Value at Risk (VaR) Backtesting

‘Evidence from a South African Market Portfolio’

Gerald Z. Katsenga

Master of Management in Finance and Investments

Dissertation Submitted in partial fulfillment of the requirements for the Degree of Master of Management in Finance and Investments. University of Witwatersrand Business School

Supervisor: Dr. Blessing Mudavanhu

20 February 2013

Declaration

I, Gerald Z. Katsenga, hereby,

1. grant the University of Witwatersrand free license to reproduce this dissertation in whole or in part, for the purpose of research;

2. declare that:

    i. this dissertation is my own unaided work, both in concept and execution,

    ii. neither the substance nor any part of this dissertation has been submitted in the past, or is being, or is to be submitted for a degree at this University or at any other university.

I am now presenting the dissertation for examination for the degree of Master of Management in Finance and Investment at the University of Witwatersrand, Johannesburg.

Gerald Z. Katsenga

Abstract

Value at Risk (VaR) has emerged as one of the most prominent risk measurement techniques in finance. It is a measure that quantifies the worst expected loss over a given confidence level and target horizon, under normal market conditions. In this thesis, the concept of VaR as an invaluable tool for financial risk management is explained, and a theoretical but detailed description of some of the methods of VaR computation are presented, with a key emphasis on the assumptions and shortcomings of these models. A discussion that is preceded by a presentation on the calculation of portfolio returns and the choice of VaR parameters as key determinants in the choice of an appropriate VaR model.

In light of the shortcomings to VaR measures, a number of backtesting techniques that examine the accuracy of VaR estimates are presented and reviewed. The review of these VaR validation methods is both from a theory and practice perspective. The unconditional coverage, independence property and the conditional coverage property are defined and their relation to backtesting methods discussed. The backtesting techniques presented are classified by whether they test for the unconditional coverage property, independence property or the conditional coverage (joint) property of a VaR measure.

The backtesting methods presented and discussed in this work are then utilized in empirically validating the accuracy of one of the most widely used VaR method in South Africa, the historical simulation. The examination of VaR estimates from this model is applied on an 'actual' interest rate portfolio. The outcome of the statistical backtests show positive results of accurate performance of the model at lower confidence levels, and an underestimation of risk at higher VaR confidence levels. However, when hedge positions are excluded from the portfolio the model's performance in accurately estimating VaR is questionable.

Acknowledgments

I would like to express my profound gratitude to Dr Blessing Mudavanhu, my supervisor, for his immeasurable support. I am also thankful to Professor Ojah for his guidance through the entire course of my study.

Of course, I am indebted to my daughter Fadzai for her support, patience and *love*. My mother and family, thank you for the encouragement.

A special thanks to Kim Buker, for the opportunity, invaluable advice, her comments and guidance. The gratitudes would not be complete without the mention of two other people, Tshepo Segone and Dr Ted Powers, your support is acknowledged.

Wits Business School, Johannesburg — Gerald Z. Katsenga
February 20, 2013

Table of Contents

List of Tables

Chapter 1

Introduction

"*All of life is the management of risk, not its elimination.*" (Wriston, 2009)

Financial markets are forever changing, and change can either be positive or negative. The prospect of gain or loss is a consequence of change, which therefore leads to risk, and risk (or more precisely, downside risk) is something that is sometimes inevitable in financial markets. The acceptance of risk does not invariably imply the elimination of risk, which clearly is impossible. Nor does it imply that one should not mitigate risks and accept consequent losses fatalistically. It means that risk requires management. Decide on what risks to avoid and how to avoid them. What risks to accept, the terms of accepting them and what new risk to take on (Dowd, 1998). This thus suggests that the two main pillars of managing risk are, firstly, to establish a risk quantification method, and secondly, to develop and implement a validation or backtesting technique for the estimated risk numbers.

1.1 Problem Statement

Value at Risk, or VaR as it is commonly known, has emerged as one of the most prominent risk measurement techniques in finance. It is presently the most used risk management technique by financial institutions, non-financial institutions, and regulators alike. VaR models were developed in order to provide a quantifiable estimate of the potential loss on a risky financial asset or a portfolio of risky financial assets (or institution) to market risk. The VaR of a portfolio (institution) quantifies the maximum loss, in value, that will not be exceeded given a level of confidence over a given target horizon, *under normal conditions* (Jorion, 2007). VaR methodologies focus on maximal potential losses of a portfolio (or institution) and derive from modern techniques in finance, developed in response to the financial failures in the financial and derivative markets. While VaR can be used by any firm to measure its risk exposure, it is commonly used by commercial and investment banks to quantify potential loss in value of financial derivatives (or traded portfolios) from adverse market movements over a given period.

Furthermore, since 1998 the Basel Committee on Banking Supervision (BCBS) regulatory guidelines have required banks with substantial trading activity to set aside capital to insure against extreme portfolio losses. The market risk capital prerequisite, or quantum to be set aside, is directly related to a measure of portfolio risk, which is measured in terms of VaR (Campbell, 2005). This current framework of regulation requires that firms or financial institutions (in particular banks) utilize their own internal risk models in computing and reporting their VaR[1]. Therefore, the risk capital requirements have a direct link to both the estimated level of portfolio risk as well as the VaR model's performance on validation (backtests). However, it is important to note that when implementing VaR systems, invariably numerous simplifications and assumptions are involved.

Notwithstanding the common acceptance and wide use of VaR as a financial risk management tool, several criticisms have arisen concerning VaR methods (and are still debated). Amongst some of the issues raised are:

(i) *Model performance*, see e.g. Berkowitz and O'Brien (2002), Engelbrecht (2003), and also the work of Bao, Lee and Saltoglu (2006);

(ii) *Unreliable risk estimates*, see for example Dowd (2001) or Wolff (2005);

(iii) *Method accuracy vs computational time*, refer to Pristker (1997);

(iv) *Choice of volatility measure in VaR*, for example Giot and Laurent (2004) or the work of Thupayagale (2010).

[1]Banks and financial institutions are required to hold capital against market risk based on VaR at 99% confidence interval and a 10 day time horizon.

It seems that the first criticism (i.e. model performance) tends to dominate the others in VaR literature. This may be as a result of the adoption of VaR by the BCBS as the international standard for external capital regulatory purposes (Pritsker, 1997). While the highlight is on model performance, there appears to be little coverage on the comparison of the estimation performance of VaR techniques. Moreover, research results tend to vary widely depending on data and VaR parameters. An illustration is for example studies by Bao, Lee, and Saltoglu (2006) and that by Mancini & Trojani (2008) which present conflicting results on the VaR performance of the RiskMetrics[2] method. The former showing a very satisfactory performance while the latter reporting discouraging results for the same model.

In addition, VaR models' comparison studies often analyze different methods using small hypothetical portfolios and not 'actual' or representative portfolios of banks or financial institutions. An example is the study by Engelbrecht (2003) who uses a small hypothetical sample portfolio consisting of interest rate swaps (IRS) and forward rate agreements (FRAs) on the South African market. The study finds the VaR Historical method lacking in accuracy despite all major banks in South Africa using this method.

To this end, it is imperative not only to consistently interrogate the various VaR methodologies but to also validate their accuracy on estimating risk on South African market 'actual' portfolios. Furthermore, as a regulatory capital holding measure, a systematic underestimation of VaR will lead to inadequate capital being held by financial institutions to cover adverse market conditions. On the other hand, an over estimation results in a conservative VaR measure and thus working capital being tied up as regulatory capital, and this may not be prudent for a firm's profitability.

Although there are some studies comparing the accuracy of various VaR models on the South African financial market, for example Engelbrecht (2003), it has been on small hypothetical sample portfolios. Moreover, we are convinced that to date, there is little research on backtesting VaR in the South African financial market, let alone on 'actual' financial institution portfolios.

[2] A J.P Morgan VaR management system unveiled in 1994 which provides freely online a variance-covariance matrix for a large number of risk factors.

1.2 Purpose of Study

The main purpose of this study is two fold. Firstly, it provides a theoretical comparison of the different methods of estimating VaR. The main aim of the theoretical analysis of VaR approaches is to ensure that the reader builds fundamental knowledge of the most common VaR methodologies. Although VaR is conceptually simple, its implementation in practice is not as straight forward. This is because there are myriad different alternatives, each with their own pros and cons, derived from the underlying hypothesis of such methodologies. More precisely, the study considers three methods, namely:

(i) *Variance-Covariance* (standard RiskMetrics model),

(ii) *Historical simulation*,

(iii) *Monte Carlo Simulation*.

The theoretical comparison of these approaches places emphasis on their shortcomings, in cognizance that their potential flaws provide motivation for the backtesting of VaR. Furthermore, the assessment of these methods will largely be in the context of regulatory supervision of financial institutions, again milieu of VaR backtesting. This has largely been a drive from the regulators rather than financial institutions themselves. Thus, the depth of the discussion of these methods is restricted to material that is germain to the empirical study of this work: the performance of different VaR backtesting approaches. In addition, a theoretical comparison of some of the fairly traditional backtesting methods is presented. Four backtesting techniques will be presented and discussed, namely; *Kupiec's proportion of failures test*, *Basel Committee's 'traffic light'* approach, *independence test* and the *joint-test*.

The second part of this work is an empirical study to provide a review of the theory and practice of VaR backtesting as a *reality check* for VaR estimates. This will be conducted on an 'actual' portfolio of a financial institution in an emerging market, in particular, South Africa. The aim is not only to ensure accurate estimation of VaR but also to ensure the prudent use of capital by financial institutions. As alluded to in the preceding section, overestimation of VaR leads to higher capital requirements from the regulator, while underestimation leaves financial institutions vulnerable in adverse market conditions. We will largely follow some of the backtesting techniques of Nieppola (2009), however, applied on an 'actual' South African market portfolio.

1.3 Questions of the Study

Financial institutions, and even regulators, face much the same problems when it comes to the management of financial market risk. Some of the key questions they have to ask and which form part of this study are listed below.

(i) How accurate is the VaR *historical simulation* method in estimating VaR of an 'actual' portfolio on the South African market?

(ii) What are the underlying assumptions in some of the VaR models?

(iii) How does one validate the accuracy of the VaR numbers?

(iv) What are some of the backtesting methods for VaR models?

(v) How do these backtesting models work, and what are the key statistical assumptions that underpin them?

(vi) How do some of the VaR backtesting techniques fare in validating the accuracy of VaR estimates of an 'actual' South African market portfolio?

1.4 Significance of the Study

The large upside returns and diversification benefits provided by emerging markets have attracted significant investors' attention which, in turn, have led to significant portfolio inflows into emerging markets (including South Africa), Thupayagale (2010). In addition, there have been significant changes in the African (including South Africa) financial market landscape with considerable growth in both the stock and bond markets (Andriaanaivo & Yartey, 2010). While these developments have motivated the study of various asset return behavior in these markets, a crucial and topical strand of contemporary theoretical and empirical research has been on the computation of VaR in these markets.

VaR models are as useful only if they can provide accurate predictions of future risks. Therefore, in order to validate that portfolio risk numbers from VaR calculations are consistent and reliable, VaR models require backtesting with appropriate statistical methods. Backtesting is a formal statistical framework that consists of validating that realized losses are consistent with projected VaR estimates (Jorion, 2007). Jorion (2007) further refers to such tests as *reality checks* that ensure that VaR forecasts are well calibrated. In the event of inaccurate VaR estimates, the models should be re-examined for either incorrect assumptions, wrong parameters or even inaccurate modeling.

It is therefore imperative to ensure that backtesting becomes an integral component of VaR reporting in contemporary financial risk management. In the absence of robust model validation (backtesting) one can never be certain about the accuracy of risk estimates from VaR models (Nieppola, 2009). The importance of VaR backtesting is especially crucial in the recent highly volatile financial market environment where market price volatility tends to drive a huge interest into portfolio risk figures from investors and regulators due to high loss accumulation. This is a very important and topical strand given that VaR is acknowledged to have stern problems with risk estimation in turbulent markets. VaR measures, by definition, expected loss only under *normal market conditions* (Jorion, 2007). Given that the birth of VaR was a result of abnormal financial disasters, this limitation presents one of the major weaknesses of VaR to date. This then makes VaR backtesting a very interesting and challenging topic. It has spanned lots of research interest by practitioners and academics alike, and this study seeks to add to that pool.

1.5 Background Literature

While the term Value at Risk did not enter the financial risk glossary until the mid 1990s, origins of the measure go further back (Dowd, 1998). The mathematics that underpin VaR measures were influenced by portfolio theory as early as the 1950s. For instance Markowitz (1952) and Roy (1952) independently published VaR measures in support of their work on equity portfolio optimization. However, it would be the crises that beset financial institutions in years to come, and the regulatory responses to these crises that would drive the impetus for the use of VaR measures.

In the period proceeding the great depression and the bank failures of the time, the first regulatory capital requirements for banks and financial institutions (some) were enacted when the US Securities Exchange Act established the Securities Exchange Commission (SEC). These included a system of 'haircuts'[3] applied on the capital of a financial institution as insurance against possible further market losses during liquidation periods (Dale, 1996). In the decades thereafter, banks and financial institutions developed their own risk measures and control systems to ensure that they obliged with the regulatory capital requirements. The increased risk emanating from the advent of the financial derivative markets and floating exchange rates in the early 1970s led to capital requirements being further refined and expanded in the SECs Uniform Net Capital Rule (UNCR) in 1975 (Dale, 1996). The UNCR categorized a firm's financial assets into twelve classes, based upon risk, and required different capital requirements for each. Some of these categories were further broken down into subcategories primarily based on maturity. To reflect hedging effects, netting was allowed for long and short positions within subcategories, but only limited netting was permitted across subcategories (Holton, 2003).

Holton (2003) argues that the first regulatory measures that evoke Value at Risk, though, were initiated in the 1980s. At this time the SEC tied the capital requirements of financial institutions to reflect a 0.95^{th} quantile of the amount of money they could lose over a thirty-day interval. While the measures were called 'haircuts' and not Value or Capital at Risk, it was evident the SEC was requiring financial institutions to embark upon a process of calculating thirty-day VaR on a 95% confidence interval and thus hold enough capital to cover potential losses (Holton, 2003). It would be a decade later that these measures were reviewed. The period between 1990 and 1995 saw numerous disastrous losses associated with financial derivatives and leverage, sparking the resurgence of VaR measures.

[3] The percentage by which the market value of a financial instrument is reduced. The size of the haircut reflects the degree of risk a lender places on the financial instruments.

The mention of the name *Nick Leeson* may draw unfavorable memories to many who work in investment banking. This may be more so for those who were directly or indirectly affected by the demise of arguably one of the largest and most conservative British banks in the early 1990s, Barings bank. Leeson, a trader working out of the Barings' Singapore subsidiary lost USD1.3 billion on derivative exposures. The huge loss wiped out the bank's entire equity capital, forcing the bank into filing for bankruptcy in February of 1995. Financial disastrous losses associated with derivatives, between 1993 to 1995, were not limited to Barings bank alone. Other noteworthy cases include:

(i) *Metallgesellschaft.* German's fourteenth-largest industrial group in 1993, which nearly went bankrupt as a result of a USD1.3 million loss from its US subsidiary's derivative positions.

(ii) *Orange County.* A US local government fund that went bankrupt in 1994 due to a USD1.8 billion loss on interest rate derivative exposure.

(iii) *Daiwa Bank.* Where a single trader managed to conceal USD1.1 billion treasury bond trades for over 11 years, only to confess in 1995.

A common element in all these financial disasters was a deficiency in systems of measurement, management and control of risk. In subsequent studies that utilize VaR, it has been estimated that the daily potential loss of Leeson's positions would exceed USD835 million five percent of the time, under normal market conditions (Jorion, 2007). This indicates that if such calculations had been in place, then probably such financial disasters might have been averted. Since Barings was viewed as a conservative bank, its bankruptcy served as a wake-up call for banks and financial institutions across the globe. Finally, financial institutions realized the need for more comprehensive risk measures.

Subsequently, in October 1994, J.P. Morgan unveiled its RiskMetrics system, which it availed for free on the internet, providing data feed for computing market risk. RiskMetrics then coined the term *Value at Risk* to describe the risk measure that emerged from their data. This marked the beginning of public access to data on the variance-covariance matrices across various securities and asset classes, information that was once proprietary. This sparked global interest in the 'new' measure, *VaR* (Jorion, 2007).

In the last decade, VaR has become the 'standard benchmark' for measuring financial risk in financial institutions and has even begun to find acceptance amongst non-financial institutions as well. Initially confined to measuring market risk, VaR is now being used to manage and control risk actively, well beyond trading portfolios. The VaR methodology is now being utilized to quantify credit risk, operational risk and liquidity risk, leading to the 'Holy Grail' of firmwide risk management (Jorion, 2007). With such widespread use, came the need to critically analyze and validate the accuracy of the numbers from this measure, especially in emerging (in particular, South African) market portfolios . And this shall be the focus of this study.

1.6 Outline of the Study

The paper consists of five chapters, with the introduction being the first one. The second chapter delves into a review of relevant literature. It provides a description of the basic idea behind VaR and gives some historical overview and background on the subject. It provides a fundamental description of the main VaR methodologies, with a key emphasis on the assumptions and weaknesses of each approach. The comparison of the VaR methods and discussion on their shortcomings leads to a discussion on some of the main VaR backtesting procedures. Four major backtest approaches are presented in some detail, however the symposium is by no means exhaustive. A discussion of all the numerous backtesting techniques and their applications is impossible in this context. The major aim is to focus on the main and widely used backtesting techniques, and rather limit the discussion to those that will be applied in practice in this work.

The third chapter forms the empirical part of this work. As such, it can be considered to be the main component of this study. It outlines details of the data, portfolio composition and the methodology of VaR calculation. The backtesting techniques presented in chapter two are then applied to data of an 'actual' South African market portfolio (VaR estimates & realized profit and loss) in a bid to validate the accuracy of the VaR model. The chapter then ends with a presentation on the actual process of backtesting the VaR estimates, which is the main focus of this work.

In chapter four, empirical results and findings of the study are presented, discussed and analyzed. In addition, a summary of the key findings and factors affecting the results are presented. The final chapter, chapter five, concludes and provides a review of the key takeouts of both the theoretical and empirical components. It also provides suggestions on some ideas and proposals for future research.

Chapter 2

Literature Review

"*Those who cannot remember the past are condemned to repeat it.*" (Santayana, 1905)

Literature on VaR model backtesting is fairly recent and relatively limited. In this section we begin by presenting a brief synopsis on the major drivers of VaR and its definition. We then present the calculation of portfolio returns as a key measure of portfolio risk before discussing the choice of parameters that go into any VaR model. Various VaR methods and their shortcomings are then reviewed. This then leads to a presentation on some of the backtesting techniques for validating the accuracy of VaR models.

2.1 Why VaR

"*Whether you love derivatives or hate them, you cannot ignore them.*" (Hull J, 2012)

The proliferation of financial derivatives has been accompanied by increased trading of cash instruments and securities as well as the proliferation of different financing opportunities. In addition, it has coincided with the growth in foreign trade and the rise in international financial linkages among companies (Linsmeier & Pearson, 1996). As a result of the sheer numbers and complexity of some these instruments, the magnitude of risks in companies' portfolios are often not obvious. As a consequence, there has been demand for portfolio level quantitative measures of market risk, and VaR has become a significant component of such risk measures (Jorion, 2007).

Over the past three decades, derivatives have become increasingly important in finance. Their use spans almost all key areas in the field. They are added to bond issues; added to capital investment opportunities; used to transfer risk in loans or mortgages from original lenders to investors; used in executive or employee compensation plans; and the list goes on. While some economic agents use them for speculation, they play a key role in hedging financial risk. This they do by transferring a wide range of risks in the economy from one entity to another. It goes without saying that the financial industry has reached a stage where it is imperative for those working in the industry, and indeed those outside of finance, to understand how derivatives work, the manner in which they are used, and more importantly, how to *quantify their risk* with such models as VaR.

2.2 Definition of VaR

VaR is a single, summary, statistical measure of possible portfolio losses due to 'normal' market movements (Campbell, 2005). According to Dowd (1996), downside risk greater than the VaR quantum can only be experienced with a specified small probability. Given the simplifying assumptions utilized in its computation, VaR aggregates all portfolio risks into a single number suitable for use in the boardroom, reporting to regulators, or disclosure in an annual report. Once one crosses the hurdle of using a statistical measure, the concept of VaR is straightforward to understand. It is simply a way to describe the magnitude of the likely losses on the portfolio (Linsmeier and Pearson, 1996). According to Jorion (2007), VaR measures the worst expected loss over a given confidence level and target horizon, *under normal market conditions*.

More mathematically, VaR describes the quantile of the projected distribution of gains and losses over a given period with a certain probability. If c is the selected confidence level, VaR corresponds to the $(1 - c)$ lower-tail level of the portfolio's return distribution (Jorion, 2007). VaR thus consists of two quantitative parameters: a *holding period* and *confidence level*. The holding period indicates how far into the future we are looking; the longer the holding period the larger the potential losses. The confidence level on the other hand determines with how much certainty the measurement is made. Higher confidence levels means higher potential losses.

For example, a R100 000, one day, 99% confidence level VaR for a portfolio means that we can be

99% certain that the daily loss of the portfolio will not exceed R100 000, under normal conditions.

An event where the return of a portfolio exceeds the estimated VaR measure is called a VaR breach. Therefore an accurate VaR measure produces a number of VaR breaches as close as possible to the number of breaches specified in the confidence level. One way of testing the accuracy of the VaR measure is to conduct backtesting, which will be the empirical focus of this study.

2.3 Measuring Returns

From the definition VaR, we can immediately establish that the VaR estimate is the downside of the portfolio return distribution[1]. Therefore we begin by defining portfolio return. We define the return on a portfolio δP as the difference between the value of the portfolio at subsequent time intervals, i.e. $\delta P = P_t - P_{t-1}$, where P_t and P_{t-1} are the portfolio values at time t and $t-1$ respectively. However, the rate of return can be expressed in two ways, namely, arithmetic and geometric. The arithmetic rate of return, R_a, is given by the difference between the current portfolio price and the previous period's price, all divided by the previous period's portfolio price, i.e.

$$R_a = \frac{P_t - P_{t-1}}{P_{t-1}}. \tag{2.3.1}$$

On the other hand, the geometric rate of return, R_g, is the logarithm of the periodic price ratio, i.e.

$$R_g = \ln\left(\frac{P_t}{P_{t-1}}\right). \tag{2.3.2}$$

Obviously here we assume all income payments such as dividends are zero or reinvested in the portfolio such that they are reflected in the portfolio price.

Given the two methods of computing portfolio rate of return, in most cases it makes economic sense to use geometric returns. If geometric returns are normally distributed, then the distribution can never lead to a price that is negative.[2] In contrast, normally distributed arithmetic

[1] This is the left tail of the return distribution.

[2] That is because the left tail of the distribution of $\ln(P_t/P_{t-1}) \to -\infty$ is achieved as $P_t/P_{t-1} \to 0$ or $P_t \to 0$.

returns could generate negative asset prices, which is meaningless since stocks or portfolios have limited downside risk.[3]

Note that $R_g = \ln(P_t/P_{t-1}) = \ln(1 + R_a)$. If R_a is small, then by Taylor expansion, $R_g = R_a - \frac{R_a^2}{2} + \frac{R_a^3}{3} - \cdots$, which implies $R_g \approx R_a$, if R_a is small. Therefore, in practice, as long as returns are small, arithmetic returns and geometric returns converge. However, in times of high volatility or long holding periods, great emphasis needs to be placed on the type of returns used as this may impact the VaR estimates (Jorion, 2007).

2.4 Choice of VaR Parameters

VaR can be applied to many different portfolios and context. In each diverse situation, different criteria apply in the selection of the holding period and confidence level. Without these two essential quantitative parameters VaR numbers are meaningless. If VaR is used to measure potential loss, parameters should be determined by the nature of the portfolio. On the other hand, if the VaR number is used to compare risks among different markets, the choice of parameters is arbitrary as long as consistency is maintained.

Dowd (1998) submits that liquidity should be used as one of the criteria for the choice of the holding period. That is, financial institutions should choose a holding period based on the amount of time required to liquidate the portfolio. For example, banks and financial institutions with actively traded portfolios, typically use one day VaR, while non-financial firms and other investors tend to choose longer horizons. However, properties of the VaR method used should also be considered. For instance, methods that assume portfolio returns are normally distributed are only accurate if short time horizons are selected. In addition, a short holding period is required for validation purposes, therefore backtesting is another consideration. For example, the Basel Committee on Banking Supervision prescribe that banks and financial institutions perform backtesting over a one day period, despite the 10 day horizon for regulatory capital (Jorion, 2007).

[3] That is, normal discrete returns $R_a = (P_t - P_{t-1})/P_{t-1} \to -\infty$ is achieved as $(P_t/P_{t-1}) - 1 < -1$ or $P_t < 0$, although it seldom occurs in most data.

When it comes to the selection of the confidence level, purpose at hand becomes crucial. That is, whether the VaR number is for capital requirements, or to provide input for internal risk management, or is for making comparisons among different portfolios/institutions (Dowd, 1998). In the assessment of capital holdings, the confidence level depends on internal risk aversion if it is for internal capital, or the regulatory prescribed levels if its for regulatory capital. On the other hand, if the purpose is for backtesting a VaR model (the focus of the empirical study of this work), then its recommended to avoid high confidence levels. These tend to reduce the number of observations in the tail of the return distribution, and thus the power of the test (Jorion, 2007). An additional aspect is to consider confidence levels for accounting and comparison purposes. Different institutions report their VaR estimates using varying confidence levels (Dowd, 1998).

2.5 VaR Methods

> "*One of the most difficult aspects of calculating VaR is selecting among the many types of VaR methodologies and their associated assumptions.*" (Minnich, 1998)

In practice, the main objective of calculating VaR is to provide a reasonably accurate estimate of downside risk at a reasonable cost. This involves selecting the most appropriate VaR method for the selected portfolio from the myriad of industry standard methods. VaR calculation approaches are normally divided into two groups, non-parametric and parametric models. Non-parametric models do not make any assumptions regarding the distributional shape of portfolio returns and therefore tend to require simulation. On the other hand, parametric models assume the return distribution belongs to a parametric family, such as the normal distribution. Thus VaR can be computed directly from the standard deviation of the portfolio's return distribution together with a multiplicative factor that is based on the VaR confidence level (Jorion, 2007).

VaR models can further be divided into two classes depending on the valuation approach. The first is *local-valuation class*. These quantify risk by valuing the portfolio once, at a point in time, and using sensitivities (or local derivatives) to infer possible changes. The second class falls under the *full-valuation class*, which measure risk by fully repricing the portfolio over a range of scenarios.

This section presents the fundamentals of the three most common VaR computation methods: *Variance-Covariance*, *Historical simulation* and *Monte Carlo simulation*. The major aim of the discussion is not to provide a detailed description of the methods, but rather, to emphasize on the strengths and weaknesses of each of the approaches. For a detailed and more comprehensive discussion on the various VaR methodologies we refer the reader to, for instance; Linsmeier & Pearson (1996), Dowd (1998) or Jorion (2007).

2.5.1 Variance-Covariance Method

The variance-covariance method is a parametric approach which departs from the arch-typical assumption that all risks of the portfolio are normally distributed, and the portfolio is a linear function of these normal risks. The normality assumption is the simplest and straightforward approach and is thus suitable for portfolios largely consisting of linear instruments[4] (Dowd, 1998). Estimating VaR is therefore attained by simply multiplying the current portfolio price/value (P_0) by the portfolio standard deviation (σ) and a multiplicative factor (α) from the normal distribution for the chosen confidence level, i.e.

$$VaR = \alpha\sigma P_0. \tag{2.5.1}$$

Using this approach, the portfolio standard deviation is assumed to be a linear combination of the volatilities and covariances of portfolio elements, and thus determined using the variance-covariance matrix, hence the name variance-covariance approach. We mention here that the famous RiskMetrics model of J. P. Morgan is an analytic variance-covariance method.

In implementing the variance-covariance method, the process involves, firstly, the 'mapping' of individual portfolio transactions into a set of standardized[5] market instruments. Secondly,

[4]Linearity means returns of the portfolio are linear functions of risk variables, thus their return distribution is assumed to be normal. On the other hand, non-linear instruments, e.g. derivative options, do not exhibit the normality property (Dowd, 1998).

[5]The importance for standardized mappings is to contain the size of the variance-covariance matrix so that it does not become unnecessarily too big to handle (Damodaran, 2007).

each portfolio transaction is then presented as a set of positions in the standardized market instruments. The next step after identifying the standardized instruments, is to estimate their variances and covariances. Usually historical data forms the basis of obtaining these statistical estimates. The last step is the estimation of the portfolio VaR using the covariance matrix (i.e. the estimated variances and covariances) and the weights on the standardized positions (Jorion, 2007). Both pros and cons of this method (variance-covariance) are consequences of the two main underlying assumptions on which it is based. Basically, the assumption about the linear relationship among market risk factors and the assumption that portfolio returns are joint normally distributed.

The main advantage of this approach is its simplicity. The normality assumption renders the calculation of VaR to be a relatively simple and easy exercise, as standard mathematical properties of the normal distribution can then be used in the estimation of VaR numbers. Furthermore, the normality assumption permits easy *translation* between different VaR confidence levels and holding periods[6] (Dowd, 1998).

The major drawback of this method emanates from the normality assumption, as well. The huge body of contemporary empirical evidence has shown that most financial assets tend to exhibit 'fat tailed' return distributions, that is, in practice extreme outcomes are highly probable than what the normal distribution assumes. As a consequence, the variance-covariance method tends to overestimate risk for small confidence levels and underestimates risk for higher confidence levels (Jorion, 2007).

Another disadvantage is the linearity assumption. This factor implies that this approach, theoretically, applies only to portfolios with linear instruments. This is a major weakness when working with portfolios that include non-linear instruments, for example derivative options, as their returns are non-linear functions of risk variables. This is a problematic trait given the increasing use of non-linear assets (especially options) in financial market portfolios.

[6]Translation between confidence levels is trivial, for instance from 95% to 99%: $\mathrm{VaR}_{0.99} = (\alpha_{0.99}/\alpha_{0.95}) * \mathrm{VaR}_{0.95} = (2.33/1.65) * \mathrm{VaR}_{0.95}$. VaR for different holding periods can simply be attained by rescaling by the ratio of the square root of time, i.e. $\mathrm{VaR}_{t_2} = \sqrt{t_2/t_1} * \mathrm{VaR}_{t_1}$, where $t_2 > t_1$.

One solution to the linearity assumption is to take first order Taylor approximations (delta-normal), and in some instances second order Taylor approximations (delta-gamma) to the returns of non-linear instruments and then utilize their linear approximation to calculate VaR (Dowd, 1998). The improvement by using first and second order approximation is obvious. However, this tends to introduce more complexity, thus loosing some of the basic simplicity of the variance-covariance method due to the additional assumptions required in light of the normality loss (Damodaran, 2007).

2.5.2 Historical Simulation

The historical simulation method is a non-parametric approach that is both easy to understand and implement. This method uses the historical distribution of asset returns in the portfolio to simulate the portfolio's VaR. This method is based on the hypothetical assumption that the portfolio is held constant over the observation holding period. The estimation of VaR can then be attained by reading the desired quantile from the distribution of the portfolio returns. The main assumption of the historical simulation method is that the historical return distribution acts as a reliable proxy for projecting returns for subsequent holding periods (Dowd, 1998).

Due to its simplicity, the historical simulation method has some undeniable advantages. It does not rely nor does it make any assumptions regarding the distribution of portfolio returns. Therefore, it allows for the capturing of 'fat tails' (as well as other non-normal properties), at the same time also obviating the need for computing and working with correlations and volatilities (Dowd, 1998). Furthermore, it is the most widely used valuation method by banks and financial institutions worldwide, and in South Africa. Consequently, it excludes the need for any linear approximations (for instance first and/or second order Taylor approximations), which tend to lead to inaccurate VaR calculations. Thus, it can be utilized on any portfolio with all kinds of instruments, both linear and non-linear (Jorion, 2007). These advantages propel it, theoretical, as superior over the variance-covariance approach, especially when dealing with non-linear portfolios, which have become a major feature of most financial market portfolios nowadays (Jorion, 2007).

However, the historical simulation approach has its own disadvantages as well, related specifically with the characteristics of using historical data. The first and major complication emanates from the inclusion into the portfolio of 'new' market instruments that don't have a robust amount of historical market data. While this could be a critique of any of the the VaR methods in this work, it is most prominent with the historical simulation method since its calculation of VaR is largely based on historical data (Damodaran, 2007). Another, and again serious shortfall of this method is the assumption that history will repeat itself. While often a reasonable assumption, it may lead to highly distorted estimates of VaR numbers in some cases (Dowd, 1998). This major disadvantage, as will be shown in the following section, is not a concern in the Monte Carlo simulation method, notwithstanding that Monte Carlo also has its own drawbacks.

Finally, there is also the challenge of selecting how far back the historical data should go. For instance, the use of very long estimation periods may lead to old market data being significantly emphasized in comparison to new market information. As a consequence, VaR estimates may have a delayed reaction to most recent changes in market data, and thus leading to inaccurate VaR estimates. Another related problem is the equal weighting of historical observations. This again leads to distorted VaR estimates when the historical data set exhibits some huge market jumps (Dowd, 1998). Dowd (1998) presents a convenient solution to the problems above by suggesting the use of a weighted historical simulation approach, which either assigns lower weightings on much older observations or raising the weights for observed stressed periods so as to accurately recognize the market jumps in past periods.

2.5.3 Monte Carlo Simulation

Monte Carlo simulation is both a parametric and non-parametric method, and is the most challenging both in comprehension and implementation. It is recommended for use when other simpler approaches are inappropriate, or when there is a need for a more sophisticated and highly accurate VaR estimates (Dowd, 1998). This VaR method can be described in two main steps. Firstly, stochastic processes for financial variables are defined, then historical or market data is then used in estimating volatilities and correlations. The second step is to simulate (thousands of times), the price paths for all the defined financial variables using random number generators. The price realizations from the simulation are then compiled to a joint distribution of portfolio returns wherein VaR estimates are then computed (Dowd, 1998).

Similar to the historical simulation, the attraction of the *non-parametric* Monte Carlo simulation method is that it does not rely on any assumption about the probability distribution of the portfolio returns. Again, this allows for the capturing of 'fat tails' (as well as other non-normal properties), and the ability to cater for portfolios with both linear and non-linear (such as options) instruments. It therefore obviates the need of computing and working with correlations and volatilities, thus negating the shortcomings of historical simulation approach as opposed to the variance-covariance matrix one (Dowd, 1998).

On the other hand, the *parametric* Monte Carlo simulation method presents theoretical superiority over the variance-covariance approach. While the former does call for the specification of parameters (in particular stochastic processes) for risk factors (i.e. introducing model risk[7]), it does not require the normality assumption and its application spans across both linear and non-linear portfolios unlike the latter approach (Jorion, 2007). In addition, the Monte Carlo method generates its own entire portfolio return distribution and therefore can also be utilized, for example, to compute expected tail loss (i.e. losses in excess of VaR).

The major drawback of the Monte Carlo method is the computational time. It is highly resource intensive with slow computational speed, especially with large portfolios[8]. Nonetheless, this shortcoming maybe of less importance in the future due to continuous computer science development. Another potential weakness is model risk, that is, if model parameter assumptions are incorrectly specified, then VaR estimates will be misleading or inaccurate (Jorion, 2007). Furthermore, the complicated procedures associated with Monte Carlo require special expertise, which might not be readily common with users of VaR numbers. Hence they may find it difficult

[7] Model risk is a result of applying wrong assumptions in the model (Dowd, 1996).

[8] For instance, increasing the accuracy of the model by 100 times requires running a 1000 times more simulations (Wiener, 1999). Therefore, Monte Carlo is highly susceptible to sampling variation, a consequence of limited simulation rounds.

to comprehend and accurately interpret the VaR estimates computed via Monte Carlo (Dowd, 1998).

2.5.4 Theoretical Comparison of Methods

From a theoretically comparison of the three VaR methods presented above, the analytic method (variance-covariance) is both faster and easier to implement. On the other hand, the simulation methods (historical and Monte Carlo) are relatively more complex and require huge computational power, albeit more precise, especially when dealing with more complex portfolios. The most obvious conclusion one can glean from a comparison of the three methods presented is that a best VaR estimation method does not exist, generally nor in absolute terms. Linsmeier and Pearson (1996) argue that the three approaches differ mainly in four broad categories:

(i) The ability of the model to capture risk of linear and non-linear (options) instruments;

(ii) The relative ease of implementation and that of interpretation by users;

(iii) Flexibility of the model to incorporate alternative assumptions;

(iv) Reliability of the results.

The last aspect is a key focus of this work as it recognizes the importance of validation (backtesting) of VaR numbers.

Ultimately, the choice of one or the other method depends largely on the type and composition of the portfolio for which we wish to quantify its risk. For portfolios dominated by linear instruments, the accuracy of VaR estimates based on the analytic approach (variance-covariance) is almost indifferent to those based on simulation methods. Moreover, the computation is faster and easier utilizing the analytic method. On the other hand, for non-linear portfolios, the analytic approach provides less accurate VaR numbers. Simulation methods (historical or Monte Carlo) would provide much more plausible VaR estimates despite being more more complex and theoretically slower (Jorion, 2007).

The variance-covariance and historical simulation approaches are best known to be the easiest to comprehend and to implement. Historical simulation while intuitively easy to understand, does however come with with slight implementation rigor. By far the most complicated implementation approach is usually the Monte Carlo method, an issue that also translates to a problem when it comes to interpretation of VaR numbers from this method by end users (Linsmeier and Pearson, 1996).

Top among the main drawbacks of historical simulation is the lack of flexibility, especially when historical standard deviation and correlation estimates do not provide reliable future projections of market conditions. On the other hand, Monte Carlo simulation and variance-covariance approaches allow the introduction of subjective views in the calculation VaR numbers. Historical simulation exhibits poor performance in this regard as its future risk estimates are a major function of historical returns data (Linsmeier and Pearson, 1996).

The most crucial issue in determining which method to use is arguably the *reliability of the results*. This is an interesting and important dimension, especially in the context of this study as the key focus will be on validation (backtesting) of VaR numbers. The results of Hendricks (1996) show that historical simulation provides exceptional performance compared to the variance-covariance approach, especially when dealing with high confidence levels. This is an observation that is to be expected given that the variance-covariance method assumes normality, and research has shown that most securities have 'fat tailed' return distributions. Other empirical studies that compare the three approaches have shown that the analytic method (variance-covariance) with linear approximations provide relatively accurate VaR numbers, however only when considering portfolios with a limited number of non-linear instruments (Campbell, 2005). The same studies allude to the superior performance of Monte Carlo over analytic models at high confidence levels as well as at longer holding periods (see e.g. Coronado (2000); Reich (2001); Gnamassou (2010).

We conclude this section by once again reiterating that, theoretical, the selection of the most appropriate and convenient VaR method depends largely on the type of portfolio as well as the desired use for the VaR estimate (Coronado, 2000).

2.6 Backtesting

> "*Disclosure of quantitative measures of market risk, such as VaR, is enlightening only when accompanied by a thorough discussion of how the risk measures were calculated and how they related to actual performance.*" (Greenspan, 1996)

VaR models are only as useful in so far as they predict risk accurately (Jorion, 2007). The preceding section presented a theoretical comparison of some of the basic VaR methods. The numerous weaknesses of the various VaR approaches were highlighted and thus provide a compelling reason as to why the accuracy of the risk estimates need to be accompanied by validation. Since the late 1990's a variety of tests have been, and continue to be, proposed that can be used to validate the accuracy of a putative VaR model.

Model validation is the process of assessing whether a model is adequate. One of the tools of conducting validation is by *backtesting*. Other forms include stress testing, independent review, and oversight, amongst others. Appropriate backtesting techniques should therefore be used in order to verify the accuracy of VaR models (Jorion, 2007). Furthermore, backtesting is a key part of the internal models approach to market risk management as laid out by the Basel Committee on Banking Supervision (1996).

Backtesting is a statistical framework that consists of verifying that realized losses correspond with projected VaR estimates. For instance, if the confidence level used for calculating one day VaR is 99%, then we expect an exception to occur once in every 100 days on average, under normal conditions. From hereon we will refer to an event where the ex-post portfolio loss exceeds the ex-ante VaR measure as an exception (or VaR violation) , following Dowd (1998). Therefore in the backtesting process, the essence is to statistically examine whether the frequency of exceptions over a given holding period correspond with the defined confidence level. Such tests are commonly known as tests of *unconditional coverage*, since they do not take into account when the exceptions occur (Jorion, 2007). When conducting unconditional coverage tests, higher than expected number of exceptions for a given confidence level and sample size would indicate that the VaR measure systematically understates the portfolio's actual risk level. On the other hand, fewer than expected VaR violations would alternatively signal an overly conservative VaR measure that overstates risk. We mention here that neither of the two extremes is desirable as they have capital implications.

Over and above measuring the number of exceptions, an ideal VaR model, theoretically, should be able to statistically measure the dispersion of the exceptions. Exceptions should be spread over time, that is, they should be independent of each other. This is often called the *independence property*. Clustering of exceptions renders a VaR model invalid as it does not accurately capture market volatility and correlations. Large losses which occur in rapid succession are more likely to lead to disastrous events such as bankruptcy (Christoffersen, 1998). VaR backtesting techniques which test for the independence property are classified as tests of *conditional coverage*. Therefore tests of conditional coverage thus also test for conditioning, alternatively time variation, in the data (Jorion, 2007). Another form of conditional coverage tests are the so called *joint tests*, which combine conditional coverage tests such as the independence property test with the unconditional tests. Verification systems should be able to satisfy both the unconditional coverage and independence properties in backtesting, Christoffersen (1998).

While a number of different testing methods have been and continue to be proposed for the purposes of backtesting VaR, this section presents four main backtesting techniques. In particular, we look at:

(i) *Proportion of Failures-Test* (POF-Test), also known as the test of Kupiec (1995),

(ii) *Basel Committees (1996) Traffic Light* approach,

(iii) *Independence Test, Christoffersen (1998)*,

(iv) *Joint-Test - Christoffersen's Interval Forecast Test (1998).*

The main aim is to present the fundamental properties of these four, fairly traditional, VaR backtesting techniques as they will be applied to validate VaR estimates in the empirical study of this work.

2.6.1 Kupiec Test or POF-Test

The earliest proposed and most widely used VaR backtests focus on the frequency of VaR limit violations. The test proposed by Kupiec (1995), also known as the proportion of failure (POF), is one of the most popular tests. The POF approach tests the *unconditional coverage* property. Using this test we validate (backtest) the accuracy of the VaR model by recording the failure rate. That is, the proportion of times VaR is exceeded in a given sample. If we denote the number of exceptions by x and the total number of observations by T, we then define the rate of failure as x/T. Suppose a VaR number is reported at the confidence interval c, then an exception occurs if realized loss exceeds the VaR number. Therefore, the expected number of exceptions x in a total of T observations is $(1-c)T$. Certainly, the number of exceptions will not be exactly $(1-c)T$. Instead, it could swing within an acceptable range. In the backtesting method, the range for x will be calculated and thus the VaR model can be accepted or rejected (Campbell, 2005).

The testing framework for realized losses that exceed VaR, x, is therefore a sequence of success or failure with probability $p = (1-c)$. Assuming all the observations are independent, it follows that this is a classic *Bernoulli trial*[9] (Bernoulli process), and follows a Binomial distribution, which is given by:

$$f(x) = \binom{T}{x}(1-c)^x c^{(T-x)}, \qquad \text{for all } x = 0, 1, 2, \cdots . \tag{2.6.1}$$

Note that for a Binomial distribution x has expected value $\mathbb{E}[x] = (1-c)T$ and variance $\mathbb{V}[x] = (1-c)Tc$. Given a large enough sample size, T, and applying the central limit theorem, we can then approximate the Binomial distribution by a normal distribution;

$$z = \frac{x-\mu}{\sigma} = \frac{x-(1-c)T}{\sqrt{(1-c)Tc}}, \tag{2.6.2}$$

where z follows the standard normal distribution $\mathbf{N}(0,1)$. Therefore, given a confidence level c, there is a range for z, say $|\ z\ | \leq \alpha$, where α is the number in the standard normal tables

[9]Bernoulli trial is a mathematical experiment where a process is repeated many times, with only two possible outcomes, success or failure.

corresponding to $(1-c)$. Hence the range for x can be calculated as;

$$(1-c)T - \alpha\sqrt{(1-c)Tc} < x < (1-c)T + \alpha\sqrt{(1-c)Tc}. \tag{2.6.3}$$

If the number of exceptions x is within the range, the model is accepted, and rejected otherwise (Dowd, 2006).

Therefore, the only parameters required for VaR model validation using the POF-test is number of exceptions (x), the total number of observations (T) and the confidence level (c). The failure rate should be equal to the expected one given the selected confidence level. Thus the null hypothesis can be defined as follows:

$$H_0 : \frac{x}{T} = \frac{x^*}{T}, \tag{2.6.4}$$

where x/T is the expected failure rate given the confidence level and x^*/T the observed failure rate. The POF-test is then conducted as a Likelihood-Ratio[10] (LR) test, which is of the form:

$$LR_{POF} = -2\ln\left[(1-p)^{(T-x)}p^x\right] + 2\ln\left[(1-x/T)^{(T-x)}(x/T)^x\right], \tag{2.6.5}$$

where $p = (1-c)$. Under the null hypothesis, the POF-test statistic given by equation (2.6.5) follows a χ^2 (Chi-squared) distribution with one degree of freedom. If the value of the LR_{POF} -statistic falls below the critical value of χ^2 (Chi-squared) distribution, p with 1-degree of freedom, the model passes the backtest. Higher values above the critical region signal an inaccurate model and should lead to a rejection of the model.

An immediate observation of this method, equation (2.6.3), is that the interval for exceptions is dependent on the test confidence level $p = (1-c)$. A larger p leads to a smaller value of α and thus a smaller interval for x. This makes it easier to reject the VaR model. To the contrary, a smaller p leads to a larger interval for x, and makes it easier to accept the current VaR model (Jorion, 2007).

[10]Likelihood-Ratio test is a test statistic which calculates the ratio of maximal probabilities of an outcome under two alternative hypotheses. The null hypothesis is rejected should the result of the test statistic be larger than the critical value of the χ^2 (Chi-squared) distribution (Jorion, 2007).

Another observation is that the interval for x is usually large. For example, for $c = 99\%$ and $T = 251$ trading days, we accept the model as long as $x < 7$. However, there is a high probability that the number of exceptions x for the 99% confidence level on a 251 day time horizon is less than seven but the model is incorrect. To measure the decision error, classically Type I and Type II errors are involved. Type I error is associated with the probability of rejecting a correct model while Type II error refers to the probability of not rejecting an incorrect model. A powerful test statistic should thus minimize the possibility of either or both errors occurring, Jorion (2007).

Below (Table 2.1) is a summary table of the non-rejection regions for the Kupiec POF test statistic for various observation periods and confidence levels (Kupiec, 1995).

Table 2.1: **Non-rejection Test Confidence Regions POF-Test**

VaR Confidence Level c	**Probability Level p=(1-c)**	**Non-rejection Region for Number of Failures x**		
		T = 251 days	**T = 510 days**	**T = 1000 days**
99.0%	0.01	$x < 7$	$1 < x < 11$	$4 < x < 17$
97.5%	0.025	$2 < x < 12$	$6 < x < 21$	$15 < x < 36$
95.0%	0.05	$6 < x < 20$	$16 < x < 36$	$37 < x < 65$
92.5%	0.075	$11 < x < 28$	$27 < x < 51$	$59 < x < 92$
90.0%	0.1	$16 < x < 36$	$38 < x < 65$	$81 < x < 120$

Source: *Adapted from Kupiec (1995):- x is number of failures and sample size is T.*

It is evident from the figures in Table 2.1 that the power of the test increases with an increase in sample size. For example, at 90% confidence level, the interval x/T for accepting the model with 251 observations is in the range $\left[16/251\right] = 0.06;\ 36/251 = 0.14\right]$ compared to $\left[81/1000 = 0.08;\ 120/1000 = 0.12\right]$ for 1000 observations, which is much tighter.

The POF-test has two main drawbacks. Firstly, the test is statistically weak with small sample sizes, for instance the Basel Committee's regulatory framework sample (250 trading days). This is a weakness that has already been acknowledged by Kupiec himself (Jorion, 2007). Secondly, the POF-test examines only the failure rate (frequency of exceptions) and not the the succession of occurrence. Therefore, it may fail to reject a model that produces serially dependent violations (i.e clustered exceptions), a common weakness of unconditional coverage models (Campbell, 2005).

2.6.2 Basel Committees 'Traffic Light' Approach

Since 1998, regulatory guidelines require banks with huge trading activities to hold capital as insurance against extreme portfolio losses. The quantum of the market risk capital requirement (or set-aside), is defined by the bank's VaR estimates. Specifically, the regulatory risk based capital requirements are a function of the larger of either the bank's current assessment of the 99% confidence level VaR over a 10 day holding period or a multiple of the bank's average reported 99% confidence level VaR over the preceding 60 day holding period plus an additional amount that reflects the underlying credit risk (c) of the bank's portfolio (Basel Committee, 1996). Mathematically the market risk capital (MRC) is defined as,

$$\mathrm{MRC}_t = \max\left[\mathrm{VaR}_t(0.99), k\frac{1}{60}\sum_{i=0}^{59}\mathrm{VaR}_{t-i}(0.99)\right] + c \qquad (2.6.6)$$

where k reflects the multiplication factor that is applied to the average of previously reported VaR estimates.

When a VaR model indicates more risk, the risk based capital requirement rises. What may be less intuitive from equation (2.6.6) above is that the risk based capital requirement also depends on the accuracy of the VaR model. Importantly, the multiplication factor, k, varies with backtesting results. The multiplicative factor k is determined by classifying the number of 99% VaR exceptions x in the previous 250 trading days, into three distinct categories as follows,

$$k = \begin{cases} 3.0 & \text{if } \; x \leq 4 \quad \text{green} \\ 3.0 + 0.2(x-4) & \text{if } \; 5 \leq x \leq 9 \quad \text{yellow} \\ 4.0 & \text{if } \; 10 \geq x \quad \text{red} \end{cases} \qquad (2.6.7)$$

Thus a VaR measure with a higher number of exceptions results in a larger multiplication factor, and accordingly a larger risk based capital requirement (Campbell, 2005).

The *green* zone defines an accurate VaR model as the observed number of exceptions fall within the expected range of VaR violations. The *yellow* zone represents a model that maybe accurate or inaccurate. The committee requires further tests to show that the more than expected failure rate in the yellow zone was due to 'bad luck', in other words not due to normal market conditions. The *red* zone, on the other hand, indicates a problematic VaR model.

Given the three categories above and $T = 250$, Table 2.2 below displays the probabilities of obtaining a given number of exceptions for a correct model with 99% coverage.

Table 2.2: **Basel 'Traffic Light' Probabilities of Obtaining Exceptions**

Zone	**Coverage = 99%, T=250**			
	Number of Exceptions (x)	**Scaling Factor (k)**	**Probability** $P(X = x)$	**Cumulative Probability** $P(X < x)$
Green zone	0	0	8.11%	8.11%
	1	0	20.47%	28.58%
	2	0	25.74%	54.32%
	3	0	21.49%	75.81%
	4	0.4	13.41%	89.22%
Yellow zone	5	0.5	6.66%	95.88%
	6	0.65	2.75%	98.63%
	7	0.75	0.97%	99.60%
	8	0.85	0.30%	99.90%
	9	1	0.08%	99.98%
Red zone	10+	1	0.02%	100.00%

Source: *Calculation. Given c=99% (p=1%) and T=250, substituting this in the Binomial distribution, one can determine the probability with which certain number of exceptions can be realized and the corresponding cumulative probability.*

It is evident from the exception table (Table 2.2) that there is a probability of 6.6% that we will have exactly five exceptions in 250 observations, and that there is a probability of 95.88% that the number of exceptions will be less or equal to five. In the event that more than ten exceptions of the 99% VaR are recorded in a 250 day horizon, corresponding to 4% of the sample period, the VaR model is deemed inaccurate and immediate steps are required to improve the underlying risk management system (Basel Committee, 2006). The 'traffic light' approach to backtesting represents the only assessment of VaR accuracy prescribed in the current regulatory framework (Campbell, 2005).

The main drawback of the Basel 'traffic light' approach is that it cannot be used to evaluate the complete accuracy of a VaR model as it does not, for instance, take into account the clustering of exceptions. This shortcoming, however, has already been acknowledged by the Basel Committee (1996) itself (Campbell, 2005). As can be seen the 'traffic light' approach is related to Kupiec's (1995) POF-test, as such suffers the same major POF weakness (unable to detect exception clustering). Due to this weakness, this backtest method is utilized mainly for internal purposes or as an initial test for VaR accuracy.

2.6.3 Independence Test (Christoffersen, 1998)

Given the major short coming of the unconditional coverage of the POF-test to detect clustering of exceptions, a number of tests have been proposed which explicitly examine the independence property of the VaR violations. One of the earliest and most widely known test of *conditional coverage* which examines the independence property, or exception clustering, is the *independence test* (or Markov test), suggested by Christoffersen (1998). The Markov test of Christoffersen (1998) tests if the probability of VaR violation (exception) on any given day depends on the outcome of the previous day (Campbell, 2005). For instance, if the likelihood of a VaR exception increased on a day proceeding a previous VaR exception, then this would point towards a need to raise VaR level estimates, as successive losses would imply higher risk exposure. The method applies the same Likelihood-Ratio statistical testing framework as Kupiec for independence of exceptions.

The test is then set out as follows. Suppose we have data of portfolio returns for T days. Each day we set a deviation indicator, and set the indicator value as follows:

$$\text{Indicator } (I_t) = \begin{cases} 0 & \text{if VaR is not breached,} \\ 1 & \text{otherwise.} \end{cases}$$

Thus we have a sequence I_t of 0s and 1s. For any two consecutive days, there will only be four outcomes; 00, 01, 10 and 11. We then define $T_{i,j}(i = 0, 1; j = 0, 1)$ as the number of days in which state j occurred in one day while it was at i the previous day. $T_{0,0}$ as the number of days that the previous day's indicator is 0 and the subsequent day's indicator is 0, and $T_{1,1}$ as the number of days that the previous day's indicator is 1 and the subsequent day's indicator is 1 (Jorion, 2007).

The test is then conducted by creating a 2x2 contingency table with all possible outcomes of the deviation indicator. Table 2.3 below illustrates the outcomes of the Markov 2x2 contingency table:

Table 2.3: **Deviation Indicator Outcomes (Christoffersen, 1998)**

	$I_{t-1}=0$	$I_{t-1}=1$	
$I_t=0$	T_{00}	T_{10}	$T_{00}+T_{10}$
$I_t=1$	T_{01}	T_{11}	$T_{01}+T_{11}$
	$T_{00}+T_{01}$	$T_{10}+T_{11}$	$\mathbf{T}$

In addition, define π_i as the probability of having an exception conditional on state i in the previous day. Let π_0 be the conditional probability of 01 occurring if the previous day is 0 and π_1 be the conditional probability of 11 occurring if the previous day is 1 (Chatfield, 2001). It follows that

$$\pi_0=\frac{T_{01}}{T_{00}+T_{01}},\quad \pi_1=\frac{T_{11}}{T_{10}+T_{11}},\quad \pi=\frac{T_{01}+T_{11}}{T_{00}+T_{01}+T_{10}+T_{11}}, \tag{2.6.8}$$

and the sum of π_0 and π_1 is π. Under the null hypothesis that exceptions are independent across days, then the probabilities should be equal, $\pi=\pi_0=\pi_1$. That is, the chance of an exception occurring after a day of no exception is the same as occurring after a day of an exception (Campbell, 2005). If these proportions differ greatly from each other, then this calls the validity of the VaR measure into question.

If the model is accurate, then a VaR violation today should not depend on whether or not a violation occurred on the previous day (Jorion, 2007). The relevant test statistic for independence of exceptions is a Likelihood-Ratio (LR) given by:

$$LR_{Ind}=-2\ln[(1-\pi)^{(T_{00}+T_{10})}\pi^{(T_{01}+T_{11})}]+2\ln[(1-\pi_0)^{T_{00}}\pi_0^{T_{01}}(1-\pi_1)^{T_{10}}\pi_1^{T_{11}}]. \tag{2.6.9}$$

Same as with the POF-test, the LR-statistic (equation (2.6.9)) follows the the χ^2 (Chi-squared) distribution with 1-degree of freedom. Similarly, if the value of the LR_{Ind} -statistic falls below the critical value of χ^2 (Chi-squared) distribution, p with 1-degree of freedom, the model passes the backtest, otherwise the model is rejected.

The major drawback of the Markov test is its limited power against clustering. It mainly tests for independence of exceptions on two consecutive days. For instance, it might be the case that the likelihood of a VaR exception depends not on whether there was an exception the previous day, but whether there was an exception one week or two weeks ago. Clearly, if this is the way in which the lack of the independence property manifests itself, then the Markov test won't have power to detect such violations of the independence property. In order to deal with this weakness of the Markov test, Christoffersen and Pelletier (2004) recently suggested a new independence test that takes into account the time elapsed between exceptions. In other words, the time between VaR exceptions should not exhibit any kind of 'duration dependence', and so they proposed a *duration test* accordingly.

However, this new test of independence (duration test) comes with an added complexity in that the resulting maximum likelihood estimates do not have closed form solutions and must be evaluated via numerical methods (Huang, 2006). While Christoffersen and Pelletier (2004) provide some evidence that this test of independence has more power than the Markov test in detecting VaR measures that violate the independence property, this new test of independence is more of a black-box (Campbell, 2005). As such, the Markov test remains the most widely used independence test for examining the amount of clustering in VaR exceptions, largely due to its simplicity.

2.6.4 Joint Test (Christoffersen's Interval Forecast Test)

Given the shortcomings of the two tests presented above, it is clear that an accurate VaR measure should take into account both the *unconditional coverage* and *independence* properties. As such, tests that jointly examine the unconditional coverage and independence properties provide an opportunity to detect VaR measures which are deficient in one way or another (Campbell, 2005). One such test of *conditional coverage* which tests for both proportion of exceptions (unconditional coverage) and the clustering of exceptions (independence property) is the Christoffersens Interval Forecast test, again proposed by Christoffersen (1998).

The method again applies the same Likelihood statistic testing framework as Kupiec, but extends the LR_{POF}-test to jointly test for both unconditional coverage and independence. By combining the Kupiec's POF-test (LR_{POF}) statistic with the independence test (LR_{Ind}), a joint test that examines both properties of an accurate VaR model is produced. The correct failure rate and independence of exceptions is then attained (Jorion, 2007). The combined test statistic for conditional coverage is then given by:

$$LR_{cc} = LR_{POF} + LR_{Ind} \tag{2.6.10}$$

The combined LR also follows the χ^2 (Chi-squared) distribution, however with 2-degrees of freedom given that there are two separate LR-statistics in the test. Should the value of the LR_{cc} -statistic fall below the critical value of the χ^2 (Chi-squared) distribution, $p = (1 - c)$ with 2-degrees of freedom, the model passes the backtest. Higher values above the critical region signal an inaccurate model and should lead to a rejection of the model (Nieppola, 2009).

The combined framework therefore may seem to suggest that joint tests are universally preferable to separate tests of either the unconditional coverage property or independence property. Joint tests have the capacity to eventually detect a VaR measure which violates either of these properties combined. However, this comes at the expense of a decreased ability to detect a VaR measure which only violates one of the two properties. Past studies have shown that in some cases it is possible that the model passes the joint test while still failing either the unconditional coverage (LR_{POF}) test or independence (LR_{Ind}) test (Campbell, 2005). Thus it is advisable to also run separate tests even when the joint test acknowledges the model as providing accurate VaR estimates.

2.6.5 Backtesting Summary

Model validation in the form of backtesting provides very important feedback about the accuracy of VaR models to the end users of VaR estimates. Consideration should be given to the VaR quantitative parameters for backtesting. Firstly, the time horizon should be as short as possible in order to have sufficient observations and also to mitigate against the effect of changes in portfolio composition. Secondly, one should avoid high confidence levels for backtesting as they decrease the effectiveness or power of the statistical test (Jorion, 2007).

This section outlined four of the commonly used approaches to VaR model backtesting. An accurate VaR model has to satisfy two equally important properties. Firstly, it should produce the expected number of exceptions in line with the confidence level. Secondly, the VaR exceptions must be serially independent of each other. The simplest forms of backtests focus only on the failure rate (number of exceptions). However, more advanced methods (conditional tests) take into account the independence between exceptions, or jointly the number of exceptions and independence of exceptions.

The most common unconditional test is the Kupiec's POF-test, it examines the failure rate (number of exceptions) over some specified holding period. If statistically too many or too few exceptions are observed, the model is rejected. Another widely used backtesting method is the Basel Committee 'traffic light' approach, which is based on the same assumptions as the Kupiec POF-test. On the other hand, a common conditional test for examining the independence (clustering) of exceptions is the Christoffersen's independence test. A more advanced, conditional test presented herein is the Christoffersen's interval forecast test. It is a joint test statistic for the number of exceptions as well as independence of exceptions.

The backtesting process is a balancing act between two types of errors: rejecting an accurate model versus accepting an inaccurate model. A statistically powerful test should therefore minimize the probability of both of these errors (Campbell, 2005). The way of increasing the power of a test is by selecting relatively low confidence levels in VaR calculation so as to realize enough exceptions for statistical testing. It is also recommended to use the large set of data possible. In practice, however, there is rarely sufficient number of observations available (Jorion, 2007).

In the backtesting process, careful attention should be placed on the selection of the backtesting method. A narrow perspective, such as heavy reliance only on the unconditional coverage tests, could lead to the acceptance of a model that does produce the expected number of exceptions, but ignoring the clustering of exceptions. Reliance should never be only on a single backtesting technique. A positive result in one test should always be confirmed with another type of test

(Nieppola, 2009).

Chapter 3

Data and Methodology

> "*One of the most important points to appreciate about any VaR system is simply that the figures it produces are subject to error. There are many sources of error. Besides sampling error, errors can arise because of data problems, inappropriate models, inappropriate implementation decisions, or just plain human error.*"
>
> (Dowd, 1998)

The empirical component of this work is to examine the accuracy of the VaR estimates of the *historical simulation* methodology on a portfolio consisting of interest rate instruments (including derivatives). Validation of the VaR estimates is conducted on a sample portfolio that is drawn and resembles an *actual* portfolio of a South African financial institution. Due to confidentiality of their trading book, here and after, the financial institution shall be referred to as *the firm*.

The performance or accuracy of a VaR method depends very much on the type of portfolio being considered. As such, the decision about the VaR simulation method and the type of portfolio used in the study has not been random, but rather carefully selected. Historical simulation is the most widely used VaR approach by South African financial institutions, and indeed the top four banks, hence the use of this approach in the study. On the other hand, the interest rate portfolio (including interest rate derivatives) comprises an important share of any of the trading portfolio of South African financial institutions and banks. More precisely, they constitute 85% of the over-the-counter (OTC) derivatives market in South Africa[1].

[1]Source: Business Day, Feb 8 2013, pg 2.

In this examination of the accuracy of VaR estimates we utilize all four of the backtesting techniques presented in this paper: *Kupiec's POF-test*, *Basel Committee 'traffic light'* approach, *Independence test* and the *Joint-test*. The backtesting procedures are conducted by comparing daily realized losses (left tail) with daily VaR estimates using a one year horizon (251 trading days). Similarly, VaR estimates are calculated using a one day moving window over a 251 day holding period.

3.1 Portfolio Composition and Data

All market and sample data is actual data for the period 1 June 2010 to 30 July 2011. The data in the study is an *actual* sample of the interest rate portfolio of the firm. We mention here that the study portfolio was not hypothetical but a smaller extract of the firm's interest rate portfolio. The portfolio consisted of both linear and non-linear instruments. However, there is a large skew towards linear instruments, a reflection of the firm's interest rate portfolio.

The portfolio comprised of term loans, interest rate swaps, Johannesburg Stock Exchange (JSE) listed government and corporate bonds, interest rate swaps, swaptions[2], FRAs[3] and Repos[4]. It also included some of the hedge positions for the book. Furthermore, in order to assess the impact of the hedge positions on the VaR estimates, we split the portfolio into two. The dataset is split into two, a *hedged portfolio* that includes all instruments plus hedge positions, and an *unhedged portfolio*, which excludes all the hedge positions from the portfolio.

Maturities for the instruments ranged from short term (0-5 years) to long term (5-30 years). The presence of long term bonds and swaptions introduces non-linearity in the portfolio from bond convexity[5] and options gamma[6], respectively. This kind of diversification in the portfolio structure enables us to effectively identify potential accuracy problems in the most commonly used VaR approach, historical simulation. In addition, it also allows us to examine the model's

[2]options on an interest rate swaps
[3]Forward Rate Agreement
[4]Repurchase agreements
[5]Refers to the non-linear relationship between bond price and interest rates
[6]Greek letter for the second order sensitivity on interest rate derivatives

ability to fully capture the interest rate risk of an 'actual' portfolio.

Daily market data for instrument valuation by both the source system and the risk management system was sourced from market data providers such as SAFEX[7], Bloomberg, Reuters and Tortem[8]. The firm has an internal method for bootstrapping the zero (bond, swap, inflation) curves that are used for interest rate valuation using observable market instruments. Absolute shifts of the bootstrapped curves were fed into the valuation system when re-valuing the current portfolio for the past 251 day holding period. This was done in order to minimize model risk.

3.2 VaR Calculation Process

The VaR of the study portfolio was calculated using *full valuation*, that is, fully repricing the portfolio over the past 251 days as if we had the same portfolio. This uses the past 251 day market data in each backward hypothetical valuation. Same as with the choice of VaR method and portfolio, the selection of parameters in VaR calculations was not arbitrary. Backtesting techniques require that you avoid higher higher confidence levels so as to attain enough observations. A confidence level of 95% is generally recommended for backtesting purposes (Jorion, 2007). With this confidence level, it is possible to observe enough VaR exceptions within the one year time period (251 trading days).

Software used in the VaR calculations was the firm's risk management system, *RiskWatch*. Portfolio valuations were conducted in the firm's source system, *Murex*. In order to make the backtesting process more effective, VaR estimates were outputted for three confidence levels, 90%, 95% and 99%. A comparison of the realized profit and loss vs the VaR estimates was then conducted in excel spreadsheets.

[7]South African Future's Exchange
[8]Data providers of market volatility data

3.3 Backtesting Process

Realized profit and loss was calculated in the firm's front office system (Murex) while VaR estimates were computed in the risk management system (RiskWatch). Results from the two systems were then transferred to a spreadsheet where the comparison of the actual profit and loss against the VaR estimates was then conducted. The number of exceptions would then be counted against the VaR estimates. These exception figures would then be used in each backtest model to validate the VaR estimates.

The profit and loss at time t was defined as r_t, i.e. the current reported profit and loss, and the corresponding VaR estimate for time t as $\text{VaR}_{(t-1)}$. That is, $\text{VaR}_{(t-1)}$ is calculated at the beginning of the period t using the closing prices of day $(t-1)$. In conducting the backtests, the number of exceptions or the test statistic values where compared to a mapping table with either the number of exceptions or the χ^2 (Chi-squared) distribution critical value, depending on the backtest method.

3.4 Problems with the Data

There were some missing data points in the data that would result in the bootstrapped curves perhaps not being as accurate.

Chapter 4

Empirical Results

“*This VaR is like an airbag that works all the time except when you have a car accident.*” (Einhorn, 2008)

Haas (2001) and Campbell (2005) argue that more than one backtesting technique should always be used to validate the accuracy of a VaR model. That is, the outcome of one test should always be confirmed by another test. As such, in this work, we utilize all four backtesting methods earlier presented in validating the accuracy of the VaR estimates.

The POF-test and the Basel Committee ‘traffic light’ approach are used to examine the *frequency of exceptions* (VaR violations), the *independence test* to check for clustering of exceptions. The Christoffersen’s Interval Forecast is used for the *joint test* for both frequency of exceptions and independence of exceptions. The application of these fairly traditional, but most common, approaches to backtesting only requires the number of exceptions, the time when the exceptions occur and the total number of observations.

Following the arguments of Campbell (2005), Niepolla (2009) and Jorion (2007), that higher confidence levels should be avoided for backtesting purposes, we utilize a fairly not so high confidence level. As such, all backtests are conducted at 95% percentile of the χ^2 (Chi-squared) distribution (Appendix A) as the critical value for the Likelihood-Ratio tests. This implies that relatively strong evidence is required in the rejection of the model, Niepolla (2009). We reiterate here that all the VaR numbers were generated using the *historical* simulation method.

4.1 VaR Data Consolidated Output

As alluded to in Section 3.3, the VaR estimate $\text{VaR}_{(t-1)}$ was compared to the realized profit and loss, r_t, at time t, in order to establish the number of exceptions as well as when they occur. Table 4.1 below presents the consolidated results of the number of exceptions from the analysis. This is the main output required for backtesting.

Table 4.1: **Consolidated Backtesting Data**

	Confidence Level (c)	Comparison Results		
		Number of Observations (T)	**Expected Number of Exceptions**	**Realized Number of Exceptions (x)**
Unhedged Portfolio	99%	251	3	13
	95%	251	13	24
	90%	251	25	33
Hedged Portfolio	99%	251	3	8
	95%	251	13	16
	90%	251	25	29

4.2 Frequency of Exceptions

Some of the earliest proposed VaR backtests focus on the property of unconditional coverage. Results from two such tests utilized in this work, that examine the number of VaR violations or frequency of exceptions (failure rate), are presented below.

4.2.1 Kupiec (1995) POF-Test

First we utilize Kupiec's POF-test to statistically test the model's accuracy in estimating the proportion of exceptions (unconditional coverage). Table 4.2 below shows the the model's performance for the split test portfolio.

Table 4.2: **Kupiec's POF-Test Results**

	Confidence Level	Kupiec's POF-Test Number of Observations	Realized Number of Exceptions	Test Statistic LR_{POF}	Critical Value $\chi^2(1;0.95)$	Test Result
Unhedged Portfolio	99%	251	13	22.23	3.84	Reject
	95%	251	24	8.78	3.84	Reject
	90%	251	33	2.54	3.84	Accept
Hedged Portfolio	99%	251	8	7.69	3.84	Reject
	95%	251	16	0.92	3.84	Accept
	90%	251	29	0.64	3.84	Accept

Note: $\chi^2(1;0.95)$ *is Chi-squared distribution values at the 95% percentile with 1-degree of freedom.*

The table (Table 4.2) is populated by substituting the relevant output data from the consolidated results table (Table 4.1) into the LR-test statistic (LR_{POF}) equation (2.6.5). For instance, the *hedged portfolio* has $x = 16$ realized exceptions at $c = 95\%$ confidence level over $T = 251$ trading days. Substituting this data into LR-test statistic equation (2.6.5) yields:

$$\begin{aligned} LR_{POF} &= -2\ln\left[(1-0.05)^{(251-16)}0.05^{16}\right] + 2\ln\left[(1-16/251)^{(251-16)}(16/251)^{16}\right] \\ &\approx 0.92. \end{aligned} \tag{4.2.1}$$

This result (0.92) is below the χ^2(Chi-squared) critical value of 3.84 (95% percentile with 1-degree of freedom), and therefore the model accuracy is accepted.

The test results highlight a serious accuracy problem with the VaR model in both the *unhedged* and *hedged* portfolios for VaR calculated at high confidence levels. The model accuracy is rejected when testing VaR estimates of the portfolio without hedge positions (unhedged portfolio) as well as when hedge positions are included (hedged portfolio) at the 99% VaR confidence level. That is to say, when estimating extreme tail losses, which is when the 'fat tails' problem appears, the model may not have the required precision to accurately estimate portfolio risk. This is in part due to a lack of normality or in part to a lack of linearity.

However, model accuracy is acceptable at lower confidence levels. For the 'actual' (hedged) portfolio, the model accuracy is accepted in both cases of VaR estimates at 95% and 90% confidence levels. On the other hand, when hedge positions are excluded, model accuracy is rejected for VaR estimates at the 95% as well as 99% level. This is a very interesting result as it reveals the effect of the hedge positions in the reduction of the entire portfolio risk, and consequently VaR.

4.2.2 Basel Committees (1996) 'Traffic Light' Approach

Like the POF-test, the Basel Committee 'traffic light' approach is an unconditional coverage test. It tests for the frequency of exceptions (failure rate). This backtesting approach is mainly used for regulatory purposes at the 99% confidence level, and exception ranges are provided for this confidence level by the regulatory framework. However, for internal backtesting purposes various confidence level exception ranges can be computed. We utilize the VaR *translation* property (see section 2.4) together with the Binomial distribution tables to compute the 'traffic light' exception ranges for the 90% and 95% confidence levels. Table 4.3 below displays the cut-off regions for the number of exceptions using the 'traffic light' approach for the three VaR confidence levels.

Table 4.3: **Basel Committee 'Traffic Light' Tables**

	Zone		
Confidence Level	**Green Zone**	**Yellow Zone**	**Red Zone**
99%	0 - 4	5 - 9	10+
95%	0 - 17	18 - 26	27+
90%	0 - 32	33 - 43	44+

Recall that the *green* zone defines an accurate VaR model, the *yellow* zone a model that maybe accurate or inaccurate and the *red* zone indicates a problematic VaR model. Results of testing for the frequency of exceptions (unconditional coverage) using the Basel Committee 'Traffic Light' test are presented in Table 4.4 below.

Table 4.4: **Basel Committee 'Traffic Light' (1996) Approach**

	Confidence Level	Basel Committee 'Traffic Light' Number of Observations	Realized Number of Exceptions	Test Result
Unhedged Portfolio	99%	251	13	Red Zone
	95%	251	24	Yellow Zone
	90%	251	33	Green Zone
Hedged Portfolio	99%	251	8	Yellow Zone
	95%	251	16	Green Zone
	90%	251	29	Green Zone

The findings show a weakness in the model accuracy at high confidence levels (99% level), *red zone* for the unhedged portfolio and *yellow zone* for the 'actual' (hedged) portfolio. This is a confirmation of the results from the POF-test. The finding makes intuitive sense since both tests are based on the same testing framework looking at the failure rate. At lower confidence levels the model accuracy is accepted, *green zone* on the 'actual' (hedged) portfolio at both 95% and 90% confidence levels.

4.3 Independence Test

The next test examines the clustering of exceptions. An accurate VaR model has to produce estimates that are serially independent (Campbell, 2005).

4.3.1 Independence Test - Christoffersen (1998)

First we present computed input data to be used in the test statistic. The input data is calculated by utilizing the the deviation indicator outcomes of the Markov 2x2 contingency tables (Table 2.3) together with the conditional probabilities (equation 2.6.8) of having an exception on the output data. Below, Table 4.5 displays the computed input data for the independence LR-test statistic (LR_{Ind}).

Table 4.5: **Independence Test Input Data**

		Independence Test Data								
	CL	**NoO**	**RnE**	T_{00}	T_{01}	T_{10}	T_{11}	π_0	π_1	π
Unhedged Portfolio	99%	251	13	226	12	12	1	5.04%	7.69%	5.18%
	95%	251	24	208	19	19	5	8.37%	20.83%	9.56%
	90%	251	33	193	25	25	8	11.47%	24.24%	13.15%
Hedged Portfolio	99%	251	8	235	8	8	0	3.29%	0.00%	3.19%
	95%	251	16	222	13	13	3	5.53%	18.75%	6.37%
	90%	251	29	199	23	23	6	10.36%	20.69%	11.55%

***NB: CL** is confidence level; **NoO** is number of observations; **RnE** is realized number of exceptions. $T_{i,j}$ is the number of days in which state j occurred in one day while it was at i the previous day, i.e. $T_{1,1}$ is the number of consecutive exceptions. And π_i is the probability of an exception conditional on state i the previous day, i.e. π_0 is the conditional probability of an exception occurring given no exception the previous day.*

The results of the statistical test of independence of exceptions utilizing Christoffersen (1998) test statistic for independence are presented in Table 4.6 below. In populating Table 4.6 input data results from Table 4.5 are substituted into the LR-test statistic (LR_{Ind}) equation (2.6.9).

For example, take the *hedged portfolio* at the $c = 95\%$ confidence level over the $T = 251$ trading days. Substituting the number of conditional $T_{i,j}$ exceptions and the probability of conditional π_i exceptions from the results data (Table 4.5) into LR-test statistic (LR_{Ind}) equation (2.6.9) gives:

$$\begin{aligned} LR_{Ind} &= -2\ln[(1-0.06)^{(222+13)}0.06^{(13+3)}] + 2\ln[(1-0.055)^{222}0.055^{13}(1-0.19)^{13}0.19^{3}] \\ &\approx 3.08. \end{aligned} \tag{4.3.1}$$

The test statistic of 3.08 is below the χ^2(Chi-squared) critical value of 3.84 (95% percentile with 1-degree of freedom), and therefore the model accuracy is accepted.

Table 4.6: **Independence Test - Christoffersen (1998)**

	Confidence Level	**Independence Test**				
		Number of Observations	**Realized Number of Exceptions**	**Test Statistic** LR_{Ind}	**Critical Value** $\chi^2(1; 0.95)$	**Test Result**
Unhedged Portfolio	99%	251	13	0.16	3.84	Accept
	95%	251	24	3.12	3.84	Accept
	90%	251	33	3.52	3.84	Accept
Hedged Portfolio	99%	251	8	0.53	3.84	Accept
	95%	251	16	3.08	3.84	Accept
	90%	251	29	2.30	3.84	Accept

Note: $\chi^2(1; 0.95)$ *is Chi-squared distribution values at the 95% percentile with 1-degree of freedom.*

The main take out from the independence test results is that exceptions are serially independent. Put in another way, there is no statistically significant results to prove dependence of exceptions in the model. While there is a statistically significant number of exceptions (POF-test results show), in some cases when exceptions occur, the test results suggest these are spread over time. In other words, there is no clustering of exceptions whether with or without hedge positions in the portfolio. This is an indication that proves the absence of successive losses, a phenomenon that tends to lead to bankruptcy if not accurately measured and monitored by risk management systems or models.

Despite acceptance of the model based on the test results, we caution on making blanket conclusions on independence of exceptions based on this test alone. The main weakness of the test statistic is that it only examines independence between two consecutive days and does not capture more general forms of dependence between exceptions such as duration based dependence.

4.4 Joint Test of Unconditional Coverage and Independence

An accurate VaR measure must exhibit both the unconditional coverage and independence properties. Such tests that jointly examine the properties together provide an opportunity to detect VaR measures which are deficient in one way or the other (Campbell, 2005).

4.4.1 Christoffersen's Interval Forecast Test

The *joint* test of Christoffersen's Interval Forecast was applied to test for both the frequency of exceptions as well as independence (clustering) of exceptions, combined. Results of the joint test on the split portfolio are presented in Table 4.7 below.

Table 4.7: **Joint Test - Christoffersen's Interval Forecast Test (1998)**

		Joint Test - Christoffersen's Interval Forecast Test						
	CL	**NoO**	**RnE**	**Test Statistic** LR_{POF}	**Test Statistic** LR_{Ind}	**Test Statistic** LR_cc	**Critical Value** $\chi^2(2;0.95)$	**Test Result**
Unhedged	99%	251	13	22.23	0.16	22.39	5.99	Reject
Portfolio	95%	251	24	8.78	3.12	11.90	5.99	Reject
	90%	251	33	2.54	3.52	6.06	5.99	Reject
Hedged	99%	251	8	7.69	0.53	8.22	5.99	Reject
Portfolio	95%	251	16	0.92	3.08	4.00	5.99	Accept
	90%	251	29	0.64	2.30	2.94	5.99	Accept

***NB: CL** is confidence level; **NoO** is number of observations; **RnE** is realized number of exceptions.* $\chi^2(2;0.95)$ *is Chi-squared distribution values at the 95% percentile with 2-degrees of freedom.*

The tabulated results of the joint LR-test statistic of *conditional coverage* LR_{cc} (Table 4.7) are attained by directly summing the frequency of exception LR-test (LR_{POF}) and the independence LR-test (LR_{Ind}) equation (2.6.10). The sum is again an LR-test against the χ^2 (Chi-squared) distribution, albeit with 2-degrees of freedom. As with the previous tests, let us consider the *unhedged portfolio* at, $c = 95\%$ confidence level, over $T = 251$ trading days. Substituting the

$LR_{POF} = 8.78$ and $LR_{Ind} = 3.12$ into equation (2.6.10) we get:

$$\begin{aligned} LR_{cc} &= 8.78 + 3.12 \\ &\approx 11.90. \end{aligned} \tag{4.4.1}$$

The test statistic of 11.90 is above the χ^2(Chi-squared) critical value of 5.99 (95% percentile with 2-degrees of freedom), and therefore the model is rejected.

Results of the joint test reject the accuracy of the model at all confidence levels when hedge positions are excluded from the portfolio (unhedged portfolio). Similarly, results of the joint test far exceeded the test statistic critical values for the 99% level of confidence even when hedge positions are included in the portfolio (hedge portfolio). This outcome is not unexpected as the POF-test affirmed the weakness of the model in accurately estimating portfolio risk at higher confidence levels. Furthermore, the POF-test showed the under estimation of risk when hedge positions are excluded. However, model accuracy for the 'actual' (hedged) portfolio is accepted at lower VaR confidence levels, i.e. 95% and 90%.

4.5 Summary Backtesting Results

We summarize results of the four backtests on the VaR model estimates for the split portfolio below. In reviewing the results we start off by reiterating the portfolio composition. That is, it was largely dominated by linear instruments despite there being a fairly reasonable amount of non-linear transactions, amongst them hedge positions. Table 4.8 below presents the summarized findings of the backtests.

Table 4.8: **Backtesting Summary Results**

	CL	NoO	RnE	Frequency of Exceptions Test		Independence Test	Joint Test
				LR_{POF}	**Traffic Light**	**LR_{Ind}**	**Christoffersen Interval Forecast**
Unhedged Portfolio	99%	251	13	Reject	Red Zone	Accept	Reject
	95%	251	24	Reject	Yellow Zone	Accept	Reject
	90%	251	33	Accept	Green Zone	Accept	Reject
Hedged Portfolio	99%	251	8	Reject	Yellow Zone	Accept	Reject
	95%	251	16	Accept	Green Zone	Accept	Accept
	90%	251	29	Accept	Green Zone	Accept	Accept

NB: CL *is confidence level;* ***NoO*** *is number of observations;* ***RnE*** *is realized number of exceptions.*

In both portfolios, unhedged and 'actual' (hedged), model accuracy in the independence test is accepted for all confidence levels. In other words, there is no evidence of clustering of exceptions whether with or without hedge positions in the portfolio. An indication that proves the absence of daily successive losses, a phenomenon that tends to lead to bankruptcy if not accurately measured and monitored by risk management systems or models.

On the other hand, for both the unhedged and 'actual' portfolios, the model's VaR estimation accuracy at the 99% confidence level are rejected based upon the frequency of exceptions and joint tests. The number of VaR violations (8 exceptions at the 99% level) for the 'actual' portfolio are more than double the expected number of exceptions (3 exceptions at the 99% confidence level).

However, the model accuracy for the 'actual' portfolio passes all backtesting (frequency of exceptions, independence and joint) tests at lower VaR confidence levels (95% and 90%). To the contrary, when hedge positions are excluded from the portfolio the model's performance in accurately estimating VaR is questionable. This is due to the fact that it fails the frequency of exceptions test at almost all confidence levels, and the joint test at all confidence levels.

Chapter 5

Discussion and Conclusions

> "*Value at Risk has become a standard measure of financial market risk that is increasingly used by financial institutions and even non-financial institutions as well.*" (Berkowitz & O'Brien, 2002)

VaR has emerged as one of the most prominent risk measurement and management techniques in finance. Presently, it is the most used by financial institutions, non-financial institutions, and regulators alike. However, it is important to note that when implementing VaR systems, a number of simplifications and assumptions are involved. Notwithstanding the wide use and common acceptance of VaR as a risk management tool, several criticisms have arisen concerning VaR methods (and are still debated). This has led to the questioning of the accuracy of VaR estimates from the various VaR models. There are a number of studies that have been conducted comparing the performance of the various VaR methods. However, most VaR models' comparison studies often compare different methods using small hypothetical portfolios and not 'actual' or representative portfolios of banks or financial institutions. In addition, very few if any, have assessed the performance of these models on an 'actual' South African market portfolio.

The first part of this work provided a theoretical comparison of some of the VaR estimation techniques. More precisely, we compared the *variance-covariance* approach, *historical simulation* and *Monte Carlo simulation*. The theoretical comparison of these methods placed emphasis on their shortcomings, in cognizance that their potential flaws provide motivation for the backtesting of VaR. In addition, a theoretical comparison of some of the, fairly traditional, backtesting methods was presented.

The fundamental properties of an accurate VaR model, *unconditional coverage*, the *independence property* and *conditional coverage* were defined and their relevance from the perspective of examining the accuracy of VaR model estimates was discussed. Tests that examine the validity of the unconditional coverage property, the independence property, or joint properties were reviewed. Four backtesting techniques were presented, namely; *Kupiec's proportion of failures-test*, *Basel Committees 'traffic light'* approach, *independence test* and the *joint-test*.

The second part of this work was an empirical study that focused on applying the presented backtesting techniques in validating the accuracy of VaR numbers. These were based on the historical simulation model, applied in estimating risk of a South African market 'actual' portfolio. VaR estimates were run for three confidence levels (90%, 95% and 99%). As a way of creating a comparative test, two portfolios were created, one that had all transactions including hedge positions (hedged portfolio) and one that excluded the hedge positions (unhedged portfolio).

Results of the backtests reveal that in both portfolios, unhedged and 'actual' (hedged), model accuracy in the independence test is accepted for all confidence levels. In other words, there is no evidence of clustering of exceptions whether with or without hedge positions in the portfolio. The empirical findings further show that the model accuracy for the 'actual' portfolio passes all backtesting (frequency of exceptions, independence and joint) tests at lower VaR confidence levels (95% and 90%). However, when hedge positions are excluded from the portfolio the model's performance in accurately estimating VaR is questionable. The model fails the frequency of exceptions test at almost all confidence levels, and the joint test at all confidence levels. For both the comparative (unhedged) and 'actual' portfolios, test results indicate a weakness in the ability of the VaR model's capability to accurately estimate portfolio risk at higher confidence levels (in particular the 99% level).

These results show that model accuracy is acceptable when hedge positions are included in the test portfolio, and also when lower VaR confidence levels are considered. This reinforces mainly two things. Firstly, the need to conduct backtests at lower confidence levels where sufficient number of exceptions can be observed in order to conduct a meaningful statistical test. Secondly, the need to conduct such tests on 'actual' portfolios with all hedge positions included. In addition, an interesting observation from the results is the accuracy of the model when hedge positions are included. This reveals the effect of hedges in reducing overall portfolio risk, and consequently VaR. This is a possible counter argument as to why Engelbrecht (2003) finds against the accuracy of the VaR historical simulation method in estimating risk for an interest portfolio on the South African market. This may provide a useful critique to Engelbrecht (2003) as their study uses a hypothetical sample that does not have hedge positions.

Despite the rather comforting positive results of accurate performance of the model at lower confidence levels, caution must be given in the interpretation. Once again, recall the major VaR assumption, that is, it quantifies risk under normal conditions. Therefore, while our results seem to pass the model as producing fairly accurate VaR estimates, even a sound VaR model may perform poorly in adverse market conditions, i.e. in times of high volatility or significant changes in asset correlations.

A key point of the study outcome is that the accuracy of VaR estimates of the examined VaR model on the tested portfolio maybe questionable at high VaR confidence levels. However, are fairly reliable at lower confidence levels. The study presented some of the shortcomings of VaR calculation, and hence the need for backtesting the numbers. Systematic backtesting should be a regular part of VaR reporting in order to constantly monitor the performance of models. To this end, it is imperative that VaR estimates are never considered to be the Holy Grail of portfolio risk management, irrespective of how sophisticated the VaR model maybe. If the users of VaR numbers comprehend the shortcomings associated with VaR models, this could lead to better interpretation of VaR numbers and consequently improved financial risk management.

The findings of this study could assist financial institutions in emerging markets (and in particular South Africa) in the determination of appropriate VaR methodologies for the management of risk of their portfolios, as well as defining the appropriate VaR model validation techniques. In addition, the study may ensure the prudent use of capital by financial institutions as VaR estimates are used for regulatory capital purposes.

In conclusion, there are several interesting avenues for further research to complement this study. The most obvious would be to look at VaR validation under stressed periods, as this work focused on backtesting VaR under normal conditions. In other words, backtesting stressed VaR. A very interesting topic but also very challenging as it requires a robust amount of data that spans stressed periods.

Appendix A

χ^2 **(Chi-squared) Distribution Tables**

df	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0	0	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.01	0.02	0.051	0.103	0.211	4.605	5.991	7.378	9.21	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.86
5	0.412	0.554	0.831	1.145	1.61	9.236	11.07	12.833	15.086	16.75
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.69	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.18	2.733	3.49	13.362	15.507	17.535	20.09	21.955
9	1.735	2.088	2.7	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.94	4.865	15.989	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.92	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.3
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.66	5.629	6.571	7.79	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.39	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.26	9.591	10.851	12.443	28.412	31.41	34.17	37.566	39.997
21	8.034	8.897	10.283	11.591	13.24	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.26	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.98	45.559
25	10.52	11.524	13.12	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.16	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.29
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.169	67.505	71.42	76.154	79.49
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.192	53.54	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.294
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Bibliography

[1] Bams, D., T. Lehnert, and L. Wolff (2005). "An Evaluation Framework for Alternative VaR Models." *Journal of International Money and Finance* **24(6)**:944-958.

[2] Bao, Y., T. Lee, and B. Saltoglu. (2006). "Evaluating Predictive Performance of Value at Risk Models in Emerging Markets: A Reality Check." *Journal of Forecasting* **25(2)**:101-128.

[3] Barone-Adesi, G., K. Giannopoulos, and L. Vosper. (1999). "VaR without Correlations for Portfolios of Derivative Securities." *Journal of Futures Markets* **19**:583-602.

[4] Basel Committee on Banking Supervision. (1996a). "Overview of the Amendment to the Capital Accord to Incorporate Market Risks." Available at www.bis.org.

[5] Basel Committee on Banking Supervision. (1996a). "Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements." Available at www.bis.org.

[6] Basel Committee on Banking Supervision. (2006). "International Convergence of Capital Measurement and Capital Standards A Revised Framework." Available at www.bis.org.

[7] Berkowitz, J., and J. O'Brien. (2002). "How Accurate Are Value at Risk Models at Commercial Banks?" *Journal of Finance* **57**:1093-1111.

[8] Brooks, C., A. D. Clare, J. W. Dalle Molle, and G. Persand. (2005). "A Comparison of Extreme Value Theory Approaches for Determining Value at Risk." *Journal of Empirical Finance* **12**:339-352.

[9] Campbell, S. (2005). "A Review of Backtesting and Backtesting Procedures." Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.

[10] Christoffersen, P. F. (1998). "Evaluating Interval Forecasts." *International Economic Review* **39**:841-862.

[11] Christoffersen, P. F. (2003). *Elements of Financial Risk Management*, 1st edn. Amsterdam: Academic Press.

[12] Christoffersen, P., and Pelletier D. (2004), "Backtesting Value at Risk: A Duration-Based Approach." *Journal of Empirical Finance* **2**:84-108.

[13] Coronado, M. (2000). "Comparing Different Methods for Estimating Value at Risk for Actual Non-Linear Portfolios: Empirical Evidence." *European Journal of Finance.*

[14] Crnkovic C., and Drachman J. (1997). "Quality Control in VaR: Understanding and Applying Value at Risk." *Risk Publications.*

[15] Damodaran, A. (2007). "Strategic Risk Taking: A Framework for Risk Management." Pearson Education, New Jersey.

[16] Danielsson, J., and C. G. de Vries. (2000). "Value at Risk and Extreme Returns." *Annales d'Economie et de Statistique* **60**:239-270.

[17] Dowd, K. (1998). *Beyond Value at Risk*, 2nd edn. Chichester: John Wiley & Sons.

[18] Dowd, K. (2000). "Assessing VaR Accuracy." *Derivatives Quaterly* **6(3)**:61-63.

[19] Dowd, K. (2005). *Measuring Market Risk*, 1st edn. Chichester: John Wiley & Sons.

[20] Einhorn, D. (2008). "Private Profits and Socialized Risk." *Global Association of Risk Professionals* Jun/Jul issue.

[21] Englebrecht, R. (2003). "A Comparison of Value at Risk Methods for Portfolios Consisting of Interest Rate Swaps and FRAs." Paper, *CiteSeerx.*

[22] Giot, P., and S. Laurent. (2004). "Modeling Daily Value at Risk Using Realized Volatility and ARCH Type Models." *Journal of Empirical Finance* **11**:379-398.

[23] Gnamassou, Y. T. (2010). "Value at Risk Prediction: A comparison of Alternative Techniques Applied to a Large Sample of Individual Stock Data." Paper, HEC Montreal.

[24] Haas, M. (2001). "New Methods in Backtesting." *Financial Engineering Research Center Caesar*.

[25] Holton, G. A. (2003). *Value at Risk: Theory and Practice*, 1st edn. Elsevier Academic Press.

[26] Huisman, R., K. G. Koedijk, C. J. M. Kool, and F. Palm. (2001). "Tail Index Estimates in Small Samples." *Journal of Business and Economic Statistics* **19**:208-216.

[27] Hull, J., and A. White. (1998). "Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk." *Journal of Risk* **1(1)**:5-19.

[28] Hull, J. (2012). *Risk Management and Financial Institutions*, 3rd edn. John Wiley & Sons.

[29] J.P. Morgan. (1996). *RiskMetrics Technical Document*, 4th edn. New York: J.P. Morgan/Reuters. Available at www.riskmetrics.com.

[30] Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd edn. New York: McGraw-Hill.

[31] Kuester, K., S. Mittnik, and M. S. Paolella. (2006). "Value at Risk Prediction: A Comparison of Alternative Strategies." *Journal of Financial Econometrics* **4(1)**:53-89.

[32] Kupiec, P. H. (1995). "Techniques for Verifying the Accuracy of Risk Measurement Models." *Journal of Derivatives* **3**:73-84.

[33] Linsmeier, J., & N.D. Pearson. (1996). "Risk Measurement: An Introduction to Value at Risk." *Working Paper*, University of Illinois.

[34] Lopez, J. A. (1997). "Regulatory Evaluation of Value at Risk Models." Staff Report 33, Federal Reserve Bank of New York.

[35] Lopez, J. A. (1998). "Methods for Evaluating Value at Risk Estimates" *Economic Policy Review* **(10)**:119-64.

[36] Nieppola, O. (2009). "Backtesting Value at Risk Models." Paper, Helsinki School of Economics.

[37] Pérignon, C., and D. Smith. (2008). "A New Approach to Comparing VaR Estimation Methods." *The Journal of Derivatives* **16(2)**:54:66.

[38] Pérignon, C., and D. Smith. (2010). "The Level and Quality of Value at Risk Disclosure by Commercial Banks." *Journal of Banking and Finance* **34**:362-377.

[39] Pritsker, M. (1997). "Evaluating Value at Risk Methodologies: Accuracy versus Computational Time." *Journal of Financial Services Research* **12**:201-242.

[40] Simons, K. (1996). "Value at Risk - New Approaches to Risk Management: Understanding and Applying Value at Risk." *Risk publications.*

[41] Stahl, G. (1997). "Backtesting Using a Generalisation of the 'Traffic Light' Approach." *Neural Network World* **4**:565-577.