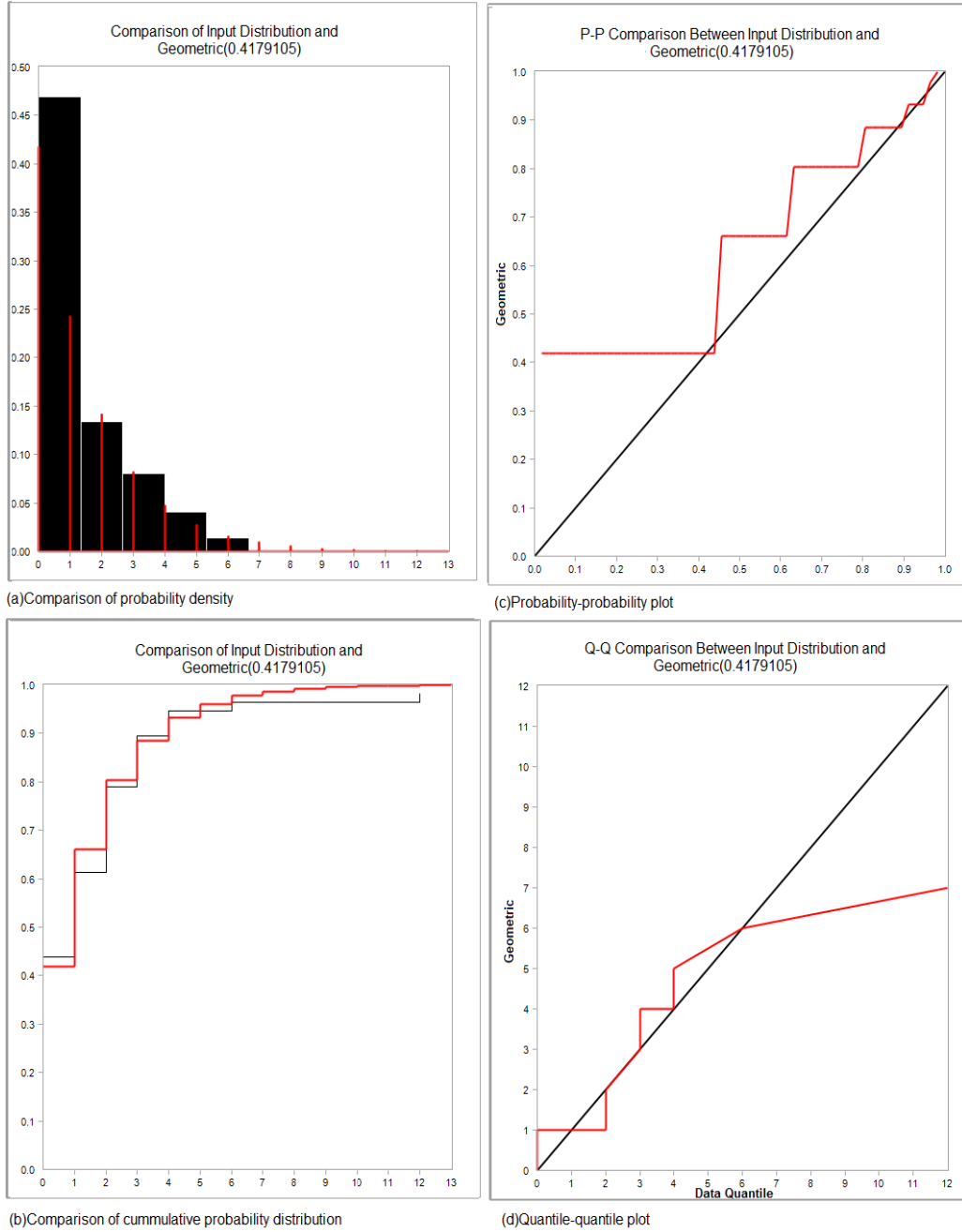


Fig. A.1: (a) Overlay of a histogram plot of the data with a density function of the fitted distribution. (b) Overlay of the cumulative frequency plots of the data and the fitted distribution. (c)(P-P) plot (d) (Q-Q) plot.

A.2 Fitting a Discrete Parametric Distribution to TBL Loss Data: Polya

Variate $Polya(\alpha, \beta)$, $\alpha = 6.557774$, $\beta = 0.7924074$

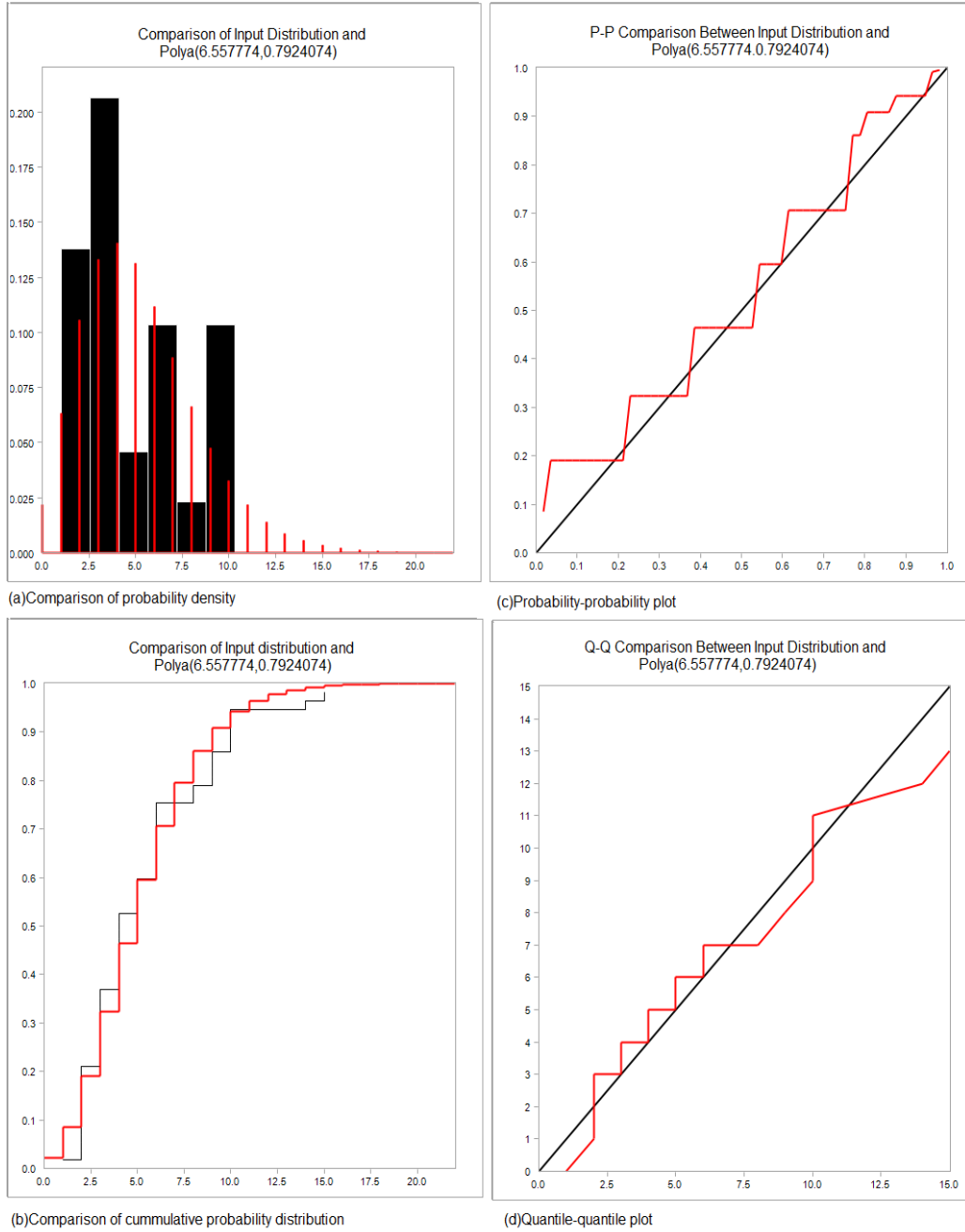


Fig. A.2: Overlay of a histogram plot of the data with a density function of the fitted distribution.

(b) Overlay of the cumulative frequency plots of the data and the fitted distribution.

(c)(P-P) plot (d) (Q-Q) plot.

A.3 Fitting a Continuous Parametric Distribution to BBL Loss Data:

$$\text{Lognormal}(\alpha, \beta), \alpha = 411736.7, \beta = 2271727$$

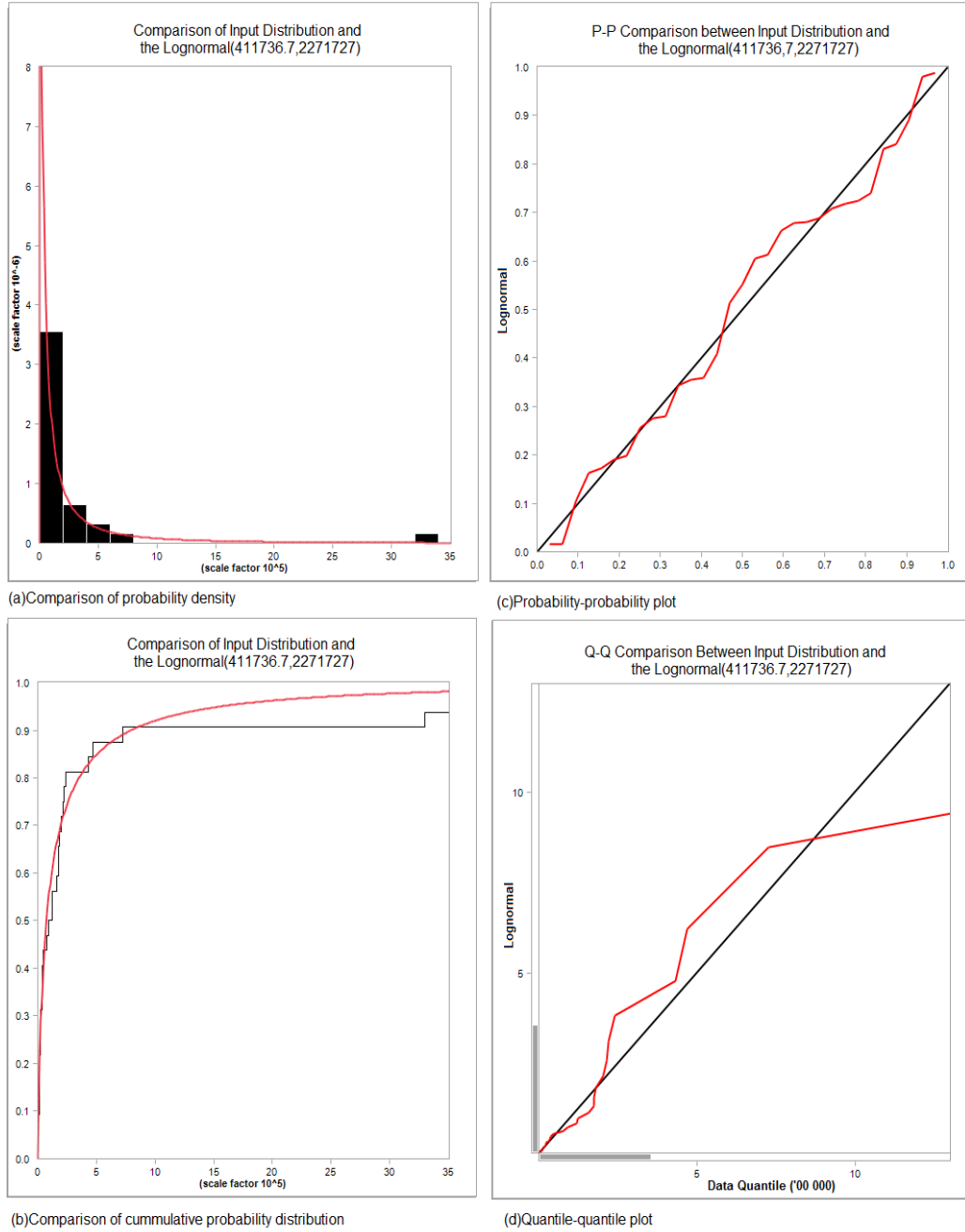


Fig. A.3: Overlay of a histogram plot of the data with a density function of the fitted distribution.

(b) Overlay of the cumulative frequency plots of the data and the fitted distribution.

(c)(P–P) plot (d) (Q–Q) plot.

A.4 Fitting a Continuous Parametric Distribution to TBL Loss Data:

$$\text{Lognormal}(\alpha, \beta), \alpha = 1043960, \beta = 6121955$$

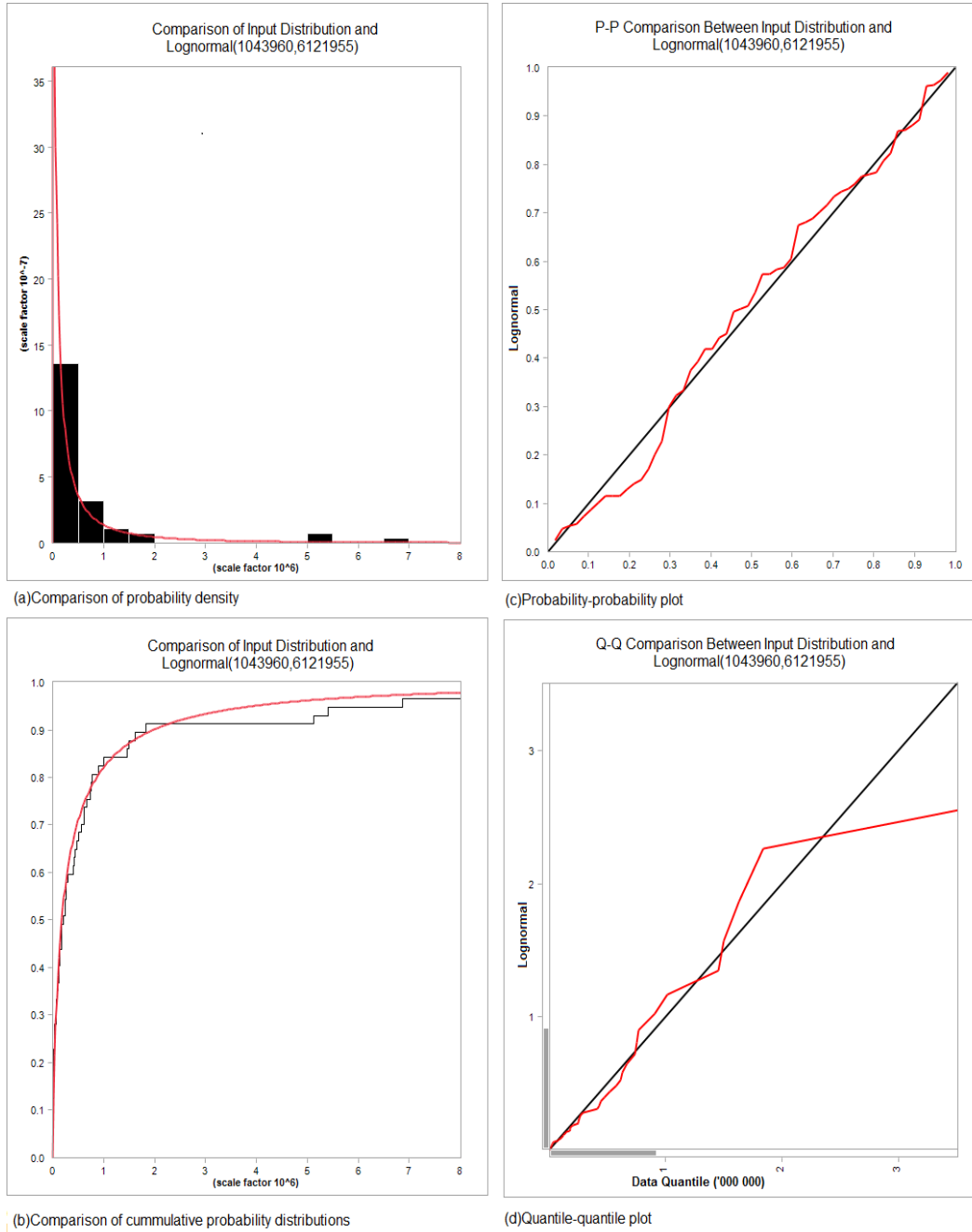


Fig. A.4: Overlay of a histogram plot of the data with a density function of the fitted distribution.

(b) Overlay of the cumulative frequency plots of the data and the fitted distribution.

(c)(P-P) plot (d) (Q-Q) plot.

A.5 Displaying information criterion for the optimally ranked (according to SIC) estimated best fit frequency and severity distributions

Fitting a range of distributions to the frequency data set and ranked by the Information Criteria statistics.				Fitting a range of distributions to the severity data set and ranked by the Information Criteria statistics.			
The table on the left shows the distributions tested against the Banking data, and to the right are distributions tested against the Trading data set :				Refer to prior description (on the left)			
The optimal fitting distribution is highlighted at the top							
Name	-SIC	-AIC	-HQIC	Name	-SIC	-AIC	-HQIC
Geometric	-186.160	-184.209	-184.920	Lognormal	-1590.617	-1586.795	-1588.11
ZIGeometric	-189.865	-186.040	-187.384	Pearson6	-831.265	-827.852	-828.366
Polya	-190.019	-186.195	-187.539	Weibull	-832.775	-830.336	-830.842
llegbin	-190.961	-187.137	-188.481	JohnsonB	-833.874	-829.676	-830.007
ZLogarithmic	-193.267	-189.443	-190.787	Gamma	-839.694	-837.255	-837.761
Delaporte	-194.045	-188.430	-190.324	Pearson5	-840.091	-837.652	-838.158
BetaBinomial	-194.045	-188.430	-190.324	Pareto	-849.817	-847.378	-847.884
ZIDelaporte	-194.407	-187.090	-189.447				
Zllegbin	-196.781	-191.167	-193.061				
ZIPoisson	-198.694	-194.870	-196.214				
PoissonUnifor	-200.587	-196.763	-198.107				
ZBinomial	-203.913	-198.299	-200.193				
Poisson	-215.899	-213.948	-214.659				
BurntFingerPr	-219.924	-216.100	-217.444				
Binomial	-219.925	-216.101	-217.445				

Name	-SIC	-AIC	-HQIC
Polya	-281.681	-277.856	-279.201
llegbin	-282.418	-278.594	-279.938
PoissonUnifor	-282.734	-278.910	-280.254
Delaporte	-284.035	-278.421	-280.315
BetaBinomial	-285.722	-280.108	-282.002
Poisson	-290.396	-288.445	-289.156
Binomial	-294.424	-290.599	-291.944
Geometric	-310.743	-308.792	-309.503
Logarithmic	-320.476	-318.525	-319.236
BurntFingerPr	-	-	-
ZBinomial	-	-	-
ZIGeometric	-	-	-
ZLogarithmic	-	-	-
Zllegbin	-	-	-
ZIPoisson	-	-	-

The best fit distributions according to the **SIC** for the separate frequency and severity data elements are shown above

The optimal fit is highlighted in blue

Higher numbers (lower absolute numbers if they are negative) indicate better fitting distributions

Fig. A.5: Ranking the best fit distribution according to SIC

Appendix B

DESCRIPTIVE STATISTICS TABLES

B.1 Table B.1 generated by Excel2LaTeX from excel sheet '1.Statistics'

B.2 Table B.2 generated by Excel2LaTeX from sheet '1.Statistics'

Range Name	Frequency Input		Severity Input	
Variable Name	Banking	Trading	Banking	Trading
Mean	1	5	415 097	1 043 233
Minimum	-	-	13	32
Maximum	24	29	1 177 155 760	2 473 823 580
St. dev.	2	3	2 463 135	6 503 148
Variance	3	9	6 067 032 552 934	42 290 932 008 098
CofV	1	1	6	6
Skewness	2	1	148	117
Kurtosis	9	4	56 819	32 209
Percentiles				
1%	-	-	970	2 151
8%	-	1	5 431	12 326
20%	-	3	15 433	35 681
35%	-	4	35 997	84 538
50%	1	5	73 581	174 974
65%	1	6	150 194	362 557
80%	2	8	350 837	856 747
95%	5	11	1 561 436	3 905 741
99%	8	14	5 579 903	14 107 303

Tab. B.1: Descriptive Statistics of Input Distributions

Range Name	Aggregate Loss Distribution Estimation	
Variable Name	Empirical (Monte Carlo)	GEV(a,b,c)
Mean	5 998 306	14 785 607
Minimum	-	-613 897
Maximum	2 007 127 700	2 441 858 287 755
St. dev.	14 106 490	2 488 330 557
Variance	198 993 047 358 932	6 191 788 962 879 870 000
CofV	2	168
Skewness	26	950
Kurtosis	1 741	928 106
Percentiles		
1%	4 219	-119 386
8%	303 004	400 089
20%	871 494	995 734
35%	1 723 883	1 805 751
50%	2 874 371	2 910 876
65%	4 602 214	4 730 948
80%	7 849 658	8 740 058
95%	120 192 540	31 129 081
99%	49 259 266	119 991 172

Tab. B.2: Descriptive Statistics of Output Distributions

Appendix C

MISCELLANEOUS FIGURES

C.1 A sample ATAT database

C.2 Data request letter

C.3 The Aggregate Multi Monte Carlo Loss Model for Process Risk

C.4 Raw Data for Process Risk

The Graduate School of Business Administration

2 St David's Place, Parktown, Johannesburg, 2193, South Africa
PO Box 98, WITS, 2050
Telephone: +27 11 717 3600, Facsimile: + 27 11 717-3514
Website: www.wbs.ac.za



Dear

Request for your assistance.

Mr Mphekeleli Hoohlo is currently enrolled as a candidate for the Master of Management in Finance and Investment (MMFI) degree at Wits Business School. The degree requires a set of core courses, and a programme thesis; with a 12-month, full time completion period.

Please be kind to support Mr Mphekeleli Hoohlo with the assistance he seeks in order to successfully complete the MMFI programme.

I am available should you require additional information.

A handwritten signature in black ink, appearing to read 'Kalu Ojah'.

Kalu Ojah, PhD

Professor of Finance

Director: Mater of Finance and Investment

Sculpting Global Leaders

Fig. C.1: Letter of motivation in support of data request

Front Arena Amendments - [Amendments]

File Edit Insert Records Window Help

Front Arena Amendments Import from XML Export to Excel Close History Display date for 2008-12-30 Apply date

Trade No	Updated	User	Portfolio	Acquirer	Trade Status	Trade date	Responsible	Type	Field	Desk Name
3873468	2008-12-30 15:24	ABLO082	VOE	NLD DESK	Void	2008-12-29 13:45:23	Front office	Voided	Other	Options
3873466	2008-12-30 15:24	ABLO082	VOE	NLD DESK	Void	2008-12-29 13:44:42	Front office	Voided	Other	Options
3873458	2008-12-30 15:24	ABLO082	VOE	NLD DESK	Void	2008-12-29 00:00:00	Front office	Voided	Other	Options
3873490	2008-12-30 15:23	ABLO082	VOE	NLD DESK	Void	2008-12-29 13:48:28	Front office	Voided	Other	Options
3873486	2008-12-30 15:23	ABLO082	VOE	NLD DESK	Void	2008-12-29 13:47:44	Front office	Voided	Other	Options
3873473	2008-12-30 15:23	ABLO082	VOE	NLD DESK	Void	2008-12-29 13:45:52	Front office	Voided	Other	Options
3873481	2008-12-30 15:23	ABLO082	VOE	NLD DESK	Void	2008-12-29 13:46:58	Front office	Voided	Other	Options
3668571	2008-12-30 15:22	ABEV024	GT Funding	Funding Desk	BO Confirmed	2008-11-11 16:37:14	Front office	Amended	Calendar	Money Market
3668570	2008-12-30 15:22	ABEV024	GROUP FUND 2476	Funding Desk	BO Confirmed	2008-11-11 16:37:14	Front office	Amended	Calendar	Money Market
2708947	2008-12-30 14:59	ABHL180	ABL7_OTC	Agria Desk	BO-BO Confirmed	2008-07-15 13:06:56	Front office	Amended	Rate	Commodities
2708363	2008-12-30 14:58	ABHL180	ABL7_OTC	Agria Desk	BO-BO Confirmed	2008-07-15 12:22:32	Front office	Amended	Rate	Commodities
2708420	2008-12-30 14:58	ABHL180	ABL7_OTC	Agria Desk	BO-BO Confirmed	2008-07-15 12:27:18	Front office	Amended	Rate	Commodities
3821387	2008-12-30 13:58	ABGM586	47274_CFD_ZERO	EQ Derivatives Desk	Void	2008-12-12 00:00:00	Front office	Amended	Notional	Equities - CFDs
3821386	2008-12-30 13:57	ABGM586	47274_CFD_ZERO	EQ Derivatives Desk	Void	2008-12-12 00:00:00	Front office	Amended	Notional	Equities - CFDs
1078400	2008-12-30 13:44	ABKCT22	CPI	IRD DESK	Terminated	2007-07-27 10:15:01	Front office	Terminated	Other	Fixed Income - IRD
681784	2008-12-30 13:15	ABEV024	LIABILITIES 9830	Funding Desk	BO Confirmed	2007-06-27 00:00:00	Front office	Amended	Other	Money Market
3870596	2008-12-30 12:47	ABGM586	47274_CFD_ZERO	EQ Derivatives Desk	Void	2008-12-29 00:00:00	Front office	Amended	Notional	Equities - CFDs
3811743	2008-12-30 12:45	ABGM586	47332_CFD_ZERO	EQ Derivatives Desk	Void	2008-12-29 00:00:00	Front office	Amended	Notional	Equities - CFDs
3870595	2008-12-30 12:45	ABGM586	47274_CFD_ZERO	EQ Derivatives Desk	Void	2008-12-29 00:00:00	Front office	Amended	Notional	Equities - CFDs
3420527	2008-12-30 12:36	ABKCT22	Swap Flow	IRD DESK	BO Confirmed	2008-10-22 15:38:51	System	Amended	Rate	Fixed Income - IRD
3863854	2008-12-30 12:31	ABLO082	PT_Cur	NLD DESK	BO-BO Confirmed	2008-12-24 10:00:00	Front office	Amended	Portfolio move	Options
512105	2008-12-30 11:48	ABKCT22	Swap Flow	IRD DESK	BO Confirmed	2005-03-31 15:44:51	System	Amended	Rate	Fixed Income - IRD
642586	2008-12-30 11:44	ABKCT22	IRP 2304	IRD DESK	BO Confirmed	2005-03-31 00:00:00	Front office	Amended	Other	Structured Notes
2383276	2008-12-30 10:47	ABDM580	LIABILITIES 2474	Funding Desk	BO Confirmed	2008-05-23 00:00:00	PCG	Amended	Other	Money Market
3772588	2008-12-30 10:18	ABKCT22	Swap Flow	IRD DESK	BO Confirmed	2008-12-08 14:04:53	Front office	Amended	Notional	Fixed Income - IRD
3787775	2008-12-30 10:15	ABEV024	LIABILITIES 2474	Funding Desk	Void	2008-12-11 10:31:54	Front office	Voided	Counterparty	Money Market
3863185	2008-12-30 09:54	ABKR173	Allocate_Ptolo_Equities	EQ Derivatives Desk	Void	2008-12-23 08:41:01	System	Voided	Other	Equities
3867326	2008-12-30 09:54	ABKR173	Allocate_Ptolo_Equities	EQ Derivatives Desk	Void	2008-12-22 10:46:35	System	Voided	Other	Equities
3871076	2008-12-30 09:53	ABKR173	Allocate_Ptolo_Equities	EQ Derivatives Desk	Void	2008-12-29 16:37:38	System	Voided	Other	Equities
3863166	2008-12-30 09:53	ABKR173	Allocate_Ptolo_Equities	EQ Derivatives Desk	Void	2008-12-23 08:42:20	System	Voided	Other	Equities
3138386	2008-12-30 09:53	ABEV024	LIABILITIES 2474	Funding Desk	BO Confirmed	2008-09-17 09:56:20	Front office	Amended	Other	Money Market

Amendment Detail

Key	Name	Updated Field	Before	After	Date Amended
3124750	Reset	read_time	11/20/08 00:00:00	12/30/08 12:34:26	2008-12-30 12:36
3124750	Reset	value	11.979	11.933	2008-12-30 12:36

Record: 14 of 54 (Filtered)

Comments/Reasons:

Related Trade: 3420527

Trade Status: BO Confirmed

Responsible: Front office

Type: Amended

Field: Calendar

Update all selected

Update selected record only

Record: 14 of 54

Portfolio move

Premium

Sales Adaptor

Fig. C.2: The automated trade amendment tracker (ATAT) loss database platform

ModelRisk Simulation for the Loss Distribution Approach Model

Date: Thursday, January 16, 2013 19:57

By: Mphekeleli

Model Inputs	No. of Days	56	
	Frequency Copula	VoseCopulaBiClayton(0.119752236751549,1)	
	Copula Values	0.672102737189410	0.862275709453428
	Frequency Severity	VoseGeometric(0.4179105) VoseLognormal(411736.613169086,2271726.50488002)	VosePolya(6.55777416696413,0.792407368004488) VoseLognormal(1043959.94607748,6121953.15802933)
Input Values	Banking Frequency		1
	Trading Frequency		7
	Banking Severity		8 546
	Trading Severity		19 038
Static Fit	AggregateMultiMC	170 796	
Overlay Fit	VoseGEV(1917993.1903)	3 693 347	
Var from Overlay Fit	99.9%	103 421 938	
	99%	15 812 446	
	95%	4 130 283	
Frequency			
Optimal Fit	VoseGeometric(0.417910438820497)	VosePolya(6.5577752128412,0.79240725879523)	
Distributions Tested	BetaBinomial	VoseBetaBinomial(2708170,0.855472600157359,1662835.34372939)	VoseBetaBinomial(11428128,6.22758321744627,13670425.4700538)
	Binomial	VoseBinomial(248002,5.6162133293302e-006)	VoseBinomial(128500,4.04363695766372e-005)
	Delaporte	VoseDelaporte(0.856451342239587,1.62631200489334,3.86472376522869e-013)	VoseDelaporte(0.938427263131033,2.39621996696091,2.94775035686402)
	Geometric	VoseGeometric(0.417910438820497)	VoseGeometric(0.161383285302594)
	NegBin	VoseNegBin(1,0.380762080881861)	VoseNegBin(7,0.557908904352968)
	Poisson	VosePoisson(1.39285712497575)	VosePoisson(5.19642861911229)
	PoissonUniform	VosePoissonUniform(6.289978953888899e-014,2.975305845499624)	VosePoissonUniform(2.065761930400693,8.523783334886481)
	Polya	VosePolya(0.856451379573607,1.62631197773167)	VosePolya(6.5577752128412,0.79240725879523)
	ZIBinomial	VoseZIBinomial(109234,2.30318871281635e-005,0.553571428571429)	Error: Unable to fit
	ZIGeometric	VoseZIGeometric(0.39743588592409,0.918693005266964)	Error: Unable to fit
	ZILogarithmic	VoseZILogarithmic(0.803874461373516,0.553571429111186)	Error: Unable to fit
	ZINegBin	VoseZINegBin(7,0.767392026544829,0.6564493648502)	Error: Unable to fit
	ZIPoisson	VoseZIPoisson(2.25126671100532,0.618699285462823)	Error: Unable to fit
Severity			
Optimal Fit	VoseLognormal(411736.613169086,2271726.50488002)	VoseLognormal(1043959.94607748,6121953.15802933)	
Distributions Tested	Gamma	VoseGamma(0.396158038050817,978370.187166381)	VoseGamma(0.405938686469785,2178349.93195117)
	JohnsonB	VoseJohnsonB(1.92524300710178,0.430116621286791,1104.7224265209,6145368.56341771)	VoseJohnsonB(2.22588385870415,0.456646391659896,3863.18998706466,21110566)
	Lognormal	VoseLognormal(411736.613169086,2271726.50488002)	VoseLognormal(1043959.94607748,6121953.15802933)
	Pareto	VosePareto(0.248606376246088,1315.06840000302)	VosePareto(0.268519685789422,4235.33443921042)
	Pearson5	VosePearson5(0.392859312207403,5377.04811758497)	VosePearson5(0.444454904731375,18190.2009765911)
	Pearson6	VosePearson6(0.89085422273348,1.01625320587668,91667.1129204916)	VosePearson6(0.803307352218347,1.00300621071693,257940.761807663)
	Weibull	VoseWeibull(0.540767795856767,185876.583144481)	VoseWeibull(0.545651725422381,452217.488016642)

Fig. C.3: Model-Process-Risk.xlsx

Input Data : 02 January 2013 - 20 March 2013				
Date	Frequency		Severity	
	Banking	Trading	Banking	Trading
2013/01/02	0	4	4 798 046	13 500 022
2013/01/03	0	2	3 300 844	6 869 039
2013/01/04	0	10	727 771	5 404 464
2013/01/07	0	3	470 116	5 125 452
2013/01/08	1	2	433 954	1 837 828
2013/01/09	2	4	241 500	1 626 940
2013/01/10	4	2	222 009	1 499 017
2013/01/11	2	10	215 153	1 454 307
2013/01/14	3	9	202 886	1 012 214
2013/01/15	2	2	182 283	901 401
2013/01/16	1	2	176 137	768 993
2013/01/17	0	9	174 372	752 782
2013/01/18	3	2	159 659	733 552
2013/01/21	2	3	125 370	667 771
2013/01/22	0	4	120 984	624 031
2013/01/23	0	6	93 671	609 064
2013/01/24	12	8	78 265	571 509
2013/01/25	1	6	48 048	519 168
2013/01/28	1	4	37 630	476 592
2013/01/29	0	5	36 877	444 169
2013/01/30	3	5	34 815	425 701
2013/01/31	0	2	24 739	413 148
2013/02/01	4	1	24 295	291 578
2013/02/04	0	6	21 842	265 845
2013/02/05	0	6	15 145	260 161
2013/02/06	1	5	14 466	249 818
2013/02/07	0	4	12 715	248 080
2013/02/08	2	5	11 792	206 959
2013/02/11	0	3	7 233	181 472
2013/02/12	3	3	1 334	177 391
2013/02/13	1	14	1 315	172 326
2013/02/14	0	3	0	138 893
2013/02/15	2	6	0	133 281
2013/02/18	1	3	0	118 890
2013/02/19	0	3	0	118 848
2013/02/20	0	2	0	105 174
2013/02/21	6	9	0	95 930
2013/02/22	3	3	0	77 837
2013/02/25	0	2	0	73 947
2013/02/26	0	2	0	65 349
2013/02/27	0	6	0	42 884
2013/02/28	1	10	0	35 814
2013/03/01	3	10	0	29 252
2013/03/04	0	6	0	24 351
2013/03/05	4	4	0	22 833
2013/03/06	2	15	0	20 881
2013/03/07	0	2	0	18 342
2013/03/08	1	3	0	18 258
2013/03/11	2	4	0	18 245
2013/03/12	0	4	0	15 880
2013/03/13	0	9	0	13 414
2013/03/14	2	6	0	11 214
2013/03/15	2	10	0	8 893
2013/03/18	0	8	0	8 449
2013/03/19	1	6	0	7 595
2013/03/20	0	4	0	4 235

Fig. C.4: Process-Risk-Data.xlsx