Royal Docks Business School

UEL University of East London

MKM227 Postgraduate Dissertation

Student Number:.........U1043853.........

	Comments	Max Mark	Actual Mark
Introduction *Identification of a valid topic, research question and objectives framed to Masters Level standard with academic rationale developed, clear industry contextualisation of the research topic*	Supplier Comments:	10%	
	2nd marker Comments:		
Critical Literature Review *Depth and breadth of*	Supervisor Comments:	25%	

literature search, engagement with seminal authors and papers, evidence of a critical approach toward the scholarly literature	2nd marker Comments:		
Research Methodology *Evaluation of research philosophies and perspectives. Justification of methodological approach, sampling strategy, data analysis and reliability and validity measures as applicable*	Supervisor Comments:	15%	
	2nd marker Comments:		
	Supervisor Comments:	35%	

Data Analysis and Interpretation *Evidence of rigor in data analysis and interpretation procedures, identification of key patterns and themes in the research data, integration of academic theory into explanation of findings*	 2nd marker Comments:		
Conclusions and Recommendations *Research question and objectives addressed with implications to theoretical and managerial concepts considered. Recommendations provided for theory, practice and future research*	Supervisor Comments: 2nd marker Comments:	10%	

Organisation, presentation and references. *Well structured and ordered dissertation with correct use of grammar and syntax. In-text citation and bibliography conforming to "Cite Them Right"*	Supervisor Comments:	5%	
	2nd marker Comments:		
Total	First Marker Total	100%	
	Second Marker Total		

Supervisor General Comments:	Agreed Mark:
2nd Marker General Comments:	

Supervisor's Name: ……………………………………….. **Signature:** …………………………

2nd **Marker's Name:** ………………………………………. **Signature:** …………………………

[Modelling the Frequency of Operational Risk Losses under the Basel II Capital Accord:
A Comparative study of Poisson and Negative Binomial Distributions]

A dissertation submitted in partial fulfilment of the requirements of the Royal Docks Business School, University of East London for the degree of [Master in Science in RISK MANAGEMENT]

[September, 2013]

[15,873]

I declare that no material contained in the thesis has been used in any other submission for an academic award

Student Number: U1043853 Date: 2nd September 2013

Dissertation Deposit Agreement

Libraries and Learning Services at UEL is compiling a collection of dissertations identified by academic staff as being of high quality. These dissertations will be included on ROAR the UEL Institutional Repository as examples for other students following the same courses in the future, and as a showcase of the best student work produced at UEL.

This Agreement details the permission we seek from you as the author to make your dissertation available. It allows UEL to add it to ROAR and make it available to others. You can choose whether you only want the ***dissertation seen by other students and staff at UEL ("Closed Access") or by everyone worldwide ("Open Access").***

I DECLARE AS FOLLOWS:

- That I am the author and owner of the copyright in the Work and grant the University of East London a licence to make available the Work in digitised format through the Institutional Repository for the purposes of non-commercial research, private study, criticism, review and news reporting, illustration for teaching, and/or other educational purposes in electronic or print form
- That if my dissertation does include any substantial subsidiary material owned by third-party copyright holders, I have sought and obtained permission to include it in any version of my Work available in digital format via a stand-alone device or a communications network and that this permission encompasses the rights that I have granted to the University of East London.
- That I grant a non-exclusive licence to the University of East London and the user of the Work through this agreement. I retain all rights in the Work including my moral right to be identified as the author.
- That I agree for a relevant academic to nominate my Work for adding to ROAR if it meets their criteria for inclusion, but understand that only a few dissertations are selected.
- That if the repository administrators encounter problems with any digital file I supply, the administrators may change the format of the file. I also agree that the Institutional Repository administrators may, without changing content, migrate the Work to any medium or format for the purpose of future preservation and accessibility.
- That I have exercised reasonable care to ensure that the Work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.
- That I understand that the University of East London does not have any obligation to take legal action on behalf of myself, or other rights holders, in the event of infringement of intellectual property rights,

breach of contract or of any other right, in the Work.

I FURTHER DECLARE:

- That I can choose to declare my Work "Open Access", available to anyone worldwide using ROAR without barriers and that files will also be available to automated agents, and may be searched and copied by text mining and plagiarism detection software.
- That if I do not choose the Open Access option, the Work will only be available for use by accredited UEL staff and students for a limited period of time.

/cont

Dissertation Details

Field Name	Details to complete
Title of thesis *Full title, including any subtitle*	Modelling the Frequency of Operational Risk Losses under the Basel II Capital Accord: A Comparative study of Poisson and Negative Binomial Distributions
Author *Separate the surname (family name) from the forenames, given names or initials with a comma, e.g. Smith, Andrew J.*	SILVER, Toni O.
Supervisor(s)/advisor *Format as for author.*	CHAN, Tat Lung (Ron)
Author Affiliation *Name of school where you were based*	University of East London
Qualification name *E.g. MA, MSc, MRes, PGDip*	MSc
Course Title *The title of the course e.g.*	MSc Risk Management
Date of Dissertation *Date submitted in format: YYYY-MM*	2013-09
Do you want to make the dissertation Open Access (on the public web) or Closed Access (for UEL users only)?	Open [X] Clos[ ]

By returning this form electronically from a recognised UEL email address or UEL network system, I grant UEL the deposit agreement detailed above. I understand inclusion on and removal from ROAR is at **UEL's discretion.**

Name: Toni O Silver

Signature: ....  .....   Date: ......2nd Sept 2013......

Modelling the Frequency of Operational Risk Losses under the Basel II Capital Accord: A Comparative study of Poisson and Negative Binomial Distributions

Dissertation submitted as part of the requirements of the degree of

MASTER IN SCIENCE IN RISK MANAGEMENT,

UNIVERSITY OF EAST LONDON

By

Toni Silver - U1043853

September, 2013

In this dissertation all quotations from other authors have been acknowledged

TABLE OF CONTENT

ABSTRACT

This study investigated the two major methods of modelling the frequency of operational losses under the BCBS Accord of 1998 known as Basel II Capital Accord. It compared the Poisson method of modelling the frequency of losses to that of the Negative Binomial. The frequency of operational losses was investigated using a cross section of secondary data published by the Banking for International Settlements (BIS) collected in the 2002 Loss Data Collection Exercise for Operational Risk. The population of the study comprised all financial institutions in the four Basel II regions of Europe, Australasia, North and South America, and Asia. The sample consisted of the entire 89 banks (census) from 19 countries worldwide that participated in the 2002 LDCE which reported a total of 47,269 individual loss events above €10,000 threshold. The following inter-related questions were investigated:

1. Is there a significant difference in the use the Poisson or Negative binomial distributions in modelling the frequency of operational risk losses?
2. Under what conditions should we adopt one for the other?

The Chi Square Goodness of fit test was carried out to test the following statistical hypotheses at 5% significant level:

1. H0 (Null hypothesis): the frequency of operational losses in banks follows the Poisson distribution.

2. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Poisson distribution.

3. H0 (Null hypothesis): the frequency of operational losses in banks follows the Negative Binomial distribution.

4. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Negative Binomial distribution.

The Poisson and the NBD models were fitted to the daily number of loss events on 3 of the 8 business lines. The models were at first fitted on the overall data aggregated daily; it then went further to fit the models on Corporate Finance, Trading and Sales, and Retail Banking. The statistical software used to fit the data was the XLSTAT 2013 and the EASYFIT 5.5 Professional edition. Findings from the study confirmed that there is a significant difference in the use of Poisson and negative binomial in modelling frequency of operational loss events; while the Poisson model fits on all data, the NBD only fits on a minority of the distributions. It went further to investigate whether there are certain conditions upon which one model can be more suitable than the other. It concluded that there is no evidence that supports the conditions upon which one model could be adopted in favour of the other. It also found no evidence to support the use of the relationship between the mean variance of a distribution in deciding between the Poisson and the Negative Binomial.

ACKNOWLEDGEMENTS

Many people contributed to the success of this project. First and foremost, I would like to acknowledge and thank my project supervisor, Dr Tat Lung CHAN (Ron) for all his formative comments, advice and constructive feedback in the course of writing this project. This project would not have succeeded without his encouragement especially when the road got tough and the only way seemed to be to give up.

I would also like to extend my gratitude to all my University tutors, for their supports through the stages of writing my assignments with success. I would also like to commend all staff of the Royal Docks Business School, University of East London especially the Programme Administrator, Mandy Steel for her kind help in sorting issues concerning administration, paperwork, parking permits, enquiries and a host of others. She greatly made my studentship with the university a hitch-free affair.

Lastly but not the least, I would like to thank my husband, Morris Silver for all his encouragement and support in the course of this work. Also, I would like to thank my children, Sharon, Sydney and Germain for their understanding and patience for missing mummy while mummy was in the library. Above all, I would like to thank the Almighty God.

CHAPTER ONE

1. INTRODUCTION

Sequel to major catastrophic operational losses (Barings Bank, 1995; Enron, 2001; Allied Irish Bank, 2002; National Australia Bank, 2004; Nationwide, 2007; Societe Generale, 2008; Lehman Brothers, 2008; Standard Life, 2009; UBS, 2009; HSBC, 2012) that resulted in the failure of a significant number of banks, the effective management of these risks have become a priority for risk managers. In response to this plight, the Central Bank Governors of the G10 countries jointly established the Basel Committee on Banking Supervision (BCBS) in 1975 which initiated the Basel II Framework in 1998 requiring banks to set aside a regulatory capital for potential operational losses. This requires the calculation of operational risk capital charge using one of the three recommended Approaches: the Basic Indicator Approach (BIA), the Standardised Approach (STA) and the Advanced Measurement Approach (AMA) which is the most sophisticated. Consequently, the quantification of a bank's operational risk exposure has become a challenging task for many banks' executives and a variety of methodologies have been suggested. More so, given that the major credit rating agencies (Moody's, S&P and Fitch) favourably rate banks that have adopted the AMA approach, specifically the Loss Distribution Approach (LDA), majority of banks have gone to adopt these advanced approaches. The problem is that the more advanced an approach is, the more sophisticated and difficult it is to measure. Under the LDA, banks are required to use historical losses to estimate the frequency, as well as the severity of operational losses which are used to calculate its operational Value at Risk (VaR).

Since the arrival of potential operational loss events is of a rather complex and chaotic nature which occur at random interval of time, it becomes paramount to examine the frequency distribution in order to understand the underlying loss arrival process. Common frequency distributions that have been used include: binomial, geometric, Poisson and negative binomial. The most common of these are the Poisson and the negative binomial which are recommended by many researchers (Cruz, 2003; Moscadelli, 2004; Chernobai et al. 2006; Perry & Dutta, 2007) to model the frequency of potential operational losses. Many articles naturally tend to focus on the severity of losses with little emphasis on its frequency. This is because severity has much more impact on capital than frequency. However, we cannot have an accurate quantification of severity events without determining how to accurately measure the occurrence of such events over a period of time; hence, the need for more research in this area of operational risk.

1.1 POISSON AND NEGATIVE BINOMIAL DISTRIBUTIONS

An essential prerequisite for developing a solid operational risk model is a systematized mechanism for data recording (Chernobai, 2007). Hence, a bank should have a consistent system with which to record every data associated with operational loss. Since observed losses arrive at irregular interval with the inter-arrival times, i.e. time interval between successive events, ranging from hours to several years; it becomes appropriate to incorporate the arrival process into the operational loss model, and to model every type of loss as a process characterised by a random frequency of events (Moscadelli, 2004; Chernobai, 2007). The Poisson and negative binomial discrete distributions are widely used to model the frequency of operational losses. The Poisson distribution is used to find the probability that a certain number of events would arrive within a fixed time interval (Ross, 2002). A special feature of this distribution that makes it easy to use is that the mean and variance is the same. The Poisson process "assumes a constant mean referred to as the intensity rate and is therefore often called a homogenous Poisson process" (Benin & Korolev, 2002).

On the other hand, the negative binomial distribution is a special generalised version of the Poisson distribution, in which the parameter λ is a gamma distribution. Hence, the assumption of a constant mean is relaxed and a greater flexibility is also allowed in the number of losses in a given period of time. The Poisson and the Negative Binomial distributions are explored in much greater details in Sections 2.4 and 2.5 respectively.

1.2 PURPOSE OF THE STUDY

The purpose of this study is to examine the two major methods of modelling the frequency of operational losses under the BCBS Accord of 1998 known as Basel II Capital Accord. It will compare the Poisson model of the daily frequency of losses to that of the Negative Binomial frequency. These losses will be investigated using a cross section of secondary data published by the

Banking for International Settlements (BIS) resulting from the 2002 Loss Data Collection Exercise for Operational Risk. Details available at: www.bis.org. Given that it is a statutory requirement under the Basel Accord on Banking Supervision for financial institutions to set aside minimum capital requirements for operational risks losses (expected and unexpected losses); hence, the aim of this study is to develop or further enhance the knowledge of risk managers to make informed decisions when deciding between the Poisson and the Negative Binomial distributions in modelling frequency of operational losses. This thesis is divided into five chapters with this section as the introductory chapter which sets the scene for the whole thesis. This chapter also introduces the underlying concepts of Poisson and negative binomial distributions. The next chapter is the literature review which examines and critically reviews academic scholars and literature on modelling the frequency of operational losses with particular reference to Poisson and negative binomial. The next chapter is the methodology which fully explains the research method upon which the key research questions will be investigated. Next is the data analysis chapter which analyses the collected secondary data for patterns with a view to drawing conclusions based on the research questions. The final chapter is the conclusion and recommendations chapter which epitomises the key research findings, as well as recommendations for further studies and research in this domain.

1.3 RATIONALE OF THE STUDY

In response to the Basel II Capital Accord which requires banks to set aside a minimum amount of capital in case of catastrophic operational losses; banks are consequently obliged to adopt certain distributions to model and estimate the capital Value at Risk (VAR). This calculated VAR estimate can be arrived by modelling the severity and the frequency of operational losses that occur in a given year. The VAR must be correctly estimated because an underestimation meant that the bank is not fully covered for operational loses. On the other hand, an over estimation implies that a bank is denying itself of investment opportunities. Hence, the rationale of this study is to research and advise managers of the most suitable distribution that can be used to model the frequency of operational losses. The Poisson and the negative random distributions will be compared and contrasted and recommendations will be made. This problem will be addressed by using data from the LDCE to fit the Observed and the Expected distributions for the Poisson and Negative binomial distributions to determine which one is a better fit. As this study is primarily concerned in modelling frequency of operational losses, details regarding modelling the severity of operational losses will be beyond the scope of this study.

1.4 RESEARCH QUESTIONS

The main objective of this study is to investigate the most suitable frequency distribution between the Poisson and Negative binomial distribution in modelling the frequency of operational risk losses in banks operating the Basel 2 Accord. It will also investigate the conditions under which each is appropriate for adoption. Hence, my research questions are:

1. Is there a significant difference from the results obtained in the use the Poisson and Negative binomial distributions in modelling the frequency of operational risk losses?

2. Under what conditions should we adopt one for the other?

The following statistical hypotheses will be tested:

1. H0 (Null hypothesis): the frequency of operational losses in banks follows the Poisson distribution.

2. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Poisson distribution.

3. H0 (Null hypothesis): the frequency of operational losses in banks follows the Negative Binomial distribution.

4. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Negative Binomial distribution.

These hypotheses will be tested using the Pearson's Chi square test for goodness of fit distribution with n-1 degree of freedom at 5% level of significance. Thus, if the value of Chi-square at n-1 degree of freedom is less than the critical value, then the Null hypothesis will be rejected and the Alternative hypothesis will be recommended for use. This Pearson's Chi Square test will determine whether the Poisson or the negative binomial can adequately predict the frequency of operational losses and will also ensure that the most appropriate distribution is used for a particular set of data. The XLSTAT and the Easy Fit 5.5 statistical software will be used to calculate the parameters for the Poisson and negative binomial distributions as well as in carrying out hypotheses testing.

1.5 SIGNIFICANCE OF THE STUDY

This study aims to equip operational risk managers with the technical knowledge and insights with which to make complex decisions regarding the appropriate discrete distributions to be adopted to estimate the frequency of losses for each of the standard Base II eight business lines. It will provide operational risk managers or anyone involved in modelling the frequency of operational risk the insights with which to carry out significant tests, e.g. Pearson's Chi square test in order to determine the most appropriate statistical distribution to be used to model loss frequencies. It is hoped that the findings of this study will be beneficial to researchers who will be urged to investigate this area further. It is also hoped that operational risk practitioners who need a handy practical approach to model complex frequency of operational loss events will find some insights in this study.

CHAPTER TWO

2. REVIEW OF LITERATURE

A literature review is an account of what has already been published or done on a research topic. It examines how the research was carried out and whether there are any issues or criticisms, as well as the conclusions drawn from past studies. More importantly, the literature review critically examines the themes of previous studies and can also refine the key research questions and the research methodologies. Hence, this literature review aims to examine previous studies as well as build on the themes of the significance of modelling the frequency of operational losses using the major discrete distributions. It will study the findings from previous studies that posits that frequency of losses are generally modelled using the Poisson and negative binomial distributions. It begins by defining and examining what constitutes operational risks, it then goes further to examine the themes of previous studies carried out on operational risk losses using the Poisson and negative binomial. Besides the Poisson and negative binomial, this literature also briefly examines other discrete distributions used to model the frequency of operational losses; these are the binomial and geometric distributions. It concluded by providing an epitome of some findings and conclusions drawn from previous studies by researchers on the Poisson and negative binomial distributions.

2.1 WHAT CONSTITUTES OPERATIONAL RISK

Operational risk has been defined by various scholars in several ways and there has been a general consensus in what constitutes operational risk. It has been defined as: "the risk arising from human and technical errors and accidents" (Jorion, 2000); a measure of the link between a firm's business activities and the variation in its business results (King, 2001); and the risk associated with operating a business (Crouhy, Galai & Mark, 2001). The generally acceptable definition of operational risk was the one given by the British Bankers Association in 2001 which was later adapted by BIS in the same year. Operational risk is "the risk of loss resulting from inadequate or failed internal processes, people or systems, or from external events" (BIS, 2001b). In the light of the above definitions, it can be inferred that operational risk is any type of risk not categorised as market or credit and can arise from four main sources namely: people, processes, systems or external events. However, BIS (2001) was quick to stipulate that their definition of operational risk included legal risk while reputational and strategic risks were excluded.

In spite of BIS (2001b) definition of operational risk, some major banks, owing to the complexities of their operation, have defined operational risk differently. Deutsche Bank defined it as "potential for incurring losses in relation to employees, contractual specifications, failures and disaster" (Chernobai, 2007). The Bank of Tokyo defined it as "the risk of incurring losses that might be caused by negligence of proper operational procedures" (Chernobai, 2007). Furthermore, in 2003 the United States Securities and Exchange Commission (USSEC) defined it as "the risk of loss due to the breakdown of controls within the firm" (Chernobai, 2007). The BCBS in 1998 also categorised risks into event type and loss type; this is explored in broader outline in the next sub section.

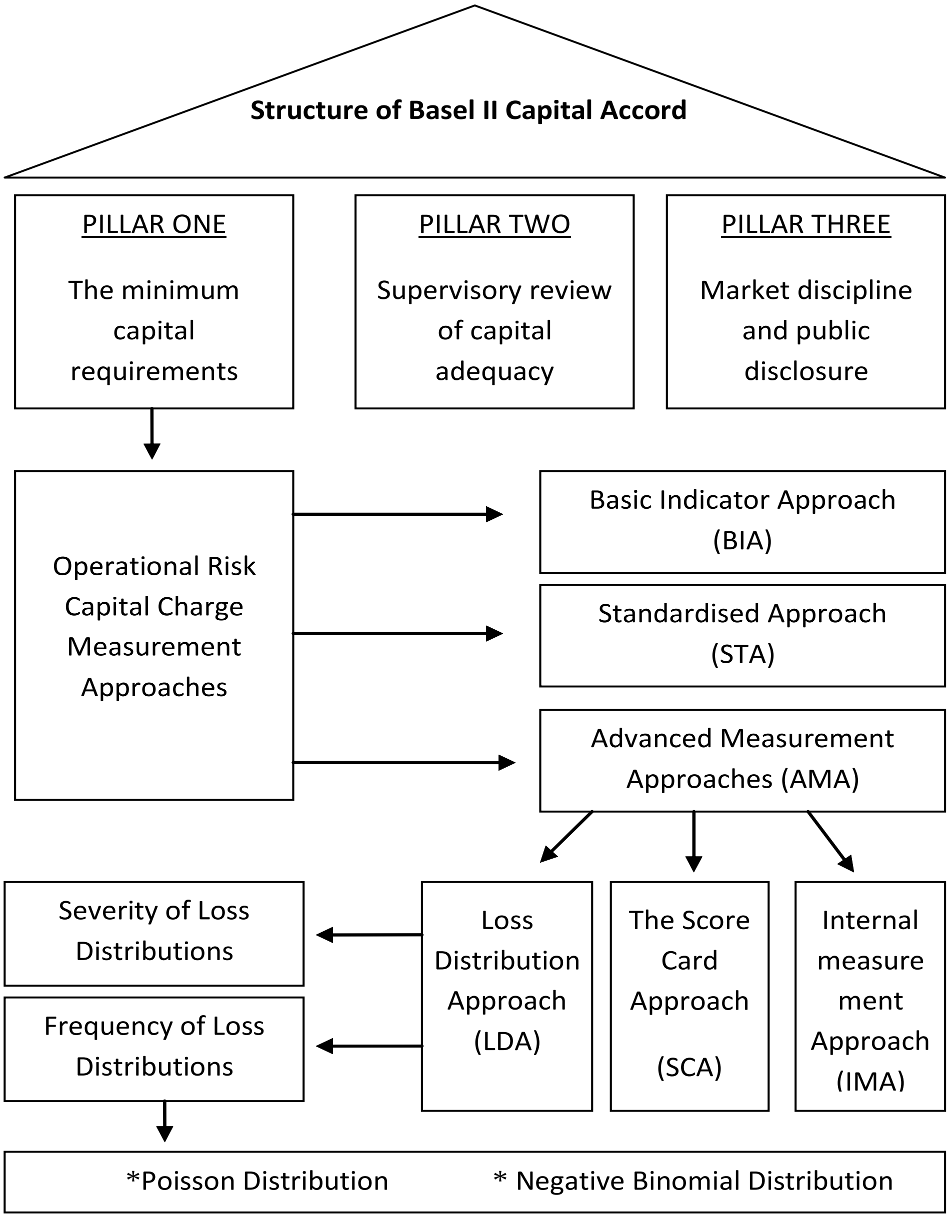
2.2 CLASSIFICATION OF OPERATIONAL RISK

While regular operational losses, e.g. minor employee errors are known as Expected Losses (EL), other infrequent losses that can arise as a result of terrorist attack, for instance, are referred as Unexpected Losses (UL). Banks however, record their losses separately and distinctly according to hazard type, event type and loss type (Mori & Harada, 2001). The distinction between hazard, event and loss is analogous to the concept of cause and effect. Hazard constitutes one or more factors that increase the probability of occurrence of loss (Alvarez, 2002). Event on the other hand, is a single incident that leads directly to one or more effects (Alvarez, 2002). Finally, Loss constitutes the amount of financial damage resulting from an event (Mori & Harada, 2001). From these definitions, it can be inferred that *hazards* potentially lead to *events*, and *events* are the cause of *losses.* Hence, an event is the effect of a hazard while loss is the effect of an event. Examples of hazard type loss include inadequate employee management, obsolete computer systems, inexperience personnel, etc. While event type includes fraud, natural disaster, system failure, etc; and finally, loss type includes restitution, legal liability, compliance, etc. The most common method of classifying losses is the one specified in the Advanced Measurement Approach of the Basel II Accord. This will be explored in broader outlines later in this chapter.

2.2.1 THE BASEL II CAPITAL ACCORD

The BCBS in 2006 finalised and published a regulatory framework known as the Basel II Capital Accord for regulating the operational risk capital charge. This brought operational risk in line with the major traditional banking risks of credit and market. This minimum regulatory capital for operational risk is known as Pillar 1. There are also Pillars 2 and 3. In order to measure the capital charge for this risk, three methodologies were proposed by the BCBS namely: the Basic Indicator Approach (BIA), the Standardised Approach (STA), and finally the Advanced Measurement Approach (AMA). The figure below is an overview of this Accord.

FIGURE 2.1 STRUCTURE OF BASEL II CAPITAL FRAMEWORK



Adapted from Chernobai, A.S. (2007)

The figure above depicts a schematic overview of Basel 2 framework for operational capital accord promulgated in 2006 which became a statutory requirement for the efficient operational risk management of banks. Pillars 1, 2 and 3 refer to the minimum risk-based capital requirements, the supervisory review of the institution's capital adequacy, and the market discipline respectively. The minimum capital charge for operational risk (Pillar 1) can be further assessed by the use of three suggested methods: the

Basic Indicator Approach (BIA) which "bases capital charge upon a fixed percentage (alpha) of average annual gross income for previous three years "(Alexander, C, 2003). This method is the simplest of the three in terms of implementation and hence, often referred to as the "top-down approach" (Chernobai, A, 2007) because the capital charge is allocated according to a fixed percentage of income currently 15%. The BIA is mostly used by smaller firms with less complex operational functions. Second in line of difficulty is the Standardised Approach (STA). This is similar to the BIA but rather than basing alpha on a single percentage, it sets separate percentages (betas) for each of the eight business lines. Table 2.2 below depicts the eight BCBS business lines with their corresponding values for beta.

TABLE 2.1 THE EIGHT BUSINESS LINES

S/no	Business Lines	Beta value
1	Corporate finance	18%
2	Trading and sales	18%
3	Retail banking	12%
4	Commercial banking	15%
5	Payment and settlement	18%
6	Agency services	15%
7	Asset management	12%
8	Retail brokerage	12%

(BIS, 2006b, Banking for Internal Supervision: www.bis.org)

The most complex and commonly used approach in measuring operational risk capital charge is the Advanced Measurement Approach which will be explored in broader outlines in the next section.

2.2.2 ADVANCED MEASUREMENT APPROACH

As the name implies, this approach is the most flexible, most advanced and most complex of the three approaches in assessing economic capital. Under this approach, banks are required to "include both qualitative and quantitative criteria for the self-assessment of operational risk" (Gregoriou, G, 2009). According to this approach, the quantitative aspect is concerned with the administration and regular review of a sound internal operational risk measurement approach (Gregoriou, G, 2009). On the other hand, the quantitative aspect includes the use of both internal and external data, stress testing and scenario analysis, Bayesian methods, business environment and internal control factors. Interestingly, under the AMA approach, a bank has to demonstrate that its operational risk measure is in parity with its internal ratings based approach for credit risk. This implies that a bank's AMA standard can be comparable to its one year holding period at 99.9th percentile confidence interval. Jobst (2007) was critical of the practicability of this confidence interval and rather suggested the use of the Extreme Value Theory (EVT) at 99.7% confidence interval in estimating VAR. More so, owing to the flexibility of the AMA, banks are allowed to literally adjust their total operational risk up to 20% of the total operational risk capital charge. Due to the popularity of the AMA, three alternative methods were proposed in 2001: these are the Internal Measurement Approach (IMA), the Scorecard Approach (SCA), and the Loss Distribution Approach (LDA).

2.2.2.1 THE LOSS DISTRIBUTION APPROACHES

A well-known AMA method is the Internal Measurement Approach (IMA) wherein the capital charge is derived by the product of three parameters. These are the gross profit, the probability of event, and the loss given the event. The product of these parameters is then used to calculate the expected loss (EL) for each business line. A second AMA method is the Scorecard Approach (SCA) which is qualitatively skewed approach. Under this approach, banks will have to determine an initial level of operational risk capital based on the BIA or TSA at the business line and subsequently modify the amounts over time on the basis of the scorecards (Chernobai, A, 2007). In other words, the scorecard approach is meant to reduce the frequency and severity of future operational losses by putting in place a proper risk control mechanism for effective risk management. A Scorecard approach is described as "simply a list of a firm's own assessment of its risks and controls, containing the risk event, risk owner, risk likelihood and control impact and control impact (Alexander, C 2003).

Paramount among the AMA methods is the Loss Distribution Approach (LDA) which is most widely used and hence, most popular among risk managers. A fundamental principle underlying the LDA is that each bank's operational losses are a reflection of its underlying operational risk exposure (Kalkbrener & Aue, 2007). Although operational risk modelling is a recent development in banking, LDA has been used by actuaries to model capital at risk calculation for many years. The LDA is considered the most robust and valid estimate for operational risk exposure (Soprano, A, 2009). Unlike the SCA, this approach uses the exact operational loss frequency and severity distributions to model economic capital. The LDA was suggested by the Basel Committee in 2001 owing to its popularity and success in the actuarial field. Under the LDA, a bank's operational losses are further divided into event types besides the business lines (see Table 2.1). This is known as business line/event type matrices. The table below is the seven event type matrix suggested by the BCBS.

TABLE 2.2 THE SEVEN EVENT TYPES

S/no	Event Types
1	Internal fraud
2	External fraud
3	Employment practices & workplace safety
4	Clients, products & business practices
5	Damages to physical assets
6	Business disruptions and systems failures
7	Execution, delivery & process management

(BIS, 2006b, Banking for Internal Supervision: www.bis.org)

The implication of splitting a bank's activities into eight business lines (Table 2.1) and seven event types (Table 2.2) is that banks have to deal with 56 paired possible risk losses (8x7=56). The fundamental task is then to estimate the loss frequency distribution and loss severity distribution for each pair. The bank can, in turn, compute the probability distribution function of the cumulative operational loss. The operational capital charge is computed as the sum of the one year value at risk (VaR) measure at 99.9%

confidence interval for each business line/event type (BCBS, 2005). The 99.9% confidence interval implies that there is only 0.1% chance that a bank will not have enough capital to cover catastrophic operational losses (DeGroot, M, 2002). Hence, the capital charge is determined by the summation of 56 paired business lines/event types capital charge. From the foregoing, it can be seen that one advantage of the LDA over the IMA is that the former assesses unexpected losses directly while the latter does so via an assumption of the existence of a linear relationship between expected and unexpected losses. Also, unlike in IMA, there is no requirement for risk managers to rescale the expected losses to with a view to determining the unexpected losses. The major shortcoming of the LDA is that it is very difficult to estimate. More so, the use of internal and external operational loss data for the past five years is also a major challenge for banks who have not kept an audit of such historical losses. The use of external data to supplement internal data is advocated by Cope & Willis (2008) who found in a study that a 20 to 30% reduction in predictive error was possible by combining the two sets of data. As stated elsewhere in this thesis, banks are required to estimate the loss frequency and loss severity distributions in order to calculate the capital charge. This process possesses a major challenge for virtually all the banks and will be explored in broader outline below with a particular emphasis on the loss frequency.

2.3 FREQUENCY AND SEVERITY DISTRIBUTIONS

The LDA is an actuarial method used in modelling the behaviour of a bank's operational losses on the basis of the frequency and severity distributions of each of the BL/ET cell matrix (Kalkbrener & Aue, 2007). While the frequency of event refers to the number of loss event that occur in a given time interval, the severity of event looks at the loss size for each event. The table below illustrates the various types of distributions that can be used to model the frequency and severity of operational losses.

TABLE 2.3 TYPES OF DISTRIBUTIONS

FREQUENCY DISTRIBUTIONS	SEVERITY DISTRIBUTIONS
Poisson	Lognormal
Negative Binomial	Exponential
Binomial	Weibull
Geometric	Gamma
Discrete Uniform	Beta
Bernoulli	Pareto
Hyper Geometric	Burr
Logarithmic	Normal
	Cauchy
	Rayleigh

In general, loss frequencies can be modelled using a discrete distribution, while loss severity uses a continuous distribution as shown in the table above. The next section will explore the frequency distributions with a particular emphasis on the Poisson and the negative binomial distributions. For details on the severity distributions which model the loss size (see Cruz, G 2002; Soprano, A, 2009; Chernobai, A, 2007). These details are beyond the scope of this thesis and will not be explored further.

2.3.1 FREQUENCY DISTRIBUTIONS

Although severity distributions which specify loss size (Table 2.3) are the most crucial component in estimating capital allocation than the frequency distributions (Alexander, C, 2003); having said that, capital charge cannot be adequately estimated without considering the loss frequencies resulting from such events. Frequent operational losses can be attributed to one of the seven event types (Table 2.2) in each of the eight business lines (Table 2.1) which can occur at irregular intervals, i.e. hourly, daily, weekly, monthly, and yearly, etc. Hence, since the inter arrival time, i.e. the intervals of time between successive losses can range from several hours to several years, it becomes appropriate to embed these random events in measuring the magnitude of operational losses. The frequency of these random events can be likened statistically to discrete random variables. A random variable X is said to have a discrete distribution, if X can take only a finite number of values (DeGroot, M, 2002). The probability of this frequency can be calculated by applying any of the frequency distribution functions in Table 2.3. Each of these discrete probability density functions will be explored in the next section.

2.3.1.1 THE BINOMIAL DISTRIBUTION

Binomial distribution represented by B(n, p) is a discrete distribution that can be applied to model the frequency of operational losses in a given interval of time. This distribution has "a fixed number of independent trials n, each of which has only two outcomes, success and failure, with probabilities p and 1-p, respectively (Dyer, G, 2003). There are four conditions under which Binomial distribution can be applied to yield a good model:

- There must be a fixed number of trails (n)
- The trials must be independent, i.e. one outcome does not preclude or affect the other outcome
- The trials must have only two outcomes (success and failure)
- The probability of success (p) is constant for each trial
  (Attwood, G, 2000)

Hence, the probability of r successes in n trials can be given by:

$$\boxed{P(X=r) = nCr\ p^r}$$

In the co uccess (r) could imply, for instance, the event that at least one operational loss has occurred in a day. The number of trials (n) could imply the total number of days in question, e.g. 5 working days in a week in which a loss is equally likely to take place. In other words, the probability that a binomial random variable X takes a value r out of n maximum possible trials, i.e. one will observe losses on r days out of n days. In binomial distribution, the mean (X) = np; while the variance (X) = np (1-p). The binomial may provide a better fit for modelling count data where the variance is less than the mean (Cruz, M, 2002). One major setback in using the binomial distribution to model frequency of operational losses is the assumption of the number of trials (n) in the calculation (Ross, 2002). This is perhaps, why this distribution is not widely used (Klugman et al (2004). More so, when p is small and n is large, the binomial can be approximated by using a Poisson distribution. More so, when n is sufficiently large, binomial can also be approximated by a normal distribution. Further details on binomial distribution can be obtained from Ross (2002), Casella & Berger (2001), Klugman et al (2004), Kingman (1993) and Grandell (1997).

2.3.1.2 THE GEOMETRIC DISTRIBUTION

The geometric discrete probability distribution is used to model the probability that "an event occurs for the first time, given that it has not occurred before" (Devroye, 1986). In statistical terminology, it models the number of failures (1-p) that will occur before a success (p). It assumes that each event is independent and that a constant probability of success. The probability density function is given by:

$$\boxed{P(X=k) = (1-p)^{k-1}\,p, \quad k=1,\ 2,}$$

The above function describes the geometric probability that an event will happen at the kth interval of time for the first time with a probability of success p. The mean (X) = 1/p; While the Variance (X) = $(1-p)/p^2$. Like the binomial distribution, the geometric distribution is not a popular choice in modelling frequency of operational losses. This might be because risk managers are not interested in modelling the first time risk occurs but on the frequency of such risk. The most popular distributions used to model risk frequency are the Poisson and the Negative binomial distributions. These two distributions are the bases for this investigation and will be fully explored in broader outlines in separate sections (2.3 and 2.4).

2.4 POISSON DISTRIBUTION

Many experiments consist of observing the occurrence times of random arrivals. Examples include arrivals of customers for service, arrivals of calls at a switchboard, occurrence of floods and other natural, arrival of staff at the office and man-made disasters. Hence, the estimation of the probability of such arrivals becomes necessary in order to study such events. Poisson distribution can be used to model the number of such arrivals that occur in a fixed period of time. In operational risk, Poisson process can be used to model the frequency of operational losses which is requisite in estimating the regulatory operational value at risk. The Poisson distribution is certainly one of the most popular frequency estimation due to its simplicity of use (Cruz, 2002). Let k be a discrete random variable with a non-negative integer, then k is said to be a Poisson distribution with probability density function given by:

$$\boxed{f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}}$$

Var (k) = λ; Mean (k) = λ, where k = number of events.

The most attractive and simplistic property of the Poisson distribution is that it assumes a constant mean. Hence, to fit a Poisson distribution to data, one only needs to estimate the mean number of events in a defined time interval. This distribution is particularly used when the mean number of operational losses is somewhat constant over time.

2.4.1 WHY CHOOSE THE POISSON DISTRIBUTION

Many researchers (e.g. Ross, 2002; Devroye, 1986; Bening & Korolev, 2002; Grandell, 1997; Kingman, 1993) have postulated the use of Poisson distribution in modelling count data. Paramount among the reasons given is its simplicity of use. To fit a Poisson distribution, only an estimate of the mean number of occurrence of such event is needed. Hence, only one parameter is required which makes it suitable to operational loss data which is not readily available. Another crucial property is that Poisson is a stable distribution and the addition of two mutually exclusive Poisson processes obeys the commutative law of mathematics. For instance, adding two Poisson processes with parameters λ1 and λ2 yields another Poisson with parameter λ1 + λ2. In other words, if a Poisson process fits in a truncated database, it will fit in the entire database. Hence, it is easy to add or include more data without structurally changing the analysis. In operational risk, this might involve the addition of a particular business line to the model as in the LDA method of AMA. A good reason for the use of Poisson distribution is the fact that the Mean and Variance is equal. This can provide a quick check on whether a Poisson might be the appropriate distribution for use (Deeks, S, 1999). If the mean and variance of the set of data is not significantly different, this is an indication that a Poisson can be used. Furthermore, another attractive consideration for applying the Poisson distribution is its scalable time-length attribute (Deeks, 1999), i.e. if that the average number of operational losses in a day is 5; then the mean and variance of the corresponding Poisson process are also 5; then it is expected that the number of losses in a 4 day period will be 4 x 5 = 20. As in the case of a truncated data earlier mentioned, this property also makes it very easy for risk managers to adjust the length of time under consideration without structurally changing the database analysis. A further consideration for the use of the Poisson process is the ability to determine its inter-arrival time between events (the length of time between two successive events). The mean of a Poisson is inversely proportional to the mean inter-arrival time. For instance, if in a 10 day interval we expect to witness 7 loss events, then we will expect the mean inter-arrival time to be 10/7 = 1.4 days between events. However, a major setback of the Poisson distribution is its assumption of a constant rate of loss occurrence over time (Soprano, et al, 2000). In reality, the frequency of most operational losses is not constant over time with the mean and variance significantly different. In such a situation, the Negative Binomial distribution may be appropriate. This will be explored in a broad outline in the next section.

2.5 NEGATIVE BINOMIAL DISTRIBUTION

In Section 2.3.1.1 on Binomial distribution, it was said that in Bernoulli with number of trials (n) with probability of success (p), the number of successes has a binomial distribution with parameters n and p. However, rather than counting successes in a fixed number of trials, it is often necessary to observe the trials until a fixed number of successes. For instance, while monitoring a piece of equipment to see when it needs maintenance, we might let it run until it produces a fixed number of errors and then repair it. This number of failures (n) until a fixed number of successes (r) has a distribution which can be modelled with the Negative Binomial Distribution (NBD). In operational risk terms, the number of failures (n) until a fixed number of successes (r) can imply the number of days (n) that elapsed before a fixed number of operational losses (r) was observed.

The NBD is given by the formula:

$$\boxed{P(X = r) = {}_{n-1}C_{r-1}\ p^{r}\ (1-p)^{n-r}}$$

Where:

n = Number of events.

r = Number of successful events.

p = Probability of success on a single trial.

2.5.1 WHY CHOOSE THE NEGATIVE BINOMIAL

Earlier applications of the NBD have been used to model animal population (Anscombe, 1949; Kendall, 1948); the number of accidents (Greenwood & Yale, 1920; Arbous & Kerrich, 1951) and in consumer spending patterns (Ehrenberg, 1988). The NBD is probably the most popular distribution in operational risk after the Poisson distribution (Cruz, G, 2002). This is because it has two parameters unlike the Poisson which has one. This availability of two parameters in NBD allows for greater flexibility in the shape of its distribution. This two-parameter property relaxes the assumption of a constant rate of loss occurrence over time assumed by the Poisson. More so, empirical studies conducted by some researchers (Moscadelli, 2004; Rosengren, 2003) have shown that the NBD is a good model for estimating the frequency of operational losses. Unlike in Poisson, the NBD does not assume that the Mean and Variance is equal; rather it assumes that the Variance is greater than the Mean. Hence, to determine whether a set of data is appropriate for the NBD, one can check whether the

Variance is greater than the Mean. This assumption further allows for over dispersion in the data (Osgood, 2000). It can be seen from the foregoing that a NBD is a special generalised case of the Poisson distribution in which the intensity rate, λ, is no longer constant but can follow a Gamma distribution with a transformed λ = m, k. Where m = mean, while k is a measure of dispersion of such distribution. This implies that λ has now been split into two parameters to consider the inherent dispersion in the data set.

2.6 EMPIRICAL STUDIES ON POISSON AND NEGATIVE BINOMIAL

It has been noted that the Poisson and the NBD are the most popular distributions of modelling operational risk frequencies and were also recommended by the BCBS. It was also noted that these two distributions differ by their mean and variance. While Poisson assumes equal mean and variance, NBD assumes that the variance is greater than the mean. Hence, choosing between these two distributions require the analyst to compute these two measures of location (mean) and the dispersion (variance) and hence, decide which one to use (DaCosta, L, 2004; Kalkbrener, 2007; Osgood, 2000). Besides, the Pearson's Chi Square goodness of fit test can also be used to determine which distribution is most appropriate for the data. This test compares the discrepancies between the observed and the expected frequencies carried out using a test of hypothesis at certain confidence interval.

Both the Poisson and the NBD have been applied in empirical studies to model loss frequencies. Cruz (2002) examines the frequency distributions of 3,338 operational losses obtained from the fraud database of a major British bank. He used both the Poisson and NBD for his modelling. He concluded that although the NBD better captures the peak of the distribution, the Poisson model better captures the overall distribution and hence, a better model (Cruz, G, 2002). In a study of the 2002 operational loss data collection exercise (LDCE), Moscadelli (2004) examines the data collected by the Risk Management Group (RMG) of the Basel Committee on the 8 business lines (Table 2.1). The Poisson and the NBD were fitted to the annual loss frequency for each of the business lines. It was drawn that the NBD is a better fit than the Poisson distribution (Moscadelli, M, 2004). In a similar vein, De Fontnouvelle et al (2005) examined the data analysed by Moscadelli (2004) but analysed the data on bank by bank basis rather than as a whole. They consider the Poisson and the NBD and concluded that the Poisson provides a better fit than the NBD. This result is also consistent with their earlier study in 2003. In a more recent study, Lewis & Lantsman (2005) examine industry-wide losses due to unauthorised lending over a period 1980 to 2001. Losses below $100,000 for smaller firms and $1,000,000 for larger firms were excluded. They considered the Poisson model and concluded that the mean frequency of loss per year is $\lambda = 2.4$ (Lewis & Lantsman, 2005). In a similar vein, Cruz (2002) simulates internal fraud data of a hypothetical commercial bank. He fitted the Poisson model and reported a mean daily frequency of loss of $\lambda = 4.88$.

Earlier studies also applied both the Poisson and NBD to model frequencies prior to the advent of operational risk. Sakamoto (1973) carried out a study to determine whether either the Poisson or the NBD better models the frequency of thunderstorm in Nevada, United States. He fitted the data using the Poisson and the NBD and concluded that the NBD is a better model (Sakamoto, C, 1973). In another earlier empirical study (Bortkiewicz, 1898) examined data collected from the Prussian army on the daily number of soldiers kicked to death by horses. He analysed the data using the Poisson and the NBD and concluded that Poisson is "an extremely good fit" for the data (Bortkiewicz, 1898). Furthermore, in a more recent study to model the measure of the influence of risk on crime incident counts, Piza, E (2012) considers the Poisson and the NBD and used the Pearson's goodness of fit test identified that the NBD is more appropriate for the data.

2.7 SUMMARY OF CHAPTER

In Sections 2.4 and 2.5, the Poisson and the NBD distributions were fully explored including their features, assumptions and their applications to operational risk. Section 2.6 further examined some conclusions drawn from previous empirical studies by key researchers in the field of operational risk. It was noted from their findings that some studies (Cruz, 2002; Fontnouvelle et al 2005 & 2003; Bortkiewicz, 1898) advocated for the use of the Poisson; while some (Moscadelli, 2004; Sakamoto, 1973; Rosengren, 2003; Piza, 2012) advocated for the use of the NBD. Furthermore, some studies (DaCosta, 2004; Kalkbrener, 2007; Osgood, 2000) advocated the use of the Poisson when the mean is somewhat equal the variance; the NBD when the variance is greater than the mean; and the Binomial distribution when the variance is less than the mean. More so, some authors (Cruz, 2002; Chernobai, 2007; Klugman, 2004) advocated the use of ratio tests, especially the Pearson's Chi Square goodness of fit test to determine the distribution by comparing observed and expected frequencies. In the light of the above, it can be deduced from the literature that the most popular distributions are the Poisson and the NBD owing to the former's simplicity of use and the latter's flexibility of use. It is also supported in the literature that there is no clear choice between the Poisson and the NBD because the choice of severity in calculating the operational value at risk outweigh that of the frequency. Hence, this answers my research questions:

1. Is there a significant difference from the results obtained in the use the Poisson and Negative binomial distributions in modelling the frequency of operational risk losses?

2. Under what conditions should we adopt one for the other?

CHAPTER THREE

3. RESEARCH METHODOLOGY

Research methods relate to the approaches or paradigms chosen to investigate a research problem. These research methods can be: descriptive, experimental, correlational, qualitative, case study, action research, policy research, to mention but a few. Irrespective of the method being used, the methodology section is the heart of the thesis which describes a detailed procedure of the steps taken to test the hypotheses or answer the research questions that are being investigated. The methodology section "describes what the researcher did" (Anderson, 2000, p86). In a research thesis, this description should be detailed enough to allow any interested party to be able to replicate the study. This detailed description normally includes: methods, procedures, sampling, research questions, data source, data analysis, instruments, ethical issues and limitation of study.

This study investigated the two major methods of modelling the frequency of operational losses under the BCBS Accord of 1998 known as Basel II Capital Accord. It compared the Poisson method of modelling the frequency of losses to that of the NBD. The frequency of operational losses was investigated using a cross section of secondary data published by the Banking for International Settlements (BIS) collected in the 2002 Loss Data Collection Exercise for Operational Risk.

3.1 POPULATION AND SAMPLING

Sampling is the method used in selecting a subset of the population for study. There are various types of sampling methods: simple random, quota, census, systematic, stratified, cluster, convenience and multi-stage sampling. On the other hand, Sample is the data selected from a population for study with a view to making generalisation about the population from which it was drawn. Whereas, population refers to the total number of objects or people that a researcher is interested in studying. For the purpose of making comparison and ease of data collections, Basel II grouped the 89 banks from 19 countries that participated into five regions. Hence, the population of this study comprises all financial institutions in the four Basel II regions of Europe, Australasia, North and South America, and Asia. In order to eliminate sample error (errors as a result of sampling), as well as increase internal validity, the sample of my study consisted of the entire 89 banks (census) from 19 countries worldwide that participated in the 2002 LDCE. Census study was considered most appropriate since secondary data would be used. Census would also ensure that any interpretation is a true representation of the population.

3.2 VALIDITY AND RELIABILITY OF INSTRUMENTS

Validity is the extent to which "an item, *sample* or instrument measures or describes what it is supposed to measure or describe" (Bell, J, 2005, p117).

Validity can be external or internal. External validity refers to generalising sample results to the entire population; while internal validity refers "to the extent to which the stated interpretation of the result is true" (Anderson, J, 2000). From these definitions, internal validity merely refers to the *validity of the sample* and does not aim to generalise the result. The term reliability, in research, refers to "the extent to which a test or procedure produces similar results under constant conditions on all occasions" (Bell, 2005). In other words, this refers to the consistency in measurement. The validity and reliability of this study was ensured by using well known theoretical frameworks for modelling count events. These are known as the Poisson distribution and the Negative Binomial distribution; for details of the Poisson and the NBD see Cruz, 2002. More so, since the LDCE was conducted by the BCBS (an international body made up of governors of prominent Central Banks worldwide), the published 2002 LDCE data was deemed to be valid and reliable.

3.3 DATA COLLECTION

The frequency of operational loss events will be investigated using a cross section of secondary data published by the Banking for International Settlements (BIS) resulting from the 2002 Loss Data Collection Exercise (LDCE) for Operational Risk. A total of 89 banks from 19 countries were asked to submit information about individual loss events up to €10,000 that occurred in the year 2001. They were also required to categorise the data into the standardised 8 business lines and 7 event types. The frequencies of loss events were also aggregated quarterly, weekly and daily by all the banks as a requirement. Overall, the 2002 LDCE reported 47,269 individual loss events above the €10,000 threshold, giving an average of 528 losses per bank (BCBS, 2003). Full detail of the data (Appendix A) is available on the Banking for International Settlements website at www.bis.org. As discussed in Section 3.1 above, these data will be the subject of my investigation with a view to answering my research questions in Section 1.4.

3.4 DATA ANALYSIS

This study analysed the distribution of daily frequencies of operational loss data collected by the Risk Management Group (RMG) of the BCBS in 2002 as mentioned in the preceding section above. The data were classified into eight business lines (Table 2.1) and seven event types (Table 2.2) pooled together across all banks not only to protect the identity of the participating banks, but also to compare the operational riskiness of each BL/ET. The Poisson and the NBD models were fitted to the daily number of loss events on 3 of the 8 business lines with a view to drawing conclusions on which model provides a better fit. The models were at first fitted on the overall data aggregated daily; it then went further to fit the models these business lines: Corporate Finance, Trading and Sales, and Retail Banking. The statistical software used to fit the data the XLSTAT 2013 and the EASYFIT 5.5 Professional edition. The XLSTAT 2013 was used mainly to carry out the Chi Square goodness of fit test while, the EASYFIT 5.5 was used to draw the graphs and charts as well as calculate the parameters of the data.

3.4.1 GOODNESS OF FIT TESTS

The process of modelling operational losses is necessarily accompanied by model risk (Chernobai, 2007). Model risk is simply the risk of selecting a wrong model. Selecting the wrong model for operational risk can affect the calculation of the operational value at risk as stated in Section 1.3. Hence, using the correct model becomes paramount in modelling operational risk losses. Consequently, some methods have been developed to test for the goodness of fit of a model. These methods can be categorised as visual tests or the more formal hypothesis tests. Common visual tests for the goodness of fit includes the Quantile-Quantile (QQ) plots which visually check whether the QQ plot coincides with an angle of 45 degree. The second visual method is the Mean Excess Plots (MEP) which compares a particular mean value against various mean values. For more details on both the QQ plot and the MEP, see Chernobai (2005) and Embrechts (1997) respectively. However, more complex models are chosen by carrying out statistical hypothesis tests which ensures that the right model is chosen at certain confidence level.

3.4.1.1. **PEARSON'S CHI SQUARE TEST**

While the Likelihood Ratio (LR) test is the most common formal test applied to continuous probability distributions, the Pearson's Chi Square (PCS) Goodness of Fit test on the other hand, is the most common test applied to discrete probability distributions such as the NBD, Binomial and the Poisson (Chernobai, 2007). The PCS test checks whether the data sample follows a hypothetical distribution. In this case, whether the observed frequencies fit the Poisson or the NBD frequencies. Hence, the PCS test was considered most appropriate for this study.

3.4.1.2 HYPOTHESES TESTING

In order to determine whether the correct model has been used in this study, the following statistical hypotheses were tested at 5% significant value using n-p-1 degree of freedom. The test that was used is the widely used Pearson's Chi Square goodness of fit test:

1. H0 (Null hypothesis): the frequency of operational losses in banks follows the Poisson distribution.

2. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Poisson distribution.

3. H0 (Null hypothesis): the frequency of operational losses in banks follows the Negative Binomial distribution.

4. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Negative Binomial distribution

Thus, if the value of Chi-square at n-1 degree of freedom is less than the critical value, then the Null hypothesis will be rejected and the Alternative hypothesis will be accepted. On the other hand, if the value of the Chi square is greater than the critical value, the Null hypothesis will be accepted and the Alternative hypothesis rejected. This Pearson's Chi square test would determine whether the Poisson or the negative binomial can adequately predict the frequency of operational losses and will also ensure that the right distribution is used for a particular set of business lines. The decision to test at 5% significance, i.e. 95% confidence interval is to reduce the likelihood of Type 1 Error (falsely rejecting H0 when in fact, H0 is true). For the Chi Square critical value table that was used, please see Appendix B.

3.5 ETHICAL CONSIDERATION

To ensure that ethical standards were met, the need to protect the identities of individual banks, as well as their respective losses became paramount in this study. To ensure anonymity, the names and addresses of participating banks were not published by the BCBS, as well as the amount of individual losses. Rather, the losses were pooled together and categorised into the standard 8 BLs and 7 ETs. Furthermore, the operational loss data collected by the Risk Management Group (RMG) of the Banking for International Supervision in the 2002 are available in the public domain at www.bis.org. No special permission needed to be sought before the use of these data. More so, ethical approval was sought from the University prior to the commencement of this thesis in the form a written proposal (Appendix C)

3.6 LIMITATION OF STUDY

The empirical data used in drawing conclusion in this study is limited to the 47,269 internal loss data on operational risk collected by the RMG in the LDCE of 2002. The scope of the study is also limited to the 19 countries and the 89 banks that participated in the loss data collection exercise. Modelling the frequency of operational loss events can be done using any discrete probability distribution function such as: Binomial, Poisson, NBD, and Geometric distributions (see Table 2.3). This study however, is limited in its analysis and modelling to the Poisson and the NBD only. Thus, this study is also subject to the prevalent assumptions and limitations in using the Poisson and NBD as theoretical frameworks underpinning the study (see 2.3.1). For further details on the assumptions and limitations of the NBD and Poisson, see Ross, 2002; Devroye, 1986; Bening & Korolev, 2002; Grandell, 1997; Kingman, 1993; Anscombe, 1949; Kendall, 1948; & Cruz, 2002.

CHAPTER FOUR

4. DATA ANALYSIS

This chapter comprises four sections. The first section will not only look at the frequency of the loss data as a whole, but will also consider the number of loss events according to each Business Line (BL) and Event Type (ET) that was reported during the year. The second section will explore the distribution of the total loss events during the year aggregated daily. It will also fit the Poisson and the NBD to the observed daily frequency distribution of the loss events. Section three tends to focus on the frequency of daily loss event according to each BL pooled together across all banks; and will also describe vital statistics and fit both the Poisson and the NBD to determine which better fits the observation. The last section concludes the chapter by making comparisons of the results obtained from fitting the Poisson and the NBD to the observed frequency in order to answer my research questions:

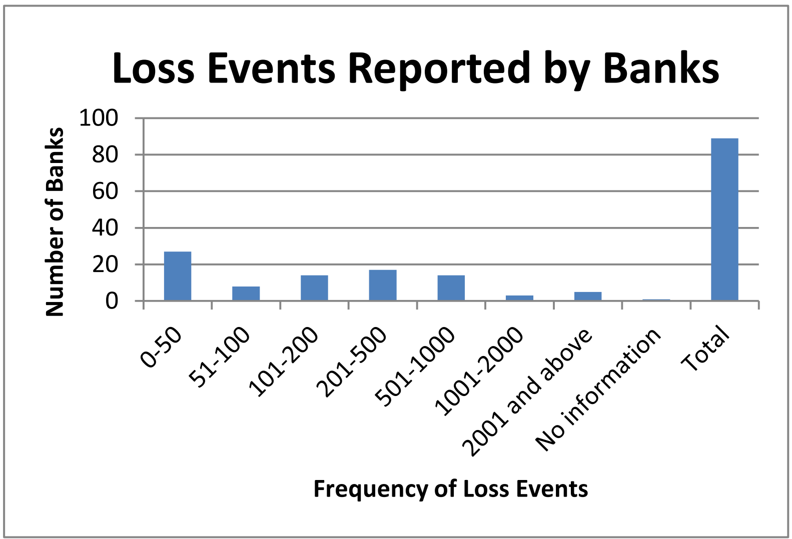
1. Is there a significant difference in the use the Poisson or Negative binomial distributions in modelling the frequency of operational risk losses?
2. Under what conditions should we adopt one for the other?

4.1 GROUPED FREQUENCY OF LOSS EVENTS

The number of individual loss events that occurred in the year 2001 as reported by the 89 banks that participated in the 2002 LDCE of the BCBS (see Sections 3.3 & 3.4). The overall distribution of this data is the subject of analysis and discussion in this section. The table below depicts this information.

FIGURE 4.1 NUMBER OF LOSS EVENTS REPORTED

TABLE 4.1 NUMBER OF LOSS EVENTS REPORTED



Number of Events/year	Number of Banks
0-50	27
51-100	8
101-200	14
201-500	17
501-1000	14
1001-2000	3
2001 and above	5
No information	1
Total	89

The figure above illustrates that the range of individual loss events reported by these banks was quite large, with values ranging from only 1 event to as large as 2,000 events. Over half of the banks (55%) reported 200 or less number of events, and the majority of these (55%) reported fewer than 50 events. On the other hand, 8 banks reported over 1,000 individual loss events, and 5 reported more than 2,000 loss events. Since the modal number of loss events per year is 0-50, *it can be concluded that the average number of loss events per year encountered by banks is up to 50.*

4.1.1 FREQUENCY OF LOSSES BY BUSINESS LINE/ EVENT TYPE

The basic features of the individual loss data submitted by all the 89 participating banks are the subject of analysis and discussion in this section. Hence, the table below illustrates the total number of individual loss events reported by each of the 89 banks in combinations of the 8 BL and 7 ET amounting to 56 separate cells. The data was classified according to BL/ET and pooled together across all the 89 banks.

TABLE 4.2 FREQUENCY OF LOSS EVENTS PER BL/ET

	Internal Fraud	External Fraud	Employment Practices & Workplace Safety	Client, Products & Business Practices	Damage to Physical Assets	Business Disruption & System Failures	Execution, Delivery & Process Management	No Event Type information	TOTAL
Corporate finance	17	20	73	73	16	8	214	2	423
Trading & Sales	47	96	101	108	33	137	4603	8	5132
Retail Banking	1268	17107	2063	2125	520	163	5289	347	26882
Commercial Banking	84	1799	82	308	50	47	1012	32	3414
Payment & Settlement	23	322	54	25	9	82	1334	3	1852
Agency Services	3	16	19	27	8	32	1381	5	1490
Asset Management	28	44	39	131	6	16	837	8	1109
Retail Brokerage	59	20	794	539	7	50	1773	26	3268
No BL information	35	617	803	54	13	6	135	36	1699
TOTAL	1564	20039	4028	3390	662	541	16578	467	47269

For the purpose of in depth analysis, the above table with values converted to percentages is shown below.

TABLE 4.3 FREQUENCY (PERCENTAGE) OF LOSS EVENTS PER BL/ET

	Internal Fraud	External Fraud	Employment Practices & Workplace Safety	Client, Products & Business Practices	Damage to Physical Assets	Business Disruption & System Failures	Execution, Delivery & Process Mgmt	No ET information	TOTAL
Corporate finance	0.04%	0.04%	0.15%	0.15%	0.03%	0.02%	0.45%	0.00%	0.89%
Trading & Sales	0.10%	0.20%	0.21%	0.23%	0.07%	0.29%	9.74%	0.30%	10.86%
Retail Banking	2.68%	36.19%	4.36%	4.50%	1.10%	0.34%	11.19%	0.73%	61.10%
Commercial Banking	0.18%	3.81%	0.17%	0.65%	0.11%	1.10%	2.14%	0.07%	7.22%
Payment & Settlement	0.05%	0.68%	0.11%	0.05%	0.02%	0.17%	2.82%	0.01%	3.92%
Agency Services	0.01%	0.03%	0.04%	0.06%	0.02%	0.07%	2.92%	0.01%	3.15%
Asset Management	0.06%	0.09%	0.08%	0.28%	0.01%	0.03%	1.77%	0.02%	2.25%
Retail Brokerage	0.12%	0.04%	1.68%	1.14%	0.01%	0.11%	3.75%	0.06%	6.91%
No BL information	0.07%	1.31%	1.70%	0.11%	0.03%	0.01%	0.29%	0.08%	3.59%
TOTAL	3.31%	42.39%	8.52%	7.17%	1.40%	1.14%	35.07%	0.99%	100.00%

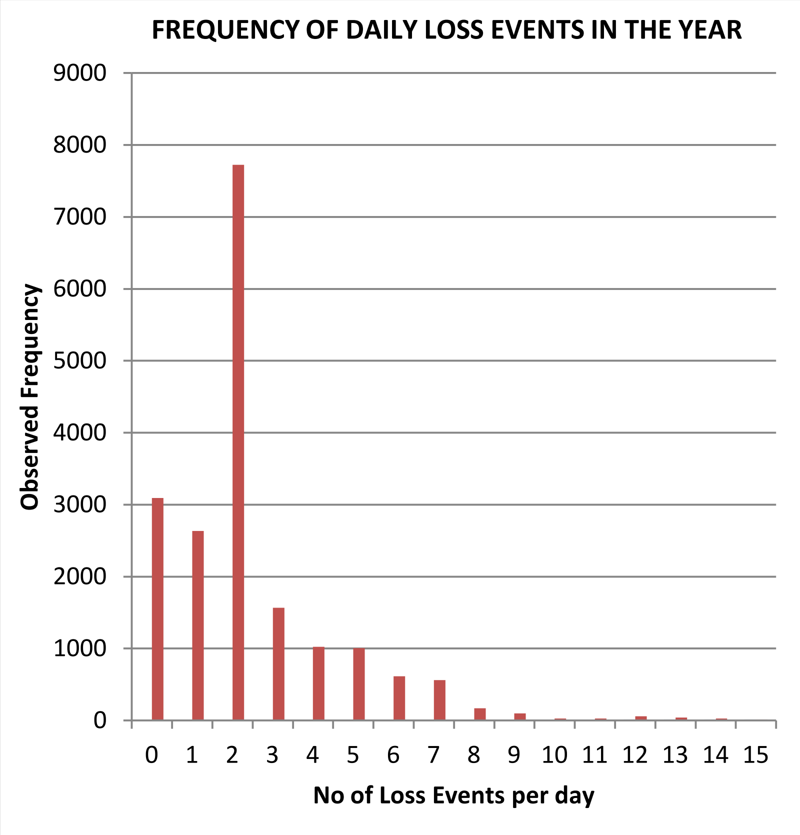
From the tables above, a total of 47,269 individual loss events categorised according to their corresponding BL/ET cells. It can be noticed that the events are not evenly spread across BL/ET. In particular, the data were clustered into half of the 8 BL, with the highest concentration in Retail Banking. This BL accounted for 61% of the total number of loss events. *Hence, it can be drawn that Retail Banking accounts for more than half the number of operational loss events*. This finding is consistent with (Cruz, 2002; Chernobai, 2007; Moscadelli, 2004). Trading and Sales accounted up to 11%, while commercial Banking and Retail Brokerage, each accounted for 7%. Thus, these 4 BL accounted for 86% of all individual loss events. On the other extreme, Corporate Finance is the fewest with just below 1%. A similar pattern of clustering is apparent in the ET category with 42% categorised as External Fraud; 35% as Execution, Delivery and Process Management; Employment Practices and Workplace Safety 9%; and Client Product and Business Practices accounting for 7%. These four categories accounted for 93% of individual loss events. On the individual BL/ET cell category, a clustering can also be found in the Retail Banking/External Fraud cell with 36% of loss events reported*. It can be drawn from the data that the number of operational loss events due to fraud in retail banking accounts for up to 40% of losses.*

4.2 DISTRIBUTION OF DAILY FREQUENCY OF LOSS EVENTS

In this section, the 47,269 individual loss events reported by the 89 banks in the 2002 LDCE but aggregated daily will be explored in broader details. The number of loss events in this case has been grouped on a daily basis, i.e. total number of loss events that occurred per day. Table 4.4 is an illustration of these daily loss events.

FIGURE 4.2 FREQUENCY OF LOSS EVENTS- AGGREGATED DAILY

TABLE 4.4 FREQUENCY OF LOSS EVENTS- AGGREGATED DAILY



No of events/day	Observed frequency
0	3094 (16.5%)
1	2633 (14.0%)
2	7726 (41.3%)
3	1568 (8.3%)
4	1022 (5.4%)
5	1008 (5.3%)
6	616 (3.3%)
7	560 (3.0%)
8	169 (0.9%)
9	98 (0.5%)
10	28 (0.015%)
11	28 (0.015%)
12	56 (0.30%)
13	42 (0.22%)
14	28 (0.15%)
15	14 (0.01%)
TOTAL	18,690

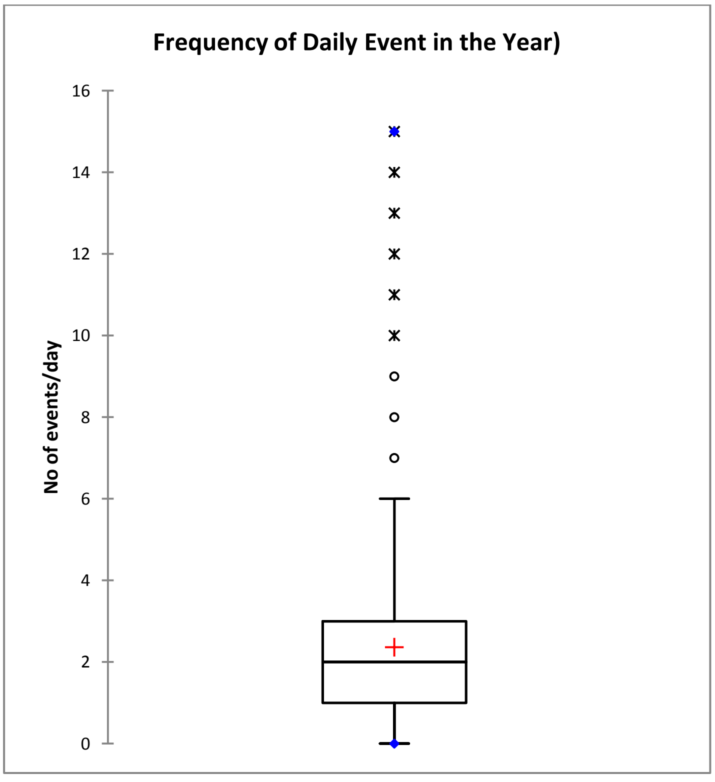
In the above table, notice that the total observed frequency is 18,690. The total number of observations of 47,269 was obtained by multiplying the number of events/day (column one) with the corresponding observed frequency (column two) which were summed and added to the number of zero events (3,094). Analysing the figure above*, It can be drawn that the average number of operational loss events is two per day* accounting for 41% of the total loss events. This modal number of 2 loss events per day is consistent with Cruz (2002) who examined the frequency of 3,338 operational loss events obtained from a major British retail bank from 1992 to 1996; he obtained a mode of 2 loss events per day (525/1311) which also accounted for 40% of number of loss events (Cruz, M, 2002, p95). This is followed by no loss events per day with 17%; then 3 loss events per day at 14% which decreases in value as the number of loss increases. A fifteen loss events per day is the smallest, as one will expect, with less than 1%. By observation, it can be seen from the figure that the distribution appears to be positively skewed as the tail of the distribution is longer on the right. It can be concluded that *the frequency distribution of operational loss events is* *positively skewed*. An exploration of the statistical parameters of this distribution will be carried out in the next section to further investigate its characteristics in broader details.

4.2.1 DESCRIPTIVE STATISTICS

A summary of the descriptive statistics and box plot for the frequency of the daily loss events in the year as shown in Table 4.4 above is shown below:

FIGURE 4.3 BOX PLOT OF FREQUENCY OF LOSS EVENTS

TABLE 4.5 STATISTICS OF FREQUENCY OF LOSS EVENTS



Statistic	Value
Sample Size	47,269
Range	15
Mean	2.3358
Variance	4.1181
Std. Deviation	2.0293
Skewness	1.9525
Kurtosis	5.9379
25% (Q1)	1
50% (Median)	2
75% (Q3)	3
Min	0
Maximum	15

The data has a skewed parameter of 1.9525 which confirms that it is positively skewed. Recall that a skew parameter of zero is a symmetrical distribution; and less than zero is a negatively skewed distribution. The above box plot further confirms this assertion. More so, since the mean (2.3358) is more than the median (2), this further confirms positive skewness. From the box plot above, the outliers (extreme values) are asterisked implying that number of events $0 \leq n \leq 6$ are not outliers; while number of events $7 \leq n \leq 15$ are outliers. However, since I am concerned here with the operational number of loss events which could exceed 15 loss events per day in very large multinational banks, I will not be considering the statistical concept of excluding outliers in this calculation. *However, it can be concluded from the above figure that any number of loss events exceeding 6 per day is too extreme and hence, "too many". While on the other hand, the number of loss events per day can be considered normal if it is between 0 and 6.*

From Table 4.5, the range is 15 and the standard deviation is 2, which shows high spread and variability in the number of loss events reported per day among the banks. The Lower Quartile (Q1-the 25% position of the data) is 1 and the Upper Quartile (Q3- the 75% position of the data) is 3. Hence, the IQR = Q3 – Q1 = 2. The IQR measures the middle 50% of the data in a bid to rid the data of any extreme values from both directions. Another important parameter for consideration is the kurtosis. This measures the spread of the values around the mean, i.e. the peakedness of the data. A high kurtosis implies a high peak in the centre of the data. A population with a high kurtosis is called *leptokurtic*. As a general rule of thumb, if the kurtosis value of a distribution is above 3, such is referred to as leptokurtic and cannot be represented by a normal distribution (Cruz, 2002, p40-43). Since my kurtosis value is 5.9379 from Table 4.5, *It is suffice to generally conclude that the daily distribution of the number of loss events has a high kurtosis and hence, leptokurtic.*

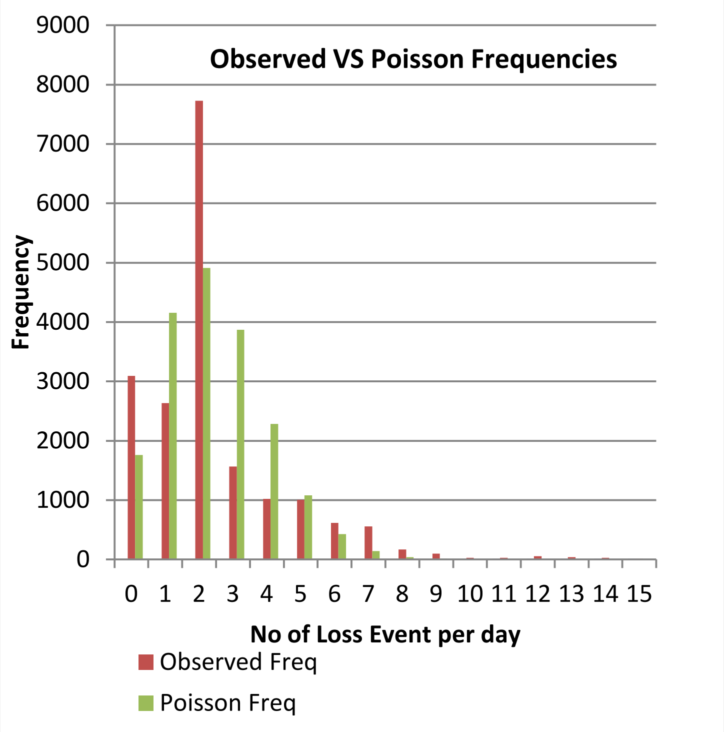
Furthermore, from Table 4.5 the mean is 2.3358 and the variance is 4.11. Since the variance is significantly greater than the mean, it is tempting for one to conclude that the NBD will fit the distribution better than the Poisson (see Section 2.6). However, such conclusion may be inconclusive without first fitting the observed frequency to the theoretical frequency on both the Poisson and the NBD. This will be the basis of the next two sections 4.22 and 4.23.

4.2.2 TESTING THE POISSON FIT TO THE DATA

The figure and table below show the observed frequency as it compares with that of Poisson frequency.

FIGURE 4.4 OBSERVED VS POISSON FREQUENCIES

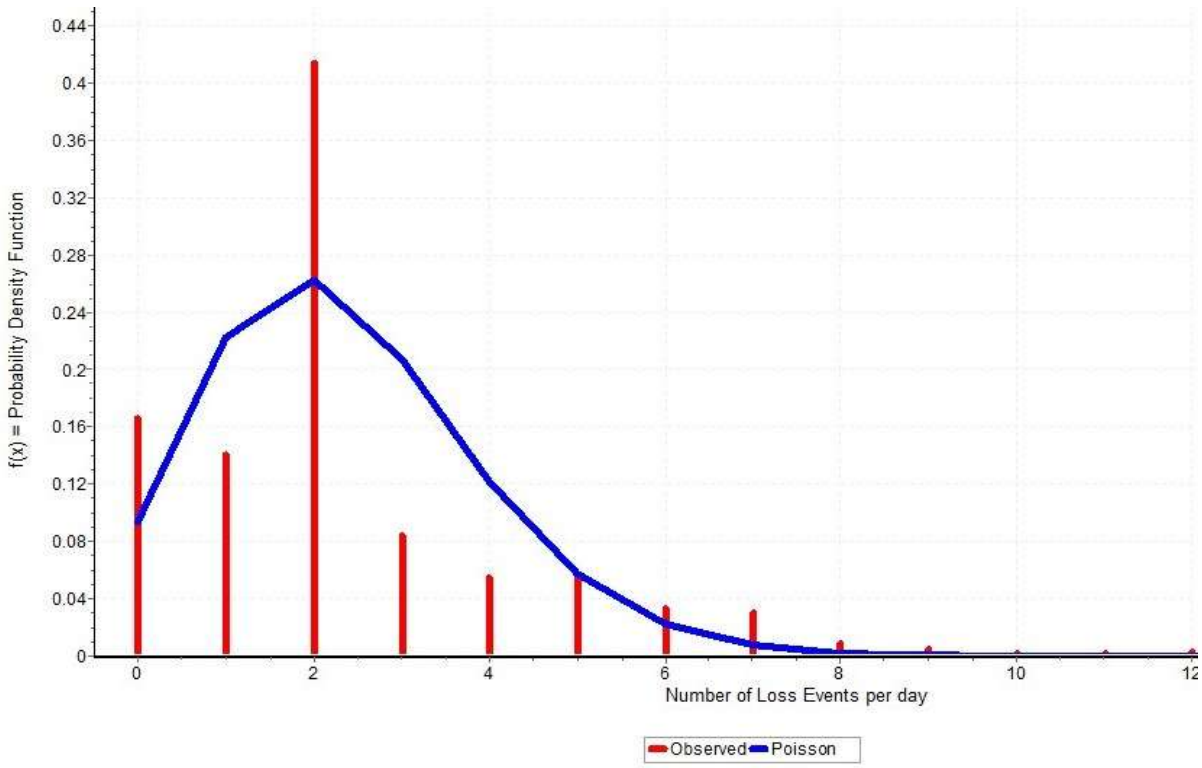
TABLE 4.6 OBSERVED VS POISSON FREQUENCIES



Events/ day	Observed Frequency	Poisson Frequency
0	3094	1758.437
1	2633	4156.177
2	7726	4911.694
3	1568	3869.700
4	1022	2286.570
5	1008	1080.891
6	616	425.792
7	560	143.770
8	169	42.476
9	98	11.155
10	28	2.637
11	28	0.567
12	56	0.112
13	42	0.020
14	28	0.003
15	14	0.001

From the above figure, observed frequencies are significantly greater when two or zero number of losses were reported. Poisson appears to fit well as the number of losses increases. However, this may not be true when you compare the corresponding value of the observed with the Poisson frequencies in Table 4.6. This will become clearer if the graph of the above distributions is drawn. The EasyFit 5.5 statistical software will be used for the statistical analysis and for drawing the graphs.

FIGURE 4.5 OBSERVED VS POISSON FIT



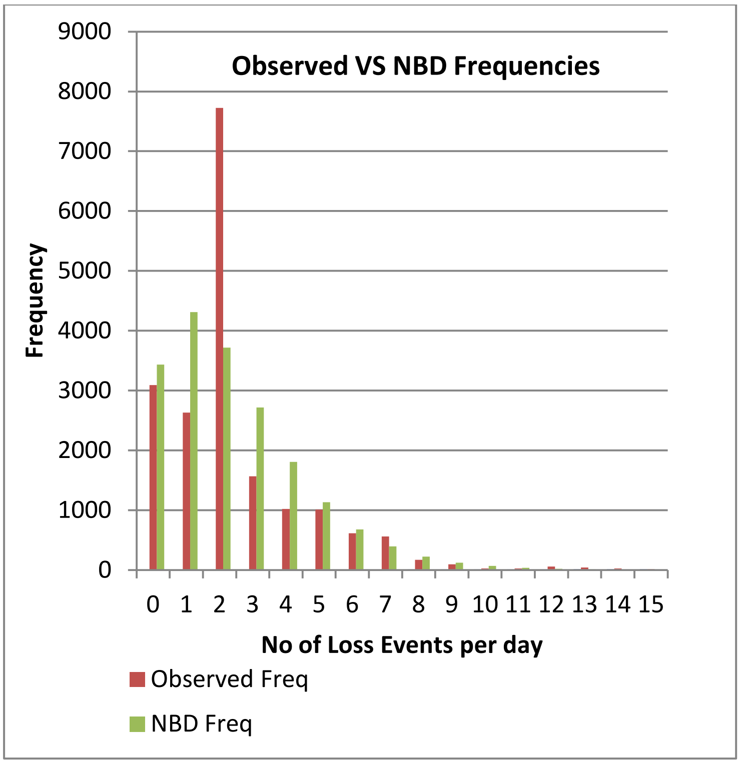
The figure above is a fitted Poisson curve to the observed frequency using the mean, parameter λ = 2.3636. Recall from the figure above that the red bars represent the observed (actual) distribution while the blue line represents the Poisson (expected) distribution. Despite the disparity when the number of losses is 2 which is significantly greater than the expected frequency, the Poisson fit captures the data overall. *T*he next section will explore the NBD to determine how the distribution compares with that of the Poisson.

4.2.3 TESTING THE NBD FIT TO THE DATA

The figure and table below show the observed frequency as it compares with that of the NBD frequency.

FIGURE 4.6 OBSERVED VS NBD FREQUENCIES

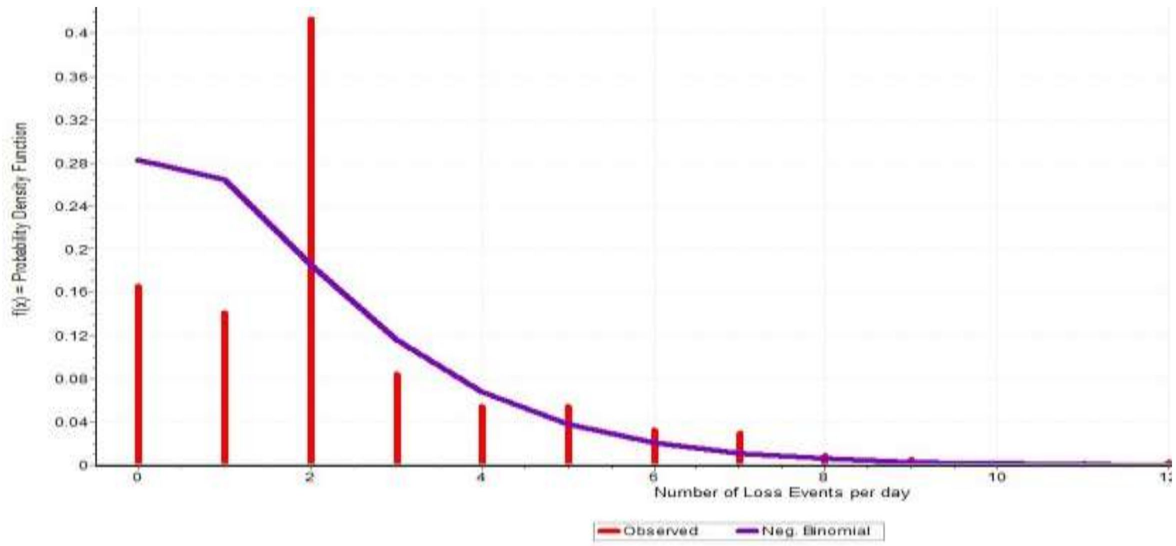
TABLE 4.7 OBSERVED VS NBD FREQUENCIES



Events /day	Observed Frequency	NBD Frequency
0	3094	3433.734
1	2633	4311.539
2	7726	3717.392
3	1568	2717.591
4	1022	1808.484
5	1008	1132.344
6	616	679.292
7	560	394.780
8	169	223.884
9	98	124.521
10	28	68.168
11	28	36.830
12	56	19.679
13	42	10.416
14	28	5.468
15	14	2.850

The observed frequency when two loss events were reported is significantly higher than that of the theoretical frequency. This pattern was also the case in Poisson frequency in figure 4.4. By visual observation, there is no any apparent difference between the fit of the NBD and that of the Poisson discussed earlier. The graph of the NBD with parameters n=2 and p= 0.53128 is shown below to further examine the distributions.

FIGURE 4.7 OBSERVED VS NBD FIT

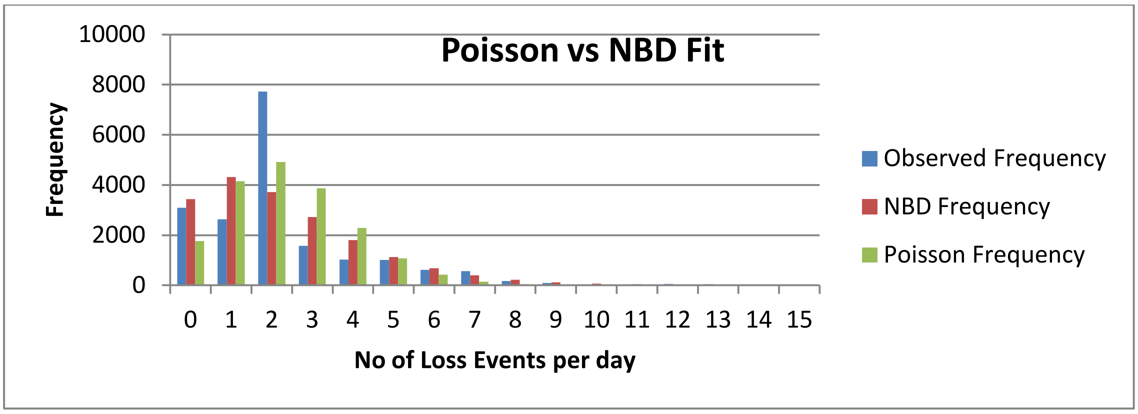


Despite the significant difference between the observed and the NBD fit when the number of losses is 2, the NBD appears to fit well the distribution especially at the upper end. However, it appears not fit the curve well at the lower end, i.e. as the number of loss events decreases. The next section will compare both distributions to determine which is a better fit.

4.2.4 POISSON AND NBD FITS COMPARED

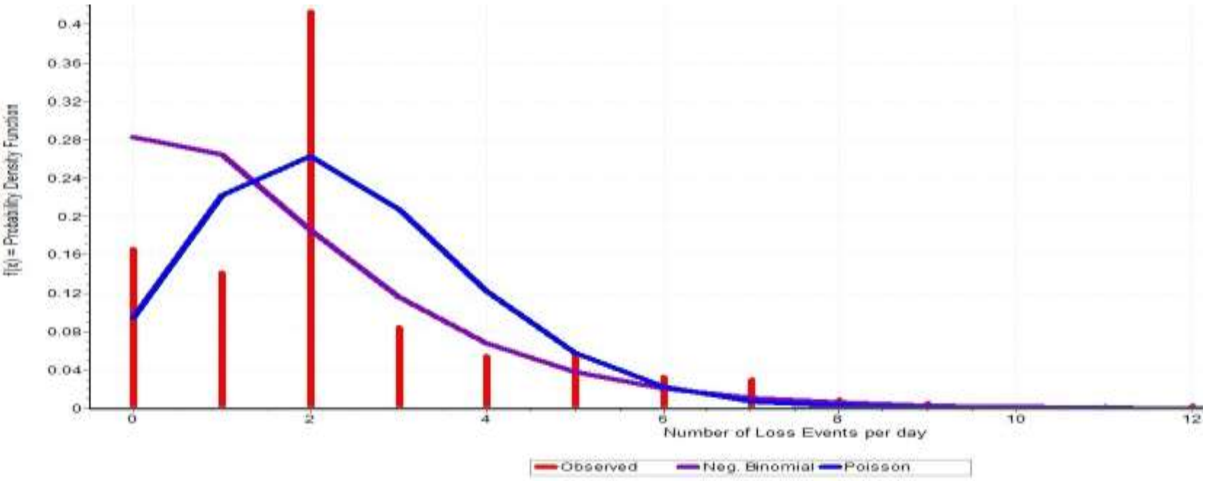
The figure below shows the observed frequency as it compares with both the Poisson and the NBD frequencies.

FIGURE 4.8 OBSERVED, POISSON & NBD FREQUENCIES



From the figure above, it appears that there are lots of variations in the frequency of the observed as it compares to the frequencies of the Poisson and the NBD. Therefore, no conclusions can be drawn to justify whether the Poisson or the NBD is a better fit. Furthermore, a new line of enquiry is to superimpose the graphs of the Poisson and the NBD distributions against the observed frequency distribution. The figure below shows the information.

FIGURE 4.9 GRAPHS OF POISSON & NBD FIT



From the above figure, on average two loss events were reported daily which was not captured by the two distributions due to its extremity. In this case, the event of 2 losses per day can be considered to be an anomaly. Besides this anomaly, it is can be seen from the graph that although both distributions captures the right hand side of the distribution, the NBD fails to capture the lower end of the data. Hence, the Poisson model seems to better capture the data overall. In the light of the above, *it can be drawn that the Poisson distribution models the overall daily operational loss events when aggregated daily better than the NBD.* This finding is also consistent with Cruz's study who modelled the frequency of 3,338 operational loss events by fitting both the Poisson and the NBD; He concluded that the Poisson is a better fit (Cruz, 2002). The next section uses a purely analytical method to determine whether the Poisson or the NBD is a better fit.

4.2.5 CHI SQUARE GOODNESS OF FIT TEST

This section will use the Pearson's Chi Square Goodness of Fit distribution to compare the actual (observed) frequency to the theoretical frequency at 5% significant level at n-p-1 degree of freedom (see Section 3.4.1.2). The chi square value will be computed using the XL STAT 2013 statistical software. The following statistical hypotheses were tested:

1. H0 (Null hypothesis): the frequency of operational losses in banks follows the Poisson distribution.

2. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Poisson distribution.
3. H0 (Null hypothesis): the frequency of operational losses in banks follows the negative binomial distribution.

4. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the negative binomial distribution.

The table below is a summary of the statistics obtained from the Chi Square test:

TABLE 4.8 CHI SQUARE GOODNESS OF FIT TEST

	NBD	POISSON
Parameters (see Sections 4.5 & 2.5)	r= 2.7; p= 0.88	λ =2.33
n (no of loss events)	16	16
Degree of freedom (df)	n- p-1 = 14	n-1 = 15
Significant Level (α)	5%	5%
X^2 (chi square value)	52.8	48.3
P (critical value)	21	25
Decision	Since X^2 > P: Hence, REJECT Ho	Since X^2 > P: Hence, REJECT Ho
Excess value (X^2 – P)	31.8	23.3

The above table shows the summary statistics obtained from the Chi square test. The parameters were estimated using the method of Moments which is the simplest and widely used method. A more advanced method is the Maximum Likelihood which maximises the likelihood of the sample. However, since my sample is large enough, the method of Moment is considered appropriate. The parameters of the test have fully explored in Sections 2.5 and 4.5. But for the purposes of emphasis:

The negative binomial type I distribution was used which is the distribution of the number x of unsuccessful trials necessary before obtaining r successes.

- r = the number of successes (NBD)
- p = the probability of success (NBD)
- λ = the mean (Poisson)
- α = Significance Level, i.e. if α is 5%, we are 95% confident that we have made the right decision.
- df = degree of freedom; since a Poisson has 1 parameter (λ), df = n-1. For NBD with 2 parameters (r, p); df = n- p – 1.
- V = variance which show the spread of the losses, i.e. how widely apart (the deviation) each loss is from the mean number of losses. This can be used to monitor the degree of variations made by each bank.
- M = mean which shows the average number of loss events for the period. This can be used to monitor the excessiveness of losses made by each bank.

Recall that in a Chi Square Hypothesis testing, Ho is rejected if the value of the Chi Square (X^2) is greater than the Critical Value (P) but accepted if X^2 is less than P. In the above table, the Chi Square values are greater than the P values in both Poisson and the NBD. Hence, the Null hypothesis is rejected in both cases. This implies that the Poisson and the NBD are not a very good fit for the distribution and hence, some advanced distribution such as the Cox distribution which is a mixture of two or more models can be applied. However, the application of these advanced statistical distributions is beyond the scope of this thesis.

However, since the objective of this study is to investigate whether the Poisson or the NBD is a better fit for the distribution of daily operational loss events in banks, it becomes necessary to determine how close the X^2 value is to the Critical value in both situations. Looking at the above again, the last row shows the Excess value, i.e. the difference between the X^2 value and the P value. Recall that Ho could have been accepted if the $P \geq X^2$; thus the closer P is to X^2 the better the distribution fits the data. Reading from the table, the

Excess value for NBD is 31.8 and that of the Poisson is 23.3. Since the Poisson's Excess value is less than that of the NBD, this implies that Poisson's P value is closer to its X^2 than in the case of the NBD. Hence, it can be drawn that *the daily operational loss events across banks fit better when modelled with the Poisson distribution than the negative binomial distribution.* This result is consistent with the conclusion drawn by super-imposing both Poisson and the NBD graphs against the observed frequency in Section 4.2.4, Figure 4.8 as shown above.

4.3 FREQUENCY OF DAILY LOSS EVENTS BY BUSINESS LINE

This section examines the loss distribution of the daily loss events classified into Basel Committee's 8 Business Line (BL) and pooled together across the 89 banks that participated in the exercise (see Section 4.1.1; Table 4.3). I will only examine the distributions of the first three BLs, i.e. Corporate, Finance, Trading & Sales and Retail Banking as examining all eight BLs is beyond the scope of this study. The decision to consider these three BLs among others is because the percentages of their loss events as shown in Table 4.3 better reflects and capture the whole data, i.e. percentages of other BLs are subsets of these three BLs. They fit the two extreme percentage values, as well as the mid percentage points of the aggregate data. Corporate Finance comprises less than 1% of data which is the least, followed by Trading & Sales at 10%. The highest is Retail Banking with 61% of total observation. Hence, these three BLs fairly represents the other five BLs in terms of their rate of occurrence and will be analysed.

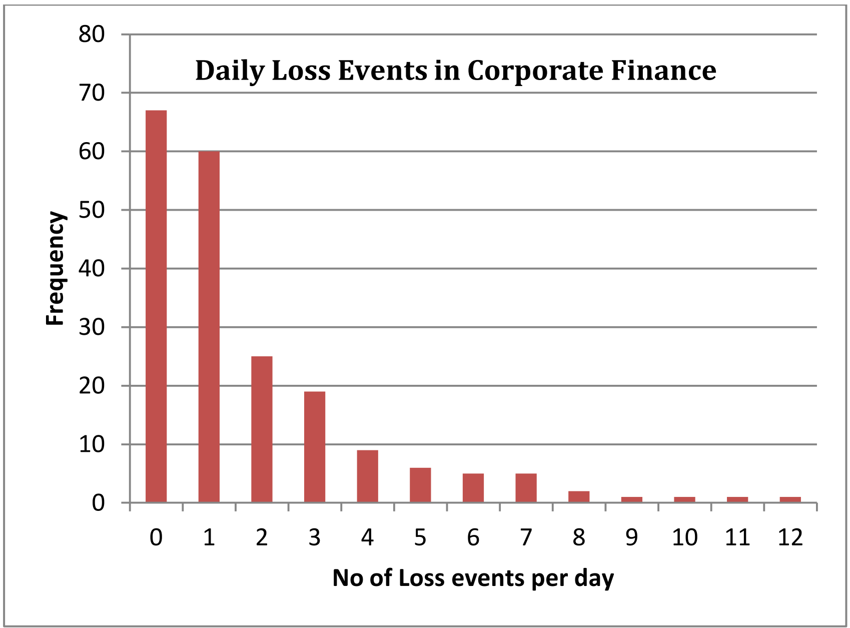
It can be seen that it was difficult to make an early conclusion on whether the Poisson or the NBD was a better fit for the distribution of frequency of loss events in Section 4.2 because the charts and graphs had too many variations when trying to interpret them. Nevertheless, the results obtained in Section 4.2.4 when the graph of the Poisson was plotted against that of the NBD and both super-imposed on the observed frequency was when the results started to emerge. More so, the Chi square goodness of fit test in Section 4.2.5 further clarified the results and reinforced the decisions and the conclusions made in determine the model that better fit the data. Hence, while considering the Poisson and the NBD fit for the three chosen BLs, only the graphs of the two models superimposed on the daily observed frequency, as well as the Chi Square tests will henceforth, be considered.

4.3.1 FREQUENCY OF DAILY LOSSES IN CORPORATE FINANCE

The table below shows total number of operational loss events that were reported in Corporate Finance in the 2002 LDCE by the 89 banks that participated in the exercise. These losses have been aggregated daily by the RMG and will be the subject of analysis and exploration.

FIGURE 4.10 FREQUENCY OF DAILY LOSS EVENTS- COPPORATE FINANCE

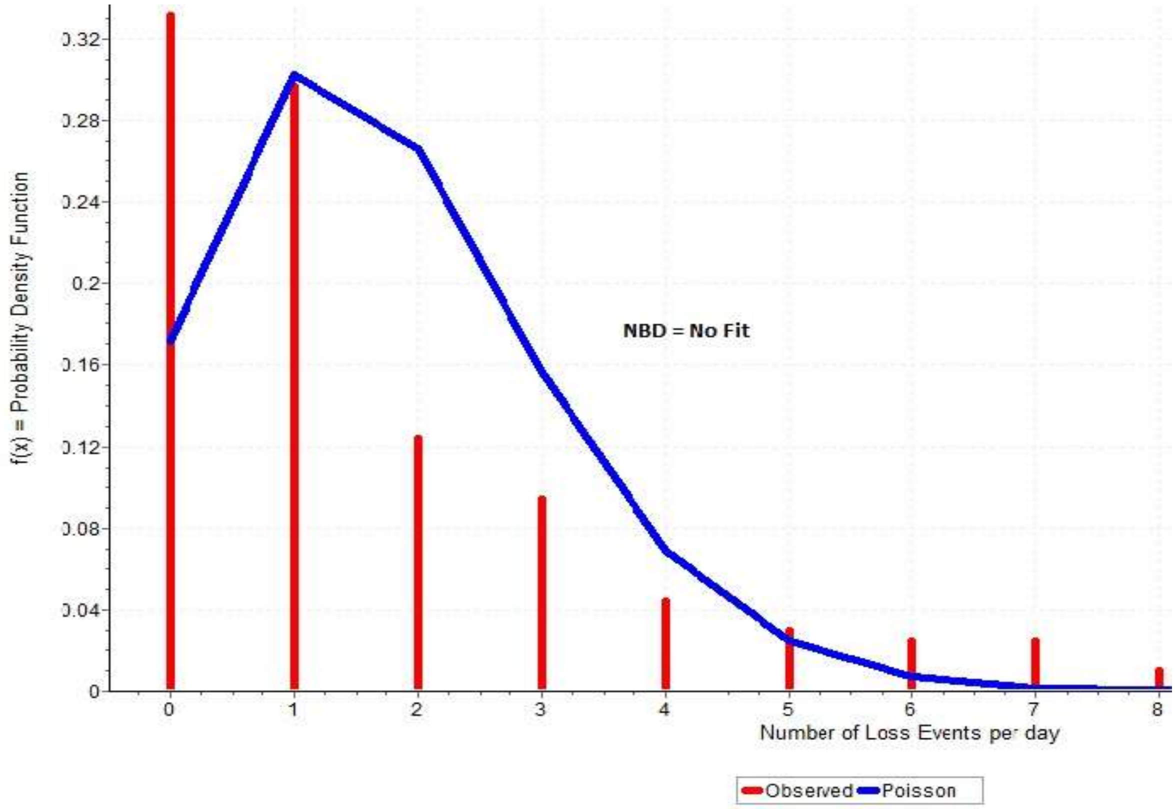
TABLE 4.9 FREQUENCY OF DAILY LOSS EVENTS- COPPORATE FINANCE



Event/ day	Observed frequency
0	67 (33%)
1	60 (30%)
2	25 (12%)
3	19 (9%)
4	9 (4%)
5	6 (3%)
6	5 (2%)
7	5 (2%)
8	2 (1%)
9	1 (0.5%)
10	1 (0.5%)
11	1 (0.5%)
12	1 (0.5%)
total	202 (100%)

Recall from Tables 4.2 and 4.3 in Section 4.1.1 that Corporate Finance comprises only 5132 (10.88%) loss events out of the total 47,269 loss events reported in the LDCE. From the diagrams above, zero loss events per day (no loss) has the highest frequency in with 33%. This is also contrary to the significantly high number of losses (2 loss events) per day when the 8 BLs were pooled together (Table 4.4). It can also be noticed that the frequency of loss events decreases as the number of loss events reported increases. This implies a negative correlation between these two random variables. Hence, *there is a negative correlation between the number of loss events per day and the frequency of losses*. The figure below shows the graphs of the Poisson and the NBD super-imposed against the daily observed frequency in Corporate Finance BL.

FIGURE 4.11 - POISSON & NBD FIT - CORPORATE FINANCE



From the figure above, the NBD did not fit the distribution hence; its graph was not shown. However, the Poisson fitted the observed frequency in Corporate Finance very well besides the peak of zero loss events. Before drawing a conclusion, it is relevant to ascertain this fitness by carrying out a chi square goodness of fit tests.

4.3.1.1 CHI SQUARE TEST - CORPORATE FINANCE

TABLE 4.10 CHI SQUARE TEST - CORPORATE FINANCE

	NBD	POISSON

Parameters (see Sections 4.5 & 2.5)	r= 2.7; p= 0.58	λ =1.762
n (no of loss events)	13	13
Degree of freedom (df)	n- p-1 = 11	n-1 = 12
Significant Level	5%	5%
X^2 (chi square value)	159.47	9.6
P (critical value)	18.31	21.03
Decision	Since X^2 > P: Hence, REJECT Ho	Since X^2 < P: Hence, ACCEPT Ho

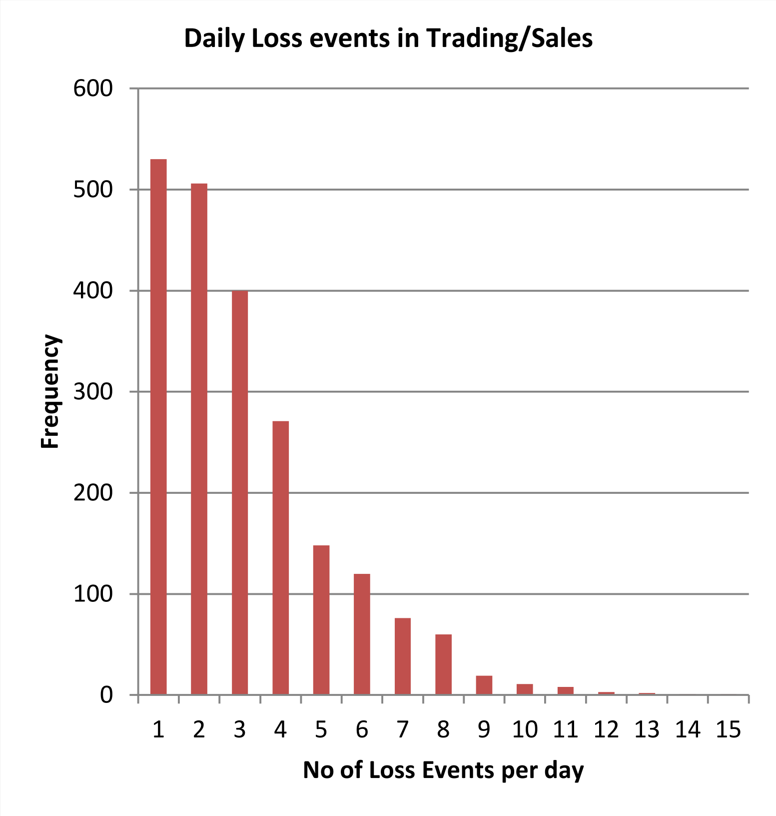
The parameters of the test have been fully explored in Section 4.2.5 under Table 4.8. From the above result, it can be seen in the last row that Chi Square accepted the Poisson and rejected the NBD at 95% confidence interval. In the light of the overwhelming evidence and the evidence obtained from the graph in Figure 4.11; it can be drawn that *the Poisson is a good distribution when modelling the frequency of daily operational loss events that occurred in the Corporate Finance line of business.* It can also be concluded conversely that *the negative binomial distribution is not a good distribution for modelling the daily frequency of loss events that occurred in the Corporate Finance line of business.*

4.3.2 FREQUENCY OF DAILY LOSSES IN TRADING AND SALES

The table below shows total number of operational loss events that were reported in Trading & Sales in the 2002 LDCE by the 89 banks that participated in the loss exercise. These losses have been aggregated daily by the RMG and will be the subject of analysis and exploration in this section.

FIGURE 4.12 FREQUENCY OF DAILY LOSS EVENTS- TRADING/SALES

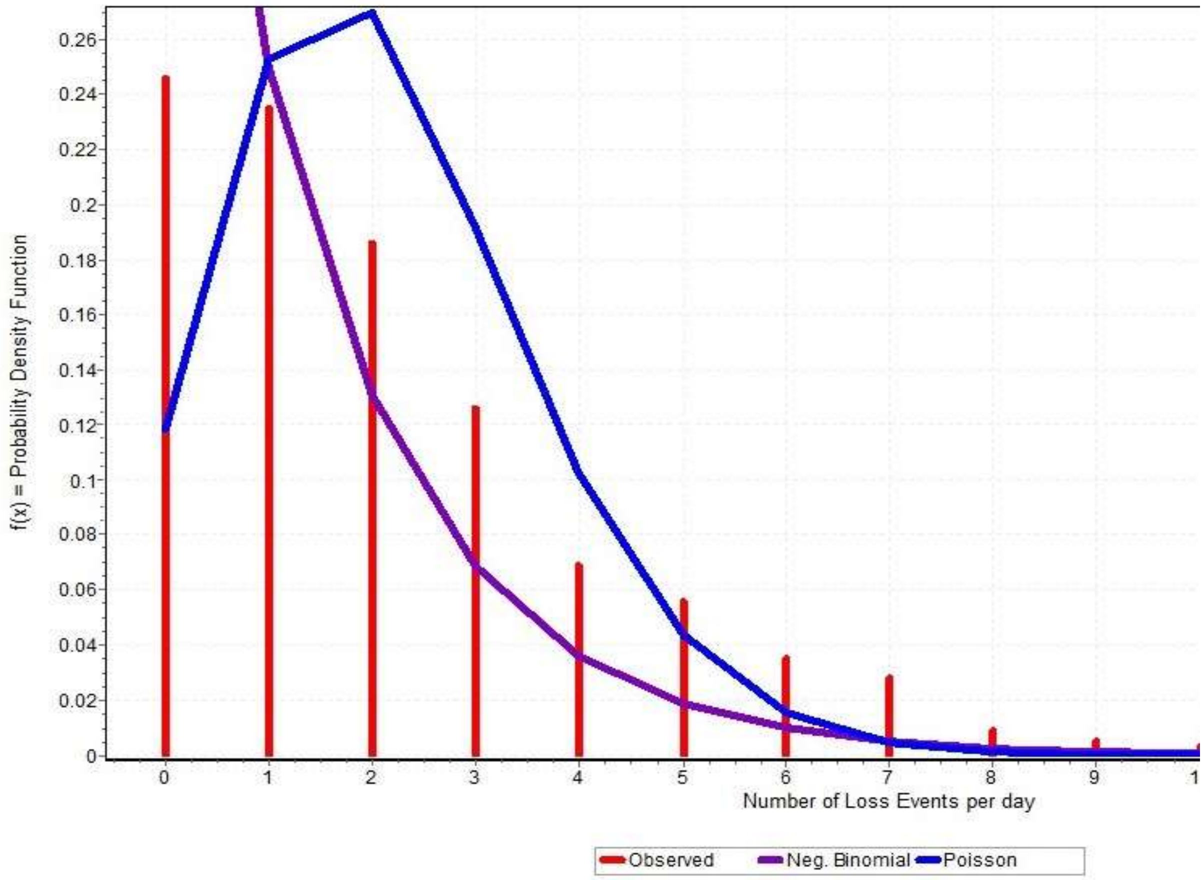
TABLE 4.11 FREQUENCY OF DAILY LOSS EVENTS- TRADING/SALES



Event /day	Observed Frequency	Percent
0	530	25%
1	506	23%
2	400	19%
3	271	13%
4	148	9%
5	120	6%
6	76	4%
7	60	3%
8	19	0.80%
9	11	0.50%
10	8	0.30%
11	3	0.10%
12	2	0.09%
13	1	0.04%
14	1	0.04%
total	2156	100%

Recall from Tables 4.2 and 4.3 in Section 4.1.1 that Trading & Sales comprises only 423 (0.89%) loss events out of the total 47,269 loss events reported in the LDCE. From the diagrams above, zero loss events per day (no loss) has the highest frequency in Corporate Finance with 67%. This is contrary to the significantly high 2 loss events per day when the 8 BLs were pooled together (Table 4.4). More so, the shape of the distribution in this BL is similar to that reported in Corporate Finance. The figure below shows the graphs of the Poisson and the NBD super-imposed against the daily observed frequency in Trading & Sales.

FIGURE 4.13 - POISSON & NBD FIT - TRADING/SALES



Examining the figure above, it appears that while the NBD seems to underestimate the observed frequencies, the Poisson, on the other hand, appears to over-estimate the observed frequencies. Although both models perfectly captured the actual frequency when one loss event was reported, both failed to capture the curve when zero events were reported which accounted for the highest frequency. Thus, no decision can be reached regarded which model better fits the data in this line of business. Hence, the Chi Square goodness of fit test will be carried out to statistically analyse the situation in broader details.

4.3.2.1 CHI SQUARE TEST AT 5%- TRADING/SALES

TABLE 4.12 CHI SQUARE TEST AT 5%- TRADING/SALES

	NBD	POISSON
Parameters (see Sections 4.5 & 2.5)	r= 2.7; p= 0.58	λ =2.135
n (no of loss events)	15	15
Degree of freedom (df)	n- p-1 = 13	n-1 = 14
Significant Level	5%	5%
X^2 (chi square value)	20.48	27.41
P (critical value)	21.03	25.0
Decision	Since X^2 < P: Hence, ACCEPT Ho	Since X^2 > P: Hence, REJECT Ho
Excess value (X^2 – P)	-0.55	2.41

The parameters of the test have been fully explored in Section 4.2.5 under Table 4.8. Interpreting the above Chi Square result, although the NBD was accepted while the Poisson was rejected, it can be seen that the Excess values of the two results are not significantly different. This implies that either of the two models will fit well for this type of observation. If the significant level is decreased from 5% to 1% thereby increasing the confidence level to 99%, it will be noticed that both Poisson and the NBD will be accepted. Notice the change in the critical values. Recall that most statistical experiments are mostly conducted at either 1% or 5%. Table 4.13 is an extract of the result obtained from Chi Square test at 1% significant level.

TABLE 4.13 CHI SQUARE TEST AT 1% - TRADING/SALES

	NBD	POISSON

Significant Level	1%	1%
Degree of freedom (df)	n- p-1 = 13	n-1 = 14
X^2 (chi square value)	20.48	27.41
P (critical value)	26.22	30.58
Decision	Since $X^2 < P$: Hence, ACCEPT Ho	Since $X^2 < P$: Hence, ACCEPT Ho

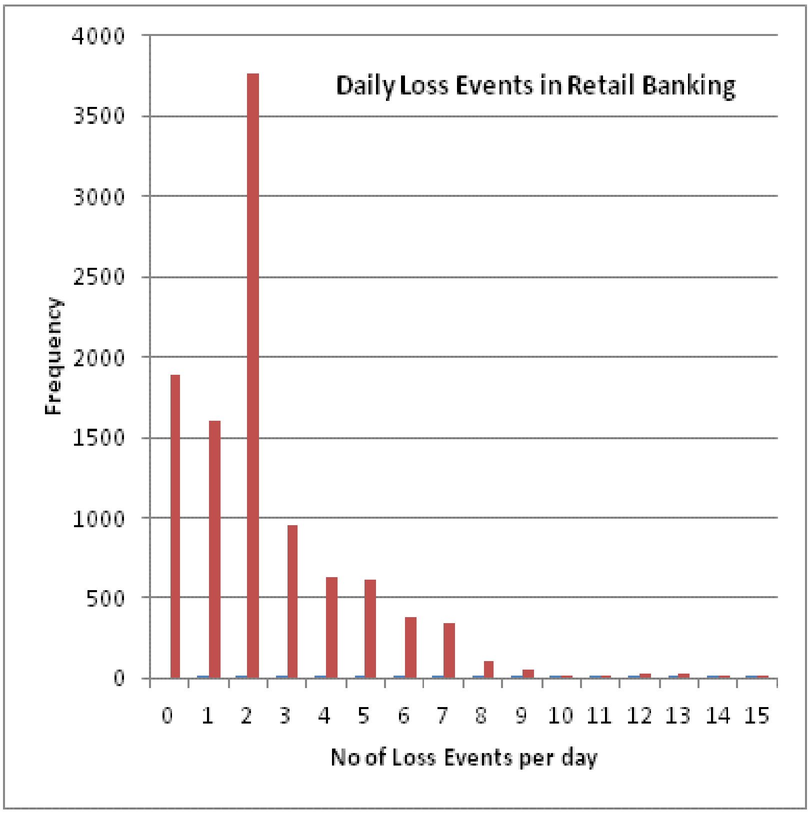
The parameters of the test have been fully explored in Section 4.2.5 under Table 4.8. At 1% significant level, both the Poisson and the NBD were accepted and at 5% significant level, only the NBD was accepted. More so, while the Poisson slightly over-estimated the actual frequencies in Figure 4.13, the NBD slightly under-estimated it. In the light of the above pieces of evidence, it can be concluded *that in modelling the number of loss events that occurred in Trading and Sales business line, the Poisson or negative binomial could be used as both fit the data.*

4.3.3 FREQUENCY OF DAILY LOSSES IN RETAIL BANKING

The table below shows total number of operational loss events that were reported in Retail Banking in the 2002 LDCE by the 89 banks that participated in the loss exercise. These losses have been aggregated daily by the RMG and will be the subject of analysis and exploration in this section.

FIGURE 4.14 FREQUENCY OF DAILY LOSS EVENTS- RETAIL BANKING

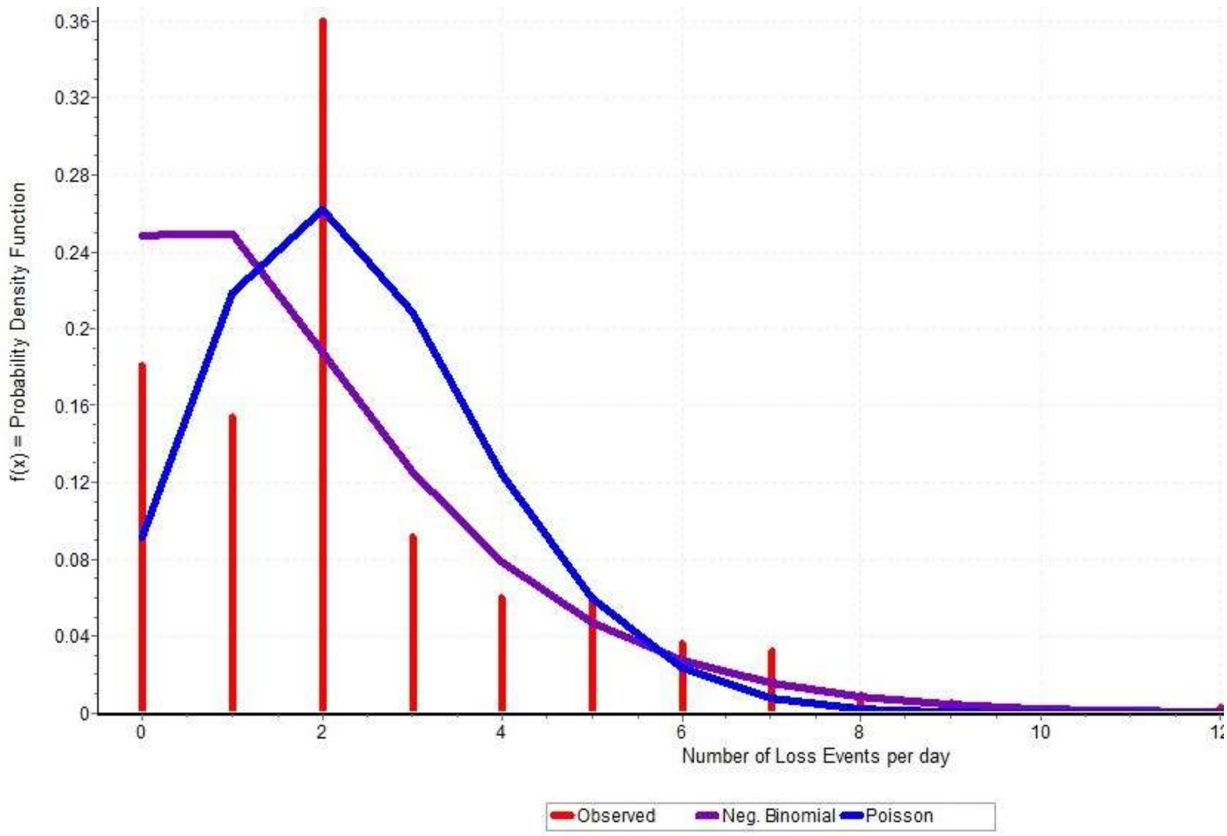
TABLE 4.14 FREQUENCY OF DAILY LOSS EVENTS- RETAIL BANKING



Event /day	Observed Frequency	Percent
0	1890	18%
1	1607	15%
2	3761	36%
3	958	9%
4	624	6%
5	615	5.8%
6	376	3.5%
7	342	3.3%
8	103	1.0%
9	59	0.6%
10	19	0.2%
11	15	0.1%
12	31	0.3%
13	24	0.2%
14	16	0.1%
15	10	0.1%
Total	10450	100.0%

The distribution of the frequency of daily loss events that occurred in Retail Banking is shown above. It comprises 26882 loss events accounting for 61% of the total number of loss events (see table 4.3). This distribution is similar to the one reported when all the BLs were aggregated in Figure 4.2. It is noticeable that the event of 2 losses per day is significantly higher than every other number of losses comprising 36% of total loss events. Besides this anomalous situation, the number of loss events decreases as the number of days increases. In other to explore this distribution further, the figure below shows the graphs of the Poisson and the NBD super-imposed against the daily observed frequency in Retail Banking.

FIGURE 4.15 - POISSON & NBD FIT - RETAIL BANKING



Undoubtedly, the fits for both Poisson and the NBD shown on the above graph is analogous to the one reported in Section 4.2.4 (figure 4.9) where the 8 BLs were pooled together and aggregated daily. This situation might be attributed to the fact that Retail Banking BL accounts for 61% of the total loss events. Hence, it is expected that its distribution will reflect that of the overall data. Details of the interpretation of this graphs has already been presented in Section 4.2.4 where it was concluded that the Poisson was a better fit. Hence*, it can also be concluded based on the fits above that Poisson is a better fit than the NBD when the daily loss events in Retail Banking is to be modelled.* It is not needed to carry out any test of hypothesis as the data is expected to support the Poisson process than the NBD.

4.4 ENSUING DISCUSSIONS

This section will discuss key issues and the conclusions drawn so far in this Chapter and will in turn, seek to answer the second research questions:

- Under what conditions should we adopt one model for the other?

We saw in Section 4.2.4 that when number of loss events that occurred in all 8 business lines were pooled together and aggregated daily, the Poisson model was a better fit than the NBD. We also saw in Section 4.3.1.1 that while the NBD did not fit into the daily frequency distribution of the loss events in Corporate Finance; the Poisson was a perfect fit. Similarly, the Poisson was also a better fit in the case of the Retail banking aspect of the BL. These findings are also consistent with Cruz, 2002; De Fontnouvelle et al, 2003 & 2005) who advocated the use of the Poisson model (see section 2.7). Conversely, we saw that in modelling the daily operational loss events that took place in Trading and Sales aspect of the BL, either Poisson or the NBD could be used as both fit the distribution. This is consistent with Moscadelli, 2004 who found no significant difference between the two models but decided to report the parameters for the NBD (see section 2.6). More so, some researchers (DaCosta, 2004; Kalkbrener, 2007; Osgood, 2000) advocated the use of the Poisson when the mean is somewhat equal the variance and the NBD when the variance is greater than the mean. This will be my next line of exploration.

4.4.1 MEAN AND VARIANCE COMPARISON TEST

It was discussed in Section 2.6 that Poisson and the NBD only differ by the assumptions made regarding their mean and variance. In Poisson, the mean and variance are equal, while in NBD the mean is less than the variance. Consequently, some researchers (DaCosta 2004; Kalkbrener 2007; Osgood 2000) have argued that Poisson should be used when the mean and variance are fairly the same; and the NBD when the mean is less than the variance. I will now use my data to investigate this recommendation because this will enable me to answer my second research question as stated in Section 4.4. The table below shows this information and the conclusions drawn from the Chi Square goodness of fit test.

TABLE 4.15 MEAN VARIANCE COMPARISON TEST

	Aggregate Data	Corporate Finance	Trading & Sales	Retail Banking
Chi Square Recommended Model	Poisson	Poisson	Poisson or NBD	Poisson
Mean (M)	2.3	1.7	2.1	2.4
Variance (V)	4.1	4.9	4.5	4.7
Condition	M < V	M < V	M < V	M < V

In all four scenarios, it can be seen that the mean is significantly less than the variance. According to this theory, this implied that the NBD would be a better fit. However, our Chi square test recommended the Poisson. Hence, it can be concluded that *there is no evidence to support the use of the mean variance comparison test in determining the correct model between the Poisson and NBD.* Conversely, *there is no condition upon which one model can be used in favour of the other.*

CHAPTER FIVE

5. CONCLUSIONS AND RECOMMENDATIONS

This chapter comprises 2 sections. It begins with a summary of the conclusions drawn from the study is presented in a table for ease of reference. It concludes by making some recommendations for further studies, as well as for practising managers of operational risks.

5.1 CONCLUSIONS DRAWN FROM STUDY

This study aimed to investigate which of Poisson or NBD is better for modelling the frequency of operational loss events in banks. The four research questions were investigated:

3. Is there a significant difference in the use the Poisson or Negative binomial distributions in modelling the frequency of operational risk losses?
4. Under what conditions should we adopt one for the other?

In order to answer the above research questions, the following Chi Square statistical hypotheses were tested at 5% significant level using n-p-1 degree of freedom:

5. H0 (Null hypothesis): the frequency of operational losses in banks follows the Poisson distribution.

6. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Poisson distribution.

7. H0 (Null hypothesis): the frequency of operational losses in banks follows the Negative Binomial distribution.

8. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Negative Binomial distribution.

Furthermore, Poisson and the NBD were fitted on the daily frequency of loss events of the following business lines: Corporate Finance, Trading and Sales and Retail Banking as well as on all eight business lines aggregated as a whole. Table 4.16 is a summary of the Chi Square goodness of fit recommended model at 5% significant level:

TABLE 5.1 RECOMMENDED MODELS

BUSINESS LINES	Aggregate Data (8 BLs)	Corporate Finance	Trading & Sales	Retail Banking
Chi Square Recommended Model	Poisson	Poisson	Poisson or NBD	Poisson

From this table, it is seen that the Poisson model fits on the three business lines including the summation of the eight business lines. Even though NBD fitted the data at both 1% and 5% levels of significant (Tables 4.12 & 4.13) on Trading and Sales, the Poisson also fitted the data at 1% significant level (Table 4.13). Additionally, looking the graphs of the daily frequency of loss events plotted against the theoretical frequencies (Figures 4.9; 4.11; 4.13 & 4.15) it can also be seen that the Poisson model fitted virtually in each of these business lines better than the NBD. Also, the NBD did not fit on the Corporate Finance business line. Hence, in the light of this overwhelming evidence, *it can be drawn that there is a significant difference in the use of Poisson or negative binomial in modelling frequency of operational loss events; while the Poisson model fits on all data, the NBD only fits on a minority of the distributions.* While this finding is inconsistent with Moscadelli, 2004 who found no significant difference between the two models but decided to report the parameters for the NBD. It is supported by many other researchers (Cruz, 2002; De Fontnouvelle et al, 2003 & 2005) who also advocated the use of the Poisson model in modelling the frequency of operational risk losses.

The second research question seeks to determine the condition on which one model can be more suitable than the other. Findings have shown so far that *there is no evidence that supports the conditions upon which one model could be adopted in favour of the other.* On the use of the mean variance relationship test to determine the model by some advocates (DaCosta 2004; Kalkbrener 2007; Osgood 2000); *there is also no evidence to support the use of the mean variance comparison test in choosing between the Poisson and NBD.*

The two tables below provide a summary of the conclusions drawn from this study. The first table is a summary of the research questions, data analysis and the main findings. The second table shows additional conclusions that ensured from data analysis and exploration. The reference sections that relate to the findings in the study are shown in the last column of the tables.

TABLE 5.2 THE SUMMARY OF MAIN FINDINGS

Item	Research Question	Data analysis	Result	Reference section
1	Is there a significant difference in the use the Poisson or Negative Binomial distributions in modelling the frequency of operational risk losses?	Chi Square goodness of fit test at 5% significant level. Graphs of observed and theoretical frequencies compared. Comparative bar charts	There is a significant difference in the use of Poisson or negative binomial model in modelling frequency of operational loss events.	4.2.4 to 4.3.3

2	Under what conditions should we adopt one for the other?	Mean Variance relationship test. Chi Square goodness of fit test at 5% significant level.	There is no evidence to support the use of the mean variance comparison test in determining the correct model between the Poisson and NBD. Hence, no evidence of any conditions.	4.4.1

TABLE 5.3 SUMMARY OF ADDITIONAL FINDINGS

Item	Findings	Reference section
1	the average number of loss events per year encountered by banks is up to 50	4.1
2	Retail Banking accounts for more than half the number of operational loss events.	4.1,1
3	the number of operational loss events due to fraud in retail banking accounts for up to 40% of losses	4.1.1
4	the average number of operational loss events is two per day	4.2
5	the frequency distribution of operational loss events is positively skewed.	4.2
6	any number of loss events exceeding 6 per day is too extreme and hence, “too many”.	4.2.1
7	the number of loss events per day can be considered	4.2.1

	normal if it is between 0 and 6.	
8	the daily distribution of the number of loss events has a high kurtosis and hence, leptokurtic.	4.2.1
9	the Poisson distribution models the overall daily operational loss events when aggregated daily better than the NBD.	4.2.4
10	the daily operational loss events across banks fit better when modelled with the Poisson distribution than the negative binomial distribution.	4.2.5
11	there is a negative correlation between the number of loss events per day and the frequency of losses	4.3.1
12	the Poisson is a good distribution when modelling the frequency of daily operational loss events that occurred in the Corporate Finance line of business	4.3.1.1
13	the negative binomial distribution is not a good distribution for modelling the daily frequency of loss events that occurred in the Corporate Finance line of business	4.3.1.1
14	In modelling the number of loss events that occurred in Trading and Sales business line, the Poisson or negative binomial could be used as both fit the data.	4.3.2
15	Poisson is a better fit than the NBD when the daily loss events in Retail Banking is to be modelled	4.3.3
16	There is no evidence to support the use of the mean variance comparison test in determining the correct model between the Poisson and NBD.	4.4.1

5.2 RECOMMENDATIONS AND FURTHER RESEARCH

The arrival of operational loss events is of a rather chaotic nature, and events occur at irregular intervals of time. It is therefore crucial to examine the frequency distribution in order to understand the underlying loss arrival process (Chernobai, 2007). This is fundamental in correctly estimating the statutory operational value at risk. Empirical studies with operational loss data mainly emphasize the use of a simple Poisson model or the negative binomial distribution to model the frequency of loss distribution over a certain time interval. This study found that the approaches used in modelling these loss events can be complex for risk managers. Nevertheless, these approaches can be strengthened by the following recommendations:

- Operational risk managers should use either the Poisson or the NBD, but preferably the former, in modelling the frequency distribution of operational loss events since it can be fitted in most databases.
- Since truncation of database, i.e. adding two separate databases or business lines together, is very frequently required, the Poisson distribution proves to be the right choice since if a Poisson fits in an entire database (e.g. the entire 8 business lines)), it will also fit a truncated database (e.g. in Retail Banking) as seen in Table 4.16.
- Following from the previous recommendations, Poisson also has the unique property of Poisson (A) + Poisson (B) = Poisson (A+B). This commutative Poisson law implies that it is easy to include more data without structurally changing the data analysis of a Poisson hence, easy to use.
- Since only one parameter (the mean) is needed to fit Poisson on a distribution, Poisson becomes very easy to use and hence, recommended.
- In choosing between Poisson and NBD that both fairly fit a distribution like in figure 4.13 (Trading & Sales), it is recommended to choose the simpler of the two, i.e. the Poisson. This view is held by Simon & Schuster (1956) in their popular theory called “Occam’s Razor Principle”. They theorized that: “descriptions should be kept as simple as possible until proved inadequate” (Simon & Schuster, 1956).

- Since a simple Excel spreadsheet has the Poisson function predetermined which is usually represented by "= Poisson(x, λ, True)", this makes it very easy for risk managers to fit the curve on a distribution without the hazards of learning the technicality of new software.
- It is not always visually easy to decipher which models better fits the actual frequencies when graphical packages are used to super-impose distributions (e.g. figure 4.13). Hence, it is recommended that risk managers should carry out hypotheses tests to determine the model that better fits the actual distribution.

On recommendations for further studies, more explorations are needed in investigating this type of problem by using recent data to determine if banks' risk appetites have changed over the years. For instance, the recent 2008 LDCE data could be used which composes of a six year published loss data from 2003 to 2007 on operational losses. This six year data can be investigated for frequency of annual rather than daily operational loss events. Besides the BCBS loss data, the Operational Risk Data Exchange (ORX) publishes similar data on operational loss events. These data be modelled for Poisson and NBD to see whether its findings are consistent with that of LDCE. The ORX data can be accessed at www.orx.org. As this investigated limited the number of business lines to three, future research could try to fit the Poisson and NBD on all eight business lines. It could even go further to fit both models on the seven event types to determine the best model for each event type and combinations of event and business type.

REFERENCES

Alexander, C. (2003). *Operational Risk: Regulation, Analysis and Management.* London, Great Britain: Prentice Hall.

Alvarez. (2002). *Operational Risk Event Classification*. Retrieved from GARP: http://www.garp.com

Anderson, J. (2000). *Fundamentals of Educational Research.* London: Routledge.

Anscombe, F. (1949). The Statistical Analysis of Insect Counts on the Negative Binomial Distribution. *Biometrics* , pg165.

Arbous, A., & Kerrich, J. (1951). Accident Statistics and the Concept of Accident Proneness. *Biometrics* (7), 358.

Attwood, G. (2000). *Statistics 2: Heinemann Modular Mathematics for AS and A Level.* Oxford: Heinemann.

BCBS. (2003). *The 2002 Loss Data Collection Exercise for Operational Risk: Summary of Data Collected.* Banking for Internal Settlements, Risk Management Group.

Bell, J. (2005). *Doing your Research Project.* London: Open University Press.

Bening, V., & Korolev, V. Y. (2002). *Generalised Poisson Models and their Applications in Insurance and Finance.* Utrecht, Boston: VSP International Publishers.

BIS. (1999). *A New Capital Adequacy Framework.* Basel Committee on Banking Supervision.

BIS. (2006b). Internal convergence for Capital Measurement and Capital Standards. *Banking for Internal Supervision*, (p. 147).

BIS. (2001b). Working Paper on the Regulatory Treatment of Operational Risk.

BoE. (1995). *The Bank of England Report into the Collapse of Barings Bank.* London: Her Majsty's Stationery Office.

Bortkiewicz, L. (1898). Dependent Events and Operational Risk. *Algo Research Quarterly , 5* (2), p5-7.

Bryce, R. (2002). Pipe Dreams: Greed, Ego and the Death of Enron. New york: Public Affairs.

Chernobai, A. (2005). *Composite Goodness of Fit Test for Left Truncated Loss samples.* Santa Barbara: Technical Report, University of California.

Chernobai, A. (2007). *Operational Risk: A guide to Basel ll Capital Requirements, Models, and Analysis.* New jersey: Wiley Finance.

Cope, E., & Willis, S. (2008). External Loss Data Helps: Evidence from ORX Database. *Operational Risk and Compliance* , 48-49.

Crouhy, M., Galai, D., & Mark, R. (2001). *Risk Management.* New York: McGraw-Hill.

DaCosta, L. (2004). *Operational Risk with Excel and VBA.* New Jersey.

de Fontnouvelle, P. (2003). *Using Loss Data to Quantify Operational Risk,.* Boston, Masachussette: Federal Reserve Bank of Boston.

Deeks, S. (1999). Modelling Counts- Poisson and Negative Binomial Regression. *Financial Econometrics* , p.750.

DeGroot, G. (2002). *Probability and Stattsitics* (3rd ed.). San Francisco: Addison Wesley.

Devroye, L. (1986). *Non Uniform Random Variate Generation.* New york: Springer-Verlag.

Dyer, G. (2003). *Edexcel GCSE Statistics.* Oxford: Heinemann.

Ehrenberg, A. (1988). *Repeat Buying: Facts, theory and Applications.* London: Charles Griffin & Co.: Oxford University Press.

Embrechts, P. (1997). *Modelling External Events for Insurance and Finance.* Berlin: Springer-Verlag.

FDIC. (1995). Regulators Terminate the US Operators of Daiwa Bank, Japan. Tokyo.

Grandell, J. (1997). *Mixed Poisson Processes.* London: Chapman and Hall.

Gregoriou, G. (2009). *Operational Risk Towards Basel III: Best Practices and Issues in Modelling, Management, and Regulation.* New Jersey: Wiley Finance.

Jarrion, P., & Roper, R. (1995). Big Bets gone Bad: Derivatives and Bankruptcy in Orange county.

Jobst, A. (2007). The Treatment of Operational Risk Under theNew Basel Framework: Critical Issues. *Journal of Banking and Regulations* , 316.

Jorion, P. (2000). *Value at Risk:The New Benchmark for Managing Financial Risk* (2nd ed.). New York: McGraw-Hill.

Kalkbrener, M. (2007). LDA at Work: Deutsche Bank's Approach to Quantifying Operational Risk. *Journal of Operational Risk , Vol.1* (4), 62.

Kalkbrener, M., & Aue, F. (2007). Loss Distribution Approach at Work: Deutsche Bank's Approach to Quantifying Operational Risk. *Journal of Operational Risk* , 51.

King, J. (2001). *Operational Risk: Measurement and Modelling.* New York: John Wiley & Sons.

Klugman, S., Panjer, H. H., & Willmot, G. E. (2004). *Loss Models: From Data to Decisions* (2nd ed.). New jersey: John Wiley & Sons.

Koernert, J. (1996). The Collapse of Barings Bank 1995. *Financial Derivatives, Banking Crises and Contagion Effects. 96/2.* Freiberg Working Papers.

Lewis, C., & Lantsman, Y. (2005). What is a Fair Price to Transfer the Risl of Unauthorised Trading? A case Study on Operational Risk. In E. Davies, *Operational Risk: Practical Approaches to Implementation* (p. 315). London: Risk Books.

Mori, T., & Harada, E. (2001). *Internal Measurement Approach to Operational Risk Capital Charge.* Technical Report, Bank of Japan.

Moscadelli, M. (2004). The Modelling of Operational Risk Experience with the Analysis of the Data Collected by the Basel Committee.

Osgood, D. (2000). Poisson-based Regression Analysis of Aggregate Crime Rates. *Journal of Quantitative Criminology* , pg 21.

Piza, E. (2012). *Using the Poisson and Negative Binomial Regression Models to Measure the Influence of Risk on Crime Incident Counts.* Newark, NJ: Rutgers Centre on Public Security.

Ross, S. (2002). *Introduction to Probability Models* (8th ed.). Boston: Academic Press.

Sakamoto, C. (1973). Application of the Poisson and Negative Binomial Models to Thunderstorm Probabilities in Nevada. *Monthly Weather Review , Vol. 101* (4), 351.

Simon, P., & Schuster, E. (1976). *The World of Mathematics, Newman J.R. ed- Occam's Razor Principle* (Vol. 2).

Soprano, A. (2009). *Measuring Operational and Reputational Risk: A Practitioners' Approach.* Chichester, West Sussex: Wiley Finance.

APPENDICES

Appendix A: BCBS 2003 LDCE Data on Operational Risk

Number of Individual Loss Events per Business Line and Event Type
Sample 1: All Bank and All Losses
89 Banks Reporting

	Internal Fraud	External Fraud	Employ-ment Practices & Workplace Safety	Clients, Products & Business Practices	Damage to Physical Assets	Business Disruption & System Failures	Execution, Delivery & Process Manage-ment	No Event Type Information	Total
Corporate Finance	17 0.04%	20 0.04%	73 0.15%	73 0.15%	16 0.03%	8 0.02%	214 0.45%	2 0.00%	423 0.89%
Trading & Sales	47 0.10%	95 0.20%	101 0.21%	108 0.23%	33 0.07%	137 0.29%	4,603 9.74%	8 0.02%	5,132 10.86%
Retail Banking	1,268 2.68%	17,107 36.19%	2,063 4.36%	2,125 4.50%	520 1.10%	163 0.34%	5,289 11.19%	347 0.73%	28,882 61.10%
Commercial Banking	84 0.18%	1,799 3.81%	82 0.17%	308 0.65%	50 0.11%	47 0.10%	1,012 2.14%	32 0.07%	3,414 7.22%
Payment & Settlement	23 0.05%	322 0.68%	54 0.11%	25 0.05%	9 0.02%	82 0.17%	1,334 2.82%	3 0.01%	1,852 3.92%
Agency Services	3 0.01%	15 0.03%	19 0.04%	27 0.06%	8 0.02%	32 0.07%	1,381 2.92%	5 0.01%	1,490 3.15%
Asset Management	28 0.06%	44 0.09%	39 0.08%	131 0.28%	6 0.01%	16 0.03%	837 1.77%	8 0.02%	1,109 2.35%
Retail Brokerage	59 0.12%	20 0.04%	794 1.68%	539 1.14%	7 0.01%	50 0.11%	1,773 3.75%	26 0.06%	3,268 6.91%
No Business Line information	35 0.07%	617 1.31%	803 1.70%	54 0.11%	13 0.03%	6 0.01%	135 0.29%	36 0.08%	1,699 3.59%
Total	1,564 3.31%	20,039 42.39%	4,028 8.52%	3,390 7.17%	662 1.40%	541 1.14%	16,578 35.07%	467 0.99%	47,269 100.00%

Legend	Greater than 20%	10% through 20%	5% through 10%	2.5% through 5%

Appendix B - Critical Value Table for Chi Square

		p				
		50%	10%	5%	1%	0.1%
	1	0.45	2.17	3.84	6.63	10.83
	2	1.39	4.61	5.99	9.21	13.82
	3	2.37	6.25	7.82	11.34	16.27
	4	3.36	7.78	9.49	13.28	18.47
	5	4.34	9.24	11.07	15.09	20.52
	6	5.35	10.64	12.59	16.81	22.46
	7	6.35	12.02	14.07	18.48	24.32
ν	8	7.34	13.36	15.51	20.09	26.13
	9	8.34	14.68	16.92	21.67	27.88
	10	9.34	15.99	18.31	23.21	29.59
	12	11.34	18.55	21.03	26.22	32.91
	15	14.34	22.31	25.00	30.58	37.70
	20	19.34	28.41	31.41	37.57	45.32
	25	24.34	34.38	37.65	44.31	52.62
	30	29.34	40.26	43.77	50.89	59.70

APPENDIX C- Approved Thesis Proposal

Royal Docks
Business School

UEL
University of East London

Student Number: U1043853

Programme: MSc Risk Management

Proposed Dissertation Title: Modelling the Frequency of Operational Risk Losses under the Basel 2 Capital Accord: A Comparative study of the Poisson and the Negative Binomial Distribution Methods

Turnitin ID: 22375218.

SMM227 Dissertation Proposal Template

Introduction to the Dissertation Topic (200 words approximately)

- *What is your topic?*

Modelling the Frequency of Operational Risk Losses under the Basel 2 Capital Accord: A Comparative study of the Poisson and the Negative Binomial Distribution Methods

- *What is your focus?*
  This study will broadly look into the two major methods of modelling the frequency of loss data under the Basel Committee on Banking Supervision (BCBS) Accord of 1998 known as Basel 2 Capital Accord. It will compare the Poisson method of modelling the frequency of losses to that of the Negative Binomial method. As the frequency of loss cannot be fully explored without mention of the severity of loss, this study will however, primarily focus on the former.
- *How will you investigate it?*

The frequency of loss will be investigated using a cross section of secondary data published by the Banking for International Settlements (BIS) resulting from the 2008 Loss Data Collection Exercise for Operational Risk. Details available at: www.bis.org. Other sources of data include the British Banking Association (BBA) Global Operational Loss Database available at: http://www.bbagold.org/Default.aspx; as well as the Operational Risk-data Exchange Association (ORX) database on operational losses available at: http://www.orx.org/orx-data.

- *Why is this research topic important?*
  It is a statutory requirement under the Basel Accord on Banking Supervision for financial institutions to set aside minimum capital requirements for operational risks losses (expected and unexpected). Hence, this study will develop or further enhance the knowledge of risk managers to make informed decisions.

Literature Review (700 words approximately)

- *What literature will you use for your project?*

The key literature will be obtained from a plethora of literature published by the Basel Committee on Banking Supervision, especially that relating to the Advanced Measurement Approach (AMA) for Operational Risks. These include the Basel Committee on banking supervision publications, its avalanche of research papers on operational risks modelling, its annual and quarterly report papers on operational risks and its internal journal of central banking publications, as well as the operational loss data collected by Basel Committee in their Quantitative Impact Study (QIS) by the Risk Management Group (RMG). Other sources of literature will include the British Banking Association (BBA) publications on supervisory guidelines on Advanced Measurement Approach (AMA); Journal of Operational Risks; Journal of Risk Management in Financial Institutions; and Chase Cooper Limited which is a private company specialising in operational risk management and compliance for the Basel 2 (see chasecooper.com). More so, key literature will be gathered through key texts such as "Operational Risk: A guide to Basel 2 Capital Requirements, Models, and Analytics"-A.S. Chernobai; "Guide to Optimal Operational Risk & Basel 2"- I. A. Akkizids; "Operational Risk Control with Basel 2"- N. Dimitris;" Measuring Operational and Reputational Risk"- A. Soprano; and a host of literature from other reputable authors in operational risks.

- *Who are the key writers?*

Prominent key writers in the area of modelling the frequency and severity of operational loss data under the Basel 2 approaches will be featured in this study. They include Cruz (2003) who studied the frequency of over 3,000 operational losses of a major British retail bank; Moscadelli (2004) who critically analysed and examined the Loss Data Collection Exercise (LDCE) by the Basel 2 Risk Management Group (RMG) in 2002; Jordan et al. (2005) who re-examined the same data analysed by Moscadelli (2005); Lewis & Lantsman (2005) whose concern was in analysing the loss data relating to unauthorised trading; Chernobai et al. (2006) who examined 10 years insurance loss data; Chernobai et al. (2005b) who examined 22 years operational risk loss data; Chernobai & Rachev (2004) who examined the inter-arrival time between frequency of consecutive losses; BIS (2009) whose committee analysed the results from the 2008 Loss Data Collection Exercise for operational risk; BIS (2011) whose committee published a supervisory guidelines for the Advanced Measurement Approaches; Perry & Dutta (2007) who examined analysis of loss distribution models for estimating operational risk capital; among others.

- *What are the key debates in the literature?*

In response to the Basel 2 Capital Accord which requires banks to set aside a minimum amount of capital in case of catastrophic operational losses. Hence, banks have to use certain distributions to model an estimate for the Value at Risk (VAR). This VAR arrived calculated using the severity and the frequency of operational losses that occur in a given year. The VAR must be correctly estimated because an underestimation meant that the bank is not fully covered for operational looses. On the other hand, an over estimation implies that a bank is denying itself of investment opportunities. The literature tries to justify the best types of distributions that can be used to model the frequency and severity of operational losses. Notable among the discrete random distributions are Poisson, Geometric, Binomial, Negative Binomial and the Hyper geometric distributions. It is however, evident from the emerging literature that the Poisson distribution is the most popular and commonly used because of its ease of use, i.e. it requires the value of only one parameter which is the average loss. Many authors also noted that the use of Poisson can be affected by the minimum amount of losses reported since most financial institutions have the minimum threshold of reported losses. More so, the Poisson also assumes a constant rate of losses occurring in a given time interval. This shortfall, according to notable literature, can be fixed by applying the Negative Binomial distribution which is more flexible with the inclusion of an additional parameter. As regards modelling the severity of operational losses, distribution used includes

Exponential, Lognormal, Weibull, Gamma, Beta, Pareto distribution, and other continuous random distributions. However, severity of distribution modelling is beyond the scope of this study.

- *What aspect of theory will you focus upon for your project?*

This study will primarily focus on modelling the frequency of operational losses using the Poisson and the Negative Binomial discrete random distributions. It will make a comparative analysis of both models with a view to drawing some emerging conclusions.

Research Methodology (400 words approximately)

- *What are your research question/objectives?*

The main objective of this study is to investigate whether the Poisson or Negative binomial distribution can be used to adequately model the frequency of operational risk losses in banks losses under the Basel 2 Accord. It will also investigate the conditions under which each is appropriate for adoption. Hence, my research questions are:

3. Is there a significance difference the results obtained between the use the Poisson and Negative binomial distributions in modelling the frequency of operational risk losses?
4. Under what conditions should we adopt one for the other?

- *How will you address them?*

This problem will be addressed by using data from the LDCE to fit the Observed and the Expected distributions for the Poisson and Negative binomial distributions to determine which one is a better fit.

- *What methodological approach will you use?*

The methodology that will be used is the concepts of the Poisson and the Negative binomial distributions which are discrete random variables that can be used to predict the frequency of operational losses over time. The following hypotheses will be tested:

5. H0 (Null hypothesis): the frequency of operational losses in banks follows the Poisson distribution.
6. H1 (Alternative hypothesis): the frequency of operational losses in banks does not follow the Poisson distribution.

These hypotheses will be tested using the Pearson's Chi square test for goodness of fit distribution with n-1 degree of freedom at 5% level of significance. Thus, if the value of Chi-square at n-1 degree of freedom is less than the critical value, then the Null hypothesis will be rejected and the Negative binomial distribution will be recommended for use. This Pearson's Chi square test will determine whether the Poisson or the negative binomial can adequately predict the frequency of operational losses and will also ensure that the right distribution is used for a particular data. The Autograph statistical software will be used to calculate the parameters for the Poisson and negative binomial distributions as well as their probabilities.

- *How do you envisage collecting your data?*

As stated elsewhere in this proposal, the frequency of operational losses will be investigated using a cross section of secondary data published by the Banking for International Settlements (BIS) resulting from the 2008 Loss Data Collection Exercise for Operational Risk under the Quantitative Impact study (QIS) publicly available for use at: www.bis.org. This publication comprised data from the 121 financial institutions that participated in the data collection exercise from 17 countries. Of the 121 participatory institutions, 119 provided their internal loss data comprising over 10.6 million frequencies of losses from the years 2002 to 2008. This includes data from all the 8 business lines and 7 event types. These data will be explored using the methodology stipulated above.

Conclusion (150 words approximately)

- *What is the significance of this project?*

This study will equip operational risk managers with the technical knowledge and understanding with which to make complex decisions regarding the appropriate discrete distributions to be employed under a given condition.

- *What insights will it provide on the topic?*

It will provide operational risk managers or anyone involved in modelling the frequency of operational risks the insights with which to carry out significant tests, e.g. Pearson's Chi square test in order to determine the most appropriate statistical distribution to be used to model loss frequencies.

- *Will the findings be useful for academics and/or practitioners?*

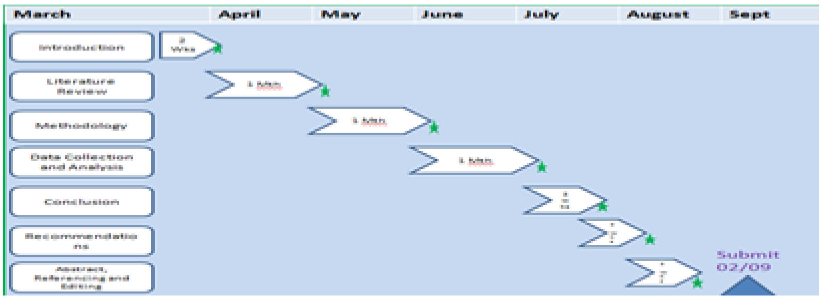
It is hoped that the findings of this study will be beneficial to researchers who will be urged to investigate this area further, and also to operational risk practitioners who need a handy practical approach to model frequency of losses.

- *How will the project benefit your future career ambitions?*

This project will equip me with the requisite knowledge and understanding to perform my job beyond expectations in my role as an Operational Risk Manager.

Time plan for completion of the project on a Gantt Chart (100 words approximately)

- What is the time frame for each chapter?



- What is your submission date?

03/09/2013

Bibliography

Alexander, C (2003) Operational Risk: *Regulation, Analysis and Management*. Prentice Hall, London, England.

BCBS (2009) Basel Committee on Banking Supervision*: Result from the 2008 Loss Data Collection Exercise for Operational Risk*, Bank for International Settlements, Basel, Switzerland.

BCBS (2011) Operational Risk: *Supervisory Guidelines for the Advanced Measurement Approaches*, Bank for International Settlements, Basel, Switzerland.

Blunden, T & Thirlwell, J (2010) Mastering Operational Risk: *A Practical Guide to Understanding Operational Risk and how to Manage It*, Prentice Hall, Gosport, Hampshire, England.

Chernobai, A.S (2007) Operational Risk: *A Guide to Basel II Capital Requirements, Models and Analysis*. Frank J Fabozzi Series, New Jersey.

Cruz, M.G (2002) Modelling, Measuring and Hedging Operational Risk. Wiley finance series, Chichester, West Sussex.

Soprano, A (2009) Measuring Operational and Reputational risk: *A Practitioner's Approach*, Wiley Publishing, Chichester, West Sussex.